DECISION FUSION APPROACH FOR MULTITEMPORAL CLASSIFICATION

Byeungwoo Jeon[1] and David A. Landgrebe[2]

[1]School of Electrical and Computer Engineering
Sung Kyun Kwan University, Korea
bjeon@yurim.skku.ac.kr

[2]School of Electrical and Computer Engineering
Purdue University, W. Lafayette, IN 47907-1285, U.S.A.
landgreb@ecn.purdue.edu

ABSTRACT

This paper propose two decision fusion-based multitemporal classifiers, namely, the jointly likelihood and the weighted majority fusion classifiers, that are derived using two different definitions of the minimum expected cost. Without any overhead incurred by multitemporal processing, a user-selected conventional pixelwise classifier makes local class decisions separately using each temporal data set, and the proposed multitemporal classifiers make the global class decisions by optimally summarizing those local class decisions. The weighted majority decision fusion classifier can handle not only the data set reliabilities but also the classwise reliabilities of each data set. Classification experiment using the jointly likelihood decision fusion with three remotely sensed Thematic Mapper (TM) data sets shows more than 10% overall classification accuracy improvement over the pixelwise maximum likelihood classifier.

Key Words : Multitemporal classifier, decision fusion, reliability, distributed hypothesis testing, distributed detection

I. INTRODUCTION

For decades, remote sensing technology has been successfully applied in many interdisciplinary applications of Earth observational data [1]. The recent advent of more powerful sensor systems enables one to extract far more detailed information than ever before from the observed data, but to realize that this goal requires the development of effective data analysis techniques which can utilize the full potential of the observed data. For example, the availability of multitemporal data sets over the same scene makes it possible to extract valuable temporal characteristics of surface cover types that may be of interest to applications requiring the monitoring of spectral or spatial characteristic changes over time [2] [3].

However, proper utilization of temporal contextual information calls for designing an appropriate multitemporal classifier. A few desirable properties of a multitemporal classifier are as follow [3];

(1) Since there are usually only a limited number of training samples available for each temporal data set, a multitemporal classifier should not require extra training samples additional to those already available for pixelwise non-temporal classification. It will be desirable if the classifier can be trained separately for each temporal data set. In this respect, it is quite common to assume class-conditional independence of features belonging to different temporal data sets.

(2) It will be also worthy if a multitemporal classifier can facilitate distribution of computation required for classification over different times by allowing easy update of the intermediate result already computed with previous temporal data sets when a new data set becomes available.

(3) A different temporal data set can have distinct properties and varying discriminating power, therefore, a multitemporal classifier would be useful if it can accommodate different reliability factors associated with each temporal data set or its class decisions.

Motivated by notion that the multitemporal classification can be thought as one example of a multisource classification problem [4] [5] where the temporal data sets are considered as separate information sources, we formulate the problem in a similar context of the M-ary distributed hypothesis testing problem [6-10]; the proposed methods are based on a fusion of "class decisions" of each separate temporal data set (we call them *local* decisions). Since only the local decisions are involved in the final step of multitemporal classification and local classifiers can make their own decisions without any additional constraints due to the fusion at later time, this approach can ease the requirements at the training stage and subsequently the computational complexity set forth above.

Different information sources can have different degrees of reliability, i.e., one data set might be more reliable than others in a specific analysis since the characteristics of sensors or data sets are not necessarily all the same. To account for this, we associate "data set reliability" with each temporal data set so that a less reliable data set has less effect on the global fusion of local decisions. Furthermore, since a certain class or a subset of classes is discriminated more successfully than the others, it will also be useful to associate a reliability factor as well to the individual class decisions which the local classifier makes. In this paper, the reliability factor associated with each class decision is called the "classwise reliability."

We propose two different multitemporal classifiers based on the decision fusion. The first one is based on the idea similar to that used by Tang et al.[10] in the M-ary detection. For the second approach, we modify the cost function used by Tang et al. to obtain the weighted majority decision fusion classifier, which can handle both the data set and classwise reliabilities.

II. MULTITEMPORAL CLASSIFICATION

Suppose there are p co-registered multitemporal remotely sensed data sets taken over the same location. The objective of designing a multitemporal classifier is to determine the optimum decision rule for a (global) class decision u_0 of a sample temporally observed as $\{x_1,..., x_p\}$ where x_k is the observation made at kth time, $k = 1, \ldots, p$. The decision u_0 is made among a set of M_0 user-defined information classes, $\Omega_0 \equiv \{\omega_1,..., \omega_{M_0}\}$. Information class is a class which is directly informational to the user according to the specific purpose of data analysis [11]. Since information classes are not necessarily separable in the feature space, data sets are usually analyzed in the training stage, for example, through a clustering, to find a mutually exclusive and exhaustive set of sub-classes or spectral (or data) classes so that each of which can be modeled by an appropriate probability density function [12]. Due to the computational complexity and a practical limitation on the requirement of training samples and so on, the data sets are assumed, in general, to be class-conditionally independent to each other (see [5] for the discussions on this assumption) and each data set is separately analyzed in the training stage. Therefore a data set at different times has generally a distinct set of spectral classes. Let us denote Ω_k as a set of M_k spectral classes in the kth temporal set, $k = 1, \ldots, p$. The local class decision made using the kth temporal set is denoted by u_k which is chosen among Ω_k.

