Pattern Recognition 44 (2011) 1023–1030

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

Fuzzy posterior-probabilistic fusion

Tuan D. Pham

School of Engineering and Information Technology, University of New South Wales, Canberra, ACT 2600, Australia

ARTICLE INFO

Article history:
Received 23 November 2009
Received in revised form
1 August 2010
Accepted 18 November 2010

Keywords:
Information fusion
Permanence of ratios
Probabilistic measures
Fuzzy measures
Fuzzy integrals

ABSTRACT

The paradigm of the permanence of updating ratios, which is a well-proven concept in experimental engineering approximation, has recently been utilized to construct a probabilistic fusion approach for combining knowledge from multiple sources. This ratio-based probabilistic fusion, however, assumes the equal contribution of attributes of diverse evidences. This paper introduces a new framework of a fuzzy probabilistic data fusion using the principles of the permanence of ratios paradigm, and the theories of fuzzy measures and fuzzy integrals. The fuzzy sub-fusion of the proposed approach allows an effective model for incorporating evidence importance and interaction.

1. Introduction

Combining information from diverse sources is a challenging and important area of research in pattern classification. There are a number of different approaches for information fusion and the literature is vast due to many developments and applications in this broad field, which pervades many scientific disciplines since the early work over many years [1–6]. Although information fusion has been reported to be useful in many practical applications of various science and engineering disciplines [7–18], it is still a challenging task which continues to be explored for effective handling of various kinds of mathematical models for measuring event uncertainty [19]. Particularly on application of fuzzy-set theory to information fusion, recent developments and applications can be found in many areas such as pattern classification, image analysis, decision making, man-made structures, medicine, biology and humanity [20–29].

Mathematical properties of various popular information fusion operators were investigated by Bloch [9]. These operators, including naive Bayesian fusion (product rule), fuzzy fusion (T-norm, T-conorm), certainty-factor fusion (MYCIN) and evidence-based fusion (Dempster–Shafer theory of evidence); can be classified into three behaviors: conjunctive (if the sources of information have low conflicting evidences), disjunctive (if the sources of information have high conflicting evidences), or compromising (if the sources of information have partial conflicting evidences). Fuzzy fusion operators can be suitable for dealing with all the above three behaviors in both context dependence and context independence, while the naive Bayesian fusion and Dempster–Shafer fusion operators were suggested to be only suitable for the conjunctive behavior in the case of context dependence. Although the naive Bayesian fusion is relatively easier to implement and often performs well in many applications, this approach has its own disadvantage on the assumption of the independence of classifiers when conflicting evidence arises. Therefore, good knowledge about the mathematical properties of different information fusion operators allows one to select an appropriate operator for effective application to a particular decision fusion problem.

Altincay [31] investigated the application of the naive Bayesian fusion (NBF) of dependent and independent classifiers and found that the combination of dependent classifiers performed better than that of independent classifiers when both using the same probabilistic approach. In the same study, the author found that the NBF may not give a better result than the best individual classifier when the individual accuracies of individual classifiers are not comparable. The finding may suggest the importance of the dependency of different classifiers in the Bayesian fusion approach. Hence, a probabilistic fusion which can address the property of classifier dependence is expected to perform better than the NBF. In the study of feature subset selection using filter and wrapper methods, Kohavi and John [32] used the naive Bayesian classifier to compute the probability of each class, assuming the features are conditionally independent. The authors found that the naive Bayesian algorithm could hardly improve the results using feature selection as this low performance was partly due to the fact that the naive Bayes was adversely affected by the conditional dependence between the features.

From the standpoint of statistics, information fusion is the study of integrating prior and preposterior probabilities into a posterior

E-mail address: t.pham@adfa.edu.au

0031-3203/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.patcog.2010.11.014

probability [30]. To alleviate the mathematical derivation of a data fusion model, the assumption of conditional independence of the data events is often imposed to conveniently calculate the posterior probabilities. This assumption may lead to inaccurate or inconsistent results when evidences about the event under study are obtained can relate to each other, and have been widely realized in data analysis in geoscience [19,33]. For example, there is no rigorous support for assuming data independence or conditional independence in various features of background and objects in images; because different features relate to either background or objects events share the same image property. As another example, a particular spatial feature extracted from one region often relates to that from another nearby location: this is due to the nature of continuity of spatial data. This realization gives rise to the challenge of the calculation of posterior probabilities without the hypothesis of any form of data independence.

It has been known that the assumption of data independence is usually employed in statistical theories, and the independence assumption is prevailed by introducing linear dependence between the data values. A simple and practical probabilistic solution for event combination is by means of weighted linear averaging of prior probabilities so that the posterior conditional probability can be conveniently calculated. For example, the calculation of probability $P(A|B,C)$ conditioned to the two data events B and C can be obtained as [19]

$$P(A|B,C) = w_1 P(A|B) + w_2 P(A|C) \tag{1}$$

where w_1 and w_2 are the positive weights associating with $P(A|B)$ and $P(A|C)$, respectively; and $w_1 + w_2 = 1$.

The above linear combination imposes convexity of the result. Such convexity is undesirable because the combined probability is bounded by the two prior probabilities $P(A|B)$ and $P(A|C)$. Thus, it prevents the possibility of data integration. Recently, Journal [33] introduced a probabilistic method, which is based on the engineering paradigm of permanence of ratios, for combining information from diverse sources where the conventional assumption of data independence is relaxed. This ratio-based probabilistic fusion method satisfies the mathematical properties of all limit conditions in the presence of complex data independence, and only requires the knowledge of the prior probability and the elementary single event-conditioned probabilities, which can be computed independently from each other.

However, the ratio-based fusion being mentioned above still does not consider the different degrees of influence of the diverse sources when trying to combine the information. Such influence factor is inherent in many practical applications and needs to be considered in the design of any information fusion methods. For example, in pattern recognition, different features of an object may give different effects in classification; and individual classifiers, which are based on different approaches, may give different solutions using the same features. It is the motivation of this paper as an attempt to incorporate the interaction between different attributes of multiple information sources into the ratio-based probabilistic fusion. The proposed fusion model is based on the mathematical frameworks of fuzzy measures and fuzzy integrals, and can be viewed as a generalized version of the ratio-based fusion.

To be self-contained in the development of the proposed approach, Section 2 presents a brief overview of the ratio-based fusion. Section 3 discusses the formulation of the ratio-based fuzzy probabilistic fusion model. Section 4 illustrates the performance of the new approach. Finally, conclusion of the finding is addressed in Section 5.

