Decision-analytic Interpretation of Remotely Sensed Data

B.G.H. Gorte
International Institute for Aerospace Survey and Earth Sciences (ITC)
P.O. Box 6, 7500 AA Enschede, The Netherlands
e-mail: `ben@itc.nl`

L.C. van der Gaag
Utrecht University, Department of Computer Science
P.O. Box 80.089, 3508 TB Utrecht, The Netherlands
e-mail: `linda@cs.ruu.nl`

F.J.M. van der Wel
Utrecht University, Faculty of Geographical Sciences — Cartography Section
P.O. Box 80.115, 3508 TC Utrecht, The Netherlands
e-mail: `f.vanderwel@frw.ruu.nl`

Abstract

Increasingly, remotely sensed data are used for taking decisions in geographical information systems. Decision making can in principle be based on a classification of such remotely sensed data into nominal information classes. Such a classification, however, typically includes an unknown amount of uncertainty. Moreover, when processing spatial data for decision making, not only the uncertainties inherent in these data but also the objectives and preferences of the decision maker have to be taken into account. This paper proposes exploiting concepts from the mathematical framework of decision analysis for integrating uncertainties and preferences. It aims to solve complex decision problems on the basis of remotely sensed data. The feasibility of the decision-analytic approach to the interpretation of spatial data is demonstrated by means of a case study.

Key words & phrases: remote sensing, GIS, decision analysis, preferences, uncertainty.

1 Introduction

To monitor, analyze and interpret developments in our changing environment, up-to-date spatial data are periodically collected and processed. Increasingly, remote sensing is considered as a valuable source for this purpose. It yields data that can be subjected to further analysis in a geographical information system (GIS) at advantageous average cost. By systematic application of spatial operations and visualisation, a GIS is able to generate, on request, derivative data sets contributing to making decisions involving characteristics of spatially-related phenomena of the environment.

Classification of remotely sensed data into qualitative information classes is useful to extract information from the spectral attributes of these data, yielding an insightful representation of the real world. Such a representation can be exploited directly as a thematic map or as part of a time series in a change detection application. Unfortunately, classification generally introduces unknown uncertainty in the information classes assigned to the spectral objects. This uncertainty propagates through the subsequent stages of the decision making process [Lunetta *et al.*, 1991]. The uncertainty can be reduced by using evidence with regard to the real world, usually derived from sources such as domain experts, maps, field work, aerial photographs, or thematic maps from former classifications. Such evidence can be exploited before, during, and after classification and hence contribute to the accuracy of the final results in various different ways [Strahler, 1980]. Despite all efforts to reduce the uncertainty introduced by classification, it always influences the results. These imperfections may seriously affect the adequacy of using classification results for taking environmental decisions. For example, the commonly used *maximum a posteriori probability classification* discards useful information that may serve to yield insight in the uncertainties. In this approach to classification, the posterior probabilities that are computed for each spatial object within an information class distinguished during sampling, are used *only* to select the most likely class. The entire probability distribution for the object, however, reflects highly valuable information about the *extent* and *distribution* of uncertainty which could be further utilised in a GIS.

If decisions are to be made on the basis of remotely sensed data, uncertainty tells only part of the story: the *objectives* to be pursued with interpretation of the data become crucial. In the presence of uncertainty, the best decisions are those that, in view of the objectives, carefully weigh the *benefits* of *correct* interpretation of the data on the one hand and the *losses* due to *incorrect* interpretation on the other hand. This idea is illustrated by an example dealing with fraud with subsidies assigned to agricultural crops by the European Union. In this example, the main objective is to detect illegal declarations of subsidised crops by taking remotely sensed images from crops on parcels, to avoid waste of public resources. From this objective alone, the number of detected illegal declarations should be maximised. However, unjust implication of fraud is highly unfavourable as it results in extra costs for verification and in loss of face. Therefore, the number of unjust implications should be kept at a minimum. In pursuing both objectives simultaneously, overlooking fraud is considered worse than over-estimating. It now depends on the probabilities computed for the various possible crops for a parcel under consideration whether or not fraud should be implied. Interpretation of remotely sensed data for decision making therefore involves both the extent and distribution of uncertainty introduced by classification and the preferences of the decision maker. These preferences concern the objectives that are being pursued with the interpretation and therefore differ from knowledge about the *subject* of the interpretation as referred to by [Strahler, 1980]. Both types of knowledge equally contribute to the interpretation, yet at different levels.

