Effective Online Detection of Task-Independent Landmarks

Martin V. Butz BUTZ@ILLIGAL.GE.UIUC.EDU
Illinois Genetic Algorithms Laboratory, University of Illinois, Urbana-Champaign, IL

Samarth Swarup SWARUP@UIUC.EDU
Department of Computer Science, University of Illinois, Urbana-Champaign, IL

David E. Goldberg DEG@ILLIGAL.GE.UIUC.EDU
Illinois Genetic Algorithms Laboratory, University of Illinois, Urbana-Champaign, IL

Abstract

One of the key problems in the development of competent adaptive autonomous agents is the learning of hierarchical cognitive structures including predictive world models. The problem may be approached by extending the reinforcement learning framework with semi-Markov decision processes (SMDPs). In SMDPs, the notion of an action is extended by *options*, that is, actions extended in time. To learn such options, states or events need to be identified that denote the potential beginning or ending of an option. We believe that most suitable beginnings or endings are states or transitions that partition the environment into relatively independent subregions. This paper is concerned with the detection of such *landmarks*. Using notions of surprise and consolidation via continued novelty, implemented by relatively simple statistics on the sensory inputs, we introduce a mechanism that reliably identifies landmarks that partition the environment appropriately. The resulting set of landmarks should be very useful for the formation of an online adaptive hierarchical problem representation enabling efficient adaptation and cognition.

1. Introduction

Recent insights in adaptive behavior research showed that flat learning architectures are only suitable in small problems. The reinforcement learning literature suggests that a scalable learning system should be able to partition its environment into useful subproblems generating higher level actions, termed *options* (Sutton et al., 1999). The resulting learning framework can be formulated in a *semi-Markov decision process* (SMDP) in which an action—then termed an option—can extend over several time steps. In general, scalable behavioral and/or environmental representations need to be hierarchically structured. The consequently smaller subtasks can be solved more effectively by, for example, reusing previously learned behavioral patterns or by encapsulating the subtasks from the rest of the environment (Drummond, 2002; Parr & Russell, 1997). The hierarchical problem representation allows faster cognitive processing potentially improving learning, planning, prediction, and adaptation.

Relatively little has been said, however, about how to select landmarks that are useful for generating options and for building up a hierarchical representation of the environment. Often, researchers assume that a good set of landmarks is available, and then concentrate on using these landmarks to do hierarchical planning (Voicu & Schmajuk, 2001; Kortenkamp et al., 1994). (Kuipers, 1998) builds a *spatial semantic hierarchy* which is bootstrapped by a landmark detection system based on the detection of distinctive states (Pierce & Kuipers, 1994). (Rizzi et al., 1997) stress the importance of landmarks as a bridge between the sub-symbolic world of control and the symbolic world of planning. However, their landmark detection system is based on matching a set of templates with a local occupancy grid built by sonar measures. This provides no constraint on how good the landmarks are at summarizing the environment.

To develop hierarchical problem representations online, effective landmark detection mechanisms are necessary that decompose the encountered problem space well. However, to the author's knowledge, no landmark detection mechanism exists that considers the

Preliminary work. Under review by the International Conference on Machine Learning (ICML). Do not distribute.

significance of landmarks for a proper environment decomposition. In the simplest case, landmarks are plainly selected after fixed time intervals (Duckett & Nehmzow, 1998; Yamauchi & Beer, 1996), or by a predefined set of feature properties that define a landmark (Rizzi et al., 1997). Slightly advanced techniques focus on perceptually significant events, making use of the notion of salience. In this method, landmarks are usually identified as unusual and/or unexpected sensory inputs (Marsland et al., 2001; Fleischer & Marsland, 2002). The notion of *surprise* makes this idea more concrete: A landmark is defined as an event in which an environmental change is observed that is unexpected. A predictive environmental model (one that allows the prediction of action consequences) is used to predict, partially action-dependent, environmental changes (Fleischer et al., 2003). If the changes are significantly unexpected, then a landmark is detected. However, again, the approach does not evaluate the quality of the landmarks in terms of effective partitioning of the environment.

In the reinforcement learning literature, several researchers have addressed the importance of accelerating RL via hierarchical structures (Sutton et al., 1999; Dietterich, 2000; Drummond, 2002; Barto & Mahadevan, 2003). While all papers emphasize the importance of automatically learning such hierarchies, most of the papers focus on what to do, when problem knowledge about suitable hierarchies is available. Drummond (Drummond, 2002) additionally addressed the detection of decision boundaries which somewhat coincide with our notion of landmarks. The approach relies on reinforcement gradients and global problem knowledge. Nonetheless, the combination with a hierarchical reinforcement learner showed a very promising performance increase. Another approach to the detection of landmarks, referred to as subgoals, can be found in (McGovern & Barto, 2001). Essentially, the approach detects bottleneck states that are always visited when a goal is reached. Again, the approach is task-dependent. A more local task-independent approach can be found in (Hengst, 2002). Hengst's HEXQ approach keeps frequency measures of feature changes and forms hierarchies with respect to those frequencies. The approach exhibits promising performance in the taxi task but appears limited to deterministic problems due to its strong feature dependency.