2. Updating information with ratio-based probabilistic fusion

Let $P(A)$ be the prior probability of the occurrence of data event A; $P(A|B)$ and $P(A|C)$ be the probabilities of occurrence of event A given the knowledge of events B and C, respectively; $P(B|A)$ and $P(C|A)$ the probabilities of observing events B and C given A, respectively. Using Bayes' law, the posterior probability of A given B and C is

$$P(A|B,C) = \frac{P(A,B,C)}{P(B,C)} = \frac{P(A)P(B|A)P(C|A,B)}{P(B,C)} \tag{2}$$

The simplest way for computing the two probabilistic models is to assume the model independence, giving

$$P(C|A,B) = P(C|A) \tag{3}$$

and

$$P(B,C) = P(B)P(C) \tag{4}$$

Thus, (2) can be rewritten as

$$P(A|B,C) = \frac{P(A)P(B|A)P(C|A)}{P(B)P(C)} \tag{5}$$

or

$$\frac{P(A|B,C)}{P(A)} = \frac{P(A|B)}{P(A)}\frac{P(A|C)}{P(A)} \tag{6}$$

However, the assumption of conditional independence between the data events usually does not statistically perform well and leads to inconsistencies in many real applications [33]. Therefore, an alternative to the hypothesis of conventional data event independence should be considered. The permanence of ratios based approach allows data events B and C to be incrementally conditionally dependent and its fusion scheme gives

$$P(A|B,C) = \frac{1}{1+x} = \frac{a}{a+bc} \in [0,1] \tag{7}$$

where a, b, c, and x are taken as the logistic-type ratio of marginal probabilities of A, B after A occurring, C after A occurring, and B and C after A occurring; respectively:

$$a = \frac{1-P(A)}{P(A)}, \quad b = \frac{1-P(A|B)}{P(A|B)}, \quad c = \frac{1-P(A|C)}{P(A|C}, \quad x = \frac{1-P(A|B,C)}{P(A|B,C)}$$

An interpretation of the fusion expressed in (7) is as follows. Let A is the target event which is to be updated by events B and C. The term a is considered as a measure of prior uncertainty about the target event A or a distance to the occurrence of A without any updated evidence. We have $a=0$ for $P(A)=1$ if target event A is certain to occur; and $a=\infty$ for $P(A)=0$ if A is an impossible event. Likewise, b and c are measures of uncertainty or the distances to A knowing about its occurrence after observing evidences given by B and C, respectively. The term x is the distance to the target event A occurring after observing evidences given by both events B and C. The ratio c/a is then the incremental (increasing or decreasing) information of C to that distance starting from the prior distance a. Similarly, the ration x/b is the incremental information of C starting from the distance b. Thus, the permanence of ratios provides the following approximate relation:

$$\frac{x}{b} \approx \frac{c}{a} \tag{8}$$

which says that the incremental information about C to the knowledge of A is the same after or before knowing B. In other words, the incremental contribution of information from C about A is independent of B. Hence, this mathematical expression relaxes the restriction of the assumption of full independence of B and C.

Expression (7) verifies the limit properties of exact decomposition [34] of the joint probability of A and the two events B and C as follows [33].

1. Consistent probability: $P(A|B,C) \in [0,1]$.
2. Closure condition: $P(\overline{A}|B,C) = xP(A|B,C) = x/(1+x) \in [0,1]$, where $P(\overline{A}) = 1-P(A)$.

3. If C is non-informative of A:$c=a$, then $x=b$, yielding $P(A|B,C) = P(A|B)$.
4. If C is fully informative about A: $c=0$ or ∞, then $x=0$ or ∞, giving $P(A|B,C)=1$ or 0, $\forall B$.

To extend the above ratio-based fusion to k data events E_j, $j=1,\dots,k$; the conditional probability provided by a succession of $(k-1)$ permanence of ratios is given as

$$P(A|E_j, j=1,\dots,k) = \frac{1}{1+x} \in [0,1] \tag{9}$$

where

$$x = \frac{\prod_{j=1}^{k} d_j}{a^{k-1}} \geq 0, \quad a = \frac{1-P(A)}{P(A)}, \quad d_j = \frac{1-P(A|E_j)}{P(A|E_j)}, \quad j=1,\dots,k$$

Once again, expression (9) satisfies all limit properties and requires only the knowledge of the prior probability $P(A)$, and the k elementary single conditional probabilities $P(A|E_j)$, $j=1,\dots,k$, which can be independently calculated.

3. Fuzzy measures and integrals

The basic concepts of fuzzy measures and fuzzy integrals are outlined as follows to allow the discussion of properties of the ratio-based fuzzy probabilistic fusion. Let X be a finite set $X=\{x_1,x_2,\dots,x_n\}$. A fuzzy measure g defined on X is a set function $g:\mathcal{P}(X)\to[0,1]$ satisfying the following axioms [1,3]:

1. $g(\emptyset)=0$, and $g(X)=1$.
2. If $A\subseteq B$, then $g(A)\leq g(B)$,

where $\mathcal{P}(X)$ denotes the power set of X.

It is noted that when the second property is not satisfied, g is called a non-monotonic fuzzy measure [39]. There are 2^n coefficients being equivalent to the cardinality of $\mathcal{P}(X)$ to compute a fuzzy measure on X. These coefficients are the values of g for all subsets of X and they are not independent since they must satisfy the property of monotonicity. Theoretically, the concept of fuzzy measures is the generalization of the classical measure theory which is restrictive on the hypothesis of additivity; where as additivity is relaxed by the theory of fuzzy measures.

Sugeno [3] defined a fuzzy measure known as the $g_\lambda-$fuzzy measure that satisfies the following additional condition, $\forall A,B\subset X$, and $A\cap B=\emptyset$:

$$g_\lambda(A\cup B) = g_\lambda(A)+g_\lambda(B)+\lambda g_\lambda(A)g_\lambda(B), \quad \lambda > -1 \tag{10}$$

To simplify the notation, let $g^i = g(\{x_i\})$ which is called a fuzzy density function. A fuzzy density g^i can be interpreted as the degree of belief or degree of importance of the corresponding attribute x_i that makes an effect or contribution towards the whole fuzzy system when all attributes are considered together.

The value of λ can be calculated follows:

$$\lambda+1 = \prod_{i=1}^{k}(1+\lambda g^i) \tag{11}$$

The following properties of the $g_\lambda-$fuzzy measure will be helpful in the computation of the parameter λ [7]. For a finite set $\{g^i\}, 0<g^i<1$, there exists a unique root $\lambda\in(-1,+\infty)$, and $\lambda\neq 0$. Based on this lemma, λ can be determined by solving $(k-1)$ degree polynomial and selecting the unique root >-1. If $\sum_{i=1}^{k} g^i < 1$, then $\lambda>0$. If $\sum_{i=1}^{k} g^i > 1$, then $-1\leq\lambda<0$.