Further elaborating on the idea that remotely sensed data can serve as a basis for decision making, the question arises whether or not it is necessary to derive a complete classification before considering viable decisions. In principle, decisions can be taken on the basis of a classification. However, classification contains uncertainty of which the extent and distribution are unknown. By making decisions directly on the data, full knowledge about the uncertainties involved can be included, thereby allowing for making better decisions. As decision making does not so much involve classification results as the extent and distribution of uncertainty introduced by classification, deriving a complete classification is no longer required and, in fact, has become obsolete (Figure 1). However, an accurate classification nevertheless serves various purposes beyond decision-making.

This paper addresses the interpretation of remotely sensed data in view of the objectives that

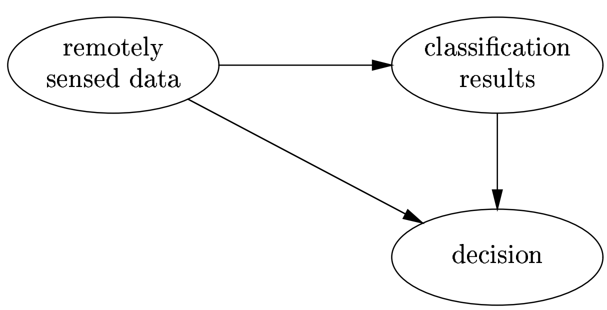

Figure 1: Founding Decision Making on Data

are to be pursued when exploiting the data for decision making. To this end, various concepts from decision analysis are introduced which allow integration of uncertainties and a decision-maker's preferences. Section 2 expresses the interpretation of remotely sensed data as a decision problem and introduces the mathematics for solving this problem. Section 3 describes the assessment of the various parameters involved in quantification of uncertainties and preferences. A case study will be presented in Section 4, demonstrating the feasibility of the decision-analytic approach.

2 Interpretation of Data: a Decision Problem

Interpretation of remotely sensed data is in essence a *decision problem*: the problem is to decide upon which decision to take for each spatial object on the basis of available data. The solution to this problem is for each object the decision that is expected to best meet the objectives that are being pursued with the interpretation. The field of *Decision analysis* provides the mathematical framework for solving complex decision problems such as the data-interpretation problem [Raiffa, 1968, von Winterfeldt & Edwards, 1986, Smith, 1988]. It offers means for structuring decision problems and for computing solutions. In this section, we express the interpretation problem and its solution in decision-analytic terms.

A decision problem involves two types of variable:

- a *decision variable* is a variable that represents viable decisions or actions that can be taken in the context of the problem at hand;
- a *chance variable* is a variable that represents the true 'state of the world'; the value of such a variable cannot be selected by the decision maker.

In the data-interpretation problem, there is only one variable of each type: a chance variable C that represents the true information class of a spatial object O and a decision variable D that represents the possible decisions that can be taken with regard to this object.

A variable in a decision problem can take its value from among a pre-defined set of values. We assume that $C_1, \ldots, C_n$, $n \geq 1$, are possible information classes of O. These classes therefore are values for the chance variable C. We further assume that the decision variable D takes its value from among the decisions $D_1, \ldots, D_m$, $m \geq 1$.