The aim of this paper is to introduce an extended landmark detection mechanism that takes the idea of effective partitioning of the environment into account. The mechanism is designed to work online in an autonomous adaptive agent. No global problem knowledge such as a complete environmental model is necessary. While we validate our approach in the abstract four-rooms problem (Sutton et al., 1999) (shown in Figure 1), we emphasize that the mechanism proposed and the basic idea behind it should be broadly applicable and essentially adjustable for many learning approaches and many problem representations. In particular, our algorithms are not limited to perceptual spaces. A robot arm, for example, could use our algorithms to explore its configuration space and identify key configurations, such as elbow locked, object gripped and so forth. These would then enable action planning at a higher level. However, we also show the omnipresent representational dependence and problem dependence of the approach so that in different representations and problems, different but similar mechanisms may be applied.

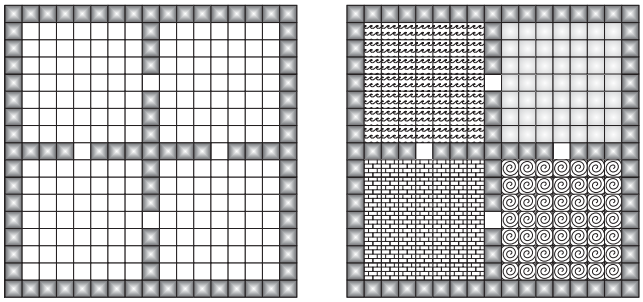

Figure 1. The four-rooms problem and the same problem with an additional color (texture) feature.

We believe that the notion of surprise is important in detecting "transitional" states, i.e. states (such as doorways) that are transition points between relatively independent sub-regions of the environment. We use simple statistics on the stream of sensory inputs, to discover significant transitions in the environment.

The remainder of this paper is structured as follows. First, we introduce our basic online landmark detection mechanism via a notion of surprise and evaluate it in the four-rooms problem. Next, we add a landmark consolidation mechanism using persistent novelty as the criterion. The consequent results show the effective and reliable detection of landmarks even in very noisy problem settings. Summary and conclusions outline the further potential of the proposed mechanism.

2. Online Landmark Detection

As discussed in the previous section, we are interested in an online landmark detection mechanism, that detects landmarks that characterize transitions between relatively independent subspaces in an environment. To do this, we proceed in two steps. First, we in-

troduce a general notion of *surprise* or *novelty* and investigate the performance of the mechanism in the four-rooms problem. Next, we enhance the mechanism with a filtering algorithm that makes the landmark detection mechanism noise robust and focuses even more on the detection of transitional states.

2.1. Detection Via Surprise

Surprise mechanisms have been implemented before in various settings. Our mechanism differs in that we use local moving averages only to detect globally significant landmark states.

We first address the four-rooms problem in which each room has a different additional property (such as a color or texture of the room) as shown in Figure 1 (right-hand side). The perception of a state consists of the x and y coordinate of the state and the additional color attribute. Movements are possible to the eight surrounding positions as long as they are not blocked by an obstacle.

The algorithm to detect landmarks based on surprise is described in Algorithm 1. Essentially, a moving average and a moving variance are kept for each perceptual feature F (in our case the X coordinate, the Y coordinate, and the color attribute). Only the characteristic of this attribute is kept, that is, if it remains the same after a movement (coded by 0) or if it changed (coded by 1). A surprise is triggered when the characteristic of the attribute is *significantly different* from its normal behavior. The significance criterion is loosely based on a 95% confidence level by defining a surprising event as an event in which the current difference is larger than the moving average plus twice the moving standard deviation. Thus, for the changing feature criterion as specified in Algorithm 1 a current change becomes significant if the animat remains long enough in one room so that the moving average and standard deviation become sufficiently small to drop below one.

Figure 2 shows that in the deterministic colored four-rooms environment, surprise-based landmark detection works perfectly. As in all experiments in this paper, the mean number of landmarks detected for each state after executing $100,000$ random steps in the environment are plotted. The results show averages over 100 independent runs. Additionally, to be able to compare the difference between the means of the states, we plot three times the standard deviation of the mean divided by the square root of the number of experiments to show the 99% confidence level comparing the means. The runs are generated with a moving average update rate $\delta = 0.1$. A landmark is considered as the transition from one position to the next so that the positions neighboring a doorway are part of the detected landmarks. Since the doorway can be reached from six positions, each neighboring position is considered as a landmark approximately one-sixth as often as the actual doorway position. Figure 3 shows the same successful detection for the case in which the doorway is not colored differently but has the same color as one of the neighboring rooms.