Having described the notion of fuzzy measures, fuzzy integrals are real nonlinear functionals with respect to a fuzzy measure. Two most popular types of fuzzy integrals are known as Sugeno integral and Choquet integral. Let g be a fuzzy measure on $\mathbf{X}$. The discrete Sugeno integral of a function $f:\mathbf{X}\to[0,1]$ with respect to g is defined by

$$S[f(x_1),\dots,f(x_k)] = \max_i^k[\min\{f(x_i),g(A_i)\}] \tag{12}$$

where $A_i=\{x_1,\dots,x_i\}$, and $f(x_1)\geq f(x_2)\geq\cdots\geq f(x_k)$ (otherwise the elements of $\mathbf{X}$ are to be permuted to ensure such an increasing order). Following the expression of A_i, fuzzy measures $g(A_i)$ can be recursively computed in terms of λ as

$$g(A_1) = g^1 \tag{13}$$

$$g(A_i) = g^i + g(A_{i-1}) + \lambda g^i g(A_{i-1}), \quad 1\leq i\leq k \tag{14}$$

Using the previously defined expression of A_i and the relation of $f(x_i), i=1,\dots,k$; the discrete Choquet integral [41] of $f:\mathbf{X}\to\Re^+$ with respect to g is defined by

$$S[f(x_1),\dots,f(x_k)] = \sum_{i=1}^{k}[f(x_i)-f(x_{i+1})]g(A_i)] \tag{15}$$

where $f(x_{k+1})=0$.

4. Updating information with ratio-based fuzzy probabilistic fusion

The approximation expressed in (8) indicates the independence of relative value (x/b) of C from B events. In order to account for formal data interdependence, a possible way is to introduce an exponential weight as a function of B and C to the prior-to-B ratio (c/a) to incorporate some B-dependence as follows:

$$\frac{x}{b} = \left(\frac{c}{a}\right)^{\tau(B,C)}, \quad \tau(B,C)\geq 0 \tag{16}$$

where $\tau(\cdot)$ is called the tau model which measures the data interaction, and is alternatively known as data redundancy (because redundancy implies the overlap of information sources). In the context of the present discussion, the meaning of data interaction is used to refer to that one source of information modifies the evidence brought by the other source.

However, the challenge for using expression (16) is the finding of a suitable value for $\tau(\cdot)$ for data events $E_j, j=1,\dots,k$. It has been reported that the values of the tau model cannot be determined analytically, and a heuristic solution has to be sought by means of training data, or $\tau(\cdot)$ is set to be 1 [35], which in turn does not take into account the aspect of data interaction.

The above limit imposed on the tau model would turn our attention to the concept of fuzzy measures based on which the operation on fuzzy information fusion is performed. Given the extension of set theory to fuzzy set theory by Zadeh [36], the notion of fuzzy measures [37] is further discussed to justify the use of fuzzy integration in the setting of the ratio-based probabilistic fusion and how it is different from similar well-known notions as follows.

Let $(\Omega,\mathcal{B},P)$ be a probability space; where Ω is an Euclidean n-space $\mathcal{R}^n$, $\mathcal{B}$ is the $\sigma-$field of Borel sets in $\mathcal{R}^n$ and P is a probability measure over $\mathcal{R}^n$. Let $A\in\mathcal{B}$, and x a point in Ω. The probability of A can be defined as

$$P(A) = \int_A dP = \int_\Omega \mu_A(x)\,dP = E(\mu_A) \tag{17}$$

where $\mu_A\in\{0,1\}$ is the characteristic function of A, and $E(\mu_A)$ is the expectation of μ_A.

Now let $\tilde{A}$ be a fuzzy event in Ω whose membership function $\mu_A\in[0,1]$ is Borel measurable. Then P can be extended to $\tilde{P}$ on $\tilde{A}$ by the definition of the Lebesgue–Stieltjes integral as

$$\tilde{P}(A) = \int_\Omega \mu_A(x)\,dP = E(\mu_A) \tag{18}$$

That is, as in (17), $\tilde{P}$ is the expectation of its membership function defined on a fuzzy set and is called the probability of a fuzzy event. Grabisch et al. [38] also showed that the extension of P to $\tilde{P}$ is simple, where $\tilde{P}:\tilde{A}\to[0,1]$ is $\sigma-$additive as $\tilde{A}$ is a $\sigma-$field of fuzzy sets. A further interest is to extend P to another property where set functions are monotone, that is, if $A\subset B$, then $g(A)\le g(B)$; but non-additive. Such functions are called fuzzy measures. An illustration of the meaning of a fuzzy measure [1] is that suppose we select an element $x\in\Omega$ but do not know which one; then for $A\subset\Omega$, we want to guess if $x\in A$. Grabisch et al. [38] pointed out that this type of information measure is analogous to the concept of Shafer's belief function [2], and brings to mind a similar statistical notion of confident coefficient in interval estimation. The interval analysis involves random samples, say, $Y_1,Y_2,\ldots,Y_n$ from a population and a random set $R(Y_1,Y_2,\ldots,Y_n)$ to estimate the mean μ of the corresponding population. Now suppose the random samples $y_1,y_2,\ldots,y_n$ are collected, which form the set $R(y_1,y_2,\ldots,y_n)$. The probability $P(\mu\in R)$ expresses the degree of confidence that $\mu\in R(y_1,y_2,\ldots,y_n)$. As for Shafer's theory of evidence, probability measures are subjectively assigned to random sets, the degree of belief that $x\in A$ expressed as $P\{x:R(x)\subseteq A\}$ is not an additive set function, which is different from the subjective evaluation of fuzzy measures.

To get closer to the argument on the use of fuzzy measures and fuzzy integrals in the extension of the permanence-of-ratio-based fusion scheme, we show a close relationship between a fuzzy measure and the concept of fuzziness in the theory of possibility [40], which provides theoretical support to the use of fuzzy integration in the extension of the hypothesis-based evidence combination. Let U be a universe of discourse; A a nonfuzzy set of U; Π_X a possibility distribution associated with variable X taking values in U; $X=u$ implying that X is assigned the value u, $u\in U$; and $\pi_X(u)$ the possibility distribution function of Π_X. The possibility measure of A, $\pi(A)\in[0,1]$, is defined as

$$\pi(A)=\sup_{u\in A}\pi_X(u) \tag{19}$$

Expression (19) can lead to the mathematical quantification of the statement "X is possibly in A". That is, $Poss(X\in A)=\pi(A)$ which induces the possibility distribution function $\pi(A)$ for X, as [40]

$$Poss(X\in A)=\sup_{u\in A}\mu_A(u) \tag{20}$$

It is obvious that the set function expressed in (20) is non-additive, but it is monotone, so it is a fuzzy measure. Now it can be argued that the evidence brought by the hypothesis of the permanence of ratio in the setting of the knowledge combination is not underlined by a random mechanism but expressed as a degree of trust or a measure of fuzziness with values in [0,1]. This evidence is considered as a subjective evaluation which identifies with the notion of fuzzy measures. Once again, this type of subjective evaluations is different from subjective probabilities as in the case of Shafer's theory of evidence in that the set functions of the first need not be additive. We now turn our attention to the application of fuzzy measures and integrals to information fusion in the context of the permanence of ratios.

