Knowledge Formulation for Supervised Evidential Classification

Derek R. Peddle

Abstract

The Dempster-Shafer Theory of Evidence provides an appropriate framework for overcoming problems associated with the analysis, integration, and classification of modern, multisource data sets. However, current methods for generating the prerequisite evidence are subjective and inconsistent. To address this, a more objective approach is presented for deriving evidence from histogram bin transformations of supervised training data frequency distributions. The procedure is illustrated by an example application in which evidential land-cover classification accuracy is increased from a kappa coefficient of 0.51 to 0.90 by appropriate use of bin transformation functions for a complex, mountainous environment in the Canadian sub-Arctic.

Introduction

New opportunities for synergy among environmental sciences, engineering, and remote sensing have emerged from the challenge to monitor and understand increasingly complex environmental processes at different scales, and as a result of concurrent advances in airborne and satellite sensor systems and computing architectures. However, for this critical evolution to occur, new approaches to image processing, analysis, classification, and modeling must be developed to help realize the full potential of these converging technologies. For example, time-honored methods of image classification such as the Bayesian maximum-likelihood algorithm were neither designed nor intended to process modern data sets which often possess (1) higher dimensions (or number of bands, e.g., hyperspectral imagery); (2) properties inappropriate for parametric statistical analyses; (3) information from different sources (i.e., multisource data) with inherent disparities, inconsistencies, errors, and uncertainty (e.g., incorporating ancillary variables, or using GIS data as an input to a remote sensing classification); and (4) data at different scales of measurement (or data levels, i.e., nominal, ordinal, interval, ratio) or with unique properties such as directionality (e.g., topographic aspect, climatological wind vectors).

To address these problems, new procedures for classifying multisource image data have been developed within the realms of pattern recognition, artificial intelligence, and knowledge-based expert systems (Argialas and Harlow, 1990; Campbell and Cromp, 1990; Tailor *et al.*, 1986). The Dempster-Shafer (D-S) Theory of Evidence (Dempster, 1967; Shafer, 1976), is one such approach that provides a framework for addressing the challenges of multisource image classification. In addition to its explicit mechanism for handling information uncertainty and conflict, a key aspect of the theory is its ability to combine, from any number of disparate sources, evidence in the form of support (information in favor of a class labeling) and plausibility (information which fails to refute that labeling) using the technique of *orthogonal summation* (denoted by $\oplus$). As an alternative approach to Bayesian theory, the D-S Theory of Evidence provides a powerful method for combining evidence into a decision using the concepts of evidential intervals and degrees of belief. However, as a result of the generality of this theory (it can be applied to any problem of statistical probability), there is no formal specification of how measures of evidence are obtained prior to the orthogonal summation process. There exists a significant gap between remotely sensed (and multisource) image data and its appropriate conversion to measures of evidence for input to the D-S approach. This gap in knowledge formulation is the basis for this contribution.

In the next section, previous applications of the D-S Theory of Evidence in remote sensing image classification will be reviewed to reveal the subjective and informal nature of current methods for deriving the necessary prerequisite evidence prior to an evidential classification, and that as a result, the full power of the D-S Theory of Evidence for multisource image analysis has yet to be realized. In the third section, the design criteria for a frequency-based approach to generating evidence from supervised training data is introduced, and in the fourth section, a bin transformation technique is described for manipulating the computed evidence to be representative over a greater range of digital values within the image domain. Prior to concluding the paper, an example application of evidential land-cover classification is presented for a mountainous environment in the southwest Yukon Territory, Canada, to illustrate how the bin transformation functions can be used to increase classification accuracy.

Background and Previous Studies

The Mathematical Theory of Evidence (Shafer, 1976) has received increasing attention in recent years for classifying multisource image data sets. However, much of the published literature to date has focused on describing and justifying the theory and its implications, with less attention directed towards exactly *how* evidence was obtained or derived. The following review is intended to summarize several previous studies involving the D-S Theory of Evidence, and to illustrate the need for a formal and more objective ap-

Earth-Observations Laboratory, Institute for Space and Terrestrial Science, Department of Geography, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada.

Photogrammetric Engineering & Remote Sensing,
Vol. 61, No. 4, April 1995, pp. 409–417.

0099-1112/95/6104–409$3.00/0

proach to generating evidence for input to an evidential classifier.

Lee *et al.* (1987) explored general methods of evidential calculus for multisource classification. Emphasis was placed on the advantages of using measures of evidence, while "... the bridge between these measures and the original data structure, whether the latter be numerical or otherwise, is left largely to the user" (Lee *et. al.*, 1987, p. 286). An example was presented for classifying a set of spectral data classes from a Landsat MSS image of an agricultural area. The visible (MSS bands 4 and 5) and infrared (bands 6 and 7) of the single image were considered as two independent sources of information, with evidence granted to various propositions using source-specific membership functions obtained from a prior statistical classification which assumed a normal distribution. In these tests, equal uncertainties were assigned for all pixels; however, they stressed the importance of determining pixel-specific uncertainty measures in future work to realize the full power of this component of evidential reasoning. The study amply demonstrated the advantages of the evidential approach; however, the implementation was restricted to ratio-level data and constrained by the use of parametric statistics to generate evidence, thereby requiring that the data conform to a normal distribution.

Moon (1990) used evidential belief functions to integrate disparate geological and geophysical data and to overcome problems of mixed data formats and different spatial resolutions. An interesting product from this study was the creation of a series of maps depicting the spatial distribution of evidential support for a series of base metal deposits. They relied on human experts to evaluate individual cells and produce qualitative assignments of evidence in support of a variety of mineral propositions. However, this assignment of partial belief functions was "... less exact and may even be arbitrary" (Moon, 1990, p. 714). Because this approach must rely on individual interpretations from a geologist, the basic framework for information representation for relating exploration evidence to mineral deposits is both difficult and subjective (Moon, 1993, p. 64). The development of a systematic and consistent technique to quantify and compute evidence for input to the evidential procedures was deemed a significant area in need of future work.