In a decision problem, there typically is uncertainty regarding the true values of the chance variables involved. In data interpretation, there is uncertainty concerning the true value of the chance variable C since the true information class of O is unknown at the time of interpretation. This uncertainty is expressed as a probability distribution $\Pr(C)$ for the variable C, specifying for every possible information class C_i the probability $\Pr(C = C_i)$ that C_i is the true class of the object. Note that this probability distribution will not be influenced by the various decisions that can be taken.

In addition to uncertainties, a decision problem involves preferences. The desirability of a decision and its consequences, with each other called a *scenario*, is quantified by means of its *utility*. In our data-interpretation problem, each combination of a decision $D = D_i$ and a true information class $C = C_j$ has associated a utility $u(D = D_i \wedge C = C_j)$. The utility expresses the desirability of the scenario where the decision D_i is taken with regard to O while it has C_j as its true information class. Actual utilities associated with the various scenarios depend upon the objectives that are being pursued with the interpretation.

Structuring all aspects of a decision problem can be done with a *decision tree*. A decision tree is a pictorial, tree-like representation of the problem. The various variables and values of the problem are organised in a (rooted) tree. Each node in the tree models a variable; the edges emerging from a node represent the values of its associated variable. The topological structure of the tree is an explicit representation of all scenarios that can possibly arise from a decision. The root node of the tree represents the initial situation before any decision is taken and each path from the root node to the tip of a terminal edge corresponds with a scenario. Figure 2 shows a tree organising the variables of our object-interpretation problem. To distinguish between the decision and chance variable, the former is depicted as a square box and the latter is shown as a circle. In the tree, the uncertainties concerning the chance variable's values are depicted with the appropriate edges; the utilities are depicted at the tips of the terminal edges of the tree.

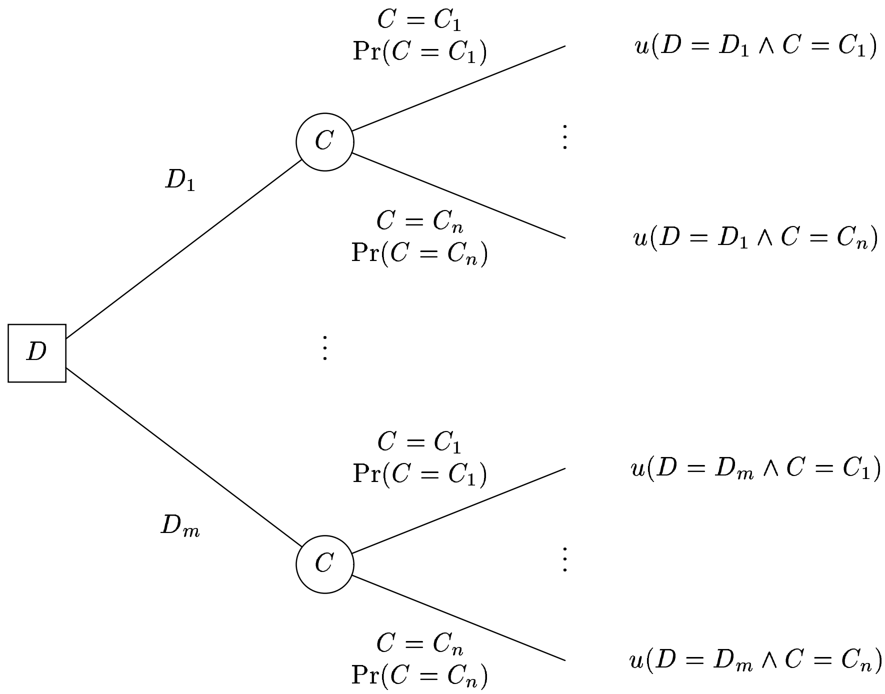

Figure 2: A Decision Tree for the Data-interpretation Problem.

