

# Machine Learning for Adaptive Image Interpretation

Ilya Levner and Vadim Bulitko

Department of Computing Science  
University of Alberta  
Edmonton, Alberta T6G 2E8, CANADA  
{ilya|bulitko}@cs.ualberta.ca

## Abstract

Automated image interpretation is an important task with numerous applications. Until recently, designing such systems required extensive subject matter and computer vision expertise resulting in poor cross-domain portability and expensive maintenance. Recently, a machine-learned system (ADORE) was successfully applied in an aerial image interpretation domain. Subsequently, it was re-trained for another man-made object recognition task. In this paper we propose and implement several extensions of ADORE addressing its primary limitations. These extensions enable the first successful application of this emerging AI technology to a natural image interpretation domain. The resulting system is shown to be robust with respect to noise in the training data, illumination, and camera angle variations as well as competitively adaptive with respect to novel images.

**Keywords:** machine learning, Markov decision models in vision, adaptive image interpretation, remote-sensing, forest mapping, natural resource inventory.

## 1 Problem formulation

Forest maps and inventories have become a critical tool for wood resource management (planting and cutting), ecosystem management and wild-life research. Unfortunately, forest mapping at the level of individual trees is a continuous and costly undertaking. Canada alone has an estimated 344 million hectares of forests to inventory on a 10-20 year cycle (Pollock 1994).

At these scales, ground-based surveys and inventories are not feasible. Researchers have therefore turned to developing automated systems to produce forest maps from airborne images and LIDAR 3D data sources. The final goal is to measure the type (species), position, height, crown diameter, wood volume and age class for every tree in the survey area.

The task of large-scale forest mapping from aerial images presents formidable challenges, including: (i) massive amounts of high-resolution data, in order to recognize and measure individual tree crowns, (ii) construction and maintenance of (and providing access to) very large databases; Canada alone has an estimated  $10^{11}$  trees, (iii)

Copyright © 2004, American Association for Artificial Intelligence (www.aaai.org). All rights reserved.

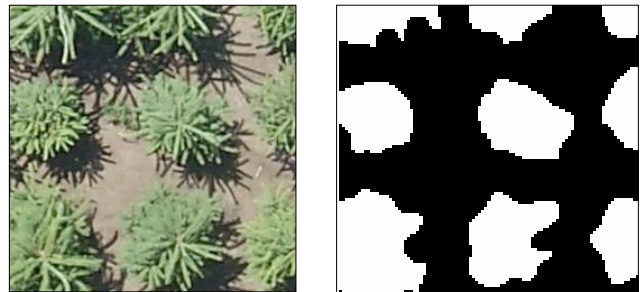


Figure 1: A fragment of an aerial image taken over a spruce plot is shown on the left. The right image is the desired interpretation with spruce canopies labeled in white. It is provided as a part of the training set.

geo-referencing of airborne images for validation purposes, (iv) orthorectification of aerial images, particularly given that elevation maps are often unavailable at the accuracy required. Of a particular interest are the challenges created by the image content, including variations in sun and camera angle and the resulting shadow and tree crown overlap. These challenges have been known to have an adverse effect on special purpose algorithms for individual tree identification (Culvenor 2002). In fact, the task is substantially challenging even to expert human interpreters resulting in up to 40% error in comparison to ground-based surveys (Gougeon & Leckie 2003).

In this paper we focus on the tree labeling problem as a first step towards an overall solution. Namely, we consider pixel-level labeling of aerial tree images (Figure 1). For each aerial image the task is therefore to identify all pixels belonging to canopies of trees of a certain kind. In the illustrations throughout the paper, we label spruce canopies. In each image such pixels are labeled in white while the rest of the image is labeled in black.

The rest of the paper is organized as follows: we survey special purpose tree canopy labeling systems and their shortcomings in section 2. Section 3 then presents a domain-independent approach in which a control policy over the space of vision operators is machine-learned. Our innovative applications of the AI technology are presented in section 4 with the empirical evaluation following in section 5. A discussion of the generalizability of the system and the current and future research directions conclude the paper.