Algorithm 1 Description of basic mechanism for detecting a surprising event—specific for perceptual features. A surprise may be directly considered as a landmark or it may undergo a further consolidation process (see Algorithm 3). The function takes as input the moving average $\bar{\mu}_F$ and moving variance $\bar{\sigma}^2_F$ of the change of the investigated feature F as well as the previous value F_{t-1} and the current value F_t of the feature. Value L_{score} indicates the significance of the surprise.

IS SURPRISING($\bar{\mu}_F$, $\bar{\sigma}^2_F$, F_{t-1}, F_t):

```
1  diff ← 0
2  if(F_{t-1} ≠ F_t)
3     diff ← 1
4  Is_surprised ← FALSE
5  L_score ← abs(diff − μ̄_F)/√σ̄²_F
6  if(L_score > 2)
7     Is_surprised ← TRUE
8  σ̄²_F ← σ̄²_F + δ((diff − μ̄_F)² − σ̄²_F)
9  μ̄_F ← μ̄_F + δ(diff − μ̄_F)
10 return (Is_surprised, L_score)
```

General Notion of Surprise Although it is imaginable that each room looks different in a real-world environment, or in general, that relatively encapsulated subspaces exhibit different perceptual features, the coloring approach is somewhat unsatisfactory. Thus, instead of using perceptual features, we now use a time-delay measure as the surprise criterion. Instead of keeping track of the change of each perceptual feature, we keep a moving average (and moving variance) of the difference between the current time and the last time of the occurrence of a perceptual pattern. To do this, we keep a list of each perceptual pattern seen so far, and an accompanying time stamp that denotes when the pattern was seen last. Thus, entering a new state, we can decide if this state is surprising dependent on the state-related time delay.

The general algorithm for a surprising event is shown in Algorithm 2. In the time-delay criterion case, the ΔX input variable would specify the current time delay with respect to the current state. In the perceptual

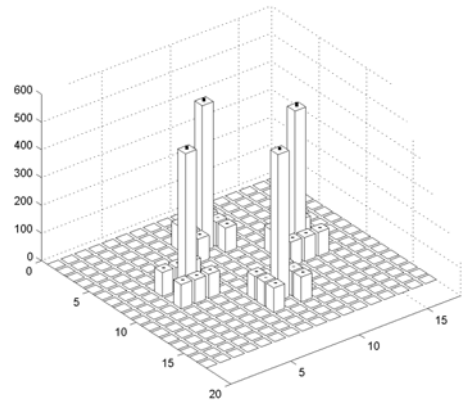

Figure 2. In the deterministic colored four-rooms problem, the landmark detection based on changing sensory information works perfectly.

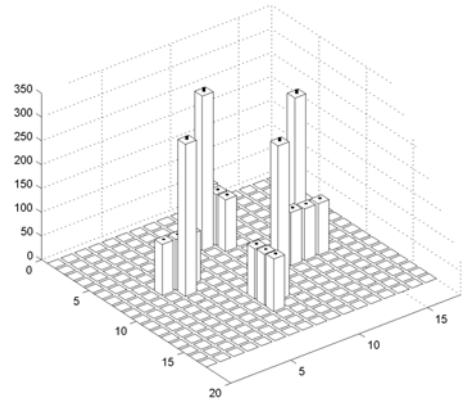

Figure 3. If the doorway is not colored differently from one of the neighboring rooms, the landmark detection still identifies the doorways reliably. However, only the transitions in which the color changes are considered.

feature case, ΔX would correspond to the `diff` variable in Algorithm 1.

Figure 4 shows that the time-delay criterion is able to reliably distinguish between correct and incorrect landmark candidates. However, the rate of detecting random states as landmarks is also quite high. When we decrease the learning rate δ to 0.01 (shown in Figure 5), the detection becomes even more noisy since the time delay measure does not adjust fast enough when remaining in one room over a short period of steps. Nonetheless, despite the simplicity of the criterion and the online nature of the mechanism, the doorways are detected more often as landmarks than the other states. Section 2.3 shows, though, that we can strongly improve this mechanism by introducing a consolidation mechanism that is based on persistent novelty.

Algorithm 2 Description of basic mechanism for detecting a surprising event—general algorithm. The function takes as input the moving average $\bar{\mu}_X$ and moving variance $\bar{\sigma}_X^2$ of the surprise criterion (in our experiments either perceptual change or time delay) and the current value of the criterion (perceptual change or time delay with respect to the current input). Value L_{score} indicates the significance of the surprise.