Wilkinson and Mégier (1990) used an evidential reasoning approach to integrate GIS data and expert system rules to resolve indecision in maximum-likelihood (ML) classification of agricultural land cover. A hierarchical class structure was used, and the evidential approach was based on a linear-time approximation to the D-S Theory of Evidence developed by Gordon and Shortliffe (1985). Supporting evidence was obtained as computed likelihoods from the ML classifier, while disconfirming evidence for the expert system rules was expressed as numeric probabilities based on qualitative relationships among GIS variables. The source of supporting evidence limits this approach to image data which adhere to ML assumptions (e.g., normally distributed data, limited dimensionality), as mentioned earlier and discussed in more detail in Peddle (1993), while the origin of the disconfirming evidence for the expert system rules was not specified. Therefore, although the general ideas put forth by Wilkinson and Mégier (1990) were valid and useful, the methods suggested for generating evidence were restrictive in their nature and lacking in objectivity. In addition, the approximation of Dempster's rule for hierarchical evidence developed by Gordon and Shortliffe (1985) was later shown to be unnecessary by Shafer and Logan (1987), based on an improved and exact algorithm which is also linear in its computational complexity.

Srinivasan and Richards (1990) provide an excellent description and evaluation of some of the advantages of the D-S Theory of Evidence cast in the spatial and remote sensing realm. In their implementation, a hierarchical class structure was also developed based on the computationally efficient algorithm described by Shafer and Logan (1987). A forward-chaining rule-based system was used, with two options for attaching evidence to rules: (1) using pre-defined functions to equate a fixed degree of evidence to rules entered in a constrained English format (e.g., for class *i*, DEFINITELY_NOT (*i*) would result in a belief of 0.9 being attached against the class labeling i); and (2) using heuristic functions to produce beliefs for and against a class (or set of classes) based on the degree to which the pre-condition of a rule was satisfied (e.g., for Landsat MSS imagery, the greater the band 4 (green) : band 5 (red) ratio value for a pixel, the more evidence would be granted in favor of a vegetation class label). However, it is often very difficult to translate heuristic knowledge into numerical degrees of belief (Srinivasan and Richards, 1990). Also, as a result of the subjectivity of human intervention, the knowledge embedded in a rule may have greater significance than the precision of the number (evidential mass) attached to it (Srinivasan and Richards, 1990, p. 516). Although this rule-based approach makes good use of the advantages and power offered by the D-S Theory of Evidence, there did not appear to be a consistent or repeatable method for translating knowledge into numerical belief values during the rule specification process.

As a final example, Goldberg *et al.* (1985) proposed the D-S Theory of Evidence as being appropriate to handle uncertainty in an expert system for updating forestry maps in western Canada based on Landsat image change detection. Within this context, the authors concluded that "Further research is required in the assignment of support and plausibility values in a consistent manner" (Goldberg *et al.*, 1985, p. 1062).

Two conclusions can be drawn from these studies: (1) the evidential approach is theoretically appropriate and shows much potential for multisource data integration and classification, and (2) an objective procedure for determining evidence for input to an evidential classifier has not been forthcoming. To address this need, the remainder of the paper describes a more objective approach to formulating knowledge as measures of evidence for input to a classification framework based on the D-S Theory of Evidence.

Knowledge Representation

Design Criteria

The supervised approach to classification is used in this implementation of an evidential classifier (Peddle, 1995) to provide the image analyst with sufficient power to address complex environmental problems which require multisource image data and the *a priori* identification of intricate physical classes for their solution. The requirements for a supervised evidential classifier include (1) a way of obtaining representative information for each class to base classification decision making, and (2) a way of converting this information into measures of evidence by class. The first requirement is satisfied using standard training data identified for each class. This approach takes advantage of existing training data acquisition modules and graphical interfaces available in most commercial image analysis systems for interactive class delineation. It also makes the evidential classifier compatible with existing training data sets used previously with other supervised classification algorithms. The second requirement poses a greater challenge due to the disparate nature of multisource data sets which preclude the use of powerful statistical models and measures of central tendency to characterize training samples (such as the Gaussian assumption in maximum-likelihood classification and many implementations of linear discriminant analysis). Therefore, a method is required which provides greater flexibility with respect to input data types and which adequately captures the increased information content available from multisource data.

To formalize these requirements, the following design criteria have been identified for converting supervised training data into measures of support and plausibility within an evidential classification framework:

- the method must be free of statistical assumptions and models;
- it must be able to handle multisource data at any scale of measurement (or data level);
- it must be able to incorporate uncertainty into the analysis;
- a mechanism must exist to grant evidence to pixel values which are representative of a class, but which do not occur within the range of training class values due to the chance location of training samples (this is essentially a question of interpolation); and
- a method is needed to determine evidence for values representative of a class but which lie outside the numeric bounds of a training sample (this is an issue analogous to extrapolation).

Deriving Evidence

The method devised to meet the design criteria for generating evidence uses training data explicitly as direct sources of evidence for class membership. Evidential support is computed with respect to the frequency of occurrence of values within training samples. The universality of this approach is based on the fact that all training data have a frequency distribution, regardless of data type, scale of measurement, or statistical properties. There are no requirements for intricate mathematical formulation, statistical processing, or reinterpretation, and, as a result, the method is relatively easy to understand intuitively and is without excessive computational burden. Two basic premises underlie the approach: (1) values found in class training samples represent that class (i.e., they provide *evidence* in support of a particular class labeling), and (2) the frequency of occurrence of a specific value within a class training sample is an indicator of the magnitude of support for that class (i.e., it *quantifies* the support for a class labeling).