## 2 Related research: special purpose systems

A number of approaches have been proposed for creating forest inventories from aerial images. Image-based (model-free) approaches use simplifying assumptions about forest images. For example, (Gougeon & Leckie 2003; Pinz 1991) use a token-based recognition approach which assumes a high level of contrast between the tree crown and the surrounding area. They deal with canopy feature extraction almost exclusively in terms of finding image features which evidence different types of tree canopies. A current example of this approach is ITC system (Gougeon & Leckie 2003) where tree canopies are detected and classified by a mixture of “valley-finding” (low intensity iso-contours), peak intensity detection (Wulder, Niemann, & Goodenough 2000b; 2000a) and texture, structure and contextual image features. This falls within traditional image segmentation and region labeling strategies where there is no explicit need to model features in terms of known tree attributes or specific 3D geometric models. Consequently, the approach is designed to apply where there are sufficiently spatially separated trees. Unfortunately, the performance can degrade significantly as such methods are applied to naturally occurring dense forests with overlapping tree crowns.

Another approach uses example-based image models. The underlying idea is to compare pre-specified tree crown image(s) with the image at hand. Typically such methods, e.g., (Murgu 1996), have a collection of example tree crowns which they match to the image. Drawbacks of such approaches include the need to collect a very large database of templates to account for differences in tree species, size, the slant of the terrain and illumination.

Model-based approaches take advantage of an explicit model of tree crowns to match with images. While minimizing the amount of image feature processing, elementary image features are used to hypothesize large numbers of regions for matching with 3D CAD tree models via computer graphics methods. For example, the STCI system (Pollock 1994) uses a template matching approach, however, unlike the example-based approaches discussed above, the crown templates are synthesized from a tree crown model. The upper part of a tree crown (known as “sun crown”) is modelled as a generalized ellipsoid of revolution and ray-tracing techniques are used to generate templates (Larsen & Rudemo 1997). Model-based approaches typically rely on detecting image features such as crown peaks and normally use pre-generated templates to match projected models with image data. The latter technique can require generation of many templates for different slant angles, tree types, etc. Additionally, model-based methods use simple shadow models and simplified 3D CAD models typically representing the canopy envelope only and are often unable to deal with natural variations in foliage and branches and the resulting irregular canopy boundaries.

## 3 Machine learning approach

All of the approaches reviewed in the previous section are promising, at least in a laboratory setting, but share some common drawbacks. First, they were carefully crafted in

a development process that required both time and expertise. More importantly, this development process exploited domain properties, such as whether the trees are separated or have overlapping canopies, whether the ground is flat or mountainous, or whether the forest has a homogeneous or heterogeneous species composition. Similarly, assumptions about the properties and position of the sensor are also integrated in the system design. As a result, these systems work within a narrow set of operating conditions, and cannot be applied under other conditions without re-engineering.

The challenges of the tree inventory problem are not unique in the realm of image interpretation. Thus, rather than crafting yet another domain-specific approach we present a general method of automated development of image interpretation systems with the following objectives: (i) rapid system development for a wide class of image interpretation tasks; (ii) low demands on subject matter, computer vision, and AI expertise on the part of the developers; (iii) accelerated domain portability, system upgrades, and maintenance; (iv) adaptive image interpretation wherein the system adjusts its operation dynamically to a given image; (v) user-controlled trade-offs between recognition accuracy and resources utilized (e.g., time spent).

These objectives favor the use of readily available off-the-shelf image processing libraries (IPL). However, the domain independence of such libraries requires an intelligent domain-specific policy to control the application of library operators. Operation of such a control policy is a complex and adaptive process. It is complex in that there is rarely a one-step mapping from image data to object label; instead, a series of operator applications is required to bridge the gap between raw pixels and semantic labels. Examples of the operators include region segmentation, texture filters, and the construction of 3D depth maps. Image recognition is an adaptive process in the sense that generally there is no fixed sequence of actions that will work well for most images (Bulitko *et al.* 2003).

The success of this approach to image interpretation therefore depends on the solution to a control problem: for a given image, what sequence of actions will most effectively and reliably detect the regions of interest?