IS SURPRISING($\bar{\mu}_X$, $\bar{\sigma}_X^2$, ΔX)*:*

```
1 Is_surprised ← FALSE
2 L_score ← abs(ΔX − μ̄_X)/√σ̄²_X
3 if(L_score > 2)
4     Is_surprised ← TRUE
5 σ̄²_X ← σ̄²_X + δ((ΔX − μ̄_X)² − σ̄²_X)
6 μ̄_X ← μ̄_X + δ(ΔX − μ̄_X)
7 return (Is_surprised, L_score)
```

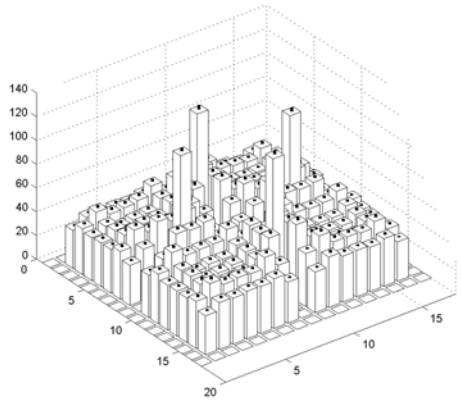

Figure 4. In the plain four-rooms problem, surprise due to significant time delay suffices to detect doorways quite reliably.

2.2. Addition of Noise

Despite the promising performance of both algorithms in the deterministic case, the algorithms are not very noise robust. Figure 6 shows that states are basically identified randomly as landmarks when we add 30% noise to the color attribute (with a 30% probability, the color attribute is set randomly to one of the five values). Due to the occurrence of an unusual perceptual change with 24% probability, the mean average and variance drop sometimes below the changing event level and consider random uncommon perceptual changes as surprising events.

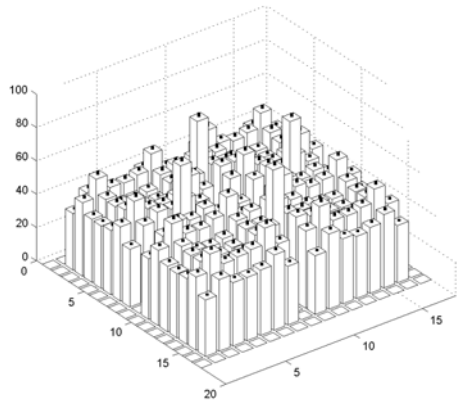

Figure 5. Decreasing the moving average update rate, the adaptation of the average values is not fast enough and the continuous novelty of the states in a new room causes additional landmark detections.

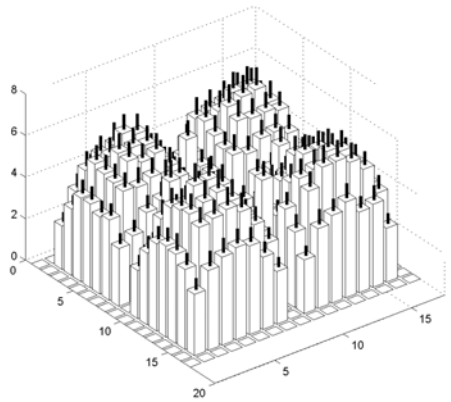

Figure 6. When the color attribute is set to a random value with 30% probability, the simple perceptual feature-based detection is not effective anymore.

Similarly, if we add only 20% noise to the color attribute in the time-delay case (the color attribute was constant before in this case), landmarks are again detected randomly (somewhat reflecting the frequency of encountering each state). Since the high time delay in the unusual perceptions for a particular state triggers the surprise criterion as often as the usual perceptions in the doorways the detection mechanism is effectively randomized.

The next section shows how to overcome this problem using a landmark consolidation mechanism based on persistent novelty.

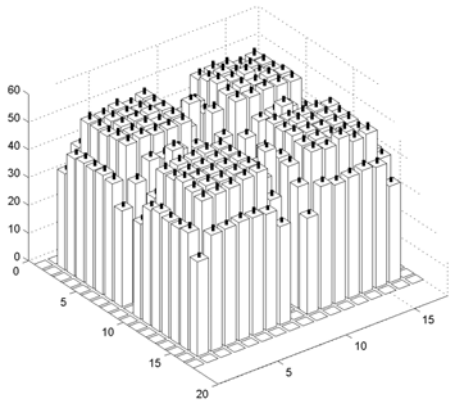

Figure 7. The simple time-delay-based surprise measure is not sufficient anymore to detect the doorways in the setting with 20% random noise in the color perception.

2.3. Consolidation by Persistent Novelty

Although the doorways were reliably identified as a landmark in the deterministic case, the mechanism (regardless if it was based on perceptual features or on time delay) broke down when a significant amount of noise was added to the perceptions. Additionally, the key criterion for a successful landmark detection mechanism was that we detect states or transitions that connect more-or-less independent subregions. Although our surprise criterion implicitly realizes the detection of such states, noise disrupts this property. We now introduce an additional criterion that filters the set of surprising states explicitly looking for surprising states that connect more-or-less independent subspaces. The key idea for this is that not only the landmark should be surprising but the successive states should be surprising as well since this is essentially the case when entering a new subspace in the environment.