The first step in this approach is to obtain frequency distributions of training samples over the entire set of classes, or *frame of discernment* (denoted by Θ). Training data are read from each data feature in sequence, and a frequency distribution of training values is compiled for each class. Thus, for i classes and k sources, there will be a total of $i \times k$ frequency distributions. During this compilation process, the training sample size (TSn) for each class is recorded. For a given input pixel value PV to be classified, the amount of evidence in support of the ith class label is computed initially as the frequency of occurrence of PV in the training data for class i divided by the number of training samples (TSn) for class i. Plausibility (P) is a measure of the extent to which the available evidence does not support the negation of a given proposition (Shafer, 1976), and is computed for a given class C_i as

$$P(C_i) = \{1 - S(\neg C_i)\} \tag{1}$$

where $S(\neg C_i)$ is the magnitude of evidential support for the negation of class C_i, computed over a total of n classes as

$$S(\neg C_i) = \sum_{\substack{j=1 \\ j \neq i}}^{n} S(C_j) \tag{2}$$

Individual measures of evidential support and plausibility lie in the range 0 to 1, inclusive, with the set of supports and plausibilities for a given input value over a frame of discernment Θ referred to here as the *evidential vector*. Uncertainty is quantified for each pixel as the amount of evidence not assigned to any particular subset (i.e., class), and it is computed as one minus the sum of supports for all classes (after Garvey *et al.*, 1981). In the rare case where this sum exceeds one, the evidential vector is normalized to unity (i.e., 1), and there will be no quantifiable uncertainty.

An example multisource training data set consisting of three sources with different properties was constructed to help explain the methodologies created for the derivation and processing of evidence. This hypothetical training data set was designed to illustrate the flexibility of the approach for classifying multisource, disparate data (e.g., remote sensing imagery together with GIS data and directional information) obtained at different scales of measurement (ratio and nominal), and which do not necessarily conform to the Gaussian distribution. Figure 1 shows the frequency distributions of the example training data set for three sources and three classes. An example input pixel {110, 6, 315} from these distributions is used to demonstrate explicitly how evidence is derived from the original training data.

From Figure 1, one can surmise the general nature of each data source and magnitudes of support for a given value over the set of classes. For example, Source 1 is at the ratio scale of measurement, and could represent a typical input feature from an 8-bit digital remote sensing image. Only training data for Class 3 have a distinct normal distribution, while Class 2 is bi-modal. Observations of these training data distributions suggest low values in Source 1 are more indicative of Class 1, while high values are more likely a member of Class 2.

Source 2 is nominal (or thematic) level data which could have been obtained from a geographic information system or from an earlier remote sensing classification. The frequency of occurrence of value i has no bearing on the frequency of value $i + 1$, because the numeric values are assigned to classes arbitrarily, and usually without any physical basis. None of the three class training samples appears to have a normal distribution. In general, the nominal values 2, 6, 8, and 9 are indicative of Class 1; values of 1, 4, and 7 provide the most support for Class 2; while values 3 and 5 are more likely to represent Class 3.

Source 3 illustrates properties of directional data (e.g., compass aspect measured in degrees from 0 to 359). Low values are more indicative of Class 1, intermediate values lend the most support to Class 3, while high values more likely represent Class 1 or 2. The directional or circular nature of these data is shown by the distribution for Class 1,

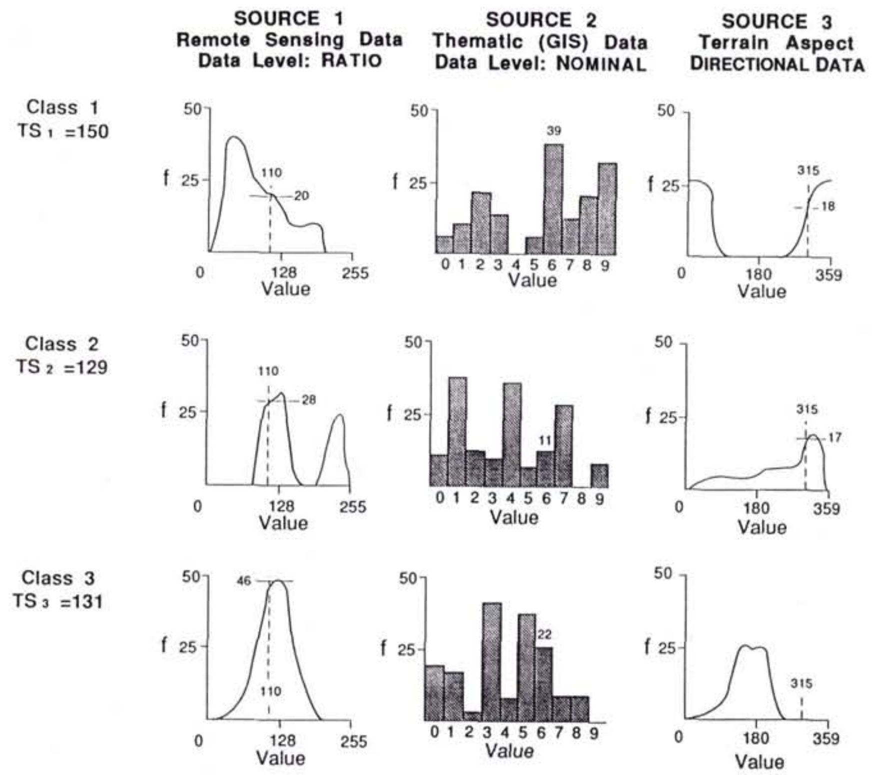

Figure 1. Example training class frequency distribution for three features from different sources in a multisource data set. Training sample size (TSn) shown in left column for each of n = 3 classes (distribution not drawn to scale). The graphic illustrates the disparate nature of the three example data sources to be integrated and classified: (1) remote sensing data at the ratio scale of measurement, (2) nominal (or thematic) GIS data, and (3) directional data (e.g., terrain aspect). The frequency of occurrence (f) of each value in a pixel vector {110, 6, 315} to be classified is shown for each distribution and used in the computation of evidence from this example, as described in the text and given in Table 1.