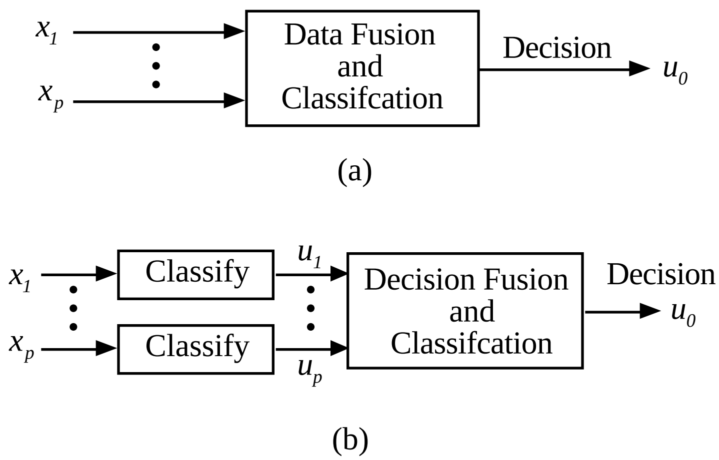

Figure 1. Classification structures. x_k is the feature vector in the kth data set and u_k is the local class decision determined using only x_k. (a). Fusion of features. (b). Fusion of decisions.

If each temporal data set is taken as a separate information source, multitemporal classification can be considered as an example of multisource data classification which has conceptually two different approaches. One category is the data fusion approach shown in Fig.1(a) in which the feature vectors of the data sources (or sensor) are given to a central decision procedure which makes the final decision u_0 among Ω_0 (see [4] for a detailed review on the works in this category). Note that the optimal Bayesian multitemporal decision rule, which is in this category, chooses a class $\omega_j \in \Omega_0$ which maximizes $P\{u_0 = \omega_j | x_1, \ldots, x_p\}$. Fleming and Hoffer [13] used the stacked vector $\{x_1, \ldots, x_p\}$ as an extended feature in the maximum likelihood classifier. However, the increased dimensionality of the feature vector requires more training samples than ordinarily required by the *p* pixelwise classifiers altogether. P. H. Swain [3] simplified the approach to derive the cascade classifier by assuming class-conditional independence of feature vectors of different data sets, thus deriving the decision rule of finding a class ω_j maximizing $P(\omega_j)P(x_1|\omega_j)...P(x_p|\omega_j)$. Note that in this approach, each data set has the same effect on the final decision of $u_0 = \omega_j$. By the way, Kalayeh and Landgrebe [14] proposed a multitemporal classifier which could utilize the temporal interpixel correlation context under the assumption that the class did not change over time. Since they assumed the same set of (spectral) classes for each temporal data set, all given multitemporal data sets should be processed together in the training stage to define the spectral classes, thus increasing the total number of necessary spectral classes. This increase is due to the constraint that the classes do not change over the time. In the multisource classification contexts, T. Lee *et al.* [5] developed a statistical multisource classifier which was later extended by J. A. Benediktsson *et al.* [4] to accommodate reliability factors associated with data sets. The global membership function in [4] is defined for $\omega_j \in \Omega_0$, $j = 1, \ldots, M_0$, as,

$$F_j(x_1,..., x_p) = P(\omega_j)\prod_{k=1}^{p}\left[\frac{P(\omega_j|x_k)}{P(\omega_j)}\right]^{\alpha_k} \tag{1}$$

A decision is made by selecting the class $u_0 = \omega_j$ which has a maximum membership function value. Eq.(1) shows how the individual weighted posterior probability affects the global membership function where α_k is the reliability factor associated with the kth data set, $k = 1, \ldots, p$. Note that the cascade classifier [3] is equivalent to eq.(1) if the data set reliabilities are all set to one. The evidential reasoning approach [15] has been also used to perform multisource classification with data set reliabilities. However, neither of these approaches utilize classwise reliabilities.