Once a decision problem has been structured in a decision tree, the best decision for the problem is easily computed. For this purpose, the tree is evaluated by *foldback analysis.* Foldback analysis starts at the tips of the terminal edges, works its way through all intermediate nodes and edges, and ends at the root of the tree. In foldback analysis, for each viable decision the desirability of taking this decision is computed. The desirability of a decision depends on the values of the chance variables modeling its consequences. However, these values are not known before the decision is taken. The desirability of a decision therefore is computed by *weighting* the utilities of the various possible scenarios that can arise from taking this decision with the probabilities that these scenarios actually do occur. For each chance variable, the *expected utility* over its values is computed, which expresses the expected utility of taking the decision corresponding with the incoming edge of the node modeling the chance variable. For each decision variable, the *maximum expected utility* over its values is computed. In a foldback analysis of the decision tree for the data-interpretation problem, the expected utility $\hat{u}(D = D_i)$ for each decision $D = D_i$ is computed as

$$\hat{u}(D = D_i) = \sum_{j=1}^{n} u(D = D_i \wedge C = C_j) \cdot \Pr(C = C_j)$$

The best decision is the decision D_k with the highest expected utility. Computing the best decision with regard to a spatial object O as outlined before will be coined *decision-analytic data interpretation.*

The statistical description of decision analysis provides a general and flexible framework for data interpretation. In fact, the framework also provides for conventional classification by taking for the decision the various possible information classes; the utilities then express the severity of different types of misclassification. As an example, we express the common *maximum a posteriori probability classification*. The only objective pursued in maximum a posteriori probability classification is to maximise the probability of correct classification: every misclassification is considered equally undesirable. This objective can be expressed in terms of utilities by taking $u(D = D_i \wedge C = C_i) = 1$, for all $i = 1, \ldots, n$, and $u(D = D_i \wedge C = C_j) = 0$, for all $i, j = 1, \ldots, n, i \neq j$, where D_i is the decision to assign class C_i to the object O.

3 Assessing Parameters

For decision-analytic data interpretation, a decision tree to model the interpretation problem is evaluated. This decision tree includes the various uncertainties and preferences involved. The accuracy of the assessment of these quantities directly determines the quality of the decision computed for the problem. This section briefly addresses the assessment of the quantities required for the decision-analytic approach.

3.1 Probability Assessment

The uncertainties involved in data-interpretation are expressed as probability distributions over the various information classes distinguished for the spatial object O under consideration. The probabilities in these distributions are computed from remotely sensed data as posterior probabilities given the spectral attributes of these data. Given a vector $\mathbf{x}$ of spectral attributes, for each information class C_i, $i = 1, \ldots, n$, the posterior probability $\Pr(C = C_i \mid \mathbf{x})$ is computed using Bayes' formula:

$$\Pr(C = C_i \mid \mathbf{x}) = \frac{\Pr(\mathbf{x} \mid C = C_i) \cdot \Pr(C = C_i)}{\Pr(\mathbf{x})}$$

where $\Pr(\mathbf{x} \mid C = C_i)$ is the probability that the vector of spectral attributes $\mathbf{x}$ occurs in the data given that the true class of the object is C_i. $\Pr(C = C_i)$ is the prior probability that the object has class C_i for its true class and $\Pr(\mathbf{x})$ denotes the probability of the vector $\mathbf{x}$ occurring in the data. $\Pr(\mathbf{x})$ is the same for every information class and does not have to be computed independently: $\Pr(\mathbf{x})$ is obtained by normalising the nominators of the right-hand side of the formula over all information classes. The probabilities $\Pr(\mathbf{x} \mid C = C_i)$ and $\Pr(C = C_i)$, however, have to be assessed explicitly for each class C_i.