Recently, a system called ADaptive Object REcognition (ADORE) became the first system to learn a complex domain-specific control policy for roof recognition in aerial photographs (Draper, Bins, & Baek 2000). As with many vision systems, it identified objects (in this case buildings) in a multi-step process. Raw images were the initial input data, while image regions containing identified buildings constituted the final output data; in between the data could be represented as intensity images, probability images, edges, lines, or curves. ADORE modelled image interpretation as a Markov decision process, where the intermediate representations were continuous state spaces, and the vision procedures were actions. The goal was to learn a dynamic control policy that selects the next action (i.e., image processing operator) at each step so as to maximize the quality of the final image interpretation. To demonstrate its general applicability, ADORE was subsequently ported to another domain (recognizing objects in office scenes) in another labo-

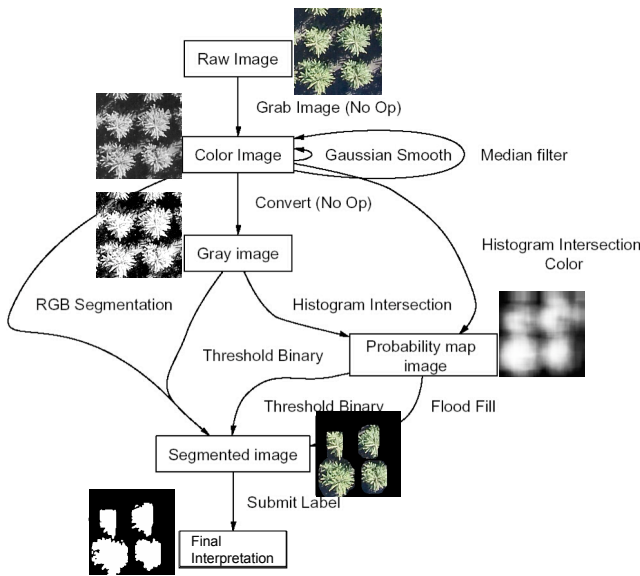


Figure 2: A fragment of the state-action graph used in our experiments. States are labeled with their vision data types and have forest samples shown next to them. Common vision operators are shown as the arcs.

ratory (Draper, Ahlrichs, & Paulus 2001).

#### 4 Innovative applications of AI technology

One of the contributions of this paper is to apply this emerging technology to the forest mapping domain, using it for the first time to recognize natural (as opposed to man-made) objects (i.e., tree canopies). More significantly, we propose and empirically evaluate solutions to ADORE’s three primary limitations. First, ADORE learned recognition strategies from training samples – but only after a human expert had selected the features that describe each intermediate level of representation. This process is effort-consuming and requires computer vision and domain expertise. Second, using the history-free Markov decision process model for selecting computer vision operators proved to be deficient in both applications of ADORE. Third, ADORE processed each image as a whole, which reduces interpretation quality while dealing with highly non-uniform images.

As in the case of ADORE, we begin with the Markov decision process (MDP) as the basic mathematical model by casting the IPL operators as the MDP actions and the results of their applications as the MDP states. Figure 2 presents a fragment of the state-action graph. The system operates in two modes as follows. During the *off-line training stage*, available subject matter expertise is encoded as a collection of training images with the corresponding desired interpretation (the so called ground truth). Figure 1 demonstrates an example of such (input image, ground truth label) pair for the forest mapping domain. Then off-policy reinforcement learning with deep backups and no bootstrapping is used to acquire a value function (Sutton & Barto 2000). At first, all feasible length-limited sequences of IPL operators are applied to each training image. The resulting interpretations are evaluated against the domain expert provided ground

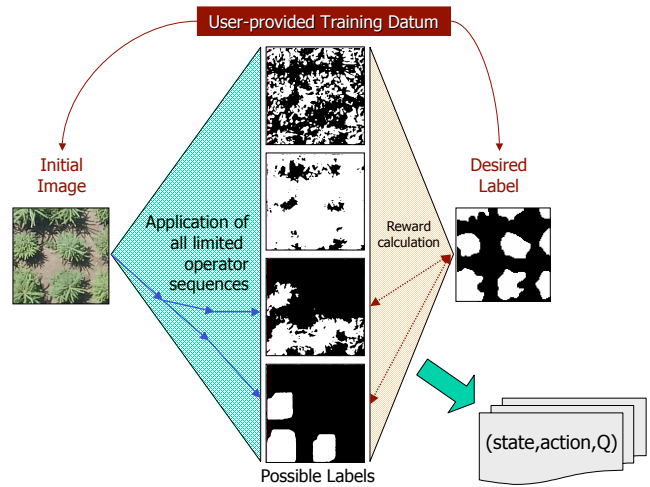


Figure 3: Off-line training stage: all limited-length operator sequences are applied to each training image. The resulting image interpretations are evaluated against the desired label. Action-state rewards are then computed.

truth as shown in Figure 3. We use a pixel-level similarity scoring metric defined as the ratio of the number of pixels labeled as the target class (e.g., spruce) by both the system and the expert to the total number of pixels labeled as the target class by either one of them. According to such a metric, an interpretation identical to the user-supplied label scores 1 while a totally disjoint interpretation will get a score of 0.

