Video Based Animal Behavior Analysis

Xinwei Xue and Thomas C. Henderson

UUCS-06-006

School of Computing
University of Utah
Salt Lake City, UT 84112 USA

June 6, 2006

Abstract

It has become increasingly popular to study animal behaviors with the assistance of video recordings. The traditional way to do this is to first videotape the animal for a period of time, and then a human observer watches the video and records the behaviors of the animal manually. This is a time and labor consuming process. Moreover, the observation results vary between different observers. Thus it would be a great help if the behaviors could be accurately derived from an automated video processing and behavior analysis system. We are interested in developing techniques that will facilitate such a system for studying animal behaviors.

The video based behavior analysis systems can be decomposed into four major problems: behavior modeling, feature extraction from video sequences, basic behavior unit discovery and complex behavior recognition. The recognition of basic and complex behaviors involves behavior definition, characterization and modeling. In the literature, there exist various techniques that partially address these problems for applications involving human motions and vehicle surveillance.

We propose a system approach to tackle these problems for animals. We first propose a behavior modeling framework, and a behavior model consisting of four levels: physical, physiological, contextual, and conceptual. We propose to explore information-based feature extraction and dimension reduction techniques, such as mutual information. Basic behavior units (BBUs) are determined from these features using the affinity graph method. A maximum likelihood approach to choose optimal parameters, such as affinity measures, and feature subsequence window size. Furthermore, we formulate a hierarchical approach and Hidden Markov Model (HMM) approaches, incorporated with our behavior models to recognize complex behaviors in laboratory animals.

1 Introduction

1.1 Motivation

As a specific problem, consider the study of the genetics of certain diseases. In one instance, this requires the determination of time the lab mouse spends grooming itself, as shown in Figure 1 The traditional way to do this is to first videotape the mouse for a period of time, and then an observer watches the video and records the behaviors of the mouse manually. This is a time and labor consuming process. Moreover, the observation results vary between different observers. Thus it would be a great help if the behaviors could be accurately derived from an automated video processing and behavior analysis system.

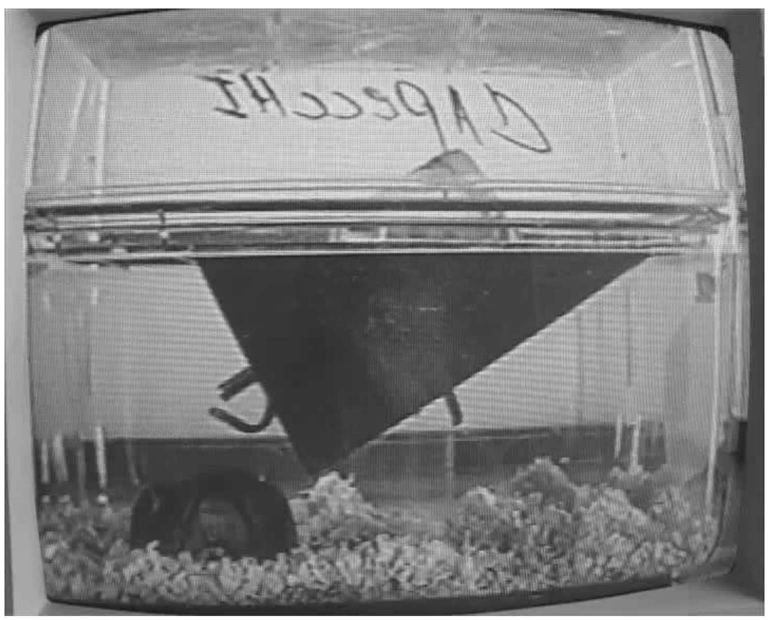


Figure 1: Mouse in a cage.

In fact, live subject behavior study has become a very important research area, in which the behavior of various animals or humans is studied for many different purposes. In the context of an animal, the behaviors may include movements (motion), posture, gestures, facial expressions, etc. Animal behavior study originates from areas including biology, physiology, psychology, neuroscience and pharmacology, toxicology, entomology, animal welfare, and so on. The animals mostly studied are mice, rats or rodents, and other animals including ants, poultry, pigs and the like. There are many reasons for studying human behavior, such as smart surveillance, virtual reality, advanced user interfaces, and human motion analysis.

It has become increasingly popular to study behavior with the help of video recordings, since video recordings can easily gather information about many aspects of the situation in which humans or animals interact with each other or with the environment. Also, the video recordings make offline research possible.

Several animal vision tracking and behavior analysis systems are commercially available. Ethovision from Noldus Company[50] is a comprehensive video tracking, analysis and visualization system for automatic recording of activity, movement and social interaction of various kinds of animals in an enclosure. It provides a range of features for video tracking and data analysis, and allows for automation of many behavioral tests. It uses color to distinguish animals from background and to analyze behavior patterns based on movement paths. However, for behaviors as complex as grooming, the user can only label it interactively. The SMART system from San Diego Instruments [1] is an animal video tracking and analysis system for behavioral tests (mazes), whose analysis is mostly based on an animal's path. The Home Cage and Open Field Video Tracking Systems from Med Associates, Inc. [2] focus on simple ambulatory and stereotypical (partial-body movement) behaviors for mice and rats. The Video Tracking System from Qubit Systems, Inc. [3] operates on the concept of contrast and tracks an animal's trajectory. The Trackit system from Biobserve Company [4] tracks the animal position and orientation in 2D and 3D for flying insects, which in turn controls the pan-tilt camera to get close-up images. The Peak Motus System from Vicon Peak Company [5] tracks human, animal and other objects automatically with markers or based on contrast. The Big Brother System from Actimetrics Company [6] tracks the path of the animal under study, which is the basis for further analysis.

These available systems have a high level of interactivity and flexibility. But in these systems, there are several major limitations to fully automatic tracking and analysis. 1) They usually employ simple image processing techniques, e.g., thresholding and background subtraction (based on obvious contrast color, or markers), to identify and track animals. 2) Usually only the animal position, or in other words, the movement trajectory is used for behavior pattern analysis. 3) Only very simple behaviors (e.g., moving, resting, etc.) can be automatically detected and analyzed. 4) No behavior modeling is incorporated in these systems.

The purpose of this research is to perform the automatic video based behavior analysis in a more systematic way. We formulate the problems of this task in a four-module framework, and incorporate behavior modeling in tracking and analyzing complex behaviors and patterns.

1.2 Problems To Solve

We propose a video-based automated behavior analysis system comprised of four modules: behavior modeling, feature extraction, basic behavior unit (BBU) discovery and complex behavior analysis, as shown in Fig 2.

Behavior modeling. This is an essential step in the automated behavior analysis system. It interacts with the other three modules. First, for a specific system, we usually have in mind the kind of behavior in which we are interested. Take the mouse-in-cage scenario as an example, where we are

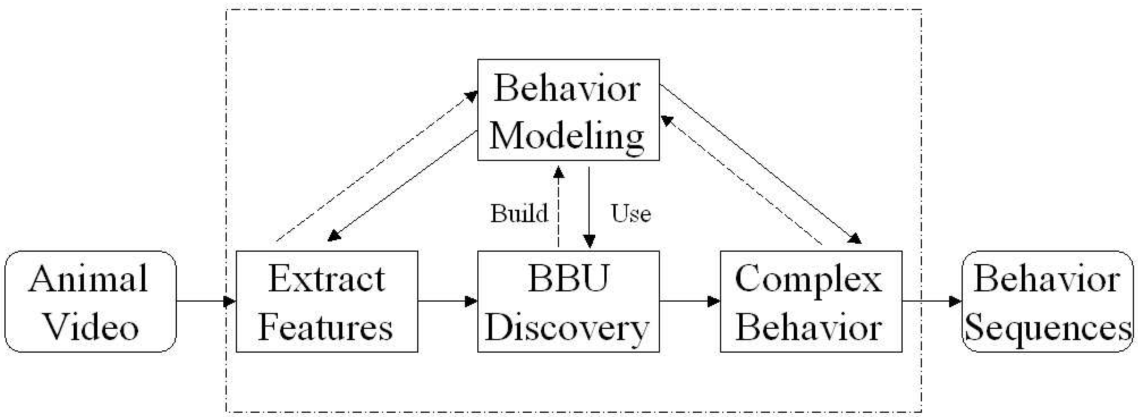


Figure 2: Work-flow for Video Based Behavior Analysis.

interested in finding the periods of the mouse resting, exploring, grooming, and eating, etc. Second, we need to define, characterize (represent), and model these behaviors in terms of three factors: physical (spatiotemporal) features; the relationship between these behaviors; and the relationship between the animal and its environment. These behaviors can then drive the task of feature extraction for basic and complex behaviors (or behavior pattern) recognition, which may in turn help the interpretation of behaviors. Third, another important component in this block is the internal model driving the behaviors of an animal.

The behavior representation and description is inherently hierarchical. The lowest level is the spatiotemporal image sequence features. The next level is the basic behavior units (BBUs) defined in terms of certain spatiotemporal image features. Then the complex behaviors are represented as a set of BBUs with certain constraints. From this level up, it may corresponds to the natural language level, i.e., a sentence consisting of BBUs and complex behaviors. Since at the natural language level, it opens up another whole research area, in this research we concentrate on the process mapping from lower level to a higher level description, up to the level of complex behaviors.