It can be realized that the product term bc expressed in (7) for two data events or the product terms of all $d_j, j=1,\ldots,k$ of distance x expressed in (9) for k data events is the performance of a sub-fusion using the multiplicative rule. One possible way to introduce data interaction is by the implementation of fuzzy measure and fuzzy integration in the sub-fusion part of the ratio-based probabilistic fusion. Thus, the ratio-based fuzzy probabilistic fusion for two data events B and C, denoted as $F(A|B,C)$, is expressed as follows:

$$F(A|B,C)=\frac{a}{a+e^2}\in[0,1] \tag{21}$$

where e is considered as the distance to A occurring after observing evidences from B and C, and raised to the power of 2 to allow numerical scaling being equivalent to the product term bc expressed in (7) for the case of two data events, and defined as

$$e=\frac{1-\phi(B,C)}{\phi(B,C)}\ge 0 \tag{22}$$

in which the functional $\phi(b,c)$ is the value of fuzzy integral of B and C with respect to a fuzzy measure of the interaction of B and C.

For k data events $E_j, j=1,\ldots,k$; the ratio-based fuzzy probabilistic fusion is given as

$$F(A|E_j, j=1,\ldots,k)=\frac{1}{1+y}\in[0,1] \tag{23}$$

where

$$y=\frac{d^k}{a^{k-1}}\ge 0$$

again

$$a=\frac{1-P(A)}{P(A)}$$

and d is modeled as the distance to A occurring after observing evidences from all k data events together by means of a fuzzy integral, raised to the power of k for the purpose of scaling being equivalent to the k events expressed in (9), and defined as

$$d=\frac{1-\phi(E_1,\ldots,E_k)}{\phi(E_1,\ldots,E_k)}$$

in which $\phi(E_1,\ldots,E_k)$ is the fuzzy integral of $E_1,\ldots,E_k$ with respect to a fuzzy measure of the interaction of $E_1,\ldots,E_k$.

It can be now seen that, based on the properties of fuzzy measures and fuzzy integrals, the value of sub-fusion term defined in (21) or (23) does not lend itself to zero if not all values of $f:\mathbf{X}\to[0,1]$, which are equivalent to preposterior probabilities $P(A|E_j)$, give completely decisive pieces of evidence about A. The multiplicative rule employed in (7) or (9) suffers from this drawback in that when the probability of any single preposterior information is fully decisive about A, then the whole multiplicative term (subfusion) becomes zero and can be biased to a single or partial evidence.

Having presented the framework of the fuzzy posterior-probabilistic fusion, the calculation of (23) requires only the knowledge the prior probability $P(A)$; the k elementary single data event-conditioned probabilities $P(A|E_j), j=1,\ldots,k$, which are equivalent to $f(x_1),\ldots,f(x_k)$ used for calculating the fuzzy integral; and k fuzzy densities (degrees of importance of data events $E_j, j=1,\ldots,k$). Thus, expression (23) takes into account prior information, preposterior evidence, and interaction of attributes in the knowledge updating process.

5. Experiments

Calculation and performance of the ratio-based fuzzy probabilistic fusion are studied using a numerical example and real datasets in the following subsections.

5.1. Numerical example

Consider the evidence assessment of an event A from three evidences $E_1=x_1=B$, $E_2=x_2=C$ and $E_3=x_3=D$. The probability of A occurring given the knowledge of B, C and D are $P(A|B)=f(x_1)=0.7$, $P(A|C)=f(x_2)=0.6$, and $P(A|D)=f(x_3)=0.5$, respectively.

Let the prior probability of A occurring be $P(A)=1/3$; and the degrees of importance of evidence obtained from data sources B, C and D be $g^1=0.1$, $g^2=0.2$ and $g^3=0.3$, respectively.

Using expression (11), the value of λ is obtained by solving the quadratic function $0.006\lambda^2+0.11\lambda-0.4=0$ for $\lambda>-1$, which yields $\lambda=3.109$. Using (14), the $\lambda-$fuzzy measure on $\mathbf{X}=\{x_1,x_2,x_3\}$ are calculated and given in Table 1.

The Sugeno integral of $f(x_1),f(x_2),f(x_3)$ with respect to the fuzzy measure on $\mathbf{X}$ using (12) is calculated as $S=\max[\min(0.7,0.1), \min(0.6,0.362),\min(0.5,1)]=0.50$. Using (15), the Choquet integral is $C=(0.7-0.6)\,0.1 + (0.6-0.5)\,0.362 + (0.5-0)\,1=0.55$.

The values for calculating $F(A|B,C,D)$ defined by (23) are: $a=(1-0.5)/0.5=1$, $d=(1-0.5)/0.5=1$ (Sugeno), $d=(1-0.55)/0.55=0.8182$ (Choquet). Thus, giving $F(A|B,C,D)=1^2/(1^2+1^3)=0.5$ for the Sugeno integral; and $F(A|B,C,D)=1^2/(1^2+0.8182^3)=0.6461$ for the Choquet integral.

Using the probabilistic fusion expressed in (9), the distances to A are: $a=(1-0.5)/0.5=1$, $d_1=(1-0.7)/0.7=0.4286$, $d_2=(1-0.6)/0.6=0.6667$, and $d_3=(1-0.5)/0.5=1$. Thus, giving $P(A|B,C,D)=1^2/[1^2+(0.4286\times0.6667\times1)]=0.7778$.

Table 2 summarizes the updated information about A using the probabilistic and fuzzy probabilistic fusion schemes.

As another example, the evidence gathered from D about A is now heavily reduced to $P(A|D)=0.001$. The combined results are recalculated as follows.

The Sugeno integral of $f(x_1),f(x_2),f(x_3)$ with respect to the fuzzy measure on $\mathbf{X}$ using (12) is calculated as $S=\max[\min(0.7,0.1), \min(0.6,0.362),\min(0.001,1)]=0.362$. Using (15), the Choquet integral is $C=(0.7-0.6)\,0.1 + (0.6-0.001)\,0.362 + (0.001-0)\,1=0.2278$.

Giving the values for calculating $F(A|B,C,D)$ defined by (23) are: $a=(1-0.5)/0.5=1$, $d=(1-0.362)/0.5=1$ (Sugeno), $d=(1-0.2278)/0.2278=3.3898$ (Choquet). Thus, giving $F(A|B,C,D)=1^2/(1^2+1^3)=0.5$ for the Sugeno integral; and $F(A|B,C,D)=1^2/(1^2+3.3898^3)=0.6461$ for the Choquet integral.

Using the probabilistic fusion expressed in (9), the distance from $D=x_3$ to A now becomes $d_3=(1-0.001)/0.001=999$. Thus, giving $P(A|B,C,D)=1^2/[1^2+(0.4286\times0.6667\times999)]=0.0035$.

Table 3 summarizes the updated information about A using the probabilistic and fuzzy probabilistic fusion schemes.