The interpretation scores are then “backed up” along the IPL operator sequences using dynamic programming. As a result, the value function  $Q : S \times A \rightarrow \mathbb{R}$  is computed for the expanded states  $S' \subset S$  and applied actions  $A' \subset A$ . The value of  $Q(s, a)$  corresponds to the best interpretation score the system can expect by applying operator  $a$  in state  $s$  and acting optimally thereafter. In reinforcement learning terms, we are representing the task as a finite horizon non-discounted problem wherein all intermediate rewards are zero except these collected by outputting an image interpretation. The latter is a positive reward proportional to the quality of the interpretation. The term  $Q$  comes from Watkins’ Q-learning (Watkins 1989).

The collected training set of Q-values  $\{[s, a, Q(s, a)]\}$  samples a tiny fraction of the  $S \times A$  space. Correspondingly, we use function approximation methods to extrapolate the value function onto the entire space. Features  $f$  are extracted off the raw (large) states  $S$  to make the approximation feasible. In this paper we use Artificial Neural Networks as the Q-function approximator.

During the *on-line interpretation stage*, the system receives a novel image and proceeds to interpret it as shown in Figure 4. Namely, the value function learned off-line now guides the control policy to apply vision operators from the IPL library. Several control policies are possible ranging from greedy policies that select the next action  $a$  so as to maximize  $Q(s, a)$  in each state  $s$  to static policies that always apply the same sequence of vision operators regardless of the input image (Levner *et al.* 2003b).

The “least-committment” control policy we use addresses

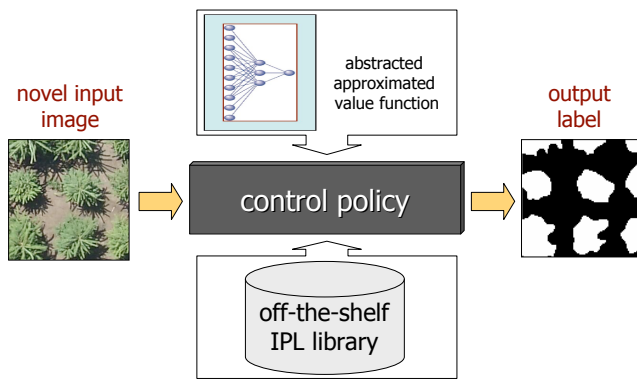


Figure 4: On-line operation: the control policy uses an approximate value function to select the best sequence of operators from an IPL library. As the result, an image interpretation label is produced.

the first two shortcomings of ADORE. First, it applies all limited feasible sequences of operators to the input image  $s_0$ . Once the set of possible image interpretations  $\{s_1, \dots, s_N\}$  is computed, the policy uses the label of each interpretation  $s_i$  to extract features from the original input image  $s_0$ . The resulting composite feature vectors  $f_{s_i}(s_0)$  are used with the machine-learned value function to select the most promising interpretation  $s_{i^*}$  as follows:  $i^* = \arg \max_i Q(f_{s_i}(s_0), submit)$ . In other words, the policy selects the interpretation  $s_{i^*}$  that is expected to bring the highest reward when submitted (i.e., output as the system's interpretation of the input image).

This technique eliminates ADORE's need to design high-quality features for every processing level as they are now required for the initial color image and the final binary interpretation only. Additionally, extracting features from the initial image provides a context for the features extracted off a candidate interpretation thereby addressing ADORE's loss of performance due to history-free Markov features.

Finally, before interpreting a novel image, we partition it into regular rectangular tiles. Each tile is processed independently by the control policy. The resulting interpretations (one per tile) are then assembled into a single interpretation of the original image. This technique greatly increases flexibility of the system by allowing it to use different operators on different parts of a non-uniform image thereby addressing the third primary drawback of ADORE.

## 5 Empirical evaluation

In the following sections we evaluate the approach with respect to several key attributes of a successful image interpretation system: overfitting-free training on real-world data, robustness to noise in the training data, and robustness to variance in illumination conditions, camera angles, and inter-image differences.