Feature extraction. To be able to distinguish behaviors, we need to be able to extract sufficient spatiotemporal physical features of the object from video sequences that represent different behaviors. The features may include: the object's position, posture, speed, contour or region pixels, dynamics, motion patterns, etc. We may also need to extract features of the environment. This process usually requires the ability to detect and track objects from video sequences. In case of a high-dimensional feature set, feature selection or dimension reduction may be necessary, to reduce computation time.

Discovery of basic behavior units (BBUs), or behavioral segmentation. BBUs are the behavior primitives and higher level analysis will be carried out in terms of these. A BBU can be defined as an activity that remains consistent within a period of time, and that can be represented by a set of spatiotemporal features. This step is based upon successful feature extraction. For the mouse-in-cage example, the BBUs of a mouse in a cage can be resting, exploring, eating, etc. The process of BBU extraction involves mapping the extracted physical features to distinctive behavior units, hence classifying subsequences of the video frames into a sequence of BBUs.

Recognition of complex behaviors. A complex behavior consists of one or multiple BBUs with spatial or temporal relationship between them or between BBUs and environment. It is a higher level in the behavioral hierarchy. Once basic behaviors are discovered, complex behaviors can be constructed and analyzed based upon the relationship between basic behaviors, the interactions with the environment, and with other objects. For example, we may find out whether the mouse always grooms after eating, or whether it always rests in the same location, or we may detect the influence of environment change to the animal, etc. We can determine the patterns of the animal's behavior, and thus interpret the causes of this behavior, given that a complete behavior sequence has been successfully recovered. This step may also help discover anomalous behaviors. If there is sudden change of pattern, then we can tell when the anomaly happens, or when the environment changes.

These four blocks are closely related to each other. In the diagram, the dotted arrows pointing to the behavior modeling module indicate the influence of the other three modules. The behavior model characterizes and represents the behavior in terms of physical features, which are then extracted in the feature extraction block. On the other hand, the features defined in the behavior model block are subject to changes and updates based upon the video quality, the reliability of the features, and the capability to distinguish behaviors. The detectable features may vary from video to video, and may be effective for one behavior but not for another, thus, the behaviors may need to be defined and characterized in different feature sets. The BBUs are directly defined in the behavior model block, and detection is based upon the extracted features from the video sequence. The BBU detection efficiency directly affects the behavior characterization and may initiate the request for feature selection or fusion. The complex behaviors interact with the behavior modeling block in a similar fashion. It depends on the efficiency of BBU detection and the accuracy of the behavior models defined (in different levels, such as physiological, contextual, and conceptual) in the behavior model block.

1.3 Applications

This study has many potential applications across a number of fields. The proposed four-module framework can be readily extended to different live animals or humans, with specific behavior models being built for specific objects and behaviors of interest. Automatic lab animal behavior analysis is the first important target, which will benefit ethology, medicine, and medical experiments. Human behavior analysis, such as office activity, consumer shopping behaviors, monitoring of children, elderly or sick people, public behaviors, crowd behaviors, etc. are other areas receiving more and more research attention. Automatic sport games analysis, say soccer, basketball, and so on, may be another application that will benefit from this study.

2 Related Work

In this section, we review the literature related to our approach to the topics mentioned in the workflow of Section 1.

2.1 Behavior Modeling

In the diagram of Figure 2, the behavior modeling block includes behavior definition, characterization, and modeling. Here we review the internal models that drive the generation of behaviors.

Behavior modeling can be found from natural physical systems, to live organisms behavior study, life-like character animation, robot motion control, and automated behavior analysis from video sequences. The study of the behavior of natural systems is the basic undertaking of science, and the general goal is to produce a description that not only explains what happens, but that can be used to predict future events. For example, a description of the change in height of an object dropped from the top of a building might be derived from Newton's laws and given as a function of height versus time. The behavior in this case is the change in position, and the resulting equation models this behavior. Such a model can be put to a variety of uses; e.g.:

- **explain behavior**: determine time or velocity of impact,
- **predict behavior**: given a desired time of impact, determine the necessary initial height and velocity, or
- **characterize behavior**: given a trajectory, determine if the object obeys the model.

The variables of such models are usually physical quantities that can be measured by well-defined instruments. The result of such measurements is called *raw experimental data*.

A similar approach may be taken in the study of living organisms as in ethology [15, 18, 29, 23, 27, 33, 34, 43, 44, 63]. Here the situation is more complicated because behavior is mediated not only by physical laws, but also by physiological conditions, internal drives and environmental context. Also complicating the issue is the interplay between success and survival at the individual and species levels.

In addition, the description of animal behavior may be couched in special variables defined by the investigator and discerned through the psychological processes of the human observer. For example, a gorilla may be watched to determine how often it displays *affection* for its young; a videotape of this would be raw experimental data, but a human produced log of *affection events* based on the video will be termed *annotated behavior* and serves as an explanation of the observed data. Such an explanation is mediated by and couched in terms of the conceptual model.

In order to produce a life-like animation, it is necessary to produce both physically and psychologically correct behavior [55]. Models for animated characters require a *body* component and a *mind* component. The latter addresses goals, drives, beliefs, etc. A motion sequence generated by such a model will be called a *generated behavior* and is a predicted sequence of events.

The mobile robot research community also produces generated behaviors [13]. However, unlike the animation characters which exist only in an electronic world, physical robots exist in the real world. Thus, these behaviors also include a control aspect in terms of the robot acting in the physical world. (While it is true that an animated character interacts with its virtual world, this again involves generated behaviors, whereas the mobile robot gets physical feedback.) In the community of intelligent multiagent systems, researchers have tried to model the functional capabilities of the brain in perception, cognition, and behavioral skills. The real-time control system (RCS) [9, 10], one of the cognitive architectures, consisting of a multi-layered multi-resolution hierarchy of computational agents each containing elements of sensory processing (SP), world modeling (WM), value judgment (VJ), behavior generation (BG), and a knowledge database (KD). At the lower levels, these agents generate goal-seeking reactive behavior. At higher levels, they enable decision making, planning, and deliberative behavior.

Finally, the area which most interests us is automatic behavior analysis. Here the goal is to combine raw experimental data (usually video) with a behavior model and produce what we term *interpreted behavior*. This corresponds to annotated behavior except that one is produced by humans and the other by computation. Interpreted behavior thus also serves as an explanation of the observations in terms of the model.

In the literature surveyed, there are a few basic approaches to behavior modeling:

- State-space models [45].
- Computing models, including:
  - Automata (finite state machines, schema, agents, etc.)[34]
  - Grammatical methods (strings[34], T-patterns [42], etc.)
- Mathematical models (e.g., dynamic state variables [23, 43], game theory [27, 36], Bayesian approaches [24], utility theory [13], etc.)
- Sequential behavior models [24].
- Analog models (e.g., neurons, electrical circuits [13])

The *State-Space model* takes a motivational approach to behavior modeling: it takes into consideration the internal and external causal factors to the behaviors, which include the physical and physiological factors. We base our model on this approach, and extend it to contextual and conceptual levels.

The automata and grammatical models allow explicit recognition of hierarchy within behavior, which coincides with our idea of BBUs and complex behaviors. The sequential behavior models basically belong to the statistical analysis category, aiming to discover behavior patterns, equivalent to our formulation of behavior patterns.

The mathematical models such as dynamic state variable models, game theory, utility theory, and optimal control theory model the decision making process of an animal from an evolutionary standpoint: "the survival of the fittest". In these models, the animal always takes the optimal move according to some optimization criterion, to enhance the "fitness score". These models can also be considered as goal-directed behaviors. On the other hand, these theories can also be interpreted as determining the best set of parameters for a behavior model or to improve the global performance of the systems. Though these models would not be used in our approach directly, the ideas of optimal parameter determination is one important task in our four-block framework.

Numerous techniques have been developed to help build various aspects of a behavior model. Typically in the animal behavior literature, data is available in the form of observations, and it is necessary to determine the behavior units for the system being modeled, as well as the relations between the behavior units. The following statistical techniques have been demonstrated useful: (auto) correlation, pattern analysis, multivariate statistics (PCA, factor analysis), cluster analysis, multi-dimensional scaling and contingency tables [15, 24, 44]. These methods are mostly used in behavior classification and pattern analysis. One of the sequential analysis methods, the Hidden Markov Model (HMM) is used as our complex behavior model.

2.2 Feature Extraction

Feature extraction usually involves two processes: (1) object detection and tracking (so as to get the spatiotemporal features), and (2) feature selection or dimension reduction.

2.2.1 Detection and Tracking

We review the literature in two categories: animal video detection and tracking, and human motion tracking and analysis. Several simple techniques exists for detection and tracking of lab animals, as used in EthoVision [50]: grayscale or color information is used to identify the animals, and thresholding is used for detection and tracking [50, 74]. Frame differencing and background subtraction is used in ant tracking in [16, 74]. Active shape models (ASM) are used in [65] to track and classify the posture of laboratory rodents. [31, 59] use simple image segmentation techniques for poultry tracking. Perner [53] uses object-oriented motion estimation techniques for tracking pigs. [21] uses face detection and tracking techniques to detect and track animal faces in wildlife videos.