The following observations can be made from the above two examples with the summary of the combined results presented in Tables 2 and 3: (1) the probabilistic fusion is sensitive to the preposterior information, which gives largest and smallest probabilities in the two examples; (2) the fuzzy probabilistic fusion using the Sugeno integral tends to stay at the lower bound of the preposterior values; and (3) the fuzzy probabilistic fusion using the Choquet integral yields the most balanced combined scores between the Sugeno-based and probabilistic fusion schemes.

Table 1
$\lambda-$fuzzy measure on $\mathbf{X}$.

Subset A_i	$g_\lambda(A_i)$
$\emptyset$	0
$\{x_1\}$	0.1
$\{x_2\}$	0.2
$\{x_3\}$	0.3
$\{x_1,x_2\}$	0.362
$\{x_1,x_3\}$	0.493
$\{x_2,x_3\}$	0.687
$\{x_1,x_2,x_3\}$	1.0

Table 2
Combination of information using ratio-based probabilistic fusion (PF) and fuzzy probabilistic fusion (FPF) methods ($g^1=0.1$, $g^2=0.2$, $g^3=0.3$).

	Prior	Preposterior (B,C,D)	Combined
PF	0.3333	(0.7, 0.6, 0.5)	0.7778
FPF (Sugeno)	0.3333	(0.7, 0.6, 0.5)	0.5
FPF (Choquet)	0.3333	(0.7, 0.6, 0.5)	0.6461

Table 3
Combination of information using ratio-based probabilistic fusion (PF) and fuzzy probabilistic fusion (FPF) methods ($g^1=0.1$, $g^2=0.2$, $g^3=0.3$).

	Prior	Preposterior (B,C,D)	Combined
PF	0.3333	(0.7, 0.6, 0.001)	0.0035
FPF (Sugeno)	0.3333	(0.7, 0.6, 0.001)	0.1545
FPF (Choquet)	0.3333	(0.7, 0.6, 0.001)	0.0250

5.2. *Molecular image data*

The proposed method was carried out using a high-content imaging data for classification of cell phases. High-throughput and high-content screening of cells generated by time-lapse fluorescence-stained microscopic imaging technology is becoming one of the most widely used research tools to assist biologists in understanding the complex process of cell division or mitosis. Its power comes from the sensitivity and resolution of automated light microscopy with multi-well plates, combined with the availability of fluorescent probes that are attached to specific subcellular components, such as chromosomes and microtubules, for visualization.

The dataset consists of 375 841 cells in 892 nuclear sequences. Each sequence consists of 96 frames over duration of 24 h. The average number of cells per sequence is 421. Imaging was performed by time-lapse fluorescence microscopy with a time interval of 15 min. Cell cycle progress was affected by drug and some or all of the cells in the treated sequences were arrested in metaphase. Cell cycle progress in the untreated sequences was not affected. Cells without drug treatment will usually undergo one division during this period of time.

After the nuclear segmentation had been performed on the raw image database, it was necessary to perform a morphological closing process on the resulting binary images in order to smooth the nuclear boundaries and to fill holes insides the nuclei. These binary images were then used as a mask on applied the original image to arrive at the final segmentation. Cell nuclei were labeled manually for validation of automated classification. From these resulting images, features were extracted. The ultimate goal is to assign correct phase to cells via the training of some classification techniques. To identify shapes and intensity differences between different cell phases, a set of seven features were extracted. These features include maximum intensity, mean, stand deviation, major axis, minor axis, perimeter, and compactness [42]. Because the feature values have different ranges, the scaling of features was therefore necessary by calculating the z-scores by subtracting the feature value with its mean and divided by its mean absolute deviation. These seven features were used as the input for training and testing various classifiers.

It was pointed out by Zhou et al. [43] that the performance of cell phase identification are highly dependent on the extracted features and the design of classifiers. There were also some cases in which individual feature-based approaches do not work well, because it is not easy to identify the phases of cell nuclear images which overlap each other. This effect would cause the feature extraction from gray-level images being partly obscured by the overlapping nucleus to be heavily influenced by noise. There are four phases to be identified: interphase, prophase, metaphase, and anaphase. The dataset was equally divided for training and testing. Training and testing sets were five times randomly selected to perform the cell-phase identification by individual classifiers and fusion methods for this four-class classification. Three individual supervised classifiers based on Bayes decision theory included in this study are the FCM-GMM which is a Gaussian mixture model (GMM) using the fuzzy c-means (FCM) for parameter estimation, Markov model (MM) coupling with maximum-likelihood-based GMM (MM-ML-GMM), and the MM-FCM-GMM [44]. The FCM-GMM was implemented by minimizing the FCM objective function for the parameter estimation, in which the FCM distance is the function of the mixture weights, mean vectors, and covariance matrices of the GMM. The MM-ML-GMM was designed by incorporating the initial and transition probabilities of the phase sequences modeled by a Markov chain with the GMM framework to determine a similarity score between an input feature vector and a GMM-based cell phase model, and the GMM parameter estimation was carried out by the maximum likelihood (ML) method. The MM-FCM-GMM classifier was designed similarly to the MM-ML-GMM but the FCM was used for the GMM parameter estimation instead of the ML.

Fusion methods include the naive Bayesian fusion (NBF), ratio-based probabilistic fusion (PF), fuzzy fusion (FF) using Sugeno and Choquet integrals, and ratio-based fuzzy probabilistic fusion (FPF) using Sugeno and Choquet integrals. For the FCM-GMM classifier, the density probability functions of unknown cell features given the corresponding GMMs for each cell phase were used in the fusion processes, then each unknown feature was assigned to the phase if the combined evidence of the feature belonging to phase is maximum. For MM-ML-GMM and MM-FCM-GMM classifiers, the normalized scores, which are the ratios of the transition probabilities and the normalized probabilistic distances and can be considered as the density values, indicating the similarity between an unknown cell feature and the cell phase models were used for information combination. The phase assignment was then performed similarly to what has been described for the FCM-GMM.