In the tests we use the least-commitment policy with an Artificial Neural Network as the function approximator over 192 color histogram based features. Cross-validation error was monitored during the training process to guard against overfitting. The operator library contained approximately 100 common vision operators similar to the ones found in

Intel OpenCV library. The parameters of each operator were tabulated in regular intervals so that invoking such an operator actually applied all parameterized instantiations of it.

Longer operator sequences have been shown to increase the interpretation quality at the cost of an exponential increase in the running time (Bulitko *et al.* 2003). In the experiments presented in this paper we capped the system at the shortest non-trivial sequence length (four).

### 5.1 Performance on spruce canopy images

In order to demonstrate the adaptive nature of the learned policy we compared its performance to the static operator sequence with the best mean performance on the training data (henceforth called 'best static policy'). Thirty four 256x256 pixel aerial images were acquired with an RGB camera over spruce plots. A fragment of a typical image is shown in Figure 1. We employed the standard leave-one-out cross-validation technique by training on a subset of 33 images and testing on the left-out image. This process was repeated  $\binom{34}{33}$  times for all combinations of the training images. The mean pixel-level similarity score of the machine-learned policy was 0.54 with the standard deviation of 0.14 while the best static policy scored  $0.46 \pm 0.13$ . Figure 5 demonstrates on-line performance of the machine-learned and best static policies relative to the perfect control policy (i.e., the policy gaining the highest possible score on each image). On most of the 34 cross-validation images, the machine-learned policy produced a nearly optimal sequence achieving  $93.6 \pm 15.5\%$  of the perfect policy's score. Substantial variation in the interpretation difficulty among the images left the best static policy at  $81.6 \pm 19.7\%$ .

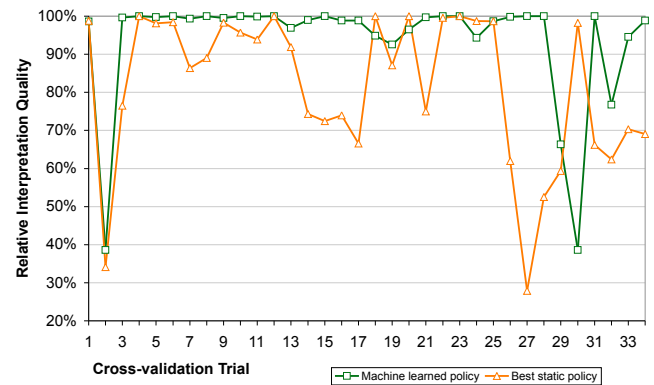


Figure 5: On-line interpretation: machine-learned policy adapts to different images frequently selecting the perfect operator sequence.

It is worth noting that the pixel-based similarity metric we use is a very strict measure penalizing every single pixel mislabeled. At the same time, missing individual pixels is not critical in the overall forest mapping framework since the desired tree parameters, such as crown diameter, position, and species, can be recovered even from an imperfect but proportional label. Since the tree parameter estimation module is not yet operational, we presented the interpretations our system produced to silviculture specialists and solicited their feedback. The current level of performance was

deemed promising for tree parameter extraction. An interesting case is shown in Figure 6 wherein the system was able to label spruce canopies “drowned” in the surrounding vegetation better than a human interpreter.



Figure 6: Shown on the left is the original image with three small spruce canopies “drowned” in surrounding young aspen. The human interpreter labeled only one of them (lower left corner of the image). The system was able to label two canopies (right image).

### 5.2 Robustness to noise in training data

The tedious and time-consuming nature of manual image labeling inevitably leads to interpretation errors. One study (Gougeon & Leckie 2003) found the tree count error to be between 10-40% relative to ground surveys. Therefore, it is important to establish the extent to which our learning model is robust with respect to noise in manually labeled training data. Pending an extensive study involving a panel of human forest interpreters and ground-based surveys, we present the robustness results with simulated training data labeling errors. Namely, we used the same 34-image spruce canopy set but this time perturbed the corresponding training labels. The two common types of human interpretation errors are labeling several neighboring tree canopies as one and missing a tree canopy completely. We simulated these errors via eroding and dilating the training labels by approximately 30%. Eroding an interpretation significantly distorts the desired canopy labels often reducing a single canopy label to a collection of disjoint pixels. On the other hand, dilation enlarges the labels sometimes merging two neighboring canopy labels into one as illustrated in Figure 7.