The second category shown in Fig.1(b) is the decision fusion approach in which a final class decision is made by summarizing only the class decisions of each data set. The key issue of this approach is two-fold; one is the design of the local classifiers and the other is the optimum fusion of local class decisions. This problem is very similar to that of M-ary distributed hypothesis testing which has recently received considerable research attention [6-10] in such fields as radar and military surveillance systems. J. Tubbs and W. Alltop [16] considered a problem of integrating classification results from multiple sensors and suggested a decision process based on a ranked lists of class decisions. R. R. Tenny and N. R. Sandell [6] first proposed a distributed detection algorithm in the case of two sensors. Z. Chair and P. K. Varshney [7] derived an optimum fusion rule when binary local decisions were given in a multiple sensor detection problem. Later, A. R. Reibman and L. W. Nolte [8] reported a system-wide optimum solution for a restricted case when the statistics and thresholds of the local detectors are assumed to be identical. Z. Tang *et al.* [10] presented a solution to the more general case of a distributed M-ary detection problem with multiple sensors.

III. DECISION FUSION IN MULTITEMPORAL CLASSIFICATION

A. Derivation of the Jointly Likelihood Decision Fusion Rule

Suppose a conventional pixelwise classifier such as the maximum likelihood(ML) classifier makes local class decisions independently using each separate temporal data set. Then the problem in the decision fusion-based multitemporal classification is how to make an optimum global decision u_0 among Ω_0 given the local decisions $\{u_1,..., u_p\}$. To find a decision fusion method which is optimum in the minimum expected cost sense, we define $J(u_0;u_1,..., u_p,\omega_j)$, the cost incurred by determining u_0 given $\{u_1,..., u_p\}$ when the true class is ω_j. The expected cost is written as,

$$E\{J(u_0;u_1,..., u_p,\omega_j)\} = \sum_{\omega_j \in \Omega_0} J(u_0;u_1,..., u_p,\omega_j) P\{u_1,..., u_p,\omega_j\} \qquad (2)$$

Tang *et al.* [10] defined a cost function $J(u_0;u_1,..., u_p,\omega_j) = J(u_0;\omega_j) \equiv [1-\delta(u_0,\omega_j)]$ where $\delta(u_0,\omega_j) = 1$ only if $u_0 = \omega_j$, and 0, otherwise. Employing the same cost function in eq.(2) leads to the decision fusion rule of choosing a class $u_0 \in \Omega_0$ maximizing $P\{u_1,..., u_p,u_0\}$. This implies finding a class u_0 that is most likely to occur jointly with the local decisions $\{u_1,..., u_p\}$.

Note that there are total $M_0 \cdot M_1 \cdots M_p$ different class combinations for $P\{u_1,..., u_p, u_0\}$. In this paper, we assume conditional independence of u_k's given u_0, that is,

$$P\{u_k|u_{k-1},..., u_1, u_0\} = P\{u_k|u_0\}, \tag{3}$$

where $P\{u_k|u_0\}$ is the probability that the local decision of a sample using kth data set is u_k given the global decision u_0 which is made by summarizing all the p local decisions. Accounting for class dependency using some models such as the Markov chain might be effective in improving the classification accuracy but with increased computational requirement [17]. Although whether u_k's are truly conditionally independent or not should be scrutinized including in the practical point of view, the assumption is made here just to make the classifier as simple as possible. Under the assumption in eq.(3), the simplified decision fusion rule optimal in Bayesian minimum cost sense is to choose $u_0 \in \Omega_0$ maximizing $H_{TP-LIK}(u_0)$ defined as,

$$H_{TP-LIK}(u_0) = P\{u_0\}\prod_{k=1}^{p} P\{u_k|u_0\} \tag{4}$$

We call this the jointly likelihood decision fusion multitemporal classifier. Since the local class decision u_k, $k = 1, \ldots, p$, can have any of M_k spectral classes in Ω_k, the total number of conditional probabilities, $P\{u_k|u_0\}$'s, to be estimated amounts to $M_0(M_1 + \cdots + M_p)$. With large number of spectral classes, their estimation is never an easy task since it is expected to be very rare in general to find all the cases of class combinations (u_k, u_0) in the training data set. One practical solution may be to assume a certain model of class transition, possibly aided by some data-specific knowledge such as development stages of crop over times, etc. For example, one might expect, possibly by analyzing the training samples, which subset of the spectral classes a sample of a certain information class is most likely classified to. From this kind of prior knowledge, one can get a practically reasonable model of the probability $P\{u_k|u_0\}$. In the experiment of this paper, a simple model of eq.(12) and (13) is used in applying the decision fusion rule in eq.(4).

B. Derivation of the Weighted Majority Decision Fusion Rule

Among the local decisions $\{u_1,..., u_p\}$, some of the decisions are more dependable in terms of reliability than others. In this case, it would be desirable if the final decision u_0 is as consistent as possible with those reliable local decisions. However, the classifier in eq.(4) cannot accommodate a disparate degree of reliabilities because the cost function $J(u_0;\omega_j)$ checks only the match of u_0 with ω_j.