where the frequency of the lowest and highest values appears to be similar. In the case of directional data, the class training set for Class 1 would resemble a normal distribution if the x-axis is relabeled starting at 180 (and continuing through 359, 0 to 179). However, in a number system which deals strictly with absolute magnitudes, this variable would be considered to have a distinctly bi-modal distribution, as shown in Figure 1. This type of data is not suitable for parametric classifiers which rely on arithmetic measures of central tendency and variance to characterize training data information (e.g., maximum likelihood, linear discriminant analysis). For example, the mean of the Class 1 distribution would be approximately 180, even though no values in Class 1 are close to that value. Similarly, the computed variance would not be representative because it would be greatly overestimated. Because these statistical models and assumptions are not used in this implementation of an evidential classifier, it is possible to process directional data together with other information at all scales of measurement.

The computation of evidential support from the example data set is provided in Table 1. The pixel vector {110, 6, 315} from Figure 1 is used to show the derivation of support values from the training data distribution of each class shown. For each source, the sum of supports for all classes does not exceed one, and therefore mass normalization to unity is not required. The importance of incorporating the training sample size into the computation of evidence from training data frequency distributions is apparent for Source 3. The pixel value 315 occurs most often in the training data for Class 1, which is also the largest training sample by class. However, the greatest support is assigned to Class 2, because the frequency of occurrence of this pixel value in that class occupies a greater proportion of a smaller training sample.

The computed measures of pixel-specific uncertainty (underlined in Table 1) represent the remaining amount of evidence which could not be ascribed to a particular class within the frame of discernment Θ (in this case, Θ is the set of classes [1,2,3]). This residual uncertainty must instead be assumed to be distributed in some unknown manner among the class propositions, and, as a result, the evidence is assigned to Θ (after Garvey *et al.*, 1981).

With reference to the design criteria identified at the be-

TABLE 1. EXAMPLE COMPUTATION OF EVIDENTIAL SUPPORT VALUES FROM THE CLASS TRAINING DATA FREQUENCY DISTRIBUTIONS SHOWN IN FIGURE 1. $P\nu$ IS THE PIXEL VALUE IN THE VECTOR {110, 6, 315} TO BE CLASSIFIED; TSn IS THE TOTAL NUMBER OF PIXELS IN EACH CLASS TRAINING SAMPLE; f IS THE FREQUENCY OF OCCURRENCE OF $P\nu$ IN EACH TRAINING SAMPLE. EVIDENTIAL SUPPORT (S) FOR EACH CLASS IS COMPUTED AS $f \div$ TSn, WHERE $0 \leq S \leq 1$. BOLD ENTRIES DENOTE THE CLASS WITH THE GREATEST AMOUNT OF EVIDENCE BY SOURCE. EVIDENCE NOT COMMITTED TO ANY CLASS (C_i) IS ASSIGNED TO THE FRAME OF DISCERNMENT ($S(\Theta) = 1 - \Sigma S(C_i)$ SHOWN AS BOLDFACE ENTRIES), AND REPRESENTS THE RESIDUAL UNCERTAINTY ASSOCIATED WITH EVIDENCE FROM EACH SOURCE (GARVEY *ET AL.*, 1981).

		Source 1 $P\nu = 110$		Source 2 $P\nu = 6$		Source 3 $P\nu = 315$	
Class	TSn	f	Support	f	Support	f	Support
1	150	20	0.13	39	**0.26**	18	0.12
2	129	28	0.22	11	0.09	17	**0.13**
3	131	46	**0.35**	22	0.17	0	0.00
Θ			**0.30**		**0.48**		**0.75**

ginning of this section, the first three have been satisfied by the method outlined. The frequency-based technique for generating support values has been shown to be (1) not restricted by statistical assumptions or models, (2) able to process multisource data at any level, and (3) equipped with a mechanism to quantify and incorporate uncertainty into the classification process. The remaining two design criteria deal with extending knowledge from training samples to encompass a greater range of values within the multisource image domain. The next section describes a transformation approach to facilitate the classification of values which do not occur in training data.

Knowledge Domain Processing by Bin Transformation

The method developed for unrestricted knowledge domain processing operates by transforming the frequency distribution of training data using weighted functions applied over a specified range, or *bin size*. This approach enables information to be both interpolated within the numerical range of class training data (Design Criterion 4), and extrapolated beyond that range (Design Criterion 5).

Knowledge from training samples of quantitative data sources is extended by propagating the evidence (frequency of occurrence) from individual data values to its neighbors, with the propagation function weighted by *proximity to the* original training data value. The approach utilizes a multiplicative linear-weighted distance decay function and is based on two premises: (1) if a value i occurs in training data for Class c, then similar values are also indicative of that class (e.g., for quantitative data, $i \pm 1 \in c$); and (2) the probability (p) that similar values represent Class c increases with proximity to i, or

$$p(i \pm 1 \in c) > p(i \pm 2 \in c) \tag{3}$$

In the current implementation, the bin size is specified by the user and can vary by individual feature. The bin size is necessarily odd so that propagation of evidence is symmetrical about the data value being considered. It should be noted that the bin transformations described here are distinct from the *a priori* division of histograms into fixed cells (sometimes referred to as bins). That process involves compressing the range of values in a histogram, with an associated loss of precision. In this implementation, no reduction or generalization of frequency distributions is applied, and therefore the original precision of training data is preserved.