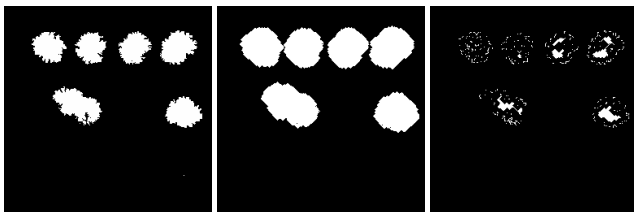


Figure 7: Noise in the training data: the true label (left), a dilated label joining neighboring canopies together (middle), an eroded label missing full canopies (right).

Training on the dilated and eroded labels reduced the performance of the machine-learned control policy to  $85.6 \pm 24.2\%$  of the perfect policy (corresponding to the pixel-level similarity score of  $0.49 \pm 0.17$ ). The static policy scored

$82.3 \pm 19.4\%$  of the perfect policy ( $0.47 \pm 0.13$  in the absolute terms). As we used the same leave-one-out cross-validation technique, the results suggest certain robustness to training data noise.

### 5.3 Robustness to changes in illumination

Large-scale forest mapping is a continuous undertaking resulting in the images acquired at different times of day and different seasons. The consequent changes in illumination can lead to significant performance degradation in systems fine-tuned for a specific set of images (Culvenor 2002). Additionally, low sun angles can result in canopies “missing” in shadows cast by taller trees. Pending a large scale experiment with multiple aerial acquisitions of the same forest plot at different times of day and different seasons, we conducted a smaller scale study using synthetic data. Namely, a scene with two types of trees (six trees of each type) was computer ray-traced at every hour between 7 o’clock in the morning and 5 o’clock in the evening (the most likely data acquisition times for real forests). As the sun traversed the sky, the rendered shadows and undergrowth illumination varied significantly (Figure 8).

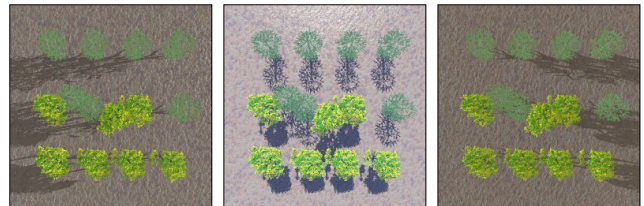


Figure 8: Effects of the sun position: 7am (left), noon (middle), 5pm (right).

We then administered the leave-one-out cross-validation procedure by training the system on 10 out of 11 images and testing it on the remaining image. The process was repeated 11 times for all sets of training images. The results in Figure 9 indicate that the machine-learned policy is robust to image variations resulting from different sun positions. On the other hand, the best fixed operator sequence achieved the optimal interpretation only on the 9am image. Cumulatively, the machine-learned policy scored  $94 \pm 9.5\%$  of the perfect policy while the best static policy was at  $63.8 \pm 21.6\%$ .

### 5.4 Robustness to camera angle variations

Another difficulty with remote sensing vision systems lies with the fact that the sensor (camera) can be at different angles when the images are acquired. This can happen due to aircraft maneuvers, turbulence, and other reasons. In forest mapping nadir angles are generally preferred as low-angle images can be plagued with overlapping canopies (Figure 10).

In this experiment we measured the robustness of machine-learned and best static control policies to increasingly off-nadir camera angles. The synthetic scene from the sun angle experiments was re-used but rather than varying the sun position we changed the camera angle from nadir to significantly slanted in five steps. Figure 11 suggests that

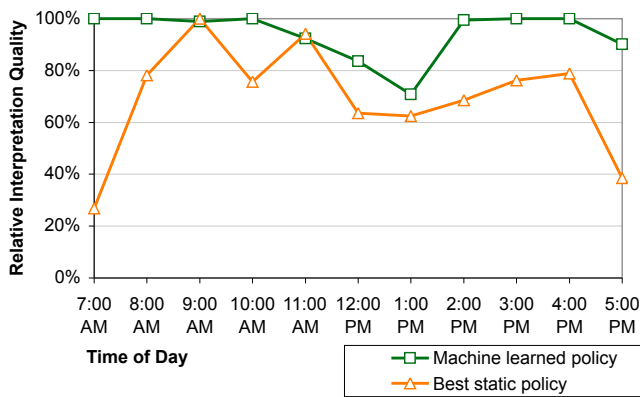


Figure 9: Performance of machine-learned and best static policies on images acquired at different times of day.