To implement the idea of honoring reliable local decisions among $\{u_1,..., u_p\}$, a slight modification is made to the cost function. Specifically, a cost function $J(u_0;u_1,..., u_p,\omega_j)$ satisfying following relation is examined.

$$J(u_0;u_1,..., u_p,\omega_j) = \sum_{k=1}^{p} J(u_0;u_k,\omega_j) \tag{5}$$

$J(u_0;u_k,\omega_j)$ is a local cost function associated with the k^th data set, and it determines the cost given to an action of selecting u_0 based on the local decision u_k. A summation of all the local costs is then the actual cost assigned to the action of selecting u_0 based on $\{u_1,..., u_p\}$. We select a cost function $J(u_0;u_k,\omega_j)$ satisfying,

$$J(u_0;u_k,\omega_j) \ = \ J(u_0;u_k) \equiv 1 - A_k(u_k)\delta(u_0,u_k) \tag{6}$$

where $0 \le A_k(u_k) \le 1$. $A_k(u_k)$ is a number associated with the local class decision u_k. It can control relative importance of consistency between u_0 and local decision u_k because the cost of selecting the global decision u_0 to match with the local decision u_k is $1 - A_k(u_k)$ while the cost in other cases is one. According to the data set and classwise reliabilities, we select appropriate values of $A_k(u_k)$ in such a way that a less reliable local decision has limited effect on making a final decision u_0 through the selected fusion rule. Using the cost function in eq.(6), the expected cost is given as,

$$E\{J(u_0;u_1,..., u_p,\omega_j)\} = P(u_1,..., u_p)\left\{ p - \sum_{k=1}^{p} A_k(u_k)\delta(u_0,u_k) \right\} \tag{7}$$

It is minimized if the second term in parenthesis is maximized with respect to the decision u_0. We define a multitemporal decision fusion classifier that chooses a class $u_0 \in \Omega_0$ maximizing $H_{TP-WHTM}(u_0)$, defined as,

$$H_{TP-WHTM}(u_0) \equiv \sum_{k=1}^{p} A_k(u_k)\delta(u_0, u_k) \qquad (8)$$

To better appreciate the role of $A_k(u_k)$'s, suppose they are all 1. Then, the classifier of eq.(8) would choose a class u_0 which is a majority class among u_k's, k=1, …, p. Therefore it is a majority rule. With distinct $A_k(u_k)$'s, then, the "vote" of each local decision u_k is weighted according to $A_k(u_k)$. Thus the classifier will select a class u_0 attaining the most weights of $A_k(u_k)$'s. For this reason, this decision fusion method will be called as the "weighted majority decision fusion rule."

C. Data Set and Classwise Reliability

Since the decision made by each temporal data set has different reliability, we define the classwise reliability, $rel(k, u_k)$, $u_k \in \Omega_k$, k = 1, …, p, as a reliability of a decision u_k using the kth temporal data set. In the same way, the data set reliability, $REL(k)$, k = 1, …, p, denotes the reliability of the kth data set as a whole. It would be very logical to assign a large cost to the case when the fusion rule fails to follow a local decision of large reliability.

Three different measures of data set reliability using class separability, equivocation, and association are introduced in [4]. Although statistical separability between classes is a good candidate for assessing data set reliability, the computation involved in evaluating separability could be non-trivial if the multivariate normality assumption about the data set is not satisfied. In the context of equivocation, the data set reliability is related to the degree that a data (or spectral) class indicates a specific information class. Since the purpose of decision fusion in this paper is classification, any data set with higher classification accuracy may be assumed more reliable than the others. Note that classification accuracy can be easily obtained irrespective of assumptions on underlining probability density functions. How to estimate or associate proper reliability to each data set is an important issue which needs more attention [4]. In the experiment of the proposed classifiers in this paper, however, instead of attempting to estimate optimal values of the data set reliabilities, we simply test several different values of them to monitor their effect on classification.