Thus, we endow our system with a short term memory in which we store the current surprising state with the additional information of where and why it was surprising in the first place. In the successive states, we update the *score* of the potential landmark, evaluating if the surprise (or novelty) is persistent. Algorithm 3 formalizes the algorithm. In the case of the perceptual-feature based criterion, measure ΔX specifies if the current perceptual feature value is different from the one in the event which triggered the surprise in the first place. The landmark score L_{score} reflects the distance between the detected difference and the mean in units of standard deviations. Initially, the score has a value of more than two since only if the current difference is larger than twice the standard deviation a

surprise is triggered in the first place. Similar to the global moving average we use a simple delta-rule learning update for the moving average of the landmark (in our experiments $\delta_E = 0.1$). Note that the landmark-specific moving average is updated using the difference between the current value and the value when the landmark was detected in the first place. Thus, it is different from the global moving average. The initial values correspond to the values determined in Algorithm 2. A surprising event is considered as a landmark only if the score stays high for an extended period of time. This time period criterion is very tunable, however our experiments showed that as long as the score is required to stay sufficiently high (in our experiments at least $LMT_{dmin} = 1.5$ and dependent on the time delay to the initial surprise event scaled by $LMT_{dd} = 2.$) for an extended period of time (at least $LMT_{min} = 10$ and for sure after $LMT_{max} = 20$), the mechanism works very reliable.

Algorithm 3 Description of the algorithm of deciding if a surprising event should become a landmark (perceptual feature specific). Essential is that the encountered event remains persistent, i.e. the observed change in an attribute persists over an extended period of time. The algorithm takes as input the current score of the potential landmark L_{score}, its age L_{age}, its moving average and variance ($\bar{\mu}_E$, $\bar{\sigma}_E^2$), as well as the current landmark-specific persistence criterion ΔX.

CHECK FOR PERSISTENT NOVELTY(L_{score}, L_{age}, $\bar{\mu}_E$, $\bar{\sigma}_E^2$, ΔX)*:*

```
1 Is_landmark ← FALSE
2 L_score ← L_score + δ_E(abs(μ̄_E − ΔX)/√σ̄²_E − L_score)
3 if(L_score < LMT_dmin)
4    Is_landmark ← FORGET_LANDMARK
5 if(L_age > LMT_max OR (L_age > LMT_min AND
       L_score > LMT_dmin + LMT_dd*
         (LMT_max − L_age)/(LMT_max − LMT_min) ))
6 Is_landmark ← TRUE
7 return Is_landmark
```

Figure 8 shows the performance of the consolidation mechanism in the feature criterion case with the settings specified above. It can be seen that the doorway is now again detected as a landmark very reliably despite the 30% randomness in the color attribute. In these runs the short-term memory size is set to one so that if another surprise is triggered while a surprising transition is in memory, the new transition is disregarded. Note that although the detection is reliable and the difference between doorways and other states is significant, very few states are detected in general (on average each landmark is detected twice in 100,000 runs). However, 30% noise in the color attribute is also very strong noise and very few states are surprising in the first place (see Figure 6). When increasing the short-term memory size to five items as shown in Figure 9, a few more random states are identified as landmarks since the random changes after a doorway are also considered for landmark candidates.

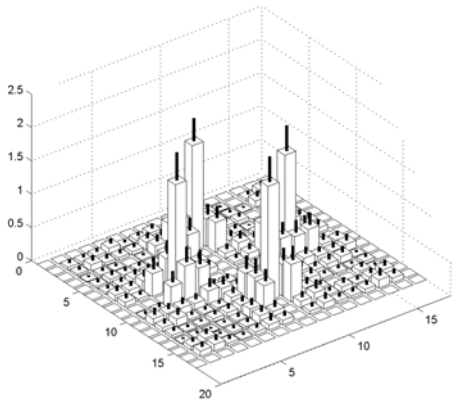

Figure 8. Requiring persistent surprise, the doorways are detected reliably even with a 30% random noise in the color attribute.

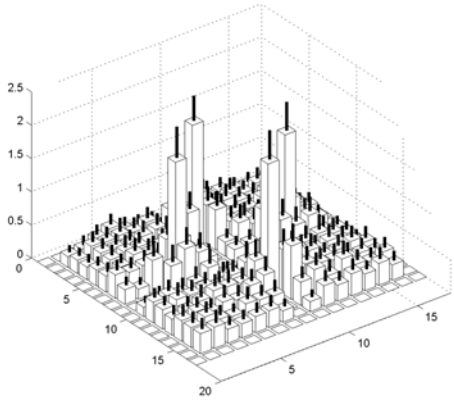

Figure 9. If the short term memory for landmark candidates is increased to five, states inside the room are also detected more frequently since the states after a surprising doorway often are surprising as well.