To assess $\Pr(\mathbf{x} \mid C = C_i)$ for C_i, $i = 1, \ldots, n$, either a parametric or a non-parametric method may be used. The *parametric method* builds on the assumption that each information class C_i yields a parametric, such as a Gaussian, distribution over the space of spectral attributes. To calculate the required probabilities, for each class the distribution is selected that best fits a set of training data; this distribution is obtained by estimating its parameters from these data, that is, for a Gaussian distribution, its mean and variance-covariance matrix. Alternatively, the *k-nearest neighbour method* is a *non-parametric method* for assessing the probabilities $\Pr(\mathbf{x} \mid C = C_i)$. It computes for the vector of spectral attributes $\mathbf{x}$ from the training data a set of neighbouring vectors. These vectors are selected by searching the space of spectral attributes at increasing distance from $\mathbf{x}$ until k vectors are included. From the set of neighbouring vectors thus selected, the number k_i of vectors corresponding with information class C_i is determined. Now, if n_i is the total number of training data corresponding with class C_i, the probability that the vector $\mathbf{x}$ corresponds with class C_i is estimated as

$$\Pr(\mathbf{x} \mid C = C_i) = \frac{k_i}{n_i \cdot V_k(\mathbf{x})}$$

where $V_k(\mathbf{x})$ is related to the distance at which $\mathbf{x}$'s k nearest neighbours are found [Fukunaga & Hummels, 1987]. In practice, the non-parametric method proves to be superior to the parametric method in the sense that it yields probabilities of higher accuracy. Especially in the context of our decision-analytic approach, the non-parametric method therefore is preferred.

The prior probabilities $\Pr(C = C_i)$ for the various information classes C_i generally are more difficult to obtain than the probabilities $\Pr(\mathbf{x} \mid C = C_i)$ as they are independent of the data to be interpreted and therefore require knowledge about the subject of the interpretation for their assessment. Depending on the available knowledge, they may be estimated by different methods with varying degrees of sophistication. The least sophisticated method builds on the assumption that any spatial object is fully characterised by its spectral attributes: this method assigns to each class C_i, $i = 1, \ldots, n$, the prior probability $\frac{1}{n}$. A more involved method assigns a prior probability to each class distinguished in the image based on knowledge about its percentage of coverage. Assessment of the prior probabilities based on local information is the most precise method. To this end, the image is sudivided into segments, for example derived from additional GIS-data. The prior probabilities subsequently are estimated per segment, either on the basis of available knowledge about processes that influence the occurence of classes in the area under consideration, or extracted from the image data, assuming accurate and representative sampling, by an iterative algorithm [Gorte, 1995].

3.2 Utility Assessment

The utilities of a decision problem are derived from the objective which is pursued and express the desirability of the various scenarios that can arise from a viable decision. In most decision problems several different objectives are pursued simultaneously. Therefore, a utility can be a complex combination of quite different commodities, such as monetary gain, status, and time. Decision analysis offers various, more or less formal, methods for performing this task [von Winterfeldt & Edwards, 1986].

The simplest, and least formal, method for utility assessment is to *visualise* all possible scenarios of a decision problem on a linear scale. The least desirable and the most desirable scenarios are identified and assigned to the ends of the scale. Every other scenario is now positioned on the scale, where the distance between two scenarios is indicative of the difference in desirability between these scenarios. Once all scenarios have been positioned, for each scenario a utility is yielded by projecting its position onto a matching numerical scale. Figure 3 illustrates the basic idea for two scenarios s_i and s_j.

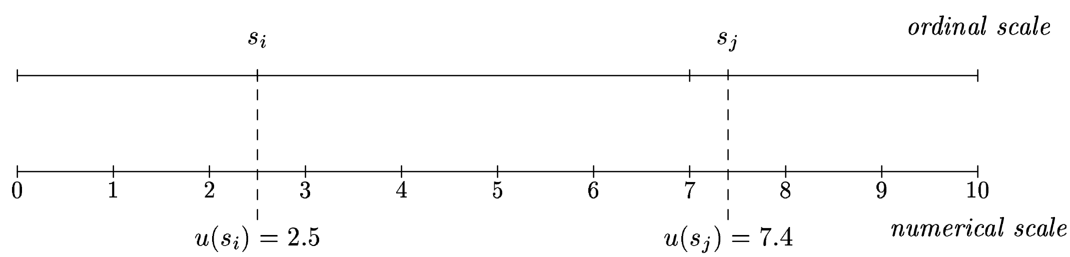

Figure 3: The Visualisation Method for Utility Assessment.