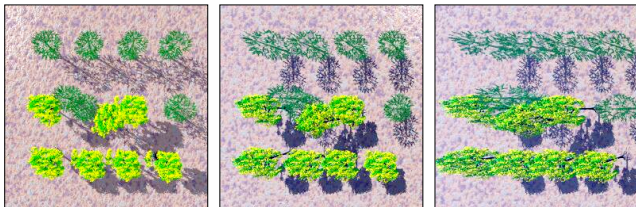


Figure 10: Effects of off-nadir camera angles.

the machine-learned policy demonstrates some robustness to varying camera angles. It scored  $93.5 \pm 14.5\%$  of the perfect policy whereas the best static policy achieved  $79.1 \pm 44.1\%$ .

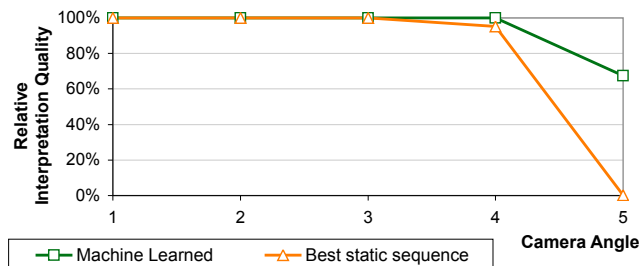


Figure 11: Machine-learned and best static policies on images acquired at increasingly off-nadir camera angles.

### 5.5 Adaptivity to novel images

Natural forests are inherently diverse in terms of tree species, canopy sizes, and foliage conditions. Consequently, even images of a simple composition stand taken within a hundred feet of each other can be vastly dissimilar as shown in Figure 12. Therefore, the ability to adapt to a novel image radically different from the training data is a crucial factor in the success of a forest mapping system.

In order to evaluate the robustness of the presented approach we compared it with a state-of-the-art system based on a trainable hierarchical hidden Markov tree (HHMT) model (Cheng, Caelli, & Ochoa 2002). We first gauged its performance on forest images with moderate variance such as parts of a single image. The HHMT-based system demonstrated an impressive mean pixel-level similarity score of 0.71 (compared to our system's score of 0.58).



Figure 12: Two diverse images acquired in a close spatial proximity.

We then selected two substantially diverse images (parts of which are shown in Figure 12). One of the images (shown left in the figure) was considerably more difficult due to the dense deciduous vegetation enveloping the spruce trees. In each of the two trials both systems were trained on one of the two images and tested on the other. The HHMT-based system had the mean pixel-level similarity score of 0.1. At the same time our approach scored 0.25. This initial experiment suggests a certain degree of adaptivity of the MDP-based approach and warrants further studies.

## 6 Discussion and future work

Conventional ways of developing image interpretation systems usually require a significant subject matter and computer vision expertise from the developers. The resulting systems are expensive to upgrade, maintain, and port to other domains.

In this paper we present an emerging AI technology with the aim of automating development of image interpretation systems. Our extensions of the state-of-the-art MDP-based machine-learned system ADORE address its three primary drawbacks. First, by partitioning input images our system can interpret uniform regions of a non-uniform image independently. Second, as high-quality interpretations are tightly coupled with the input image thereby violating the history-free Markov assumption, our control policy extracts features off the initial image as well as the candidate interpretation. Finally, our least-commitment policy does not rely on high-quality features for all processing levels thereby increasing the interpretation quality as well as portability of the system.

Cumulatively, the three extensions enabled the first application of an MDP-based vision system to interpretation of natural (as opposed to man-made) objects. The preliminary experiments presented in this paper suggest that the approach is robust with respect to noise in the training data, variable illumination and camera angles as well as high variation between training and testing images.

The initial success encourages further automation of the system. In particular, we are presently investigating automated methods of feature selection (Levner *et al.* 2003a). When successful, they will eliminate computationally expensive application of all operator sequences as some sequences will be discarded dynamically. We have also had initial success with automated techniques for operator parameter tuning and selection (Bulitko, Lee, & Levner 2003).

This is expected to result in significant computational savings as only a fraction of the operator set will have to be executed on- and off-line. Ensemble learning and boosting methods have recently seen success in classification and regression. We are presently developing extensions of such methods to control policies (Li *et al.* 2003). Finally, a more adaptive image splitting and interpretation merging procedure is under development and is expected to replace the current regular tiling scheme.