For an individual training data value i, the method works by first multiplying its frequency of occurrence f by a specified constant equal to the bin size. The constant is then decremented by 2, multiplied by f, and added to the frequencies of occurrence of the two next adjacent values ($i + 1$, $i - 1$). This is continued over the entire bin. Therefore, values which lie within a given bin are transformed as follows: for Class c, given a training sample value i with a frequency of occurrence $f(i) = a$ and a specified bin size b, the evidence for value j ($f(j) \geq 0$) belonging to Class c would be incremented as

$$f(j) = f(j) + a \times (b - 2 \times |i - j|) \tag{4}$$

where $|i - j| < b \div 2$ (i.e., j lies within the bin).

The specified bin size is applied to each value of all class training samples for all valid sources (i.e., quantitative data selected for bin transformation). Once all frequency distributions have been processed, evidence is computed for a given input value as discussed earlier. The training sample size (TSn) used in the denominator of the evidential support equation is adjusted to reflect the increased frequency total produced through the bin transformation process.

This approach is illustrated by example in Table 2. A bin size of 5 is applied to a training data set containing four samples (TSn = 4) for an arbitrary Class c. This small bin size and a small number of samples was chosen to simplify

TABLE 2. EXAMPLE FREQUENCY TRANSFORMATIONS OF TRAINING DATA USING A BIN SIZE OF 5. THE ORIGINAL FREQUENCY OF TWO TRAINING SAMPLE VALUES (70 AND 72, SHOWN IN BOLD IN ROW 1) ARE TRANSFORMED INTO LOCAL DISTRIBUTIONS IN TABLES 2A AND 2B, RESPECTIVELY. ADJACENT VALUES INCLUDED IN THE TRANSFORMATION ARE SHOWN IN BRACKETS IN ROW 1 OF EACH TABLE. TABLE 2C SHOWS THE ADDITION OF THE TWO TRANSFORMED DISTRIBUTIONS, THE UPDATED FREQUENCIES FOR EACH VALUE, AND THE RESULTING COMPUTATION OF EVIDENTIAL SUPPORT.

A:

1. Pixel Values:		(68)	(69)	**70**	(71)	(72)
2. Original Frequency:				1		
3. Bin Transformation:	×	1	3	5	3	1
4. Transformed Frequencies:	=	1	3	5	3	1
5. + Original Frequency:	+			1		
6. Transformed Local Frequency Distribution:	=	1	3	6	3	1

B:

1. Pixel Values:		(70)	(71)	**72**	(73)	(74)
2. Original Frequency:				3		
3. Bin Transformation	×	1	3	5	3	1
4. Transformed Frequencies:	=	3	9	15	9	3
5. + Original Frequency:	+			3		
6. Transformed Local Frequency Distribution:	=	3	9	18	9	3

C:

1. Pixel Values:	68	69	70	71	72	73	74
2. Transformed Frequencies: (from Table 2A; $P\nu$ = 70)	1	3	6	3	1		
3. Transformed Frequencies: (from Table 2B; $P\nu$ = 72)			3	9	18	9	3
4. Total Frequencies; f = (adjusted TSn = 56)	1	3	9	12	19	9	3
5. Evidential Support: $f \div$ TSn:	0.018	0.054	0.161	0.214	0.339	0.161	0.054

the illustration (in practice, these values will usually be considerably larger). In the example, there is one occurrence of the value 70 in the training sample, and three occurrences of the value 72 (Tables 2A and 2B, respectively). The bin transformation is applied to each sample value, and the frequencies are accumulated in Table 2C. The TS*n* value is adjusted and evidential support is computed as shown in the bottom line. The resulting frequency spread is better suited to recognize an input value in the range 68 to 74 as being a member of Class *c*. For example, without applying the bin transformation process, there would be no evidence for assigning an input value of 71 to the class which contained 70 (once) and 72 (three times) in its training data. The fact that 71 was not in the training set is almost certainly due to the chance location of the sample data, and not because the value is not representative of that class. This is an example of interpolation within training data (Design Criterion 4). Evidence can also be extrapolated beyond the range of training sample data. This occurs when the transformation process is applied to extreme values within the training data frequency distribution. In the example (Table 2), evidence generated for the values 68, 69, 73, and 74 would be considered as extrapolated support (Design Criterion 5). In this research, a linear relationship is assumed between numerical proximity and probability of membership during the bin transformation process; however, more complex relationships (e.g., square-root functions, logarithmic decay) may provide improved results when larger bin sizes are applied.

Bin Size Considerations

The selection of bin size is an important parameter in this implementation of an evidential classifier. Different bin sizes may be specified for different sources, with a given bin size selected with reference to the nature of the data being classified and the precision of the classes under consideration. Ideally, all valid data values which represent a given class would be incorporated into the expanded training sample through the bin transformation procedure, with no values exceeding known class limits. In practice, however, training data from different classes often are not mutually exclusive for all data sources, and training data overlap among classes is inevitable. Although bin transformation increases the likelihood of overlap, the effects are minimal. This is because, for a given feature, the same bin size is used for all class training samples, and, as a result, the overlap from Class *a* to Class *b* caused by the bin transformation would have no consequence because the frequency of the Class *b* training value would be increased by the full amount within its own bin. This serves to cancel out any negative impact on the ability to discriminate Class *b*. It also suggests that there may not be a maximum bin size beyond which classification accuracy would be expected to degrade (i.e., overall accuracy will stabilize at and beyond a sufficiently large bin size, provided the bin sizes do not exceed the numeric range of the data). Therefore, providing sufficient memory exists, a general rule is to select large bin sizes to ensure sufficient expansion of training data values to be representative of its class. This also permits classification based on fewer training samples, because the information provided by each sample value is propagated over a wider range of digital values. However, because larger bin sizes require more memory and computation time, this method of bin size selection may be limited in practice by available resources. In this case, the theoretical optimal bin size would be the minimum bin size which can still yield the maximum level of classification agreement (i.e., the bin size at which classification accuracy begins to stabilize). A more refined approach is required to determine this region of stability. Early results from empirical analyses of histograms, descriptive statistics, and moments of training sample frequency distributions suggest that a reasonable initial bin size can be determined as one-fifth of the sample range for a given class training set (for multi-modal training data, the range of individual data groupings identified in the histogram should be used). However, this will likely vary by data source, application, and training sample size; therefore, some iteration and experimentation with bin sizes may be necessary to achieve optimal results. For example, with larger training sample sizes, one could expect maximum classification accuracy to be reached using smaller bin sizes. Conversely, if the training sample size is small, larger bin sizes would likely be required.