In case of the classwise reliability, one can specify that it must be also large for the reliable local decision so that the global decision can be biased as much as possible to the reliable local one. Although there is no definitive measure of how much reliable a specific local decision is, we consider two measures; one is based on the detection probability, and the other on the classwise probability of correct classification. The first one is defined as, for $u_k \in \Omega_k$ and $\omega_j \in \Omega_0$, $k = 1, \ldots, p$ and $j = 1, \ldots M_0$,

$$rel(k, u_k = \omega_j) \equiv P(u_k = \omega_j | x_k \in \omega_j) \, . \quad (9)$$

It is nothing but the detection probability of class ω_j which is the probability of correctly classifying a sample x_k as belonging to a class ω_j. Its rationale is that any class with high detection probability should be reliable. By the way, there can be a problem in using this measure as explained in following hypothetical example. Suppose a local classifier which is very poorly designed or, whose feature vectors of a certain data set are of very poor quality assigns all samples to a particular class. In this case, the measure of eq.(9) will assign the highest reliability of 1 to that particular class, although the decision to this class is meaningless. On the contrary, the second measure defined as,

$$rel(k, u_k = \omega_j) \equiv P(x_k \in \omega_j | u_k = \omega_j) \quad (10)$$

does not have this kind of problem. It is the probability that a sample x_k is truly from the class ω_j when the local decision u_k is ω_j, that is, eq.(10) is the probability that the local decision is correct. In the experiments, we have tested these two measures of classwise reliability to observe that the measure of eq.(10) is more effective. It is something expected since high value of eq.(10) implies that, statistically speaking, the local decision is most likely correct, therefore the local decision u_k is better to strongly influence the global class decision of x_k .

There still remains a problem in associating the data set and classwise reliability measures to actual values of weights $A_k(u_k)$'s. Since it appears difficult to do optimally, at least for now, a seemingly simple way as given by eq.(11) is used in the experiment.

$$A_k(u_k) = REL(k) \times rel(k, u_k) \quad (11)$$

In implementing the fusion rule in eq.(8), one needs $A_k(u_k)$'s as prior information, equivalently, the data set reliabilities $REL(k)$'s, and the classwise reliabilities $rel(k,u_k)$'s. Note that compared to the jointly likelihood fusion rule in eq.(4), the weighted majority rule is applicable with much reduced prior information. The classwise reliabilities either in the form of eq.(9) or (10) can be estimated easily from the classification results of representative training samples.

VI. EXPERIMENTS AND DISCUSSIONS

A. Multitemporal Data Sets and Training

To test the proposed decision fusion approaches in multitemporal classification, experiments are carried out using three multitemporal Landsat Thematic Mapper (TM) data sets acquired over the same agricultural areas in Tippecanoe County, Indiana respectively in April, July, and September, 1986. Ground truth data were gathered in July. All seven bands are used in the classification. Based on the available ground truth data, four information classes in Ω_0 = {corn, soybean, wheat, alfalfa/oat} are defined in the July and September data sets. Several sub-classes are developed, *separately* for each data set, for each of the information class in Ω_0 to satisfy the multivariate normality assumption of the sub-classes. Since the class "wheat" is the only green crop type observed in the April data set, and green vegetation has a substantially different spectral reflectance compared to the soil [3], it can be identified with a relatively high accuracy. For this reason, only two information classes -wheat and "the others"- are defined in the April data set. About 15,000 samples are chosen for test in each data set, and about 10% of the samples are randomly selected for training.

Initially, the pixelwise maximum likelihood(ML) classifier classified each temporal data set separately, and the *overall* (OVA) and *class-averaged classification accuracy* (CAG) are used as references in evaluating the classification performance of the proposed temporal classifiers. The final (global) decision using all three multitemporal data sets are made among Ω_0. Note that the jointly likelihood decision fusion rule in (4) requires the class transition probabilities $P\{u_k|u_0\}$'s. Since u_k indicates a class among the set of spectral classes, Ω_k, and the global decision u_0 is made among the set of information classes, Ω_0, there are cases where more than one classes of u_k correspond to the same information class u_0. In the experiment of this paper, the probabilities $P\{u_k|u_0\}$'s are determined heuristically in such a way that the probability $P\{u_k|u_0\}$ of a local decision u_k which belongs to the same information class as the global one u_0 has a higher probability than the other cases: suppose the local decision u_k corresponds to an information class $u_0 = \omega_j$, and there are total n sub-classes(spectral classes) in Ω_k which corresponds to the information class ω_j, then, we define, for $u_k \in \Omega_k$, $\omega_j \in \Omega_0$, k= 1, ---, p,

$$P\{u_k|u_0 = \omega_j\} = P_o / n \tag{12}$$

When the local decision u_k does not belong to the information class indicated by u_0, then,

$$P\{u_k|u_0 = \omega_j\} = (1 - P_o) / (M_k - n) \tag{13}$$

where P_o is a user defined number between zero and one: P_o being one means no allowance of a class transition to a different information class. If P_o is zero, class transition is permitted only to a different information class. Several values of P_o are tested and some of the results are shown in Table 1. The values of P_o which achieve the best performance are chosen for comparison with other classifiers. In the case of classifying July data with April data, P_o is set to one.

Table 1 Bi-temporal Classification of July Data with September Data with Different Class Transition Probabilities(Equal Data set reliability).