Two cases involving the prior knowledge about the cell phases were designed for the study of information fusion. In case 1, equal probabilities of occurrences of the cell phases were assumed. In case 2, the prior probabilities for interphase = 0.92, prophase = 0.02, metaphase = 0.04, and anaphase = 0.02. These prior probabilities were estimated based on the number of phases found in the database. The classification results including statistical variances (σ^2) for case 1 by individual classifiers as well as fusion methods where equal prior probabilities were assumed are presented in Table 4. The classification results including statistical variances (σ^2) for case 2 by individual classifiers and fusion methods where the prior probabilities were estimated are presented in Table 5. It can be seen that by taking into account the prior information in the combination of diverse knowledge in the fuzzy probabilistic fusion scheme, the results can be improved in both cases. The probabilistic and Sugeno fusion models gave similar results but lower than the Choquet integral. The naive Bayesian fusion performed only slightly better than the best individual classifier (MM-FCM-GMM). For the equal prior probability case as shown in Table 4, the fusion schemes performed by the probabilistic fusion and fuzzy fusion schemes using Sugeno and Choquet integral are found to be equivalent although most results given by the Sugeno and Choquet integrals were marginally higher those given by the PF. The naive Bayesian fusion could improve the results but still lower than other fusion schemes. The fuzzy probabilistic fusion (FPF) using the Choquet integral gave the best performance. For the case of unequal prior probabilities as shown in Table 5, the probabilistic fusion gave a marginally higher performance than the fuzzy fusion using Sugeno integral, but marginally lower performance than the Choquet fuzzy fusion. By taking into account both prior information and degrees of importance of classifiers, the fuzzy probabilistic fusion method using either the Sugeno or Choquet integral outperformed the ratio-based probabilistic fusion. To measure the strength of the evidence against the claims of the average classification results using the proposed FPF given in Tables 4 and 5; we randomly selected other fifty testing datasets and applied the z-test, where a 5% significance level is adopted. The p-values obtained from the z-tests, which are the probabilities of observing the given results if the assertions are true (small values of p cast doubt on the validity of the assertions). The p-values obtained from the z-tests using the FPF based on Sugeno and Choquet integrals for the case of equal prior probabilities are 0.8370 and 0.805, respectively. For the case of unequal prior probabilities, the p-values obtained from the z-tests using the FPF based on Sugeno and Choquet integrals for the case of equal prior probabilities are 0.8510 and 0.8327, respectively. The results of the hypothesis tests indicated that the average performance data are statistically significant at the 5% significance level, suggesting the average performance of the FPF would occur just by chance no more than 5% in the repeated samples.

Table 4
Classification rates (%) on molecular image data, where equal prior probabilities applied to information fusion (FCM-GMM: fuzzy c-means based Gaussian mixture model, MM-ML-GMM: maximum-likelihood-based Gaussian mixture models coupling with Markov models, MM-FCM-GMM: fuzzy-c-means-based Gaussian mixture models coupling with Markov models, naive Bayesian fusion (NBF), ratio-based probabilistic fusion (PF), fuzzy fusion (FF), ratio-based fuzzy probabilistic fusion (FPF), σ^2: statistical variance).

Method	Interphase	Prophase	Metaphase	Anaphase	σ^2
FCM-GMM	91.24	72.55	80.11	82.45	0.0046
MM-ML-GMM	92.37	74.00	78.32	75.68	0.0057
MM-FCM-GMM	94.89	77.28	84.41	82.19	0.0054
NBF	95.01	77.54	84.90	82.79	0.0002
PF	95.52	78.60	86.49	84.77	0.0002
FF (Sugeno)	95.55	79.00	85.44	84.21	0.0002
FF (Choquet)	95.57	79.81	86.60	85.11	0.0002
FPF (Sugeno)	96.48	79.21	87.32	85.34	0.0002
FPF (Choquet)	98.49	82.50	90.22	90.90	0.0002

Table 5
Classification rates (%) on molecular image data, where unequal prior probabilities applied to information fusion (naive Bayesian fusion (NBF), ratio-based probabilistic fusion (PF), fuzzy fusion (FF), ratio-based fuzzy probabilistic fusion (FPF), σ^2: statistical variance).

Method	Interphase	Prophase	Metaphase	Anaphase	σ^2
NBF	95.38	78.65	85.11	83.52	0.0002
PF	97.87	80.60	87.50	85.82	0.0003
FF (Sugeno)	97.51	80.10	86.27	85.71	0.0003
FF (Choquet)	97.80	81.55	87.73	86.39	0.0002
FPF (Sugeno)	97.90	80.98	88.00	86.34	0.0003
FPF (Choquet)	98.71	84.50	92.22	90.37	0.0002

5.3. Microarray data

Microarray-based measure of gene expressions is one of the most recent breakthrough technologies in experimental molecular biology [45]. The utilization of microarrays allows simultaneous study and monitoring of tens of thousands of genes. One of the most useful quantitative analyses and interpretations of microarray-based data is diseased-state classification.

The first microarray data used to test the proposed algorithm is the microarray-based hereditary breast cancer data which were first studied by Hedenfalk et al. [46]. The data consist of 22 cDNA microarrays with 3226 genes. The 22 breast tumor samples were collected from the biopsy specimens of seven patients with germ-line mutations of BRCA1 (breast cancer susceptibility gene 1), eight patients with germ-line mutations of BRCA2 (breast cancer susceptibility gene 2), and seven patients with sporadic cases. BRCA1 and BRCA2 are human genes that belong to a class of genes known as tumor suppressors. In normal cells, BRCA1 and BRCA2 help ensure the stability of the genetic material (DNA) of the cell and help prevent uncontrolled cell growth. Mutation of these genes is linked to the development of hereditary breast cancer. The microarray data can be represented in matrix notation as $\mathbf{X} = [\mathbf{x}_{ij}]$, $i = 1, \ldots, N$, $j = 1, \ldots, M$, where N and M are the numbers of tumor samples and genes, respectively. Thus, this is a three-class problem for classification: BRCA1, BRCA2, and sporadic.

The leukemia dataset [47], which has become a benchmark for evaluation and comparisons of different algorithms for classification of gene expression cancer data, was used as another dataset to test the performance of various information fusion models. This dataset consists of 6817 human genes, where bone marrow or peripheral blood samples were taken from 72 patients with either acute myeloid leukemia (AML) or acute lymphoblastic leukemia (ALL). The gene expression levels were obtained using Affymetrix high density oligonucleotide microarrays. Thus, this dataset involves two classes for the classification: AML and ALL.

Various models of the fuzzy c-means clustering were used to classify the classes involved with the two respective microarray datasets. A critical information for minimizing the objective function of the FCM is the modeling of algorithmic matrix in the calculation of the distance between cluster centers and data points, which is expressed as $(d_{ik})^2 = \|\mathbf{x}_k - \mathbf{v}_i\|^2_{\mathbf{A}} = (\mathbf{x}_k - \mathbf{v}_i)^T \mathbf{A}(\mathbf{x}_k - \mathbf{v}_i)$, where $\mathbf{x}_k = (x_{k1}, x_{k2}, \ldots, x_{kp})$ is the feature vector of data point k, $\mathbf{v}_i = (v_{i1}, v_{i2}, \ldots, v_{ip})$ the feature vector of cluster i, and $\mathbf{A}$ is the matrix defining the norm for space $\mathcal{R}^p$ and called the algorithmic variable for the FCM methods [48]. Variations of $\mathbf{A}$ in the FCM clustering allows to model various shapes of the clusters and therefore may capture more accurate topological information of the data structures under study. The Euclidean norm (N_E), Mahalonobis norm (N_M), and fuzzy-scatter matrix norm (N_S) were implemented in the FCM approach for classifying the two microarray datasets. For the implementation of fuzzy-scatter matrix norm [48], the distance $d_{ik} = \|\mathbf{x}_k - \mathbf{v}_i\|_{\mathbf{A}_i}$, where $\mathbf{A}_i = [\rho_i \det(\mathbf{S}_i)^{(1/p)}(\mathbf{S}_i^{-1})]$; in which $\mathbf{S}_i = \sum_{k=1}^{n}(u_{ik})^m(\mathbf{x}_k - \mathbf{v}_i) - (\mathbf{x}_k - \mathbf{v}_i)^T$, $\rho_i > 0$, and $m \in [1, \infty)$ is the weighting exponent of the FCM objective function. In this study, $\rho_i = 1, \forall i$; and $m = 2$.