Instead of first visualising the differences in desirability among scenarios, these differences can be quantified directly, by using a *standard reference gamble*. A standard reference gamble serves for comparing three scenarios with regard to their desirability. Let s_i, s_j, and s_k be scenarios such that s_i is less desirable than s_j, and s_j in turn is less desirable than s_k. In assessing utilities for these three scenarios, a probability p is found such that scenario s_j is as desirable as a gamble that yields scenario s_k with probability p and scenario s_i with probability $1 - p$. Through this probability p, the utilities $u(s_i)$, $u(s_j)$, and $u(s_k)$ have now been assessed to satisfy

$$u(s_j) = p \cdot u(s_k) + (1 - p) \cdot u(s_i)$$

By using the standard reference gamble for appropriate three-tuples of scenarios, a system of equations is obtained from which a set of utilities is computed. The use of a standard reference gamble tends to yield better calibrated utilities than the visual method; the method, however, is more time-consuming.

If the utilities of a decision problem are composed of various commodities that are hard to compare, utility assessment can be especially cumbersome. The assessment then often is simplified by decomposing the utilities into their separate commodities. In terms of these separate commodities, *marginal utilities* are assessed, for example using one of the techniques outlined above. These marginal utilities subsequently are combined to yield overall utilities [von Winterfeldt & Edwards, 1986].

4 A Case Study

The decision-analytic approach to data interpretation has been applied to a case study. Although the situation described in the study in itself is hypothetical, it emerges from a real-life issue. The study concerns fraud with subsidies provided by the European Union to support the cultivation of certain agricultural crops. These subsidies are paid on the basis of declarations submitted by farmers. A fraud detection mechanism can make use of remotely sensed data. For each parcel, the viable decisions to consider on the basis of the data concern approval of the declaration on one hand, and an implication of fraud followed by further investigation on the other hand.

The study area is located around the village of Biddinghuizen in the province of Flevoland, the Netherlands. A Landsat Thematic Mapper image of the area is available (we used spectral bands 3, 4 and 5) from June 1987, as well as crop maps from 1986 and 1987. Seven different land-cover classes are distinguished: grass, wheat, potatoes, sugar beets, peas, beans, and onions. The crop maps, originating from an initial survey that included interviews with farmers, likely contain errors and uncertainties. In our study, we have used the 1986 map to calculate local prior probabilities. In the calculation, crop rotation cycles have been taken into consideration; so, the land-cover classes in successive years are not independent. Part of the 1987 crop map has been used for training sample selection, in combination with a colour composite of the image. From the 1987 map we have subsequently extracted the fields with peas or beans, and considered them as farmers' declarations for subsidy on those two crops.

To investigate viable decisions, various utilities have been assessed. The decision to imply fraud and suggest further investigation is very advantageous if the farmer's declaration specifies peas or beans and there is a different agricultural crop in reality: this scenario uncovers an illegal declaration. The scenario is assigned a utility of 10. The decision to not inspect such a field is extremely bad. This scenario is assigned a utility of 0. If a declaration turns out to be legal after further investigation, we have put ourselves (or the farmer) through unnecessary trouble. However, an investigation that turns out superfluous is not so bad as overlooking a false declaration. This scenario therefore is assigned a utility of 3. Avoiding superfluous investigations is more advantageous anyway: we assign a utility of 8. These utilities are summarised in Table 1.