(%)

P_o in eq.(10)	Corn	Soybeans	Wheat	Alfalfa/Oats	CAG	OVA
0.80	91.02	59.42	66.34	78.55	73.83	75.20
0.99	92.03	60.28	64.12	78.92	73.84	75.78
1.00	89.84	60.03	61.40	79.43	72.68	74.55
Best[1]	91.80	63.08	69.88	73.35	74.53	76.67

[1] P_o = 0.99 for corn and soybeans, P_o = 0.8 for wheat, P_o = 1 for alfalfa/oats.

B. Multitemporal Classification with Data Fusion

For comparison, the conventional multitemporal classifier based on the data fusion in eq.(1) classified the July data with April and September data and its results are shown in Table 2. Since the ground truth was gathered in July and matched best with July data, all comparisons are made with respect to the July data set. Several different data set reliability factors (from 0.6 to 1.0 at a step of 0.1 for each data set) are tested to see their effect on classification. The overall classification accuracy, however, is observed not to vary much; the maximum deviation due to different data set reliabilities is less than 1% (the result reported in Table 2 used REL(APR)=REL(JUL)=REL(SEP)=0.6). The multitemporal classification based on the data fusion in eq.(1) generally attains better results than any of the single pixelwise maximum likelihood classification; when the September and April data sets are used together with the July data set, the increase of overall classification accuracy over the pixelwise ML classifier is about 6%. Including April data improves the classification accuracy of wheat and alfalfa/oats significantly. The September data set is also helpful in classifying the class soybeans in July data, but there is a slight degradation in classification accuracy of the alfalfa/oats (see JUL+SEP case). The improvement due to including September data in classifying July data is seen to be marginal in the data fusion approach. Note that the classification accuracy in September data is generally very low except for the class corn.

Table 2. Classification Accuracy Comparison of the Multitemporal Classifiers

Data Sets	Percent Classification Accuracy (%)						
	Corn	Soybeans	Wheat	Alfalfa/ Oats	CAG	OVA	Increase in OVA[1]
	Separate maximum likelihood classification of each temporal data set						
April	89.59[2]		90.29		89.94	89.65	-
September	82.59	55.06	51.28	47.07	59.00	65.28	-
July	90.18	57.72	68.72	77.89	73.63	74.37	-
	Classifier based on data fusion in eq.(1)						
JUL+APR	90.29	56.42	86.50	83.16	79.09	76.16	1.79
JUL+SEP	92.30	64.63	69.22	67.72	73.47	76.80	2.43
JUL+APR+SEP	92.52	65.56	88.07	79.50	81.41	80.23	5.86
	Jointly likelihood decision fusion rule in eq.(4)						
JUL+APR	90.18	57.72	89.96	80.82	79.67	76.67	2.30
JUL+SEP	94.19	75.63	68.72	73.79	78.08	82.24	7.87
JUL+APR+SEP	95.79	77.08	88.89	71.52	83.32	85.10	10.73
	Weighted majority decision fusion rule in eq.(8)[3,4]						
JUL+APR	92.61	57.72	84.53	73.50	77.09	76.43	2.06
JUL+SEP	96.83	76.89	64.44	51.46	72.41	81.08	6.71
JUL+APR+SEP	97.18	76.77	75.23	64.86	78.51	83.60	9.23

[1]increase of OVA over the pixelwise ML classification of July data set only.

[2]In classifying April data with the pixelwise ML classifier, there were only 2 information classes {wheat, others}. Classification accuracy of "others" is given under corn.

[3]Equal data set reliabilities, REL(APR)=REL(JUL)=REL(SEP)=1.0 are used.

[4]Eq.(10) is used for the classwise reliability.

C. Multitemporal Classification with the Jointly Likelihood Decision Fusion

Table 2 also shows the classification accuracy of the proposed two multitemporal classifiers - the jointly likelihood decision fusion rule in eq.(4) and the weighted majority decision fusion rule in eq.(8). In the decision fusion approach, only limited information of local class decisions are combined compared to the data fusion which combines posterior probabilities. However, the jointly likelihood decision fusion rule is seen to perform much better than the data fusion approach in eq.(1). Therefore the *a priori* information of the joint probability, $P\{u_1, u_2, ..., u_p, u_0\}$ required by the jointly likelihood decision fusion rule is found to be effective in combining information for classification. Under the conditional independence assumption of $P\{u_k|u_{k-1},..., u_1,u_0\} = P\{u_k|u_0\}$, it is sufficient to estimate M_k times M_0 conditional probabilities of $P\{u_k|u_0\}$'s for the k^th data set where M_k is the number of classes in the k^th data set and M_0 is the number of classes in Ω_0. As stated before, although these class transition probabilities can be estimated from the training samples, we choose to use the probability values calculated according to the model in eq.(12) and (13) in the experiment to avoid additional complexity of their estimation.