Human motion analysis has received much more attention in the research communities. The following survey papers give a good review of this subject: [7, 8, 22, 37, 47, 67]. Besides the techniques described for animal behavior analysis systems, more sophisticated methods have also been employed in human and other object detection and tracking, such as Kalman filters, the Condensation Algorithm, Bayesian networks, model-based, kernel-based, feature-based, contour-based, and region-based tracking, spatiotemporal information based tracking, etc. Many advanced pattern recognition methods have been applied to the action and pose recognition problem, such as different flavors of template matching techniques, correlation, hidden Markov models (HMMs), neural networks, principle components analysis (PCA), state-space approaches, etc. Recently, semantic descriptions of actions are becoming increasingly popular, which applies the grammatical concept of natural languages to vision systems.

The model-based method models the 2D or 3D appearance of the object, and tracks it through the video sequence. This applies well to rigid body detection and tracking, such as vehicles. For non-rigid objects like animals, whose body shape changes non-rigidly, the model based technique fails miserably. The kernel-based (mean-shift) technique is a robust video tracking technique, but it can only track the object in a specific form: either a box, or an ellipse. It is best for tracking the position of an object, but not able to track the orientation of the object. Feature-based techniques track objects based on features extracted from the video images, such as corners. Again the major tracking result is the object trajectory. Contour-based and region-based techniques try to track the object contour or silhouette. By tracking the contour, we are able to calculate several features, such as shape, posture, aspect ratio, etc. beyond the position information. Snakes (active contours), deformable models and level set methods are the major contour tracking techniques found in the literature. The implementation usually takes a level-set approach, which is topology change free.

In our application, we need not only the trajectory of the animal, but also postures, speed, motion, object contour, etc. Thus more features are needed to extract the BBUs. We adopt the level set framework to track the silhouettes of the animal in the video, utilizing the spatiotemporal information, which will be introduced in the next section.

2.2.2 Feature Dimension Reduction

The image features, simple or complex, extracted either directly from video images or from detection and tracking results, may be high dimensional. Directly using high dimension data is computationally expensive, hence impractical. Thus dimension reduction is necessary.

The are two major categories of methods for dimensionality reduction: feature selection, and feature transformation. Feature selection methods keep only useful features, and feature transforms construct new features out of the original variables. Most of the techniques focus on feature transforms.

The classical techniques for dimension reduction, Principle Components Analysis (PCA) and Multidimensional Scaling (MDS) are simple to implement, efficiently computable, and guaranteed to discover the true structure of data lying on or near a linear subspace of the high-dimensional input space. Linear discriminant analysis (LDA), independent component analysis (ICA), heteroscedastic discriminant analysis (HDA), and support-vector machines (SVM) has also been proposed for linear dimension reduction. Torkkola [64] developed a non-parametric method for learning discriminative feature transforms using mutual information as the criterion.

For input data with intrinsic non-linearity, Tenenbaum et al. [62] proposed the complete isometric feature mapping (Isomap) algorithm for non-linear dimensionality reduction, which efficiently computes a globally optimal solution. [40, 57] incorporated statistical methods and extended this algorithm to natural feature representation and place recognition.

We are using object region and contour pixels in each frame as one of the features, which is high dimensional. Hence we use the PCA technique to extract the major axis of the region and contour pixels, calculate the distribution along the principle axis, and use that as a representative feature.

2.3 Basic Behavior Discovery

Most of the techniques extract basic behaviors (or actions) directly based upon one or more features extracted (trajectory, motion, posture, etc.) from the detection and tracking results. Pattern recognition techniques (template matching, clustering analysis) are used to classify the video sequence into actions or behavior units, as discussed in the survey papers [7, 8, 22, 37, 47, 67]. These methods are effective in their specific applications. The idea is to utilize all the available distinguishing features to perform classification. Recently, new approaches based on data (or feature) variance or similarity analysis have been developed for discovering BBUs: PCA-related techniques, and affinity graph-based techniques.

PCA is a classical data analysis tool. It is designed to capture the variance in a dataset in terms of principle components, which is a set of variables that define a projection that encapsulates the maximum amount of variation in a dataset and is orthogonal (and therefore uncorrelated) to the previous principle component of the same dataset. This technique first calculates a covariance matrix from the data, then performs the singular value decomposition (SVD) to extract the eigenvalues and eigenvectors. The eigenvector corresponding to the largest eigenvalue is the *principle component*.

The affinity graph method is also an eigenvalue decomposition technique, or spectral clustering technique. It captures the degree of similarity between the data sequences. Different from the PCA technique, it computes an affinity matrix based upon an affinity measure (e.g., distance, color, texture, motion, etc.) instead of a covariance matrix. The eigenvectors extracted by SVD go through a thresholding step to segment out the first cluster. Then it goes on to process the next eigenvector to find the second cluster, and so on.

PCA-related techniques. Jenkins [39] employs a spatiotemporal nonlinear dimension reduction technique (PCA-based) to derive action and behavior primitives from motion capture data, for modularizing humanoid robot control. They first build spatiotemporal neighborhoods, then compute a matrix D of all pairs' shortest distance paths, and finally perform PCA on the matrix D. Barbic et al. [17] propose three PCA-based approaches which cut on where the intrinsic dimensionality increases or the observed distribution of poses changes, to segment motion into distinct high-level behaviors (such as walking, running, punching, etc.).

Affinity graph method. The affinity graph method has mostly been applied in image segmentation, as summarized in [68]. Recently, this method has been applied to event detection in video [54, 72]. Though not exactly the same approach, the concept of similarity matrix for classification is applied in gait recognition [19] and action recognition [28].

Different affinity measures have been proposed to construct the affinity matrix. In image segmentation, distance, intensity, color, texture and motion have been used [32]. In video-based event detection, as in [72], a statistical distance measure between video sequences is proposed based on spatiotemporal intensity gradients at multiple temporal scales. [54] uses a mixture of object-based and frame-based features, which consist of histograms of aspect ratio, slant, orientation, speed, color, size, etc., as generated by the video tracker. Multiple affinity matrices are constructed based on different features, and a weighted sum approach is utilized for constructing the final affinity matrix.

The most closely related methods to our work are [54] and [72]. [72] constructs an affinity matrix from temporal subsequences using a single feature, while the former constructs the affinity matrices for each frame based upon weighted multiple features.

We are particularly interested in discovering animal behaviors from video sequences. We propose a framework for discovering basic behaviors from temporal sequences based on multiple spatiotemporal features. In our approach, we combine the advantages of the approaches from [54] and [72]: 1) We construct one affinity matrix based on a feature vector consisting of a set of weighted features. The combined features provide us with more information. 2) We construct the affinity matrix on a subsequence of the frame features (multiple-temporal scale), instead of on one frame. Thus we can encode the time trend feature into the problem, and capture the the characters of the temporal gradual changes. We also investigate the choice of affinity measures as well as the optimal length of the temporal subsequence.

2.4 Complex Behaviors and Behavior Pattern Analysis

The term *complex behaviors*, is also called complex *events* or *scenarios* in the video event mining literature [12, 46, 48, 49]. A BBU in our definition, corresponds to one type of *event*. The representation and analysis of complex behaviors or events follow a hierarchical structure.

State-based representations have been employed to represent temporal sequences or trajectories. The Hidden Markov Model (HMM) [26, 41, 56] is a popular state-based model. Complex behaviors are decomposed into a number of states, and recognized by analyzing the transition probabilities.

Variants of HMMs have been used for activity representation and recognition, such as entropic-HMMs [20], coupled-HMMs [51], and parametric-HMMs [71]. These approaches usually apply to low-level interpretations. In [30], HMM and data fusion are used to detect drowning behaviors in a swimming pool. Finite state machines is another state-based models, which has been used to represent a human's state in an office environment [14].

In [48, 49], complex behavior is divided into two categories: single-thread composite events, and multiple-thread composite events. The former events correspond to a consecutive occurrence of multiple primitive events, while the latter is the composition of multiple single-thread events with some logical, temporal or spatial relations between them. They represent single-thread events using Bayesian networks, and multiple-thread events with finite-state machines (FSM).

Another popular approach is the grammatical method, where grammars and languages are used. The work of [38] recognizes visual activities and interactions by stochastic parsing. It employs a two-level event abstraction: HMMs are used to model simple events at the lower level and a stochastic context free grammar in the higher level. In [49], a language-based representation of events in video streams was developed.

A traditional artificial intelligence approach using interval temporal logic is proposed in [11]. In [58], the video sequence interpretation involves incremental recognition of states of the scene, events and scenarios, and the recognition problem is translated into a constraint-solving problem.

In [12], an explicit representation formalism for behavioral knowledge based on formal logic is presented that can be used in the tasks of understanding and creation of video sequences. The common sense knowledge is represented at various abstraction levels in a Situated Graph Tree (SGT).