Based on these utilities, we have applied our decision-analytic method to the decision for each

	inspection	
crop	yes	no
grass	10	0
wheat	10	0
potato	10	0
sugar beet	10	0
pea	3	8
bean	3	8
onion	10	0

Table 1: Utilities for the Detection of Illegal Farmer Declarations.

pixel. The result is a binary raster map, indicating the decision per pixel. Subsequently, a majority criterion has been applied to identify the *fields* that have been indicated for further investigation.

These results are shown in Figure 4. Of 81 fields with a declaration of peas or beans, 22 will be inspected.

Now consider a slightly different (perhaps less realistic) situation in which the subsidies paid are rather small and the fraud detection agency is under-staffed. In this situation, farmers generally will be given the benefit of the doubt and only very suspicious looking declarations will be inspected. The utility assigned to the scenarios for this situation are shown in Table 2. After applying our decision-analytic method to the same data with these new utilities, the number of fields to be investigated has decreased from 22 to 16 as expected.

5 Conclusions

Remotely sensed data are exploited to an increasing extent for decision making. For processing spatial data for this purpose, the objectives and preferences of the decision maker have to be taken into account. In principle, decisions may be taken on the basis of a complete classification of the data at hand. However, as taking the best decision involves the full extent and distribution of the uncertainty in the data, decision making is better founded directly on the data themselves. Decision-analytic interpretation, proposed in this paper, provides such an approach by integrating preferences and uncertainties in a mathematically well-founded way. The aim of the method is to assist a decision maker in taking the *best* decision and not so much to reconstruct reality, thereby contrasting conventional classification.

The decision-analytic approach to the interpretation of spatial data has been illustrated by means of a simple case study. Because of the simplicity of the presented study, it does not serve to fully demonstrate the potential power of the approach. However, it illustrates the issue of *customisation*: from a single set of spatial data, various results can be obtained tailored to a decision maker's objectives, by interpreting the data with different sets of utility assessments. An interesting issue that remains to addressed is the performance of the decision-analytic approach to data interpretation at a level beyond pixels. The approach is suitable for decision making for spatial objects instead of for individual pixels, as the concepts involved remain the same; however, an image segmentation pre-processing step is required. Applying the approach to spatial objects is expected to benefit from (probabilities of) geometrical and topological properties of objects for decision making.

To conclude, we would like to emphasise that our approach is based on a well-known and long-established mathematical framework from decision analysis for solving complex decision problems. The rich field of decision analysis provides a wealth of methods, for example for assessing probabilities and utilities, that can be applied to the problem of interpreting spatial data. Thanks to its flexibility and mathematical well-foundedness, the framework has the potential to become an integral part of geographical information systems.

	inspection	
crop	yes	no
grass	10	6
wheat	10	6
potato	10	6
sugar beet	10	6
pea	0	20
bean	0	20
onion	10	6