The bin transformation process cannot be applied to qualitative data (i.e., nominal and ordinal level data) because, at these scales of measurement, magnitude of difference between data values is either inappropriate or unknown. However, this generally does not limit the utility of the frequency-based approach for classifying qualitative data because these variables often possess a limited range of values, for which their original frequency distributions are usually representative. For the same reason, frequency transformations are not always necessary for quantitative data which possess a limited dynamic range or a very large training sample. In these cases, the user would specify that no bin transformation of training data is to be performed for that feature.

Additional Functionality

In addition to bin size, several other options have been incorporated into the knowledge representation scheme to permit greater control of the classification process if additional information about the multisource data is known *a priori*. For example, it is possible to assign weights of importance to each input data variable if the relative quality of these variables is known with respect to the classes of interest. Reliability specifications such as these are often not available in conventional statistical classifiers, despite the importance of reliability measures for multisource data which, by their nature, are more likely to contain variables with varying degrees of relevance to a particular application (Benediktsson *et al.*, 1990).

The weights of evidence concept has also been extended in this implementation to allow, for a given data feature, different weighting factors to be assigned for different classes. This option permits the user to include detailed information about how individual features and classes are related based on information such as field surveys, aerial reconnaissance, or laboratory analyses. Both of these optional evidential weighting capabilities provide the user with ways to capture more fully the additional information content available in multisource image data sets for optimizing classification results.

As a result of the disparate nature of multisource data sets, individual data variables sometimes possess missing values, undefined data fields, or information with different properties. These inconsistencies create problems which should be dealt with explicitly by a classifier. In this implementation, missing data values can be identified by a data flag and excluded from the analysis. Undefined values can be similarly flagged; however, in this case the user may choose to include this information if it is warranted (e.g., flat terrain has an undefined topographic aspect; however, this information is useful for classification). No bin transformations are permitted on undefined data.

TABLE 3. EVIDENTIAL LAND-COVER CLASSIFICATION ACCURACY WITH INCREASING BIN SIZE. THE BIN TRANSFORMATION OPTION WAS NOT USED FOR THE FIRST ENTRY OF THE TABLE (ø); THE LAST ENTRY SHOWS THE HIGHEST ACCURACY OBTAINED USING THE FEATURE-SPECIFIC BIN SIZES LISTED IN TABLE 4. ACCURACY IS EXPRESSED IN TERMS OF PERCENT AGREEMENT AND THE KAPPA COEFFICIENT (κ) FOR 455 INDEPENDENT TEST PIXELS.

Bin Size	%	κ
ø	56.2	0.51
3	57.6	0.53
7	67.9	0.65
11	81.0	0.79
15	85.0	0.82
19	85.7	0.84
23	86.8	0.85
27	86.4	0.84
31	86.6	0.84
35	86.7	0.85
39	86.6	0.84
Variable (see Table 4)	91.2	0.90

The bin transformation process can also be controlled to permit the correct incorporation of directional data (e.g., wind direction, solar azimuth data, terrain aspect) into the evidential classification process. This is achieved by specifying the actual or theoretical range of data values which is used to modify the bin transformation process accordingly through the use of a wrapping function to ensure that the appropriate frequency values are incremented.

Example Application

The approach described here for deriving evidence from multisource image data has been implemented in the C-programming language as part of the MERCURY⊕ evidential classifier (Peddle, 1995), which runs under the UNIX, ULTRIX, VAX/VMS, MS-DOS, and Apple Macintosh operating systems. This software was shown to provide significantly higher classification accuracies than traditional maximum-likelihood and linear discriminant analyses in an extensive comparison of high-relief land-cover classifications (Peddle, 1993). It was also crucial for the complex environmental application of permafrost active layer depth classification (Peddle and Franklin, 1993) through its ability to handle disparate, multisource data which otherwise could not be processed by conventional means. As a follow-up to those detailed studies, a series of experiments is presented in this section to study the effects of different bin sizes on land-cover classification accuracy. A brief description of the experimental design is given here; however, for a full account of the study area, data sets, and processing strategies, the reader is referred to Peddle (1993) and Peddle and Franklin (1993).

The sub-Arctic study area is located in mountainous terrain of the Ruby Range, southwest Yukon Territory, Canada. Vegetation and land cover vary through an elevation range of 900 m, and have been generalized into nine classes as follows: white spruce forest, woodland, upland shrub, alpine shrub, alpine tundra, alpine barrens, organic terrain, exposed slopes, and water (after Franklin, 1987). The digital data sources for this study include a cloud-free multispectral SPOT HRV image acquired 21 July 1990 and a co-registered dense grid digital elevation model (DEM). Spectral image texture was processed from each SPOT image band using a spatial co-occurrence algorithm, with noise removal procedures and geomorphometric processing (slope, aspect, curvature, and relief) applied to the DEM. Spectral texture and geomorphometric image processing have been shown to provide the additional information necessary for increasing land-cover classification accuracy in complex, mountainous areas such as the Ruby Range (Franklin, 1987; Peddle and Franklin, 1991). However, the higher dimensionality of the data set and the fact that many of the new variables do not conform to a normal distribution complicated the use of conventional classifiers, and resulted in a need for the new classifier presented here.