The literature has shown that HMM and FSM are pretty good approach for complex behavior recognition. The grammatical and temporal logic approaches have also been applied to small-scale problems. However, when the grammar and logic become very complex (as in the case of very complex activities), with a large vocabulary, the implementation of the systems may become prohibitively difficult. Hence here in our research, we take the HMM-flavored approach.

To summarize, the techniques mentioned above have mostly been applied to human motion or vehicles, from an observer point of view; i.e., the observer breaks the observations into several states, and then the computer recognizes the complex behaviors from these observed states. As we are interested in animal behaviors, we take an object-centered approach, and combine the complex behavior analysis task with a behavior model that considers the influences of factors from physical, physiological, contextual, and conceptual levels. We explore the HMM approach, incorporated with our behavior models and spatiotemporal constraints.

3 Goals and Proposed Work

The goal of this research is to develop algorithms corresponding to the four modules for automatic animal behavior analysis from video sequences shown in Figure 2. We first propose a behavior modeling framework and a model consisting of four levels to help analyze animal behaviors. Then we extend existing techniques and propose new methods for feature extraction, basic behavior unit determination, and complex behavior analysis. We apply these techniques to synthetic data and the real mouse-in-cage video. The following sections describe our goals and proposed work in detail. (Also, see [35])

3.1 Behavior Model

3.1.1 Goals and Proposed Work

Our goal is to construct an object-centered behavior model that will help with the analysis of behavior from observed data. Here we propose a behavior modeling framework and a four-level animal behavior model.

3.1.2 Behavior Modeling Framework

To better understand the various manifestations of behavior and the emphases of different disciplines, we propose the general framework for scientific investigation shown in Figure 3. The world (Box 1) signifies the object of study which may be, for example, gravitational force, or the foraging behavior of army ants. Typically, direct access to the world is not possible and the world must be understood through observation and measurement (Box 2). Such observations may arise through human perception or the use of measuring instruments.

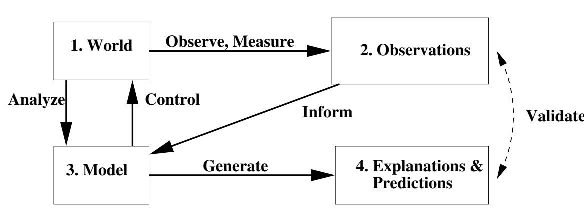


Figure 3: General Framework for Scientific Explanations.

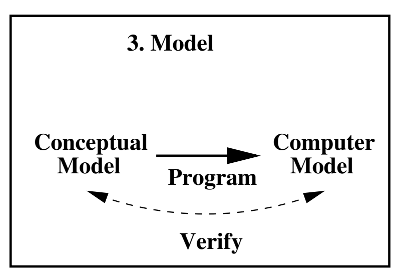


Figure 4: Computer Models Must Undergo Verification with Conceptual Model.

A model (Box 3) of the object of interest is developed based on measurements and observations of the object. Modeling and observation are highly coupled in that the observations provide desiderata for model creation, while the model itself informs the experimental framework for data acquisition. The model serves two major purposes; first, it should explain the observations; second, it should predict new phenomena. These explanations and predictions (Box 4) can be compared to the observations in order to **validate** the model. Finally, the model can provide guidelines to control the object of study; this can either be to define or improve observation conditions, or can be done with the goal of achieving a certain predicted result.

Another level of detail is required to distinguish computer models from other formal frameworks; this is shown in Figure 4. The conceptual model is converted to a computer model by programming an implementation. To ensure the equivalence of the two models requires **verification**. This includes, understanding and eliminating algorithmic errors, numerical errors, coding errors, etc.

It is interesting to see that the various research domains of interest map directly onto this framework. For example, an ethologist produces explanations of observed behavior based on the conceptual model which provides descriptions of units of behavior. The production of behaviors for animated characters requires the generation of action sequences that are then validated against how realistically they capture real world behavior. Robot behaviors involve models of various aspects of the world, and thus usually have multiple components. For example, there may be a physical world component, a homeostasis, self-regulatory component (i.e., robot physiology), and a high-level conceptual model. Each of these interacts with the others either by providing its own explanations and predictions as observations to the other components or by comparing explanations and predictions directly. Automatic behavior analysis produces an explanation of behavior which can only be validated against the human explanation of the observed data; Figure 5 shows this situation in which results from human conceptual models are compared to machine results in order to validate the machine model.

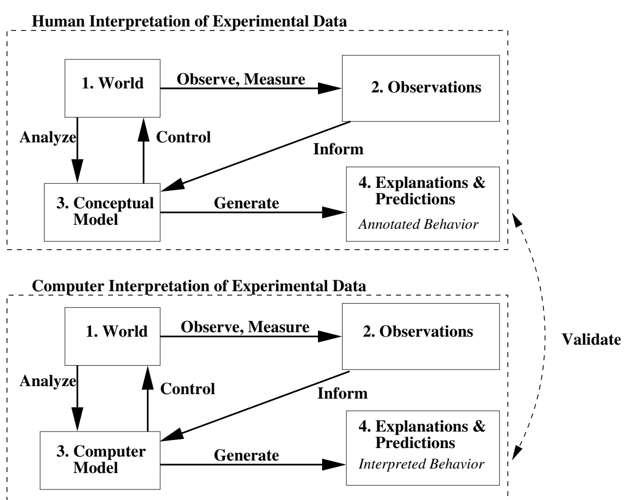


Figure 5: Validation of Automatic Behavior Analysis by Machine Requires Human Explanatory Data.

3.1.3 Behavior Model of Four Levels

We propose that behaviors of live animals should be modeled in four different levels: physical, physiological, contextual, and conceptual. In the physical level, the image structure, geometric features, textures, and motion can be used to distinguish behavior units. In the physiological level, behavior is related to the physiological state of the animal or human. For example, as time goes on, the degree of hunger increases. In the contextual level, the relationship between environment and object, or between objects can be used to distinguish behaviors. The relative position, posture, motion, etc., can be used as parameters. In the conceptual level, the live object may exhibit goal-driven behavior.

Our approach to modeling follows that of McFarland [36, 45]. The typical model for a purely *physical system* behavior is shown in Figure 6. The model is divided into two major parts: (1) the

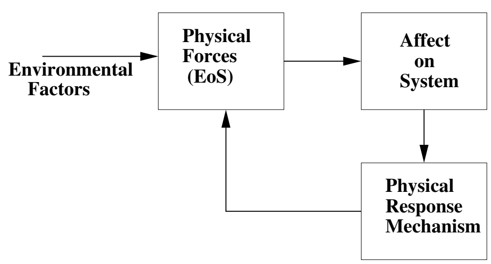


Figure 6: Physical System Behavior Model.

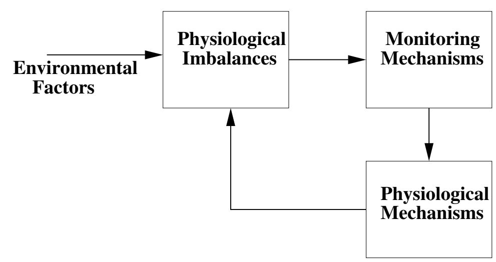


Figure 7: Physiological System Behavior Model.

Equations of State (EoS) which describe all forces of interest at work in the system, and (2) the specific characteristics of the particular object under study. For example, (1) will usually elaborate $F = ma$ while (2) specifies mass, initial position velocity, etc., as well as any other local constraints (e.g., gravitational constant, existence of floors, walls, etc.). In the physical model of the animal, the two major parts can be modeled as: 1) The physical energy (which enables it to perform various kinds of activities) state of the animal (e.g., $E(t) = f(t, E(t-1), a(t), w(t))$, where $E(t)$ refers to the energy state at time t, $a(t)$ is the current activities, and $w(t)$ is random noise); 2) The influence of the various activities to its energy state. I.e., the exploring activity will decrease its energy gradually, while eating will increase its energy instantly.

The next level of our animal behavior model describes the *physiological system*. Figure 7 shows the basic scheme for this. The overall mechanism is similar to the physical system, but we can model the animal in various physiological aspects. For example, we can model hunger and thirst as internal drives related to time passed and the activities performed, and this approach allows an appropriate conceptualization of hunger. Similar equations as in the physical model can be used here: $D(t) = f(t, D(t-1), a(t), w(t))$, where D represents the internal drives.

In the contextual level, the environmental factors act as external contextual stimulus to the behavior of the animal, which reflects the interaction between the animal under study and its environment.

The environmental context includes environmental changes, the appearance or disappearance of another animal of the same kind, etc. The conceptual level of the system models the motivational or goal-directed behavior.

