Active Planning, Sensing and Recognition Using a Resource-Constrained Discriminant POMDP

Zhaowen Wang[†], Zhangyang Wang[†], Mark Moll[‡], Po-Sen Huang[†], Devin Grady[‡]
Nasser Nasrabadi[§], Thomas Huang[†], Lydia Kavraki[‡], Mark Hasegawa-Johnson[†]
[†]Beckman Institute, University of Illinois at Urbana-Champaign, Urbana, IL 61801
[‡]Dept. of Computer Science, Rice University, Houston, TX 77005
[§]U.S. Army Research Laboratory, Adelphi, MD 20783
{wang308, zwang119, huang146, huang, jhasegaw}@ifp.uiuc.edu
{mmoll, devin.grady, kavraki}@rice.edu, nasser.m.nasrabadi.civ@mail.mil

Abstract

In this paper, we address the problem of object class recognition via observations from actively selected views/modalities/features under limited resource budgets. A Partially Observable Markov Decision Process (POMDP) is employed to find optimal sensing and recognition actions with the goal of long-term classification accuracy. Heterogeneous resource constraints – such as motion, number of measurements and bandwidth – are explicitly modeled in the state variable, and a prohibitively high penalty is used to prevent the violation of any resource constraint. To improve recognition performance, we further incorporate discriminative classification models with POMDP, and customize the reward function and observation model correspondingly. The proposed model is validated on several data sets for multi-view, multi-modal vehicle classification and multi-view face recognition, and demonstrates improvement in both recognition and resource management over greedy methods and previous POMDP formulations.

1. Introduction

In real-time object recognition applications, it is often preferred to sequentially obtain the most informative sensory data in order to reduce the current recognition uncertainty. Such data acquisition scheme, generally called active sensing [14], is useful especially when a huge amount of data are available from various sensors and modalities while we do not have the luxury to capture and process all of them.

Most early works on recognition with active sensing rely on a greedy strategy that selects the next best sensor based on some information theoretic criteria. For example, sensor scheduling algorithms have been proposed which greedily select the view angle to observe object that leads to the maximum entropy reduction in class hypothesis [3] or maximum expected mutual information between the class label and observation [5]. Unfortunately, the sensors selected greedily may not be optimal for long-term recognition. Moreover, entropy-based criteria usually involve high computation and are not robust to model estimation error.

On the other hand, Partially Observable Markov Decision Process (POMDP), which can deal with the active recognition problem on an arbitrarily long time horizon, has been widely studied and applied in gesture recognition [4], mine detection [11] and image object detection [12]. Information measure on the class posterior can be used to guide the policy learning in POMDP [16], and reinforcement learning algorithms are employed to learn object model and planning policy simultaneously [16, 12, 13]. In POMDP, the objective is represented by a single reward function, which makes it difficult to balance between improving recognition accuracy and preserving sensing resources. Also, conventional POMDP has a generative formulation which limits its recognition performance especially for high-dimensional multimedia data with insufficient training samples.

In this paper, we address the problem of object recognition via active sensing with limited budgets on motion and sensing resources. A novel POMDP formulation is proposed that incorporates heterogenous resource constraints and discriminative classification models. The consumption of each resource is explicitly monitored in our state variable, and a prohibitive penalty is used to prevent any resource depletion so that policy learning can focus on the recognition task. In addition, by introducing a single recognition action, we decouple the learning of classification model and sensing policy so that more powerful discriminative classifiers can be used within the POMDP

framework. Our reward function and observation model are also customized for classification purpose correspondingly. The proposed model is simple but effective. It is applied to multi-view, multi-modal vehicle classification and multi-view face recognition on several data sets, and outperforms greedy methods and previous POMDP formulations in both recognition accuracy and resource management.

The remainder of this paper is organized as follows. After reviewing related work on POMDP and active recognition in Sec. 2, we present our models with resource constraints and discriminative classifiers in Sec. 3. Extensive evaluation of the proposed method is reported in Sec. 4. Sec. 5 concludes the paper.

2. POMDP for Active Recognition

A POMDP is defined by a tuple $<\mathcal{S}, \mathcal{O}, \mathcal{A}, T, O, R>$, where $\mathcal{S}$, $\mathcal{O}$ and $\mathcal{A}$ denote a finite set of discrete states, observations and actions, respectively. The state S is modeled as a Markov process, whose transition from time t to $t+1$ driven by action A taken at t follows the transition distribution:

$$T(s, a, s') = p(S_{t+1} = s' | S_t = s, A_t = a). \quad (1)$$

The state S is hidden, and its value can only be inferred from its observation O according to the observation distribution:

$$O(s', a, o) = p(O_{t+1} = o | A_t = a, S_{t+1} = s'). \quad (2)$$

An action a taken in a particular state s results in a reward which is described by a function $R(s, a)$. Our goal is to find a policy that decides what is the optimal action to take based on the belief in current state so that maximum is achieved for the expected accumulated reward: $\mathrm{E}[\sum_{t=0}^{K-1} \gamma^t R(S_t, A_t)]$, where K is the length of time horizon to consider, and $\gamma \in (0, 1]$ is the discount factor for future rewards. A more detailed description of POMDP and algorithms to solve for the optimal policy can be found in [15].