To establish statistics of the performance of the fusion methods, one sample of a class was removed while samples of other two classes remained the same, and the data were used for classification. The leave-one-out procedure was then repeated for each class to carry out the classification. Equal prior probabilities were used in the relevant fusing models. The results obtained from the three FCM methods (N_E, N_M, N_S), naive Bayesian fusion (NBF), probabilistic fusion (PF), and two fuzzy probabilistic fusion (FPF) methods (Sugeno integral, Choquet integral) using the breast cancer and leukemia datasets were shown in Tables 6 and 7, respectively. The results show that the FPF using the Choquet integral gave superior classification rates to other methods in both datasets. Being similar to the experiment on molecular imaging data, the BNF could improve the results (slightly better than the best individual classifier) but its performance was found to be lower than the other fusion schemes. Results given by the PF and FF (Sugeno) were more or less similar. For individual fusion methods, the Choquet integral performed better than both the Sugeno integral and ratio-based probabilistic fusion. Once again, the combined results provided the fuzzy posterior-probabilistic methods were found to be better than those obtained from individual classifiers (FCM with different norms) as well as individual fusion models (naive Bayes, probabilistic fusion, Sugeno integral, and Choquet integral).

Table 6
Classification on breast cancer data using information fusion (naive Bayesian fusion (NBF), ratio-based probabilistic fusion (PF), fuzzy fusion (FF), ratio-based fuzzy probabilistic fusion (FPF), σ^2: statistical variance).

Method	Correction (%)	σ^2
FCM (N_E)	77.61	0.2952
FCM (N_M)	75.46	0.2436
FCM (N_S)	81.82	0.2100
NBF	82.41	0.1451
PF	82.63	0.2045
FF (Sugeno)	80.36	0.1268
FF (Choquet)	83.47	0.1207
FPF (Sugeno)	88.19	0.1026
FPF (Choquet)	90.91	0.1000

Table 7
Classification on acute leukemia data using information fusion (naive Bayesian fusion (NBF), ratio-based probabilistic fusion (PF), fuzzy fusion (FF), ratio-based fuzzy probabilistic fusion (FPF), σ^2: statistical variance).

Method	Correction (%)	σ^2
FCM (N_E)	84.72	0.6214
FCM (N_M)	87.17	0.6835
FCM (N_S)	89.63	0.3333
NBF	89.71	0.1466
PF	90.17	0.1528
FF (Sugeno)	90.56	0.1703
FF (Choquet)	92.91	0.1524
FPF (Sugeno)	95.14	0.1327
FPF (Choquet)	97.11	0.1022

6. Conclusion

A fuzzy posterior-probabilistic fusion scheme has been presented and discussed in the foregoing sections. The novel feature of the proposed method is the embedding of the subfusion of the ratio-based probabilistic fusion with the fuzzy integral. Thus, the new approach is able to take into account the (1) prior knowledge, (2) preposterior information as well as (3) interaction of the preposterior sources of knowledge; whereas the permanence-of-ratio based fusion considers only the first two information factors.

The experimental results on various datasets have shown the effective application of the proposed approach for pattern classification either the prior information of the data events are equally assumed or estimated from training data. The Choquet-integral-based probabilistic fusion appears to yield better results than the Sugeno-integral-based probabilistic fusion. A more effective approach for deriving the fuzzy measure of multiple attributes (preposterior information events) would help further improve the combined results.

References

[1] M. Sugeno, The Theory of Fuzzy Integrals and its Applications, Ph.D. Thesis, Tokyo Institute of Technology, 1974.
[2] G.A. Shafer, Mathematical Theory of Evidence, Princeton University Press, Princeton, 1976.
[3] M. Sugeno, Fuzzy measures and fuzzy integrals: a survey, in: Fuzzy Automata and Decision Processes, Amsterdam, North-Holland, 1977, pp. 89–102.
[4] M.A. Abidi, R.C. Gonzalez, Data Fusion in Robotics and Machine Intelligence, Academic Press, San Diego, 1992.
[5] D.L. Hall, J. Llinas, An introduction to multisensor fusion, in: Proceedings of the IEEE, vol. 85, 1997, pp. 6–23.
[6] J. Kittler, M. Hatef, R.P. Duin, J.G. Matas, On combining classifiers, IEEE Transactions on Pattern Analysis and Machine Intelligence 20 (March) (1998) 226–239.
[7] H. Tahani, J.M. Keller, Information fusion in computer vision using the fuzzy integral, IEEE Transactions on Systems, Man, and Cybernetics 20 (1990) 733–741.
[8] S. Benediktsson, Consensus theoretic classification methods, IEEE Transactions on Systems, Man, and Cybernetics 22 (1992) 688–704.
[9] I. Bloch, Information combination operators for data fusion: a comparative review with classification, IEEE Transactions on Systems, Man, and Cybernetic 26 (1996) 52–67.
[10] T.D. Pham, H. Yan, Fusion of handwritten numeral classifiers based on fuzzy and genetic algorithms, in: Proceedings of the North America Fuzzy Information Processing Society, 1997, pp. 257–262.
[11] A.R. Mirhosseini, H. Yan, K.M. Lam, T. Pham, Human face image recognition: an evidence aggregation approach, Computer Vision & Image Understanding 71 (1998) 213–230.