Finally, the full model for the animal (or autonomous agents) is formulated in a feedback control-like system, as shown in Figure 8. The physical and physiological systems are integral components

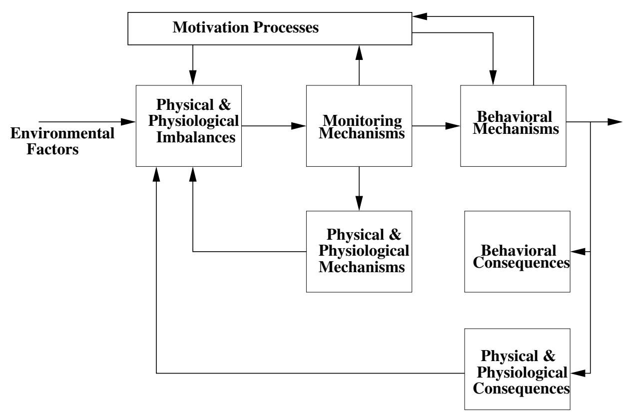


Figure 8: Motivated System Behavior Model.

of this model, with the environmental factors influencing the behavior in the contextual level, and motivational processes in the conceptual level. The behavioral mechanisms (i.e., the action generating processes) give the possible responses of the system. Such actions have consequences both in terms of the behavioral state of the agent, as well as in terms of the physical and physiological state. For example, if the selected action is *eat*, then there are required physical motions, and there are physiological consequences such as decrease in hunger and increase in thirst.

The *motivational processes* are those that play a role in creating goals, shifting attention, influencing drives, etc. and which are not strictly physical or physiological. Such processes may interact intimately with lower level processes; for example, the agent may choose to ignore pain in order to obtain food.

Mapping this behavior model to the automated behavior analysis problem is an important engineering task. The model interacts with the modules of feature extraction, BBU discovery, and complex behavior recognition as follows: the physical and physiological models (as mentioned earlier) can be used to predict behaviors with constraints from contextual and conceptual factors. The monitoring mechanism observes (or monitors) the animal's internal states, high-level conceptual states and the contextual states. Then it make corresponding perceptual and behaviorial decisions. For instance, the monitoring mechanism may choose to use different feature sets to measure the environment, and the animal's current behavior. These measurements can also be used to verify the prediction, and then make proper update to the model parameters.

3.2 Feature Extraction

3.2.1 Goals and Proposed Work

The two major questions to answer are:

1. What features to extract.
2. What detection and tracking method to use to extract these features.

Our goal here is to be able to extract the necessary features from video sequences for further analysis of BBUs, and select optimal features. Object trajectory is a feature many tracking methods try to extract from the video sequence, as reviewed in the related work. Due to the non-rigid nature of animals, we need more features than just the trajectory, e.g., posture, shape, orientation, motion, etc. We adopt the level set framework to track the silhouettes of the animal in the video, utilizing the spatiotemporal information.

The level set method [60] is a curve propagation technique that has been widely used in image denoising [70], segmentation [73], reconstruction [69], and registration [66]. It also has been applied in video object tracking [52, 61]. Our work is an extension of the work of [61, 73].

The level set method is generally formulated as a curve $\phi(x(t), t) = 0$ ($x(t)$ is the curve location at time t), and the curve propagates in the curve normal direction with a speed:

$$\begin{aligned} \phi_t(x) &= -F|\nabla\phi| \\ \phi_{t=0}(x) &= C_0 \end{aligned}$$

The speed function F, in our framework, consists of a smoothing term (the curvature κ), a background and foreground competition term bg, a spatial gradient term g_s, and a temporal gradient term g_t:

$$F = w_1\kappa + w_2 bg + w_3 g_s + w_4 g_t$$

The weights w_i need to be chosen according to the image quality and the foreground region (the animal) to track.

Features can then be extracted from the tracking result. Here we use the following features: x_c, y_c, vx_c, vy_c, pa, ma, r, where x_c, y_c are the coordinate of the centroid of the bounding box of the tracked region, and vx_c, vy_c are the corresponding speed, pa, ma are the principle and minor axes of the tracked region, and r is the ratio of the area of the tracked region to the bounding box. The principle and minor axes of the tracked region can be calculated using the PCA technique, where the dimension of the region pixels is reduced from n (the number of pixels) to two.

3.3 Basic Behavior Discovery

3.3.1 Goals and Proposed Work

Our goal here is to be able to detect BBUs from the features obtained from previous step, at a reasonable accuracy. We propose to apply the affinity graph method to the *subsequences* of features and we study the optimal selection of parameters, such as affinity measures and subsequence window size, etc.

A behavior model is built in terms of basic behavior units (BBUs). An appropriate BBU set must be found for the particular modeling approach. For example, suppose that we wish to model a mouse in a cage. Then, a set of BBUs of interest might include: resting, eating, drinking, grooming, and exploring. If we adopt the state-space approach, then the observable variables we use are: position ($p(t)$), speed ($s(t)$), and acceleration ($a(t)$). This is the physical level model, and we intend to explore the other levels in the future; e.g., a thirst variable at the physiological level, an avoidance variable at the contextual level, and various goal variables at the conceptual level.

It is possible to make general functional characterizations of the BBUs in terms of the temporal variation of these variables. For example:

- resting: $p(t) =$ ground level; $s(t) = 0$; $a(t) = 0$
- eating: $p(t) =$ raised body; $s(t) = 0$; $a(t) = 0$
- drinking: $p(t) =$ raised body; $s(t) = 0$; $a(t) = 0$
- grooming: $p(t) =$ any; $s(t) = sin(t)$; $a(t) = square(t)$
- exploring: $p(t) =$ any; $s(t)$ varies randomly; $a(t)$ varies randomly.

However, this is difficult since it involves high level notions about motion (random, sine, square wave, etc.), and in fact, should consider the motions of the limbs and head separately. Another approach is to obtain video data of the BBUs of interest, and then calculate time sequences of position (e.g., of the center of mass), speed, and acceleration, and determine whether these allow discrimination of the distinct BBUs.

3.3.2 Affinity Graph Method

We propose to use the affinity graph method, an unsupervised learning method to discover basic behavior units. We take a subsequence (of length T, the number of frames or seconds) of the video images as an element, calculate the affinity measure between all elements and construct the affinity matrix.

This is done by choosing an *element* for consideration. Next a matrix is constructed in which each (i, j) entry gives an affinity (or similarity) measure of the i^{th} and j^{th} elements. The eigenvalues and eigenvectors of the matrix are found, and the eigenvalues give evidence of the strength of a cluster of similar elements. As described by Forsyth and Ponce [32], if we maximize the objective function $w_n^T \mathcal{A} w_n$ with affinity matrix $\mathcal{A}$ and weight vector w_n linking elements to the n^{th} cluster, and requiring $w_n^T w_n = 1$, then the Lagrangian is:

$$w_n^T \mathcal{A} w_n + \lambda(w_n^T w_n - 1)$$

which leads to solving $\mathcal{A} w_n = \lambda w_n$. Therefore, w_n is an eigenvector of $\mathcal{A}$. Thus, the eigenvectors of the affinity matrix determine which elements are in which cluster. We use this to extract basic behavior units in terms of their position, velocity, etc. of various state variables of interest.

Our approach differs from the closest literature [54, 72] as described in the related work in two aspects: 1) We construct one affinity matrix based on a feature vector consisting of a set of *weighted* features, instead of calculating affinity matrices for each feature. The combined features provide us with more information. We also propose a sequential hierarchical BBU segmentation based upon the distinguishing power of the features. 2) We construct the affinity matrix on a *subsequence* of the frame features (multiple-temporal scale), instead of on one frame. Selecting the optimal affinity measure, and time scale (length of the subsequence) is our next step.

3.3.3 Optimal Parameter Selection

In applying the affinity graph method, the affinity measure and optimal subsequence window size are two important parameters to choose. These parameters can be determined by applying optimization techniques to the training data or updated dynamically.

The BBU discovery process can be considered as a function $f : \{V, \theta\} \rightarrow S$ that maps the video sequence V to a behavioral segmentation, i.e., a behavior sequence S, with a parameter vector $\theta \in \Omega$. Our goal is to find the optimal $\hat{\theta}$ that generates S with minimal segmentation error. The optimal behavior $\hat{S}$ sequence must be chosen from the set of all computable segmentations $\{S_i\}$, obtained from varying segmentation algorithm parameters θ, where i denotes the ith possible segmentation result. If the prior can be assigned to each segmentation, then the maximum a posteriori (MAP) method can be applied. This requires maximizing the posterior possibility $P(S_i|V) = P(V|S_i)P(S_i)/P(V)$ as

$$\hat{S} = \underset{S_i}{arg}\, max(P(V|S_i)P(S_i))$$

which is equivalent to mimimizing the following

$$\hat{S} = \underset{S_i}{arg}\, min(-log_2 P(V|S_i) - log_2 P(S_i))$$

In order to adopt this approach, we need to properly define the term $P(V|S_i)$ in our context.

3.4 Complex Behaviors and Behavior Patterns

3.4.1 Goals and Proposed Work

Once the BBUs are detected, we now can perform the recognition of complex behaviors, which are built on the BBUs with spatiotemporal constraints. Our goal is to combine the complex behavior recognition task with our behavior model (from physical, physiological and contextual aspects), for better recognition and behavior interpretation. We are going to explore two approaches: (1) Hierarchical structure with spatiotemporal constraints, and (2) Hidden Markov Model (HMM) to analyze and recognize complex behaviors.