The problem of recognition with active perception naturally fits into the framework of POMDP, and existing work can be found in [4, 2, 11, 12]. In this problem, we want to recognize the class category X of a target, which is modeled as part of the hidden state S. To achieve this goal, observations $\{O_t\}$ of the target with different sensing parameters are sequentially obtained using appropriate sensing actions $\{A_t\}$ until a classification action predicts a class label for the target. A positive reward is given if the target is correctly labeled, and a negative reward is given otherwise. Additional rewards can be assigned to each sensing action to model sensing costs [2]. A good policy will select actions dynamically based on the current belief in X inferred from previous observations, either making another observation using the most informative sensor or making a final classification with the most likely class.

Most existing POMDP formulations for active recognition, unfortunately, suffer from two problems. First, the objectives to improve recognition performance and preserve sensing resources are wrapped in a single reward function, which makes it difficult to balance their relative importance. The issue becomes even more complex if we have multiple heterogenous resources which cannot be compared on the same scale. Second, due to its generative nature and Markov property, POMDP mainly relies on naive Bayesian classifier for recognition task, which usually gives unsatisfactory results compared with more advanced discriminative classifiers. We will address these two problems by introducing novel POMDP formulations in the next section.

3. Proposed Models

We consider an active recognition scenario in which a mobile sensing platform observes a target from different view angles using different sensing parameters including sensor modality, feature extractor, *etc*. Therefore, navigation, sensing and recognition are all considered in our POMDP models, whose basic components are specified in the following.

The state $\mathcal{S}$ is defined to include all the combinations of target class label X and sensing platform position Z, which are all discrete variables. Here we only consider the finite number of positions where observations can be taken as the possible values for Z. However, the motion planning between different Z's is done in the continuous space using a separate model as will be described in Sec. 4.2. In addition, a special terminal state s_T is used to represent the state after recognition is done.

The action $\mathcal{A}$ can be divided into three categories. A *move action* $a \in \mathcal{A}_m$ drives the sensing platform to the specified view position where the target can be observed from a particular view angle. An *observe action* $a \in \mathcal{A}_o$ makes observation of target using the specified sensing parameter from the current view position. A *classify action* $a \in \mathcal{A}_c$ predicts label X with the specified class.

Since the class label X never changes and we assume perfect navigation control over position Z, the transition model $T(s, a, s')$ is actually deterministic. It is designed according to the three action types as

$$T(s, a, s') = \begin{cases} 1, & \text{if } a \in \mathcal{A}_m, x' = x, z' = z_a \\ 1, & \text{if } a \in \mathcal{A}_o, s' = s \\ 1, & \text{if } a \in \mathcal{A}_c, s' = s_T \\ 0, & \text{otherwise} \end{cases}, \quad (3)$$

where z_a is the position specified by *move action* a. Unless otherwise noted, we use $x(x')$ and $z(z')$ to denote the class label and view position represented by $s(s')$, respectively.

The observation $\mathcal{O}$ consists of quantized feature values as well as a dummy observation o_d. For $a \in \mathcal{A}_o$, the observation distribution $O(s', a, o)$ assigns a probability to any $o \neq o_d$ according to the likelihood that o is generated from class x' under view position z' with the sensing parameter specified by a. For other actions, only o_d will be observed.

The incorporation of multiple resource constraints and the design of reward function $R(s, a)$ will be detailed in the next.

3.1. POMDP with Resource Constraints

In practice, we often have limited budgets to execute either motion or sensing actions. A straightforward way [2] to take this into consideration in POMDP is to have the reward function assign a cost (negative reward) of $-\alpha_i \cdot n$ to any action that consumes the i-th resource by an amount of n. Let R_c denote the expected reward for the final classification action, and we can express the goal of POMDP as

$$\max R_c - \alpha_1 \cdot n_1 - \alpha_2 \cdot n_2 - ... \tag{4}$$

where n_i is the total amount of consumption in the i-th resource. Unfortunately, there is no explicit way to balance the relative importance between recognition reward R_c and resource cost α_i. Moreover, if the recourses are of heterogenous types, such as navigation distance and sensing energy, making trade-off between all the α_i's becomes another question.