In classifying July data with September and April data, the jointly likelihood decision fusion rule achieves approximately 5% overall classification accuracy increase over the best data fusion multitemporal classifier. The increase of classification accuracy is especially significant for the soybeans class (~ 11%), however, the class "alfafa/oats" experiences about 8% loss of classification accuracy, thus the increase of the class-averaged classification accuracy over the data fusion amounts only to 2%. Compared to the ML pixelwise classification of July data only, the jointly likelihood decision fusion achieves 10.73% of overall classification accuracy increase.

D. Multitemporal Classification with the Weighted Majority Decision Fusion

Classification results with the weighted majority decision fusion rule are presented in Table 2 as well (the results shown in Table 2 are obtained with the same data set reliabilities - this corresponds to the simple majority rule; see also Table 3). Note that the weighted majority fusion rule of eq.(8) further reduces the requirement for prior information: only M_k different classwise reliability factors, $rel(k,u_k)$'s, are sufficient for the k^th data set. In the experiment, instead of estimating the data set reliabilities, several different values (0 ~ 1.0) are assigned to each data set to see only minor differences in overall classification accuracy (less than 1.3%). However, the class-averaged accuracy is observed to be somewhat sensitive to the data set reliability as can be seen in Table 3 (see the classes, wheat and alfafa/oats).

Table 3. Weighted Majority Decision Fusion with Different Data Set Reliabilities

Data Set Reliability			Percent Classification Accuracy (%)					
APR	JUL	SEP	Corn	Soybeans	Wheat	Alfalfa/Oat	CAG	OVA
1.0	0.6	0.6	92.50	78.11	74.49	72.04	79.28	82.77
1.0	0.7	0.7	96.85	77.23	74.07	72.04	80.05	84.27
0.6	1.0	0.6	96.55	76.24	87.49	62.45	80.68	84.03
0.7	1.0	0.7	96.87	76.24	78.11	64.86	79.02	83.54
0.6	0.6	1.0	92.50	78.11	74.49	72.04	79.28	82.77
0.7	0.7	1.0	96.85	77.23	74.07	72.04	80.05	84.27
1.0	1.0	1.0	97.18	76.77	75.23	64.86	78.51	83.60

In case of the data fusion scheme in eq.(1), maximum difference in the classification accuracy (both in CAG and OVA) with different data set reliabilities is less than 2%. However, the data set reliability determine $A_k(u_k)$ in eq.(8) and there is no other data-dependent quantity except the classwise reliability, therefore, the classification accuracy seems to be more sensitive to the data set reliability in the weighted majority rule. This indicates that estimation of optimum data set reliability (and the classwise one) is an important issue in applying the weighted majority decision fusion scheme. As for the classwise reliability, the measures in eq.(9) and (10) show significant differences in their performance; classwise reliability of eq.(10) produces significantly better result (about 6~8% improvement in OVA; 3~4% improvement in CAG) than that of eq.(9). This can be easily understood since the classwise reliability in eq.(10) indicates more directly the possibility of a local decision being true. The results in Table 2 are those using eq.(10). Note that although the weighted majority decision fusion rule requires much less prior information than the jointly likelihood decision fusion and it is much simpler than the data fusion based rule in eq.(1), it performs almost comparably with them at least in terms of overall classification accuracy, which suggests its usefulness as a simplified multitemporal classifier. But the experimental result also suggests need for further research on deciding optimum data set reliabilities to improve class-averaged classification accuracy as well.

Although there is further need for research on an optimum selection of data set and classwise reliabilities, multitemporal classifiers based on decision fusion proposed in this paper are observed to perform quite successfully compared to the non-contextual ML classifier, or the multitemporal classifier with feature level fusion. Note that data fusion-based multitemporal classifiers combine posterior probabilities of each data set and therefore, all data sets must be describable with statistical probabilities. If data sets are very diverse in terms of their statistical properties, combination of the posterior probabilities might not be able to produce desirable results since one data set with large ranges of probability values can easily dominate the global decision process. On the contrary, the decision fusion-based approach can be applied without such problems. With data set and classwise reliability, or the information about conditional probability $P\{u_k|u_0\}$'s, it is very straightforward to control the relative importance of a specific data set, or particular class decisions on the final global decision. Note that decision fusion approaches are computationally very simple and always applicable to classifying multitemporal data sets whenever the class decisions of each temporal data sets are available.