4 Preliminary Results

In this section, we present some preliminary results on synthetic data and the mouse-in-the-cage video data.

4.1 Synthetic Data

We consider two problems:

1. A simple physical system.
2. A 2-state mouse problem.

4.1.1 Physical System

Behavior Model
Here we consider a bouncing ball with no friction and no energy loss. Figure 9 shows the height and velocity as a function of time. In this example, the only governing physical factor is the gravity, and no other behavior levels are involved.

Basic Behavior Discovery
Here we take the position and velocity functions (versus time), as the basic data, and obtain behavior

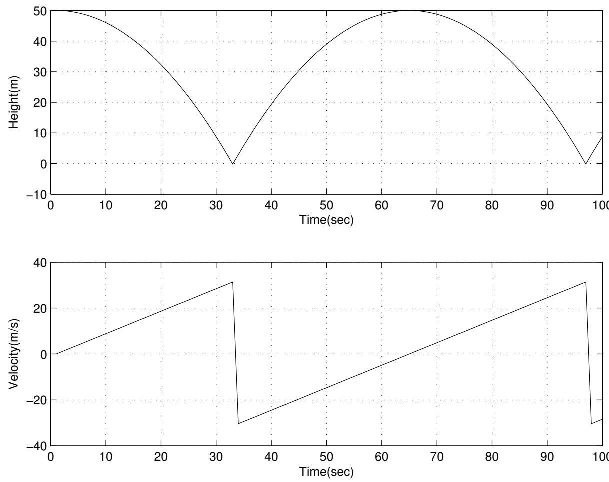


Figure 9: Bouncing Ball Position and Velocity Trace.

units by selecting a time interval and have these short sequences serve as elements. The affinity matrix is produced by running a correlation function pairwise on the elements. Next the eigenvalues are found, and only 3 are significant. Table 1 shows the absolute values of the eigenvectors and the clusters found by thresholding at 0.2 (each row corresponds to a 5-sec trace). Three behavior units were found (corresponding to going up, going down, and reversing direction).

eigenvectors			clusters		
0.2875	0.0000	0	1	0	0
0.2889	0	0	1	0	0
0.2888	0.0000	0	1	0	0
0.2888	0.0000	0	1	0	0
0.2888	0	0	1	0	0
0.2888	0.0000	0	1	0	0
0.0000	0.0000	1	0	0	1
0.0000	0.4084	0	0	1	0
0.0000	0.4085	0	0	1	0
0.0000	0.4085	0	0	1	0
0.0000	0.4086	0	0	1	0
0.0000	0.4087	0	0	1	0
0.0000	0.4068	0	0	1	0
0.2886	0.0000	0	1	0	0
0.2889	0	0	1	0	0
0.2888	0	0	1	0	0
0.2888	0.0000	0	1	0	0
0.2888	0.0000	0	1	0	0
0.2888	0.0000	0	1	0	0

Table 1. Eigenvectors (cols 1-3) and Clusters (cols 4-6).

4.1.2 A Two-State Problem

Behavior Model
Consider the simulation and analysis of a very simple two-state mouse-in-cage scenario.

For data synthesis we assume that the physiological, contextual and conceptual models may be expressed as probabilistic functions of some S-curve form (sigmoidal, hyperbolic tangent, etc.). As the Basic Behavior Units, suppose the mouse can either rest (BBUrest) or wander (BBUwander). Furthermore, suppose that the transition between these two behaviors is characterized by two functions, $F_{rest \to wander}(t)$ and $F_{wander \to rest}(t)$:

$$F_{rest \to wander}(t) = 1/(1 + e^{K_{rest} - t})$$

$$F_{wander \to rest}(t) = 1/(1 + e^{K_{wander} - t})$$

where K_{rest} is a parameter specifying the length of rest periods, and the function gives the likeli-

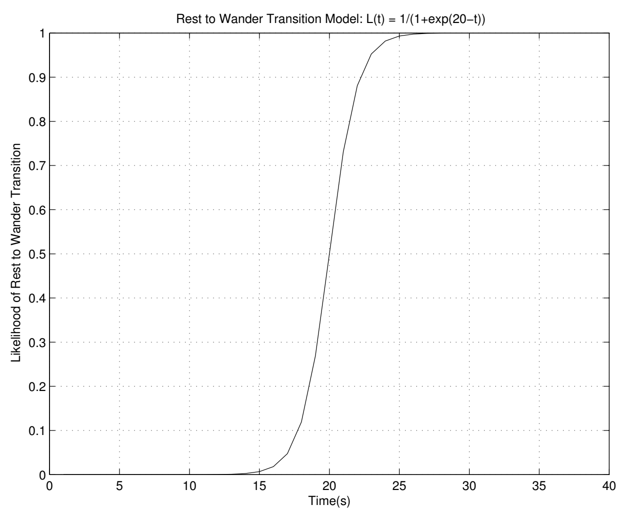


Figure 10: Likelihood of transition from rest to wander behavior function (sigmoid).

hood as a function of time that the mouse will wake up and start to wander. K_{wander} is a parameter specifying the distance wandered, and gives the likelihood that after moving a distance d the mouse will stop wandering and begin to rest. Figure 10 shows the transition likelihood for the sigmoid models used here to synthesize data sequences. Behavior sequences are synthesized over 20,000 time steps using fixed values of K_{rest} = 40 and K_{wander} = 40, and the observables are: $x, \dot{x}, y, \dot{y}, a, \dot{a}$, where a is the mouse heading angle.

Basic Behavior Discovery
First, the basic behavior units (rest, explore) are determined using the affinity graph method. Figure 11 shows the segmentation of part of the data sequence (the actual behavior sequence is shown as positive values and the segmented as negative to enhance the visual effect). A critical parameter in this temporal sequence analysis is the subsample time period, T (set to 3 here). As can be seen in the figure, there is a little error at the onset of each behavior segment. The error in segmentation (i.e., number of time steps incorrectly labeled) is about 3%. We are going to develop a statistical algorithm that can choose the optimal T parameter.

Complex Behaviors
The time spent in individual behavior units is used to develop a statistical model of the transition probability. These probabilities may be used to form different types of models (HMM, etc.); here we recover the same form of model as was used to generate the data in order to allow a straightforward comparison of the results. The best parameters for $F_{rest \to wander}(t)$, and $F_{wander \to rest}(t)$ are then determined. Let C_{rest} be the total number of resting BBUs and C_{wander} be the total number of wandering BBUs. Transition likelihoods are calculated using the length of time spent in each BBU;

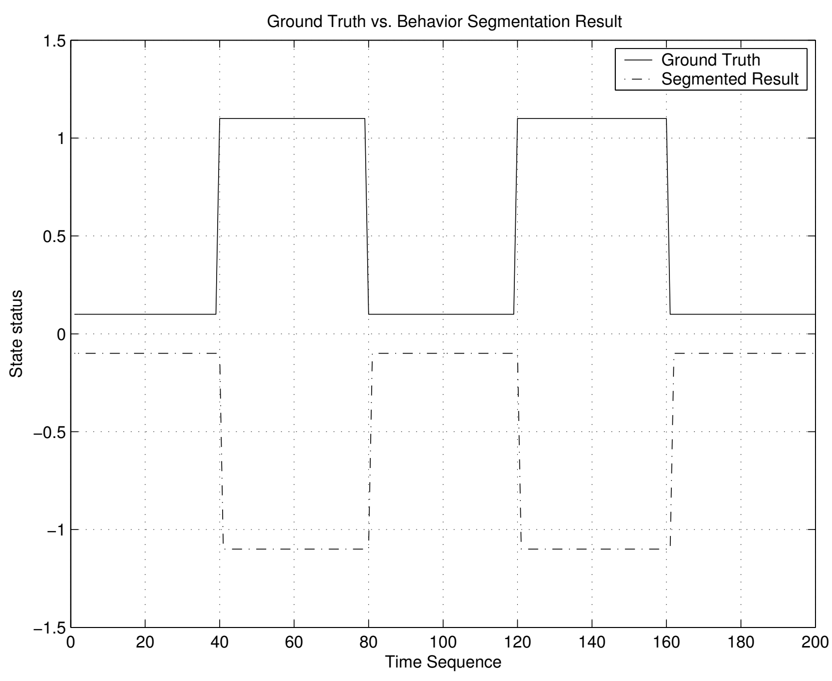


Figure 11: Comparison of segmented result with ground truth.

let the length of the i^{th} BBU be $|BBU_{state,i}|$; then:

$$L_{rest \to wander}(t) = \frac{|\{|BBU_{rest,i}| < t\}|}{C_{rest}}$$

$$L_{wander \rightarrow rest}(t) = \frac{|\{|BBU_{wander,i}| < t\}|}{C_{wander}}$$

Figure 12 shows the histogram of the times spent at rest, and Figure 13 shows the cumulative likelihood of transition curve derived from the histogram (i.e., its integral). $\hat{K}_{rest}$, the estimated value of K_{rest}, is 37.5 (versus 40). Next consider the role of physiological, contextual or conceptual