Instead of solving the resource-regularized problem (4), we propose to use a resource-constrained objective:

$$\max R_c \quad \text{s.t. } n_1 \leq \beta_1,\ n_2 \leq \beta_2,\ ...\ , \tag{5}$$

where β_i is the budget for the i-th resource which limits the total consumption in this resource. Since β_i has a specific physical meaning, its value is easier to determine than α_i. Eq. (5) also decouples the recognition objective from all the resource constraints so that the policy learning for POMDP becomes more focused on the recognition task.

To implement the objective in (5), we propose POMDP with Resource Constraints (POMDP-RC) which extends our basic formulation discussed earlier in the following ways.

First, we augment the state space $\mathcal{S}$ with a set of variables $\{B_i\}$ where each B_i keeps track of the remaining budget for the i-th resource. B_i is initialized to the total budget β_i.

In the transition model $T(s, a, s')$, when a *move action* $a \in \mathcal{A}_m$ or an *observe action* $a \in \mathcal{A}_o$ is taken, the amount of consumption in the i-th resource will be deducted from the corresponding remaining budget b'_i in s'. If the deduction leads to a negative b'_i (the resource is used up), we set s' to be the terminal state s_T.

To prevent any resource from being depleted, the reward function $R(s, a)$ is designed to assign a prohibitive penalty $r_p < 0$ in such cases. When a *classify action* $a \in \mathcal{A}_c$ is taken, a recognition reward $r_c > 0$ will be assigned if the class prediction is correct, and zero reward will be given otherwise. Specifically, our reward function is defined as:

$$R(s, a) = \begin{cases} r_p, & \text{if } a \notin \mathcal{A}_c, T(s, a, s_T) = 1 \\ r_c, & \text{if } a \in \mathcal{A}_c, x_a = x \\ 0, & \text{otherwise} \end{cases}, \tag{6}$$

where x_a is the class predicted by the *classify action* a. It can be seen that in POMDP-RC a nonzero reward will be given only when the terminal state is reached, implying that a long-term goal on recognition performance is emphasized.

3.2. POMDP with Discriminative Classifier

In conventional POMDP, the class label X is inferred from all the observations $\{O_t\}$ through the observation model. Since each observation O_t is conditionally independent given the state S_t, the maximum a posteriori estimation of X is essentially the same as the Naive Bayesian classification:

$$\hat{x} = \arg\max_c p(X{=}c) \prod_t p(O_{t+1} | X{=}c, Z_{t+1}, A_t). \tag{7}$$

When the observation comes from a high-dimensional space, *e.g.* image and audio, a good estimation of the observation likelihood in (7) requires a large number of samples which are often unavailable in real applications. In addition, the strong independency assumption used in the Naive Bayesian classifier leads to the loss of high-order statistics which may contain useful discriminative information.

To address this problem, we propose a POMDP with Discriminative Classifier (POMDP-DC) in which recognition decision is made by an external classifier instead of by a POMDP policy. Discriminative classifiers, such as Support Vector Machine (SVM) and logistic regression, show better generalization capability than naive Bayesian method in many cases. Suppose a classifier is trained independently for each combination of view Z and sensing parameter A, and it assigns a score $s_{c,Z,A}(O)$ to class c when observation O is acquired. The final recognition can be done by fusing the scores from all the classifiers:

$$\hat{x} = \arg\max_c \sum_t s_{c, Z_{t+1}, A_t}(O_{t+1}). \tag{8}$$

High order statistics between multiple observations can also be incorporated by training classifiers in the joint observation space:

$$\hat{x} = \arg\max_c s_{c, \{Z_t\}, \{A_t\}}(\{O_t\}), \tag{9}$$

where $s_{c, \{Z_t\}, \{A_t\}}$ denotes the classifier trained on the joint space of observations under view/sensory combinations

$\{(Z_{t+1}, A_t)\}$. To implement (8) and (9), we further extend POMDP-RC as discussed below. We notice the recent work in [12, 13] has a similar idea to integrate external classifier with Markov Decision Process (MDP), but with the goal of anytime performance instead of the accuracy upon termination.

In POMDP-DC, only one *classify action* is defined in $\mathcal{A}_c$, which is used to stop sensing and make recognition according to (8) or (9).

Since the goal of POMDP-DC is to select the best views and sensing parameters for classification using an external classifier, the reward function $R(s, a)$ for *classify action*, inspired by the objective of SVM classifier, is changed to maximize the classification score margin between the correct class x and any other class c:

$$R(s, a) = \min_{c \neq x} \left[\min(\delta, s_x - s_c)\right], \quad a \in \mathcal{A}_c, \qquad (10)$$

where $\delta > 0$ is the minimum required margin, and s_c is the total score for class c defined by (8) or (9).