Observations of land cover from field work and aerial photointerpretation were compiled for 1693 pixel sites. These pixels were identified in the registered data sets, divided randomly into a mutually exclusive set of 1238 training and 455 test pixels, and written to disk as independent attribute table files. These attribute tables were used in a series of 11 tests of the MERCURY⊕ evidential classifier using different bin sizes. The full complement of available SPOT image bands, image texture, and geomorphometric variables was used in each classification to conduct a rigorous test under conditions of high data variability and maximum available data volume. This set of 12 variables has also been shown to possess the highest amount of information content in a series of empirical analyses (Peddle, 1993). Classification accuracy was determined with respect to ground data and expressed using percent agreement and the Kappa coefficient for the 455 independent test pixels. The experiments were controlled as follows: (1) in each test, the same bin size was used for all sources; (2) all sources and classes were weighted equally in each test; (3) separate test data were used for all classification assessments to avoid overestimating accuracy; (4) the same training and test samples were used in all classifications; and (5) all other parameters were kept constant throughout the experiment. Bin size was the only parameter altered between individual tests.

Table 3 shows the percent accuracy and kappa coefficients obtained as the bin size was increased from 3 to 39, in increments of four. When the bin transformation option was not used, classification accuracy was 56.2 percent (Kappa coefficient: κ 0.51). This was the lowest accuracy obtained in the experiments, and suggested immediately that using only the frequency of occurrence of original training data without any transformation was insufficient for acceptable classification accuracy. This would be expected because training samples from higher level, quantitative data rarely possess the full range of values representative of a given class. A minimal increase in accuracy was found with a bin size of 3. Again, many pixels which represent a given class are still not being included in the transformed sample. However, with larger bin sizes, significant increases in classification accuracy were observed. The accuracy increased by 10 percent and then by a further 13 percent using bin sizes of 7 (67.9 percent, κ 0.65) and 11 (81 percent, κ 0.79), respectively. This illustrates the positive effect of including a greater range of pixels in the bin transformation process for generating evidence within the MERCURY⊕ classifier. Classification accuracy increased further to 85 percent (κ 0.82) and to 85.7 percent (κ 0.84) with bin sizes of 15 and 19, after which the accuracy stabilized at ≈86 percent. From these results, there appears to be a bin size threshold beyond which classification accuracy reaches a maximum and remains constant. However, because computational requirements and the amount of memory needed increase with bin size, it would be desirable to avoid using unnecessarily large bins. These results confirm the earlier notion that the optimal bin size should be set as the smallest bin size for which classification accuracy is maximized. In this case, an optimal bin size of 19 was found for this data set when the bin size was held constant for all features tested. Clearly, however, if memory resources or processing speed are significant controlling factors, a trade-off exists between classification accuracy and available computing resources.

Additional increases in classification accuracy are possible using the feature-specific bin size option available in the MERCURY⊕ software. After experimentation with different bin sizes for each feature, the highest classification accuracy achieved was 91 percent (κ 0.90). The bin sizes used in that classification are shown in Table 4, and ranged from 11 to 27. In general, the bin sizes were larger for features with a greater dynamic range of digital numbers. However, these selections of bin sizes are data dependent, and therefore the optimal bin sizes found in this study (Table 4) cannot be readily generalized to other study areas or data sets. Also, in this example, all sources were assigned equal weights for these experiments. Even higher classification accuracies may be possible by weighting each feature according to relative information content, or by introducing different weights for each class, as discussed earlier.

TABLE 4. FEATURE-SPECIFIC BIN SIZES USED TO OBTAIN THE HIGHEST LEVEL OF EVIDENTIAL LAND-COVER CLASSIFICATION AGREEMENT OF 91.2 PERCENT, KAPPA COEFFICIENT 0.90

Feature	Bin Size
SPOT Band 1	17
SPOT Band 2	17
SPOT Band 3	21
Image Texture: Band 1	23
Image Texture: Band 2	27
Image Texture: Band 3	27
Elevation	23
Aspect	11
Slope	23
Down Slope Convexity	25
Cross Slope Convexity	25
Relief	17

Conclusion

The Dempster-Shafer Theory of Evidence provides an established and mathematically sound framework for consolidating evidence from multisource data for image classification. However, current methods to first derive these measures of evidence from image data prior to invoking the orthogonal summation process have been shown to be arbitrary, subjective, and inconsistent, and have prevented the full power and versatility of evidential classification from being realized. To overcome these problems, a more objective method for generating evidence from supervised training data has been presented as an interface to a Dempster-Shafer multisource image classifier. Evidence is computed from transformed frequency distributions of training data, and without reference to restrictive mathematical models or statistical assumptions. The approach permits the integrated classification of data at all scales of measurement (i.e., nominal, ordinal, interval, ratio), from different sources (e.g., thematic GIS data together with remotely sensed imagery and ancillary information), and with different or unique properties (e.g., directional data, or information sources which include undefined or missing data points). The user may control the bin transformation process and also has the option to specify individual feature weights and class weights as appropriate.

The approach to deriving evidence was illustrated in an example application of alpine land-cover classification using the MERCURY⊕ evidential classification software (Peddle, 1995). Information available from 12 multisource input variables comprised of SPOT imagery, image texture, and geomorphometry from a digital terrain model was used in a series of experiments to test the effect of increasing bin size on classification accuracy. Classification accuracy increased steadily from a kappa coefficient (κ) of 0.51 using no bin transformation functions, to κ 0.84 using a bin size of 19 for each data feature. The results stabilized at this level using larger bin sizes, from which it may be concluded that the optimal bin size in terms of both memory requirements and overall classification performance is the smallest bin size for which the highest level of classification agreement is attained. In each test, the bin size was kept constant over the entire data set; however, when the bin size was varied by feature, a higher overall classification agreement of κ 0.90 was achieved. Additional increases in classification agreement are also possible through the use of individual feature weights, and by utilizing the class weighting option. These tests illustrate the utility of the approach taken for representing training knowledge as evidence for input to an evidential classifier, and the importance of applying bin transformation procedures to achieve higher levels of classification accuracy for multisource data sets. The continued development of improved procedures for image processing and analysis will be vital if we are to realize the full potential of powerful new computational techniques for extracting the increasingly rich information content available from multisource data. The interface developed for a Dempster-Shafer formalism and presented here is one such example.