V. CONCLUSION

In this paper, two different multitemporal classifiers based on decision fusion are derived. The first one, namely the jointly likelihood decision fusion rule is found to give better result compared to the classifier based on the data fusion by about 5% in the overall classification accuracy and about 2% in the class averaged classification accuracy. The second one, the weighted majority decision fusion rule is shown to perform almost comparably with the data fusion and the jointly likelihood decision fusion rule at least in terms of the overall classification accuracy, even if it is supplied with much reduced a priori information compared to the jointly likelihood decision fusion rule, and it is much simpler than the data fusion rule. But the experimental result also suggests need for further research on deciding optimum data set reliabilities to improve the class-averaged classification accuracy as much as the overall classification accuracy.

In addition to its simplicity, the proposed weighted majority decision fusion rule has a feature of handling not only the data set reliabilities but also the classwise reliabilities. Between the two different assessments of classwise reliabilities, the one based on the probability of correct classification in eq.(10), as expected, is found to be far more effective than the other. This decision fusion approach in multitemporal classification is very attractive since it satisfies all three requirements of a multitemporal classifier stated in the Section I, and one can apply the method even to the problems in which a certain data set cannot be modeled by known probability density functions.

REFERENCES

[1] D. A. Landgrebe, "The evolution of Landsat Data Analysis," Photogrammetric Engineering and Remote Sensing, vol.63, No.7, pp.859-867, July 1997.

[2] F. Wang, "A knowledge-based vision system for detecting land changes at urban fringes," IEEE Trans. Geosci. Remote Sens., vol.31, No.1, pp.136-145, Jan. 1993.

[3] P. H. Swain, "Bayesian classification in a time-varying environment," IEEE Trans. Systems, Man, and Cybernetics, Vol.8, No.12, pp.879-883, Dec. 1978.

[4] J. A. Benediktsson, P. H. Swain and O. K. Ersoy, "Neural network approaches versus statistical methods in classification of multisource remote sensing data," IEEE Trans. Geosci. Remote Sens., vol.28, No.4, pp.540-552, July 1990.

[5] T. Lee, J. A. Richards and P. H. Swain, "Probabilistic and evidential approaches for multisource data analysis," IEEE Trans. Geosci. Remote Sens., vol.GE-25, pp.283-293, May 1987.

[6] R. R. Tenny and N. R. Sandell, Jr., "Detection with distributed sensors," IEEE Transactions on Aerospace and Electronic Systems, AES-17, pp.501-510, July 1986.

[7] Z. Chair and P. K. Varshney, "Optimum data fusion in multiple sensor detection systems," IEEE Transactions on Aerospace and Electronic Systems, AES-22, pp.98-101, 1986.

[8] A. R. Reibman and L. W. Nolte, "Optimal detection and performance of distributed sensor systems," IEEE Transactions on Aerospace and Electronic Systems, AES-23, pp.24 -30, 1987.

[9] I. Y. Hoballa and P. K. Varshney, "Distributed Bayesian signal detection," IEEE Transactions on Information Theory, Vol.35, No.5, pp.995-1000, Sept. 1989.

[10] Z. Tang, K. R. Pattipati and D. L. Kleinman, "A distributed M-ary hypothesis testing problem with correlated observations," Proceedings of the 28th IEEE Conference on Decision and Control, Tampa, Florida, pp.562-568, Dec. 1989.

[11] P. H. Swain, "Fundamentals of Pattern Recognition in Remote Sensing," Remote Sensing - The Quantitative Approach, edited by P. H. Swain and S. Davis, McGraw-Hill Book Company, New York, 1978.

[12] K. Fukunaga, *Introduction to Statistical Pattern Recognition,* 2nd edition, Academic Press, New York, 1990.

[13] R. M. Hoffer, M. D. Fleming, L. A. Brrtolucci, S. M. Davis, R. F. Nelson, "Digital processing of Landsat MSS and topographic data to improve capabilities for computerized mapping of forest cover types," LARS Technical Report 011579, Purdue University, West Lafayette, IN 47907, 1979.

[14] H. M. Kalayeh and D. A. Landgrebe, "Utilizing multitemporal data by a stochastic model," IEEE Trans. Geoscience and Remote Sensing, Vol.24, No.5, pp.792-795, Sept. 1986.

[15] Hakil Kim and P. H. Swain, "Evidential reasoning approach to multisource data classification in remote sensing," IEEE Trans. System, Man, and Cybernetics, vol.25, pp.1257-1265, Aug. 1995.

[16] J. D. Tubbs and W. O. Alltop, "Measures of confidence associated with combining classification results," IEEE Trans. System, Man, and Cybernetics, vol.21, No.3, pp.690-692, May/June 1991.

[17] B. Jeon and D. A. Landgrebe, "Classification with spatio-temporal interpixel class dependency contexts," IEEE Trans. Geoscience and Remote Sensing, Vol.30, No.4, pp.663-672, July 1992.