Eq. (9) only enables modeling high-order statistics for the classifier. Such information can also be utilized in POMDP-DC by introducing additional *observe actions* to $\mathcal{A}_o$, each of which can acquire more than one observation, or a *meta-observation*, at once with different sensing parameters from the current view position. The observation space $\mathcal{O}$ and distribution $O(s', a, o)$ are augmented correspondingly, and the resources to obtain a meta-observation are aggregated and deducted from state s' in the transition model $T(s, a, s')$.

Lastly, we want to note that although there seem to be many variables included in our model, the only thing unknown in state S is the class label X, and all the other variables can either be observed or transit deterministically. This ensures our model can be solved efficiently.

4. Experiments

4.1. Data Sets and Implementations

The proposed POMDP-RC and POMDP-DC models are tested on three data sets in this section.

The first two data sets, MSTAR [17] and Civilian Vehicle Domes (CVDOME) [8], are both multi-view radar image sets for vehicle classification. The MSTAR contains airborne X-band Synthetic Aperture Radar (SAR) images for 10 classes of military vehicles, with sample images shown in Fig. 1. The vehicles were captured from various angles. 4785 images with depression angles 17° and 30° are used for training, and 4351 images with depression angles 15° and 45° are used for testing. The azimuth angles are quantized into 12 discrete values as if the image could be acquired from one of the 12 view positions. 12 images from the same class but with different views are randomly selected and combined as a training/testing instance, and each instance is also assigned with an initial view position. The same approach to generate sample instances is used for other data sets unless otherwise stated.

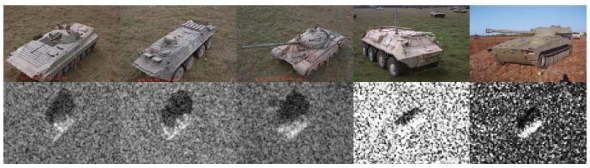


Figure 1. Sample images for 5 out of 10 classes in the MSTAR data set. First row shows the illustrative real life images and second row shows the SAR images from the data set.

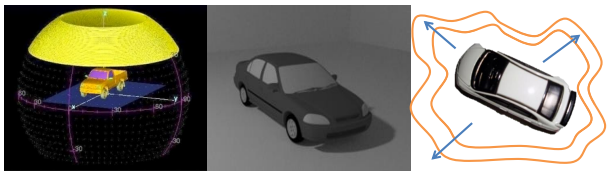


Figure 2. The image synthesis model (left), sample image (center), and acoustic attenuation model (right) for the CVDOME data set.

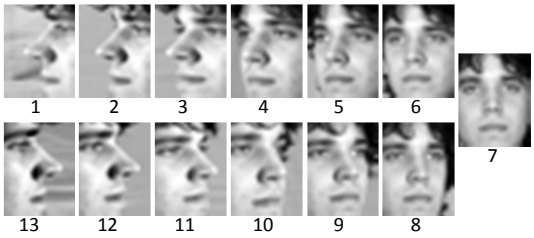
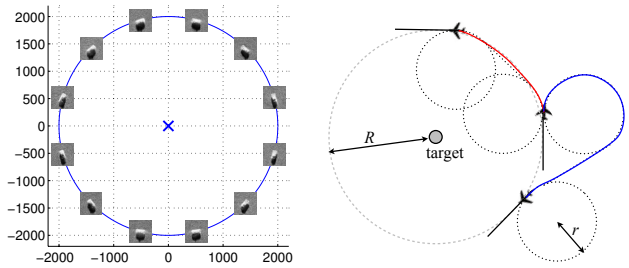


Figure 3. Sample face images of one subject under 13 poses (indexed below) in the Multi-Pie data set.

The CVDOME contains simulated X-band scattering images for 8 classes of civilian vehicles, with simulation layout and a sample image shown in Fig. 2. The azimuth angles of vehicles are quantized into 6 values. The images for each class under each view are randomly divided into 70 images for training and 30 images for testing. We also extend CVDOME with audio data for multi-modal sensor selection by collecting the engine sounds for the 8 vehicle classes from the Youtube. The sounds are attenuated differently in 6 view directions according to the vehicle shapes (illustrated on the right of Fig. 2).

We also test on the CMU Multi-Pie data set [10] for multi-view face recognition. There are 15480 images of 129 subjects collected in 3 sessions for training, and 5160 images in another session for testing. Face images with 13 poses are used as data collected from 13 different views, shown in Fig. 3. The images of one subject in one session under one illumination are combined as a sample instance.

In all the experiments, we use the quantized PCA projections of raw pixels and Mel-Frequency Cepstral Coef-



(a) (b)

Figure 4. (a) The layout of view positions along a circle centered at the target location. The image plotted at each view point corresponds to an observation of the target at the view angle. (b) The Dubins curves for an airplane to travel from one view point to two other view points.