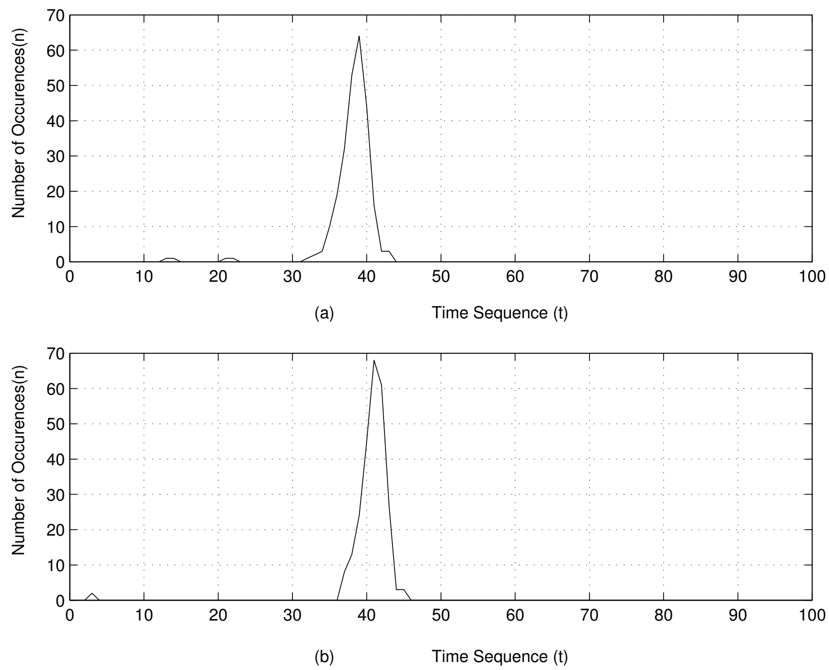


Figure 12: Histogram of Action Transition Intervals for (a) Rest⇒Explore Transition and (b) Explore⇒Rest Transition.

variables in determining behavior. Our premise is that these variables change the parameter or form of the behavior likelihood functions. For simplicity of the demonstration, we assume that only the function parameter changes with the change in physiology, context or conceptual frame of mind. For example, suppose that the mouse tends to rest for longer periods and wander for shorter periods when it is dark; then the resting transition likelihood function shifts to the right, and the wander function to the left. If a behavior sequence is available which includes periods of dark and light, then this is readily determined by the appearance of multiple peaks in the transition time histogram (see Figure 14).

Functions with the appropriate respective parameters for light and dark can then be found. Figure 15 shows this with the shifted versions of the transition likelihood functions. It is also possible to determine the causal role of light if the observed data includes some measure of the phenomenon (e.g., light intensity as a function of time).

For physiological and conceptual variables, there will be no corresponding observable data. However, it is still possible to detect multiple peaks in the behavior time histogram and infer hidden

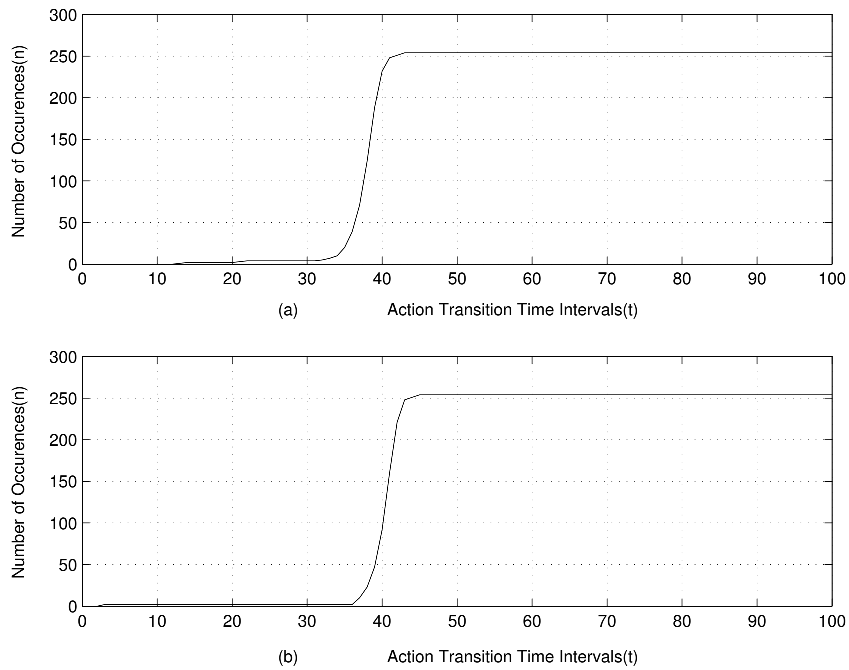


Figure 13: Cumulative Histogram of Action Transition Intervals for (a) Rest⇒Explore Transition and (b) Explore⇒Rest Transition.

variables.

4.1.3 Mouse-in-Cage: Synthetic Video

Behavior Model We synthesized several clips of the mouse-in-cage scenario with a simple mouse shape built from ellipsoids and using four behaviors: resting, exploring, eating (reaching up to reach the ̈food)̈, and grooming (standing on tail with two front legs brushing the head with slight body motion), as shown in Figure 16. The little sphere in the center of image represents ̈food.̈

This 2000-frame synthetic video sequence consists of 8 rest segments, 4 segments of reaching up, 2 grooming segments, and the rest is exploring segments. The synthetic behavior generation follows similar equations as the rest, explore transition probability functions described in the two-state problem. The labeled behavior sequence is shown in Figure 18.

This synthetic video (which makes tracking easier)is very helpful in studying the effectiveness of the proposed technique. The mouse moves around in 3D space, sometimes closer to the viewer, and sometimes farther away (thus smaller). This requires that the features be scale invariant. Also, note the mouse moves in random directions, which increases the technical challenge of BBU discovery.

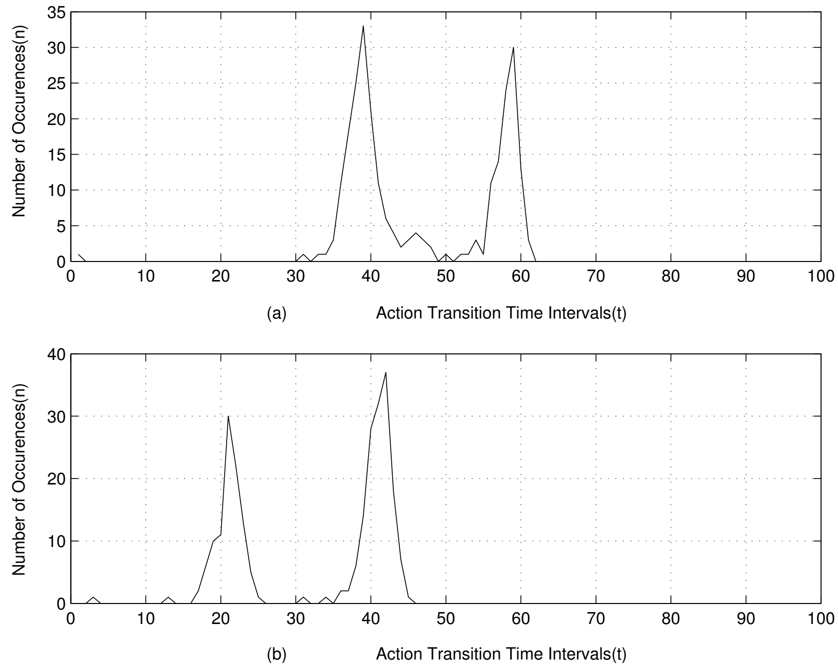


Figure 14: Histogram of Action Transition Intervals with Light Context, for (a) Rest⇒Explore Transition and (b) Explore⇒Rest Transition.

Feature Extraction
The tracking result using the level-set tracking framework is show in Figure 17.

Basic Behavior Discovery We experimented with the following features extracted from the silhouette, as result of the contour tracking: position (centroid of the blob), speed (of the blob centroid), principle axes of the blob, principle axes change, aspect ratio (width/height), and similar features of the motion history image (MHI) [25]. We used a subsequence of length 10 and slides one frame at a time in the experiments. We have tried two approaches: one using combined weighted features in the BBU detection step, the other using a sequential inference approach. The experiment results show that the global motion of the blob is a good feature for segmenting out the frames with no or slight motion. The change of principle axes, and features of MHI are good to separate the grooming (slight global motion, with local motion) from resting behavior, and separate the reaching up behavior from the exploration behavior. Based upon this observation, we come up with the idea of sequential hierarchical BBU segmentation with the affinity method:

1) Select the feature set with most distinguishing power, and perform affinity method with these features. This segments the image sequence into several segments.

2) Select the next feature set with most distinguishing power, and perform BBU segmentation with these features on the segments produced by previous step.