Again we can apply the same idea for other features and for other novelty (or surprise) criteria. We consequently apply the consolidation via persistent novelty mechanism also for the time delay criterion so that the ΔX is set to the difference between the current time and the time the current state was seen last. Figure 10 shows that with the addition of the consolidation via persistent novelty method the doorways can be de-

tected as landmarks much more reliably than in the simple surprise-based detection case.

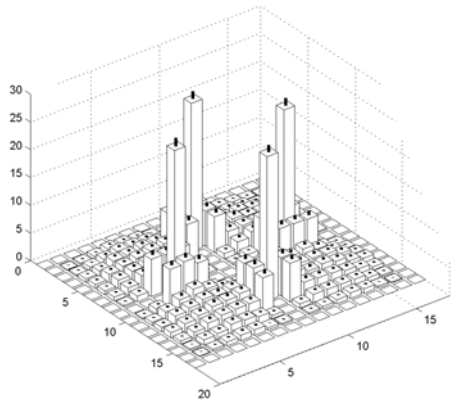

Figure 10. With the addition of the consolidation via persistence, the detection of doorways in the time-delay-based landmark detection mechanism is strongly improved.

Also in the case with 20% noise in the color attribute, the consolidation mechanism detects the doorways reliably (Figure 11). The number of landmarks detected drops significantly since the moving variance is very high and thus few transitions succeed in remaining persistent over an extended period of steps. This hypothesis is confirmed in the case when we add 100% noise to the color attribute. In this case, less states are filtered and the detection mechanism works actually slightly more reliably (Figure 12). This is the case since the moving average of the variance is actually lower in this case. A simple calculation confirms this fact: Considering only one state that is continuously revisited, in the 0.2 noise case, the mean occurrence of the normal perception in that state is $1/0.84 = 1.19$ steps on average. The mean occurrence of one of the four other perceptions in that state is $1/0.04 = 25$. Thus, the mean delay between seeing a state is $0.84 \cdot 1.19 + 0.16 \cdot 25 = 5$ (as expected since there are five possible perceptions). This is also equal to the case in which each perception is equiprobable. However, in the case in which each state is equiprobable the expected variance is actually zero whereas in the 0.2 noise case, the expected variance equals $0.84(5-1.19)^2 + 0.16(5-25)^2 = 76.19$. Of course, the variance is higher in our four-rooms problem since there are actually 200 states and each state has five perceptual codes (thus actually 1000 possible perceptual input patterns) and we use a moving average and moving variance instead of the actual average and variance. However, the approximation shows that the variance is higher on average in the 0.2 noise case which explains the less frequent landmark detection in this case.

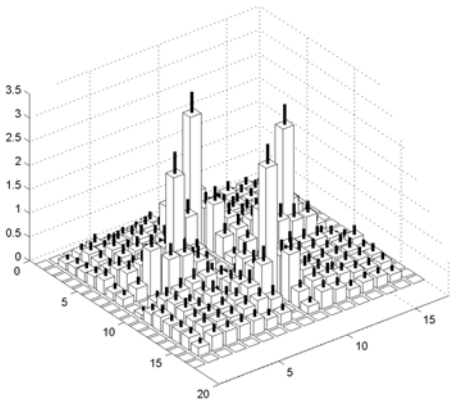

Figure 11. Although the persistent measure filters lots of states in the setting with 20% random noise in the perceptual feature, doorways are reliably identified as landmarks using the time-delay criterion.

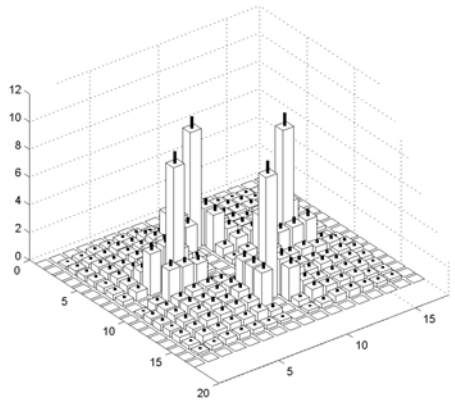

Figure 12. Even when adding 100% noise to the color attribute (all five values are equally probable) the doorways are still detected reliably using the time-delay criterion.

In sum, the results confirm that the general notion of surprise plus consolidation via persistent novelty work very reliably and very robustly in the investigated problems. Although the doorways as the landmarks seem to be very obvious when looking at the maze from a global perspective, for the local perceptual perspective of the system the detection is very hard. The investigated system uses no notion of neighborhood, dimensionality, or integer values (the perceived coordinates are treated like simple nominals). In the time delay case, the color attribute provides no information about the environment and serves only as a parameter that confuses the mechanism (in the event of noise, each state is perceived as one of five possible perceptual patterns). Thus only the implic-

itly perceived topology of the maze and the inherent independence of the four rooms were exploited by our mechanism. This indicates that the mechanism should be applicable in many other scenarios possibly using many other surprise and persistence criteria detecting many other landmarks that are transitional states between more-or-less independent subregions in the environment.