Table 1. Geometry settings for three scenes.

scene parameter	S1	S2	S3
target range	4,350	800	2,000
straight line length	234	234	234
turning radius	500	1000	800
total distance	15,000	35,000	20,000

Table 2. Recognition accuracies (%) on the MSTAR data set under three scene settings.

method	S1	S2	S3
Infor [5]	57.6	87.0	79.4
Nearest	77.5	80.0	82.2
Infor+Nearest	72.8	86.3	84.4
POMDP [2]	81.9	89.6	84.0
POMDP-RC	**86.7**	**90.9**	**88.2**

ficients (MFCCs) as the observations from visual sensors and acoustic sensors, respectively. In vehicle classifications, 20-dimensional PCA and 8-level quantization are used. In the more challenging face recognition task, 50-dimensional PCA and 1024-level quantization are used.

An efficient Monte-Carlo Value Iteration (MCVI) algorithm [1] is used here to find an approximated solution for POMDP. With 10,000 particles, MCVI can typically learn a satisfactory policy around 1 or 2 hours on a server with 16 Xeon 2.40GHz cores. The policy is learned in a tree structure which traverses between state beliefs upon different observations. The rewards of $r_p = -5$, $r_c = 10$ and discount factor $\gamma = 1$ are found to work well in general cases. SVM classifiers are used for POMDP-DC, which are trained with the LibLinear package [9].

4.2. Evaluation of POMDP-RC

We first evaluate the performance of POMDP-RC on multi-view vehicle image classification with constrained motion and sensing resources. The scenario is illustrated in Fig. 4 where a mobile agent can navigate and observe the target from several view positions evenly located on a circle

Table 3. Recognition accuracies (%) on the CVDOME image data set under three scene settings.

method	S1	S2	S3
Infor [5]	78.1	83.3	74.5
Nearest	88.5	77.6	77.4
Infor+Nearest	88.8	80.5	80.4
POMDP [2]	87.8	84.3	80.9
POMDP-RC	**90.7**	**84.7**	**83.3**

centered at the target. The agent travels between two view positions with the shortest trajectory defined by Dubins curve [7], and move along a straight line when acquiring the SAR image to form a synthetic aperture. At each view position, the agent can observe one dimension of the PCA feature in a way similar to compressive sensing [6]. The total traveling distance and total number of observations are the resources under budget control. We consider three scenes (S1, S2, S3) with different geometric settings and total distance constraints as summarized in Table 1. The total observations are constrained not to exceed 12 in all the scenes.

Our POMDP-RC is compared on the two vehicle data sets with several baseline methods including: "Infor" [5], a greedy approach based on information measure; "Nearest" which always observes the nearest view first; "Infor+Nearest", a weighted combination of the previous two; and POMDP [2] with a conventional formulation. From Table 2 and 3, we can see most baseline methods can perform well in some scenes but fails in others. On the contrary, POMDP-RC can adapt its strategy according to available resource budgets, and therefore achieves the highest accuracies in all the cases.

We further investigate multi-view, multi-modal classification with the audio-visual CVDOME data. In this experiment, we assume the agent can choose to obtain each observation from either an audio or a visual sensor. Generally, images contain more information about target class than audio data but also consume higher memory and bandwidth. We further assume each audio(visual) observation takes 1(4) units of memory respectively, and the system has a total of 8 units available. The geometric settings are the same as in S2. Table 4 gives the performance comparison between different methods, including "Lowest mem" which always selects the memory-efficient audio sensors from the nearest view. The first column shows POMDP-RC has the highest accuracy, and the second and third columns show that POMDP-RC makes good use of available resources in the sense that the average remaining budgets at the time recognition is done are low for both memory (rm mem) and distance (rm dist). The last column shows the average number of observations made with audio and visual sensors. POMDP-RC makes only one memory-costly visual observation and allocates the remaining memory for more audio

Table 4. Recognition accuracies (%), remaining resources and observation allocations on the audio-visual CVDOME data set.

method	acc	rm mem	rm dist	#obs (a+v)
Lowest mem	72.6	3.00	2814.6	5.00+0.00
Infor [5]	70.0	0.28	12699.4	0.78+1.74
Nearest	73.1	0.00	6461.0	4.00+1.00
POMDP [2]	76.0	1.61	1896.6	2.39+1.00
POMDP-RC	**77.9**	0.24	1640.3	3.76+1.00

Table 5. Recognition accuracies (%) and average number of observations on the 3-class toy data with different quantization schemes.

quantization	acc	#obs
1-bit uniform	33.65	6
2-bit uniform	63.51	3
3-bit uniform	83.33	2
POMDP-RC adaptive	**92.57**	3.8

observations.