[12] D. Dubois, M. Grabisch, F. Modave, H. Prade, Relating decision under uncertainty and multicriteria decision making models, International Journal of Intelligent Systems 15 (2000) 967–979.
[13] T.D. Pham, An image restoration by fusion, Pattern Recognition 34 (2001) 2403–2411.
[14] V. Dasigi, R.C. Mann, V.A. Protopopescu, Information fusion for text classification—an experimental comparison, Pattern Recognition 34 (2001) 2413–2425.
[15] N.W. Park, K.H. Chi, A probabilistic approach to predictive spatial data fusion for geological hazard assessment, in: Proceedings of the IEEE Geoscience and Remote Sensing Symposium, 1993, pp. 2425–2427.
[16] R.R. Yager, Fusion of ordinal information using weighted median aggregation, International Journal of Approximate Reasoning 18 (1995) 35–52.
[17] A. Temko, D. Macho, C. Nadeu, Fuzzy integral based information fusion for classification of highly confusable non-speech sounds, Pattern Recognition 41 (2008) 1814–1823.
[18] J. Synnergren, B. Olsson, J. Gamalielsson, Classification of information fusion methods in systems biology, in: Silico Biology, vol. 9, 2009, p. 0007. 〈http://www.bioinfo.de/isb/2009/09/0007/〉.
[19] E.I. Polyakova, A.G. Journel, The nu expression for probabilistic data integration, Mathematical Geology 39 (2007) 715–733.
[20] S. Auephanwiriyaku, J.M. Keller, P.D. Gader, Generalized Choquet fuzzy integral fusion, Information Fusion 3 (2002) 69–85.
[21] C. Labreuche, M. Grabisch, The Choquet integral for the aggregation of interval scales in multicriteria decision making, Fuzzy Sets and Systems 137 (2003) 11–26.
[22] K.-C. Kwak, W. Pedrycz, Face recognition using fuzzy integral and wavelet decomposition method, IEEE Transactions on Systems, Man, and Cybernetics 34 (2004) 1666–1675.
[23] A. Soria-Frisch, Unsupervised construction of fuzzy measures through self-organizing feature maps and its application in color image segmentation, International Journal of Approximate Reasoning 41 (2006) 23–42.
[24] Y.M. Kim, C.K. Kim, S.G. Hong, Fuzzy based state assessment for reinforced concrete building structures, Engineering Structures 28 (2006) 1286–1297.
[25] N.P. Bidargaddi, M. Chetty, J. Kamruzzaman, Fuzzy measures and integrals in profile hidden Markov models for protein sequence analysis, Journal of Intelligent and Fuzzy Systems 17 (2006) 541–556.
[26] S. Feng, L. Li, P. Wang, C. Wang, Q. Yue, S. Guo, Fuzzy modeling of the medical treatment effects of superoxide dismutase, Expert Systems 23 (2006) 323–329.
[27] S. Srivastava, M. Singh, V.K. Madasub, M. Hanmandlu, Choquet fuzzy integral based modeling of nonlinear system, Applied Soft Computing 8 (2008) 839–848.
[28] A. Gavrilut, A. Croitoru, Non-atomicity for fuzzy and non-fuzzy multivalued set functions, Fuzzy Sets and Systems 160 (2009) 2106–2116.
[29] M.F. Anderson, D.T. Anderson, D.J. Wescott, Estimation of adult skeletal age-at-death using the Sugeno fuzzy integral, American Journal of Physical Anthropology 142 (2010) 30–41.
[30] J. Caers, T.B. Hoffman, The probability perturbation method: an alternative Bayesian approach for solving inverse problems, Mathematical Geology 38 (2006) 81–100.
[31] H. Altincay, On naive Bayesian fusion of dependent classifiers, Pattern Recognition Letters 26 (2005) 2463–2473.
[32] R. Kohavi, G.H. John, Wrappers for feature subset selection, Artificial Intelligence 97 (1997) 273–324.
[33] A.G. Journel, Combining knowledge from diverse sources: an alternative to traditional data independence hypotheses, Mathematical Geology 34 (2002) 573–595.
[34] M. Spiegel, Probability and Statistics, in: Schaum's Outline Series, McGraw Hill, New York, , 1975.
[35] A. Boucher, K.C. Seto, A.G. Journel, A novel method for mapping land cover changes: incorporating time and space with geostatistics, IEEE Transactions on Geoscience and Remote Sensing 44 (2006) 3427–3435.
[36] L.A. Zadeh, Fuzzy sets, Information and Control 8 (1965) 338–353.
[37] L.A. Zadeh, Probability measures of fuzzy events, Journal of Mathematical Analysis and Applications 23 (1968) 421–427.
[38] M. Grabisch, H.T. Nguyen, E.A. Walker, Fundamentals of Uncertainty Calculi with Applications to Fuzzy Inference, Dordrecht, Kluwer, 1995.
[39] M. Grabisch, The representation of importance and interaction of features by fuzzy measures, Pattern Recognition Letters 17 (1996) 567–575.
[40] L.A. Zadeh, Fuzzy sets as a basis for a theory of possibility, Fuzzy Sets and Systems 1 (1978) 3–28.
[41] G. Choquet, Theory of capacities, Annales de l'Institute Fourier 5 (1953/1954) 131–295.
[42] X. Chen, X. Zhou, S.T.C. Wong, Automated segmentation, classification, and tracking of cancer cell nuclei in time-lapse microscopy, IEEE Transactions on Biomedical Engineering 53 (2006) 762–766.
[43] X. Zhou, F. Li, J. Yan, S.T.C. Wong, A novel cell segmentation method and cell phase identification using Markov model, IEEE Transactions on Information Technology in Biomedicine 13 (2009) 152–157.
[44] T.D. Pham, D.T. Tran, X. Zhou, S.T.C. Wong, A microscopic image classification system for high-throughput cell-cycle screening, International Journal of Intelligent Computing in Medical Sciences and Image Processing 1 (2007) 67–77.
[45] A. Brazma, J. Vilo, Gene expression data analysis, Federation of European Biochemical Societies Letters 480 (2000) 17–24.
[46] I. Hedenfalk, D. Duggan, Y. Chen, M. Radmacher, M. Bittner, R. Simon, P. Meltzer, B. Gusterson, M. Esteller, O. Kallioniemi, B. Wilfond, A. Borg, J. Trent, Gene-expression profiles in hereditary breast cancer, The New England Journal of Medicine 344 (2001) 539–548.
[47] T.R. Golub, D.K. Slonim, P. Tamayo, C. Huard, M. Gaasenbeek, J.P. Mesirov, H. Coller, M.L. Loh, J.R. Downing, M.A. Caligiuri, C.D. Bloomeld, E.S. Lander, Molecular classification of cancer: class discovery and class prediction by gene expression monitoring, Science 286 (1999) 531–537.
[48] J.C. Bezdek, Pattern Recognition with Fuzzy Objective Function Algorithms, Plenum Press, New York, 1981.

Tuan D. Pham received his PhD degree in 1995 from the University of New South Wales. His current research interests include image processing, molecular and medical image analysis, pattern recognition, bioinformatics, biomedical informatics, fuzzy-set algorithms, genetic algorithms, neural networks, geostatistics, signal processing, fractals and chaos. His research has been funded by the Australian Research Council, academic institutions, and industry. Dr. Pham is an editorial board member of several journals and book series including Pattern Recognition (Elsevier), Current Bioinformatics (Bentham), Recent Patents on Computer Science (Bentham), Proteomics Insights (open access journal, Libertas Academica Press), Book Series on Bioinformatics and Computational BioImaging (Artech House), International Journal of Computer Aided Engineering and Technology (Inderscience Publishers). He has been serving as chair and technical committee member of more than 30 international conferences in the fields of image processing, pattern recognition, computational intelligence, and computational life sciences.