3. Summary and Conclusions

Landmarks suitable for a hierarchical decomposition of an environment are those which mark *transitional states* in the environment, that is, states that mark a transition from one subspace to the next. In the investigated four-rooms problem, these are the doorways. Other transitions could be characterized by touch, grip, influence range, change in movement speed or direction etc. For example, once a robot arm grips a tool, it is able to work with the tool. Thus, a successful grip could serve as a landmark indicating the transition from no interaction to a potential interaction. Again, only if the grip was successful (surprise) and the item stays in the robot hand (persistent novelty) the landmark should be detected, effectively filtering perceptual errors.

This paper showed that it is possible to detect such transitional states online with simple mechanisms based on notions of surprise and consolidation via persistent novelty. Although our investigations focused on the four-rooms problem, the basic principle appears to be applicable within many other environments potentially detecting many other types of transitional states. The detection criteria are not limited to perceptual features or complete perceptual states. Perceptual states or other perceptual features may be used in a similar fashion. Any pattern extracted from the environment could be endowed with a time stamp and a running time delay measure. Thus, the mechanism is readily applicable in, for example, neural-net type learning algorithms, since the hidden layer usually evolves a representation of relevant perceptual features (Baluja & Pomerleau, 1995). The same idea could also be used with kernel methods, which extract environmental features more explicitly.

Besides the further landmark detection evaluation and application of the concept of surprise and persistent novelty in other environments, learning frameworks, and surprise criteria, the integration of this mechanism into reinforcement learning and other adaptive behavior mechanisms is necessary. Of particular interest is the extension to the semi Markov decision process framework (SMDP) in which actions extend in time. As mentioned in the introduction, such higher level actions, called *options*, are suggested for the navigation between landmarks (Sutton et al., 1999). With the introduced landmark detection mechanism, it should be straight forward to combine detected landmarks with a higher level reinforcement learning system. Such a system may incrementally introduce higher level options that specify the possible behaviors in the subspaces between identified landmarks.

Also hierarchical predictive environmental models may be learned forming online hierarchical models of the environment suitable for, for example, model-based reinforcement learning (Kaelbling et al., 1996). The combination of such a mechanism with online generalizing model learning mechanisms, such as the anticipatory learning classifier system (Butz, 2002; Butz & Hoffmann, 2002), is also imaginable.

With respect to subspace identification, distance metrics may be used to cluster neighboring landmarks. Landmarks that are located next to each other may be combined into macro landmarks that would consist of a small set of neighboring states or features. For example, a short hallway may then be characterized by one single landmark.

Also, the identified subspaces may be characterized by the means of the neighboring landmarks. It is imaginable that due to the proper identification of subspaces by the means of the detected landmarks, non-Markov environments could be separated and solved as separate Markov environments by endowing each subspace with an internal identifier. This relates to the addition of internal states to the online learning system XCS (Lanzi, 2000) but the internal states would be set in a more direct way based on the detected landmarks.

Due the observed strong noise robustness we suggest the integration of the learning mechanism in real robot systems. A robot arm could detect different stages of activity as suggested above. A moving robot may detect doors or other transitional states such as a coffee automaton online (due to a significant persistent change in available activity).

Due to the general task-independence of the detected landmarks, they may also be useful in life-long learning tasks since the property of a transitional state may be useful for many different tasks. In particular, the integration of model-learning capabilities and multi-task representations with notions of motivations and emotions may be effective using the proposed landmark detection mechanism for effective navigation and planning between subtasks.

Finally, we want to mention the strong relation of

our landmark detection mechanism to actual cognitive aspects. As has been argued so many times before (e.g. (Goldberg, 2002)) our environment has an inherent hierarchical structure. Thus, also cognitive processes need to be predisposed to develop hierarchical structures mimicking the structures in the environment encountered during interaction with it. Straightforward examples of such hierarchical cognitive representations are the hierarchical structure of language (phonemes, morphemes, syllables, words...) or the general relation between words and meaning where words often relate to particular substructures in the world such as objects, transitions, encapsulated subspaces etc. By the means of our simple mechanism based on surprise and consolidation, modified in the appropriate form and enhanced to different kinds of representation and notions of novelty, various kinds of substructures may be identifiable in the encountered world. Due to the possible encapsulation, the substructures can be represented distinctively and consequently learned and adapted faster. Additionally, the learned representations and behavioral patterns for each substructure may be recombined in suitable possibly task-dependent and goal-oriented manners to allow hierarchical innovative cognitive processing.