POMDP-RC can also be used for adaptive quantization when observation comes from continuous features, which is demonstrated below through a synthetic example in Fig. 5. As many POMDP solvers work on discrete observation $\mathcal{O}$, uniform quantization is often applied to the features as preprocessing. However, observations useful for discriminating different classes may come from a small range (*e.g.*, the left part of the 1-D distributions in Fig. 5). In this way, a very small quantization step has to be used in a uniform quantizer in order to capture all the discriminative information (*e.g.*, we have to use the 3-bit uniform quantizer shown in the top middle of Fig. 5 to distinguish all the three classes).

This problem can be solved by introducing quantization actions $\mathcal{A}_q$ in POMDP-RC, which specify all kinds of quantization functions with different quantization levels. In this example, we assume a limited sensing bandwidth is imposed and there are only 6 bits available to encode the quantized observations. We use POMDP-RC to select the quantization action that both requires very few bits to encode and preserves discriminative information as well (as shown on the right of Fig. 5). With those learned quantizers, POMDP-RC can make multiple informative observations and achieves a good recognition result. The accuracy of adaptive quantization using POMDP-RC is compared with several uniform quantizers in Table 5. The 1-bit or 2-bit uniform quantizer cannot capture sufficient class information, and the 3-bit uniform quantizer wastes too much bandwidth on unlikely or noninformative observations. POMDP-RC achieves the highest accuracy and acquires more observations than 2-bit and 3-bit uniform quantizers with the same bandwidth budget.

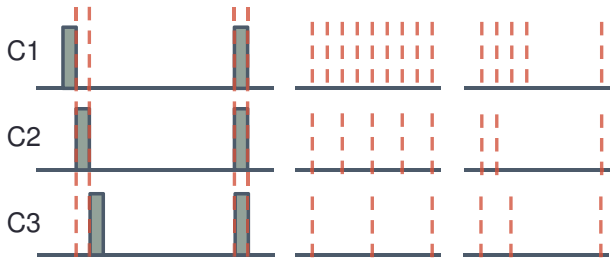


Figure 5. Distributions of 1-D toy data from 3 classes (left), with uniform quantizers (middle), and the non-uniform quantizers learned by POMDP-RC (right).

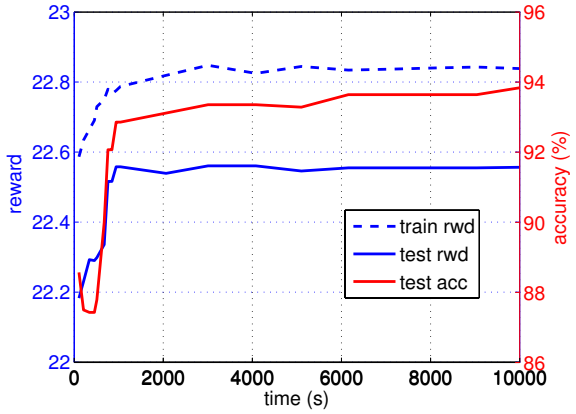


Figure 6. Reward function and test accuracy during the policy learning for POMDP-DC on the MSTAR data set.

4.3. Evaluation of POMDP-DC

In the following, we focus on evaluating the recognition performance of POMDP-DC with only the number of observations or features being constrained.

We first use the MSTAR data and train SVM classifiers for each view. The classification margin reward in (10) is optimized during policy learning for POMDP-DC, which is plotted in Fig. 6 versus training time. It can be seen that the reward value and classification accuracy on test data both increase with the training reward, indicating (10) is an effective objective for classification. The effect of margin parameter δ in (10) on test accuracy is studied in Fig. 7. It is observed that a too small δ leads to lower accuracy as it cannot ensure sufficient safe margin between classes; while a too large δ may disrupt the goal of classification and has poor performance. We set δ=0.8 which gives the highest accuracy.

In this experiment, we want to make observations from 3 actively selected views to achieve best recognition. Different combinations of classifiers and view planning methods are evaluated, with accuracies reported in Table 6. SVM classifiers trained for each view independently achieve much higher accuracy than naive Bayesian classifiers. POMDP-DC can further improve over the static

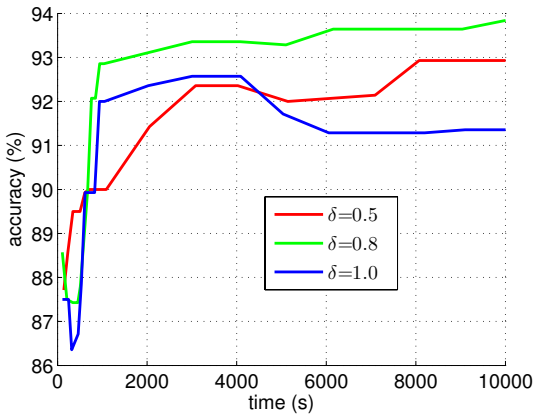


Figure 7. Test accuracy during the policy learning for POMDP-DC with different values of margin δ on the MSTAR data set.