On the application side, the system is presently considered for deployment at an Alberta government natural resources office within the next several months. It will be evaluated with respect to identification of forest fire burn boundaries and validation of commercial thinning practices. Being domain-independent, our approach is already being tested for brain tumor detection in magnetic resonance imaging (MRI) scans in cooperation with the Alberta Cross Cancer Institute.

## Acknowledgements

Bruce Draper participated in the initial system design and contributed to an early draft of the related research section. We appreciate help of Li Cheng, Terry Caelli, Lihong Li, Greg Lee, Dave McNabb, Ken Greenway, Rongzhou Man, Joan Fang, Dorothy Lau, Russ Greiner, Jane Hilderman, Guanwen Zhang, and Albert Murtha. The funding was generously provided by NSERC, iCORE, the University of Alberta, and the Alberta Ingenuity Centre for Machine Learning.

## References

- Bulitko, V.; Lee, G.; Levner, I.; Li, L.; and Greiner, R. 2003. Adaptive image interpretation : A spectrum of machine learning problems. In *Proceedings of the ICML Workshop on The Continuum from Labeled to Unlabeled Data in Machine Learning and Data Mining*, 1–8.
- Bulitko, V.; Lee, G.; and Levner, I. 2003. Evolutionary algorithms for operator selection in vision. In *Proceedings of the 7th Joint Conference on Information Sciences : the 5th International Workshop on Frontiers in Evolutionary Algorithms*, 359–362.
- Cheng, L.; Caelli, T.; and Ochoa, V. 2002. A trainable hierarchical hidden markov tree model for color image segmentation and labeling. In *Proceedings of the International Conference on Pattern Recognition*.
- Culvenor, D. 2002. Tida: an algorithm for the delineation of tree crowns in high spatial resolution remotely sensed imagery. *Computers & Geosciences* 28(1):33–44.
- Draper, B.; Ahlrichs, U.; and Paulus, D. 2001. Adapting object recognition across domains: A demonstration. In *Proceedings of International Conference on Vision Systems*, 256–267.
- Draper, B.; Bins, J.; and Baek, K. 2000. ADORE: adaptive object recognition. *Videre* 1(4):86–99.
- Gougeon, F., and Leckie, D. 2003. Forest information extraction from high spatial resolution images using an individual tree crown approach. Technical report, Pacific Forestry Centre.
- Larsen, M., and Rudemo, M. 1997. Using ray-traced templates to find individual trees in aerial photos. In *Proceedings of the 10th Scandinavian Conference on Image Analysis*, volume 2, 1007–1014.
- Levner, I.; Bulitko, V.; Li, L.; Lee, G.; and Greiner, R. 2003a. Automated feature extraction for object recognition. In *Proceedings of the Image and Vision Computing New Zealand conference*, 309–313.
- Levner, I.; Bulitko, V.; Li, L.; Lee, G.; and Greiner, R. 2003b. Towards automated creation of image interpretation systems. In Gedeon, T., and Fung, L., eds., *Lecture Notes in Artificial Intelligence (LNAI)*, volume 2903. Berlin, Heidelberg: Springer-Verlag. 653–665.
- Li, L.; Bulitko, V.; Greiner, R.; and Levner, L. 2003. Improving an adaptive image interpretation system by leveraging. In *Proceedings of the 8th Australian and New Zealand Conference on Intelligent Information Systems*, 501–506.
- Murgu, D. 1996. Individual tree detection and localization in aerial imagery. Master's thesis, Department of Computer Science, University of British Columbia.
- Pinz, A. 1991. A computer vision system for the recognition of trees in aerial photographs. In *J.C.Tilton (ed.), Multisource Data Integration in Remote Sensing*, 111–124.
- Pollock, R. 1994. A model-based approach to automatically locating tree crowns in high spatial resolution images. In Desachy, J., ed., *Image and Signal Processing for Remote Sensing*.
- Sutton, R., and Barto, A. 2000. *Reinforcement Learning: An Introduction*. MIT Press.
- Watkins, C. 1989. *Learning from Delayed Rewards*. Ph.D. Dissertation, Kings College, Cambridge University.
- Wulder, M.; Niemann, K.; and Goodenough, D. 2000a. Error reduction methods for local maximum filtering. In *The Proceedings of the 22nd Symposium of the Canadian Remote Sensing Society*.
- Wulder, M.; Niemann, K.; and Goodenough, D. 2000b. Local maximum filtering for the extraction of tree locations and basal area from high spatial resolution imagery. *Remote Sensing of Environment* 73:103–114.