Acknowledgments

This research was supported by Scholarships and Fellowships from the Natural Sciences and Engineering Research Council of Canada, the Eco-Research Tri-Council Secretariat, the Alberta Heritage Scholarship Fund, and Forestry Canada. I acknowledge with gratitude the contributions made by Dr. Steven E. Franklin, Department of Geography, The University of Calgary, who created and supported many of the opportunities and the broader research context within which the work reported here evolved as one component. I am grateful to Dr. Peng Gong, University of California at Berkeley, Dr. S.E. Franklin, and also to Dr. Michael Hodgson (*PE&RS* Associate Editor) and the anonymous reviews for helpful comments on this work. Readers interested in obtaining the MERCURY⊕ classifier or other software developed in this research may contact the author.

References

Argialas, D.P., and C.A. Harlow, 1990. Computational image interpretation models: An overview and a perspective, *Photogrammetric Engineering & Remote Sensing*, 56(6):871-886.

Benediktsson, J.A., P.H. Swain, and O.K. Ersoy, 1990. Neural network approaches versus statistical methods in classification of multisource remote sensing data, *IEEE Transactions on Geoscience and Remote Sensing*, 28(4):540-552.

Campbell, W.J., and R.F. Cromp, 1990. Evolution of an intelligent information fusion system, *Photogrammetric Engineering & Remote Sensing*, 56(6):867-870.

Dempster, A.P., 1967. Upper and lower probabilities induced by a multivalued mapping, *Annals of Mathematical Statistics*, 38: 325-339.

Franklin, S.E., 1987. Terrain analysis from digital patterns in geomorphometry and landsat MSS spectral response, *Photogrammetric Engineering & Remote Sensing*, 53(1):59-65.

Garvey, T.D., J.D. Lowrance, and M.A. Fischler, 1981. An inference technique for integrating knowledge from disparate sources, *Proceedings, Seventh International Conference on Artificial Intelligence*, Vancouver, Canada, pp. 319-325.

Goldberg, M., D.G. Goodenough, M. Alvo, and G. Karam, 1985. A hierarchical expert system for updating forestry maps with landsat data, *Proceedings of the IEEE*, 73(6):1054-1063.

Gordon, J., and E.H. Shortliffe, 1985. A method for managing evidential reasoning in a hierarchical hypothesis space, *Artificial Intelligence*, 26:323-357.

Lee, T., J.A. Richards, and P.H. Swain, 1987. Probabilistic and evidential approaches for multi-source data analysis, *IEEE Transactions on Geoscience and Remote Sensing*, 25(3):283-292.

Moon, W.M., 1990. Integration of geophysical and geological data using evidential belief function, *IEEE Transactions on Geoscience and Remote Sensing*, 28(4):711-720.

———, 1993. On mathematical representation and integration of multiple spatial geoscience data sets, *Canadian Journal of Remote Sensing*, 19(1):63-67.

Peddle, D.R., 1993. An empirical comparison of evidential reasoning, linear discriminant analysis and maximum likelihood algorithms for alpine land cover classification, *Canadian Journal of Remote Sensing*, 19(1):31-44.

———, 1995. MERCURY⊕: An evidential reasoning image classifier, *Computers & Geosciences* (in press).

Peddle, D.R., and S.E. Franklin, 1991. Image texture processing and data integration for surface pattern discrimination, *Photogrammetric Engineering & Remote Sensing*, 57(4):413-420.

———, 1993. Classification of permafrost active layer depth from remotely sensed and topographic evidence, *Remote Sensing of Environment*, 44(1):67-80.

Shafer, G., 1976. *A Mathematical Theory of Evidence*, Princeton University Press, Princeton, New Jersey, 297 p.

Shafer, G., and R. Logan, 1987. Implementing Dempster's rule for hierarchical evidence, *Artificial Intelligence*, 33:271-298.

Srinivasan, A., and J.A. Richards, 1990. Knowledge-based techniques for multi-source classification, *International Journal of Remote Sensing*, 11(3):505-525.

Tailor, A., A. Cross, D.C. Hogg, and D.C. Mason, 1986. Knowledge-based interpretation of remotely sensed images, *Image and Vision Computing*, 4(2):67-83.

Wilkinson, G.G., and J. Mégier, 1990. Evidential reasoning in a pixel classification hierarchy — A potential method for integrating image classifiers and expert system rules based on geographic context, *International Journal of Remote Sensing*, 11(10):1963 -1968.

(Received 24 May 1993; revised and accepted 9 May 1994; revised 14 June 1994)

Derek R. Peddle

Derek R. Peddle received the Bachelor of Science Degree with Joint Honours in Computer Science and Geography from Memorial University of Newfoundland in 1987. He then worked for two years as a Remote Sensing Scientist at the Newfoundland Oceans Research and Development Company (Nordco Limited). He received the Master of Science Degree in Geography from The University of Calgary in 1991 and is presently completing a Ph.D. degree in Geography and Environmental Studies at the University of Waterloo. In 1994 he was a Visiting Scientist at the NASA Goddard Space Flight Center and he is presently a Research Scientist in the Earth-Observations Laboratory, Institute for Space and Terrestrial Science, working on the BOREAS Project. His current research involves developing remote sensing image processing algorithms for studies of environmental change in forested and alpine ecosystems.