Table 6. Recognition accuracies (%) on the MSTAR data with different classification and view selection models.

classifier	view selection	acc
naive Bayesian	single view average	49.89
	static views 2+6+7	81.00
	Infor [5]	89.93
	POMDP [2]	92.21
single-view SVM	single view average	52.51
	static views 2+6+7	89.86
	POMDP-DC	93.86
multi-view SVM	static views 2+6+7	90.07
	POMDP-DC	94.29
	POMDP-DC-MO	**95.57**

selection of top 3 views (2+6+7) with the best single-view performances, and also outperforms baseline methods Infor [5] and POMDP [2]. We also train SVMs in the concatenated feature space of each pair of views to model between-view high-order statistics, which are referred to as multi-view SVMs. Correspondingly, meta-observations collecting features from two views at once are added to POMDP-DC, and the resulting model is denoted as POMDP-DC-MO. As can be seen from the bottom rows in Table 6, both SVM classifier and POMDP-DC-MO view planner can benefit from the high-order information between views, leading to a much improved recognition rate than baseline approaches.

For a better understanding of the behavior of POMDP-DC-MO, we conduct another experiment in which a total number of 6 features, *i.e.* PCA dimensions, can be selected from the first view of MSTAR data. In each observation made by POMDP-DC-MO, up to 3 features can be collected as a meta-observation. Baseline POMDP-DC approaches collecting a fixed number of (1 and 3) features in each observation are used for comparison. Fig. 8 shows how the average recognition accuracy for each method increases with the number of features obtained. The single-feature

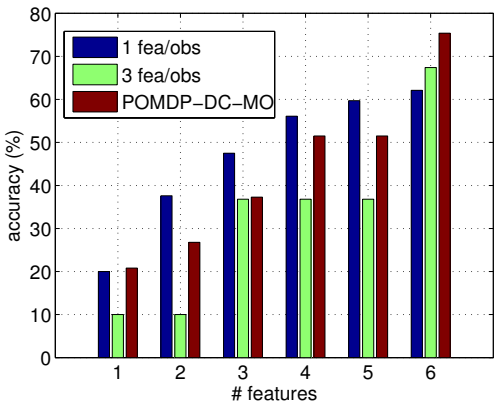


Figure 8. Recognition accuracies on the MSTAR data with different meta-observation selection methods.

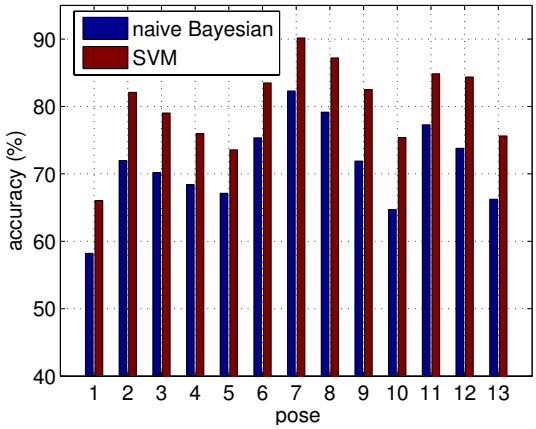


Figure 9. Recognition accuracies on the Multi-Pie data under different poses using naive Bayesian and SVM.

observation allows the most flexible sensing strategy which is adapted upon each feature observed, but it falsely assumes independence between all the features. On the other hand, the meta-observation with 3 features can model the high-order feature correlation, but it limits the number of opportunities to actively adjust sensing options. POMDP-DC-MO makes a tradeoff between the two and achieves the highest accuracy after observing all the 6 features. Note that the performance of POMDP-DC-MO is not the best before all the features are observed, which indicates our policy is optimized for a long-term goal instead of an immediate goal.

POMDP-DC is also tested on the Multi-Pie data, where we are restricted to select 3 poses for face recognition. The performance of SVM and naive Bayesian classifiers on each single pose are compared in Fig. 9. On this data set, the accuracies under different views vary a lot; nevertheless, SVM is consistently better than naive Bayesian classifier. Table 7 also confirms the advantage of SVM classifier, and shows that POMDP-DC can further improve over static selection of the 3 best single views (6+7+8). We also try to

Table 7. Recognition accuracies (%) on the Multi-Pie data with different classification and view selection models.