Table 2: Modified Utilities for the Detection of Obvious Illegal Declarations

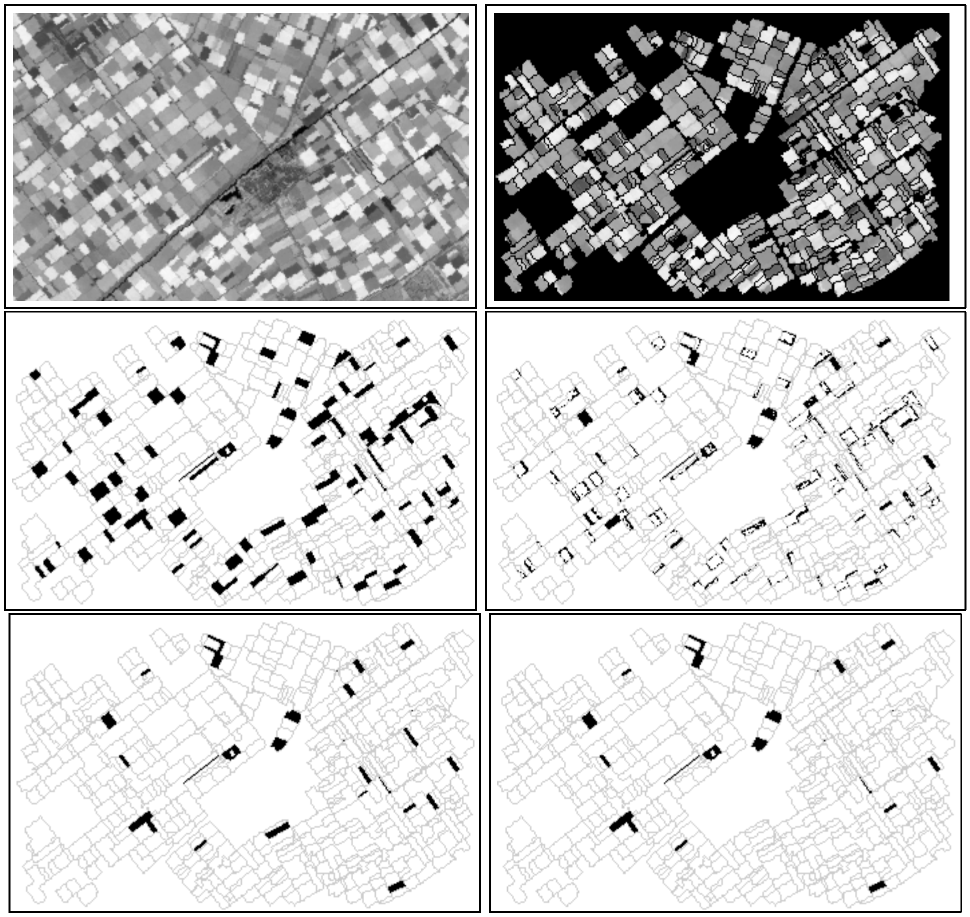

Figure 4: Experimental results. Upper Left: Landsat TM Image (band 4). Upper Right: Fields under consideration. Centre Left: Fields with declaration for 'peas' or 'beans'. Centre Right: Pixels with positive 'inspection' decision. Lower Left: Fields to be inspected. Lower Right: Fields to be inspected after modification of utilities.

References

[Lunetta *et al.*, 1991] R.S. Lunetta, R.G. Congalton, L.K. Fenstermaker, J.R. Jensen, K.C. McGwire, and L.R. Tinney (1991). Remote sensing and geographic information system data integration: error sources and research issues. *Photogrammetric Engineering & Remote Sensing*, vol. 57, no. 6, pp. 677 – 687.

[Fukunaga & Hummels, 1987] K. Fukunaga, D.M. Hummels (1987). Bayes error estimation using Parzen and kNN procedures, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 9, pp. 634 – 643.

[Gorte, 1995] B.G.H. Gorte (1995). Improving spectral image classifications by incorporating context data using likelihood vectors. *Proceedings of IPA 1995*, IEEE Conference Publications, no. 410, pp. 251 – 255.

[Morrison, 1995] J.L. Morrison (1995). Spatial data quality. In: S.C. Guptill & J.L. Morrison (ed.). *Elements of Spatial Data Quality*. Pergamon, Oxford, pp. 1 – 12.

[Raiffa, 1968] H.A. Raiffa (1968). *Decision Analysis: Introductory Lectures on Choices Under Uncertainty.* Addison-Wesley, Reading, Massachusetts.

[Smith, 1988] J.Q. Smith (1988). *Decision Analysis. A Bayesian Approach.* Chapman and Hall Ltd., London.

[Strahler, 1980] A.H. Strahler (1980). The use of prior probabilities in maximum likelihood classification of remotely sensed data. *Remote Sensing of Environment*, no. 10, pp. 135 – 163.

[von Winterfeldt & Edwards, 1986] D. von Winterfeldt and W. Edwards (1986). *Decision Analysis and Behavioral Research.* Cambridge University Press, New York.