References

Baluja, S., & Pomerleau, D. A. (1995). Using the representation in a neural network's hidden layer for task-specific focus on attention. *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence*, 133–141.

Barto, A. G., & Mahadevan, S. (2003). Recent advances in hierarchical reinforcement learning. *Discrete Event Dynamic Systems*, *13*, 341–379.

Butz, M. V. (2002). *Anticipatory learning classifier systems*. Boston, MA: Kluwer Academic Publishers.

Butz, M. V., & Hoffmann, J. (2002). Anticipations control behavior: Animal behavior in an anticipatory learning classifier system. *Adaptive Behavior*, *10*, 75–96.

Dietterich, T. G. (2000). Hierarchical reinforcement learning with the MAXQ value function decomposition. *Journal of Artificial Intelligence Research*, *13*, 227–303.

Drummond, C. (2002). Accelerating reinforcement learning by composing solutions of automatically identified subtasks. *Journal of Artificial Intelligence Research*, *16*, 59–104.

Duckett, T., & Nehmzow, U. (1998). Mobile robot self-localization and measurement of performance in middle scale environments. *Robotics and Autonomous Systems*, *24*, 57–69.

Fleischer, J., & Marsland, S. (2002). Learning to autonomously select landmarks for navigation and communication. *From Animals to Animats 7: Proceedings of the Seventh International Conference on Simulation of Adaptive Behavior*, 151–160.

Fleischer, J., Marsland, S., & Shapiro, J. (2003). Sensory anticipation for autonomous selection of robot landmarks. In M. V. Butz, O. Sigaud and P. Gérard (Eds.), *Anticipatory behavior in adaptive learning systems: Foundations, theories, and systems*, 201–221. Berlin Heidelberg: Springer-Verlag.

Goldberg, D. E. (2002). *The design of innovation: Lessons from and for competent genetic algorithms*. Boston, MA: Kluwer Academic Publishers.

Hengst, B. (2002). Discovering hierarchy in reinforcement learning with HEXQ. *Proceedings of the Nineteenth International Conference on Machine Learning*, 243–250.

Kaelbling, L. P., Littman, M. L., & Moore, A. W. (1996). Reinforcement learning: A survey. *Journal of Artificial Intelligence Research*, *4*, 237–285.

Kortenkamp, D., Huber, M., Koss, F., Belding, W., Lee, J., Wu, A., Bidlack, C., & Rodgers, S. (1994). Mobile robot exploration and navigation of indoor spaces using sonar and vision. *Proceedings of AIAA/NASA Conference on Intelligent Robots in Field, Factory, Service and Space (CIRFFSS'94)* (pp. 509–19). Houston, Texas.

Kuipers, B. (1998). A hierarchy of qualitative representations for space. In C. Freksa, C. Habel and K. F. Wender (Eds.), *Spatial cognition: An interdisciplinary approach to representing and processing spatial knowledge*, vol. 1404 of *Lecture Notes in Artificial Intelligence*, 337–350. Berlin Heidelberg: Springer Verlag.

Lanzi, P. L. (2000). Adaptive agents with reinforcement learning and internal memory. *From Animals to Animats 6: Proceedings of the Sixth International Conference on Simulation of Adaptive Behavior*, 333–342.

Marsland, S., Nehmzow, U., & Duckett, T. (2001). Learning to select distinctive landmarks for mobile robot navigation. *Robotics and Autonomous Systems*, *37*, 241–260.

McGovern, A., & Barto, A. G. (2001). Automatic discovery of subgoals in reinforcement learning using diverse density. *Proceedings of the 2001 International Conference on Machine Learning (ICML2001)*.

Parr, R., & Russell, S. (1997). Reinforcement learning with hierarchies of machines. *Advances in Neural Inforamtion Processing Systems* (pp. 1043–1049). MIT Press.

Pierce, D., & Kuipers, B. (1994). Learning to explore and build maps. *Proceedings of the Twelfth National Conference on Artificial Intelligence (AAAI-94)*. Cambridge, MA: AAAI/MIT Press.

Rizzi, S., Maio, D., & Golfarelli, M. (1997). A hierarchical approach to sonar based landmark detection in mobile robots. *Proceedings of the 5th Symposium on Intelligent Robotics Systems* (pp. 77–84). Stockholm, Sweden.

Sutton, R., Precup, D., & Singh, S. (1999). Between MDPs and semi-MDPs: A framework for temporal abstraction in reinforcement learning. *Artificial Intelligence*, *112*, 181–211.

Voicu, H., & Schmajuk, N. (2001). Spatial navigation using hierarchical cognitive maps. *Proceedings of the 4th International Conference on Cognitive Modeling (ICCM'01)*. Fairfax, USA.

Yamauchi, B., & Beer, R. (1996). Spatial learning for navigation in dynamic environments. *IEEE Transactions on Systems, Man and Cybernetics B*, *26*, 496–505.