classifier	view selection	acc
naive Bayesian	static views 6+7+8	89.53
	Infor [5]	91.09
	POMDP [2]	90.31
single-view SVM	static views 6+7+8	95.35
	POMDP-DC	96.24
	POMDP-DC w/ DiscObs	96.94
multi-view SVM	POMDP-DC-MO	**97.33**

improve the discriminative power of our observation model by defining $O(s', a, o) \propto \exp\{s_{x',z',a}(o)\}$, where $s_{x',z',a}(o)$ is the SVM score function. The resulting discriminative observation model brings a 0.7% improvement in accuracy compared with the generative one, as shown in the row for "POMDP-DC w/ DiscObs" in Table 7. Moreover, as in the previous experiment on MSTAR, POMDP-DC is configured with multi-view SVM and meta-observation on pairs of views, and the highest accuracy of 97.33% is obtained in this setting. By examining the views selected by POMDP-DC, we find that almost 40% of the time the view combination of 2+3+7 or 2+7+8 is selected, which suggests that some side-view poses can provide complimentary information to the most discriminative frontal view poses.

5. Conclusions

We present a novel POMDP model for active object recognition under limited motion and sensing resources with two key components introduced. First, heterogeneous resource constraints are explicitly monitored in the state variable rather than indirectly penalized in the reward function, leading to a new reward function with more focus on long-term recognition performance. Second, discriminative classifiers with high-order class information are incorporated in place of conventional generative classifiers, and the reward function and observation model are customized accordingly. The proposed model proves to be effective in terms of both recognition accuracy and resource management in multi-view/multi-modality/multi-quantizer active sensing tasks for vehicle classification and face recognition.

In future, we are interested in extending the model for a wider range of pattern recognition problems in dynamic and interactive environments.

Acknowledgement

The work is supported by the U.S. Army Research Laboratory and U.S. Army Research Office under grant number W911NF-09-1-0383.

References

[1] H. Bai, D. Hsu, W. S. Lee, and V. A. Ngo. Monte carlo value iteration for continuous-state POMDPs. In *Algorithmic foundations of robotics IX*, pages 175–191. Springer, 2011.

[2] D. Blatt and A. O. Hero. Optimal sensor scheduling via classification reduction of policy search (CROPS). In *International Conference on Automated Planning and Scheduling*, 2006.

[3] H. Borotschnig, L. Paletta, M. Prantl, and A. Pinz. Appearance-based active object recognition. *Image and Vision Computing*, 18(9):715–727, 2000.

[4] T. Darrell and A. Pentland. Active gesture recognition using partially observable markov decision processes. In *13th International Conference on Pattern Recognition*, volume 3, pages 984–988, 1996.

[5] J. Denzler and C. M. Brown. Information theoretic sensor data selection for active object recognition and state estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(2):145–157, 2002.

[6] D. L. Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52(4):1289–1306, 2006.

[7] L. E. Dubins. On curves of minimal length with a constraint on average curvature, and with prescribed initial and terminal positions and tangents. *American Journal of mathematics*, pages 497–516, 1957.

[8] K. E. Dungan, C. Austin, J. Nehrbass, and L. C. Potter. Civilian vehicle radar data domes. In *SPIE Defense, Security, and Sensing*, pages 76990–76990, 2010.

[9] R.-E. Fan, K.-W. Chang, C.-J. Hsieh, X.-R. Wang, and C.-J. Lin. Liblinear: A library for large linear classification. *The Journal of Machine Learning Research*, 9:1871–1874, 2008.

[10] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker. Multi-PIE. *Image and Vision Computing*, 28:807–813, 2010.

[11] L. He, S. Ji, and L. Carin. Application of partially observable markov decision processes to robot navigation in a minefield. *ICAPS 2006*, page 14, 2006.

[12] S. Karayev, T. Baumgartner, M. Fritz, and T. Darrell. Timely object recognition. In *Advances in Neural Information Processing Systems*, pages 899–907, 2012.

[13] S. Karayev, M. J. Fritz, and T. Darrell. Dynamic feature selection for classification on a budget. In *International Conference on Machine Learning (ICML): Workshop on Prediction with Sequential Models*, 2013.

[14] C. Kreucher, K. Kastella, and A. O. Hero. Sensor management using an active sensing approach. *Signal Processing*, 85(3):607–624, 2005.

[15] W. S. Lovejoy. A survey of algorithmic methods for partially observed markov decision processes. *Annals of Operations Research*, 28(1):47–65, 1991.

[16] L. Paletta and A. Pinz. Active object recognition by view integration and reinforcement learning. *Robotics and Autonomous Systems*, 31(1):71–86, 2000.

[17] T. D. Ross, S. W. Worrell, V. J. Velten, J. C. Mossing, and M. L. Bryant. Standard SAR ATR experiments using the MSTAR public release data set. *Proc. SPIE Algorithms for Synthetic Aperture Radar Imagery V*, 3370:566–573, 1998.