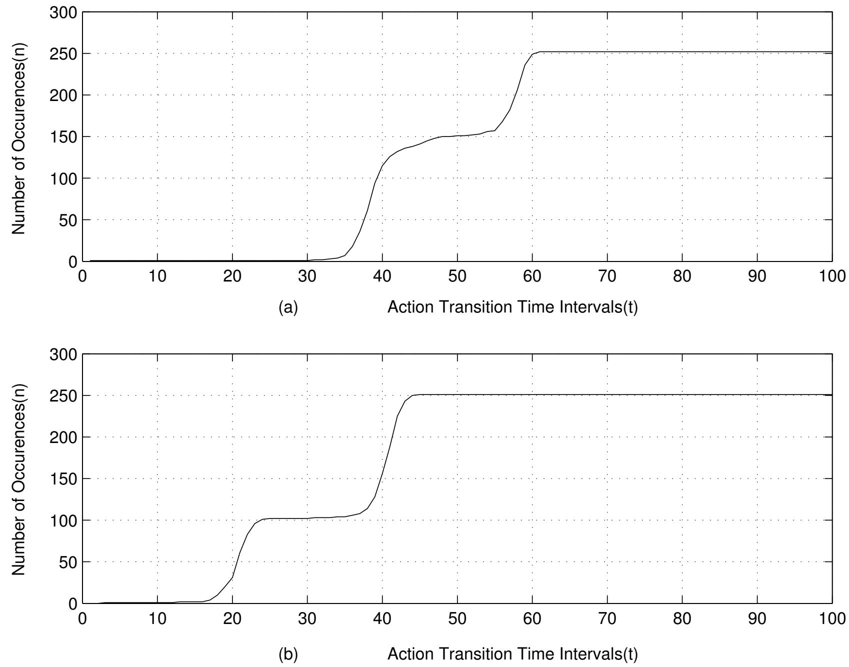
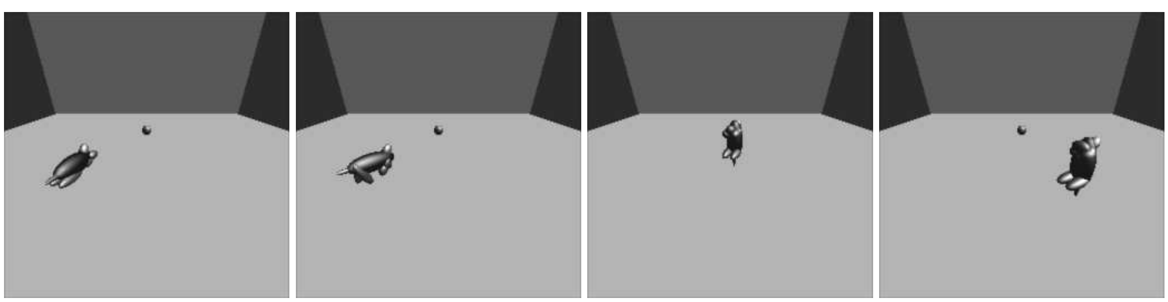


Figure 15: Cumulative Histogram of Action Transition Intervals with Light Context, for (a) Rest⇒Explore Transition and (b) Explore⇒Rest Transition.



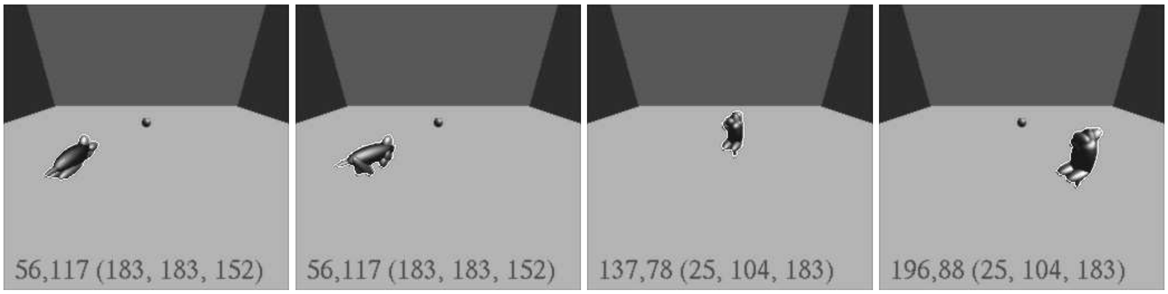
(a) Rest (b) Explore (c) Eat (d) Groom

Figure 16: Synthetic Mouse-in-Cage Scenario Video Clips.

3) Repeat step 2) with the rest of the features.

Here we segment the video into static and dynamic sequences using the affinity measure on speed feature in step 1. Then the rest of the features are used to segment the *groom* behavior from the *rest* behavior, and segment the *reachup* behavior from the *explore* behavior.

The BBU detection results separating the static and dynamic sequences is shown in Figure 19, and the error rate is about 4%. Notice that one of the *reachup* behaviors (at around frame 1300) was mostly misclassified as a static sequence, this is because the mouse's back is to the viewer, which



(a) Rest (b) Explore (c) Eat (d) Groom

Figure 17: Synthetic Mouse-in-Cage Scenario Tracking Result.

makes the reaching up and down action not obvious. The BBU detection result for the next step is shown in Figure 20. The overall misclassification rate is about 8%, which includes false positive rate and missing detection.

The errors come from two major sources, one of which is the selection of features. The other is the affinity measure and the optimal choice of parameters (such as subsequence length, skip length, weights of features, etc.). Another improvement can be made by incorporating the behavior model in BBU detection, which will help predict the behavior in next frame, and verify against the measure.

4.2 Mouse-in-Cage: Real Video

We got this video from Prof. Mario Capecchi in the University of Utah Medical School. Some snapshots of the mouse resting, exploring, standing up, and grooming are shown in Figure 21.

4.2.1 Feature Extraction

Tracking The tracking result using the level-set tracking framework is shown in Figure 22.

4.2.2 Basic Behavior Discovery

We are going to use same set of features used in the synthetic video in the real mouse video, and explore new features for good BBU detection. Similar combined feature and sequential inference techniques used for the synthetic video data will be applied here.

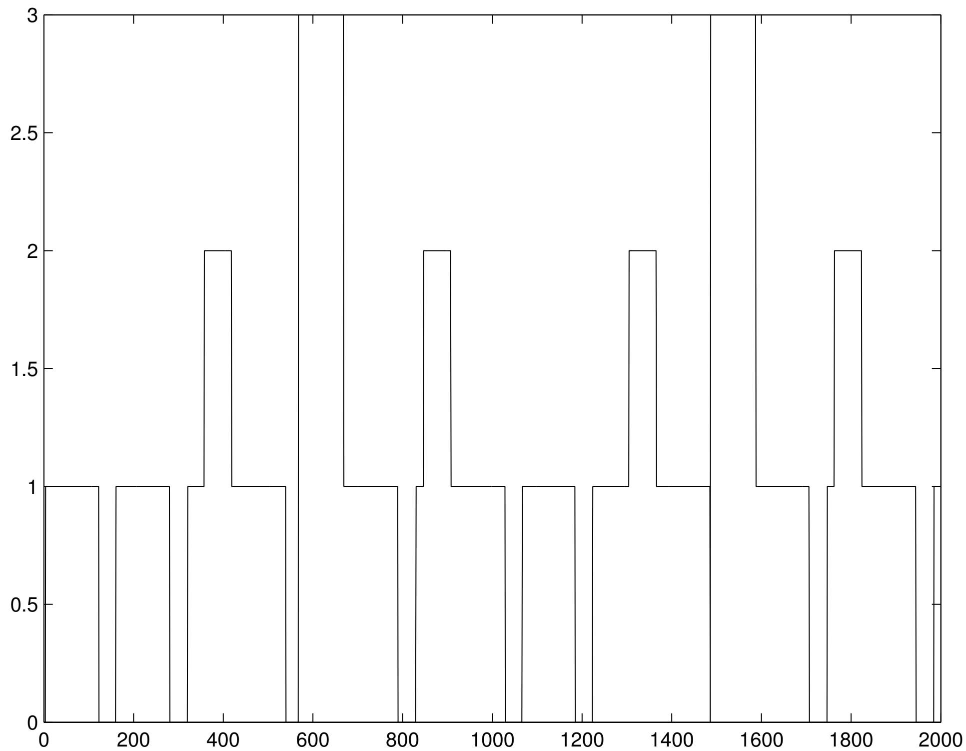


Figure 18: Behaviors in the synthetic video sequence. (Rest = 0, Explore = 1, Reachup = 2, Groom = 3)

5 Conclusions and Future Work

We have proposed a framework for automatic video based behavior analysis system, consisting of behavior modeling, feature extraction, BBU discovery and complex behavior recognition modules. We presented our methodology for implementing each module. We presented a four-level (physical, physiological, contextual, and conceptual) animal behavior model, and we showed preliminary results on feature extraction and BBU discovery with the synthetic mouse video. The low BBU discovery error rate indicates that the affinity method is a promising BBU grouping method. In addition, we simulated the construction and recovering of complex behaviors through a simple two-state mouse problem. The advantage of using synthetic data or simulation is that we can study the effectiveness of the proposed methods and isolate the problems in each stage by comparing with the ground truth.

In the future, we are going to apply the proposed feature extraction (e.g., object contour tracking) BBU discovery method to the real mouse video. We are going to explore these methods to solve the challenges posed by the low-quality real mouse video (where some portion of the background and mouse are indistinguishable in color) and explore new spatial-temporal features.

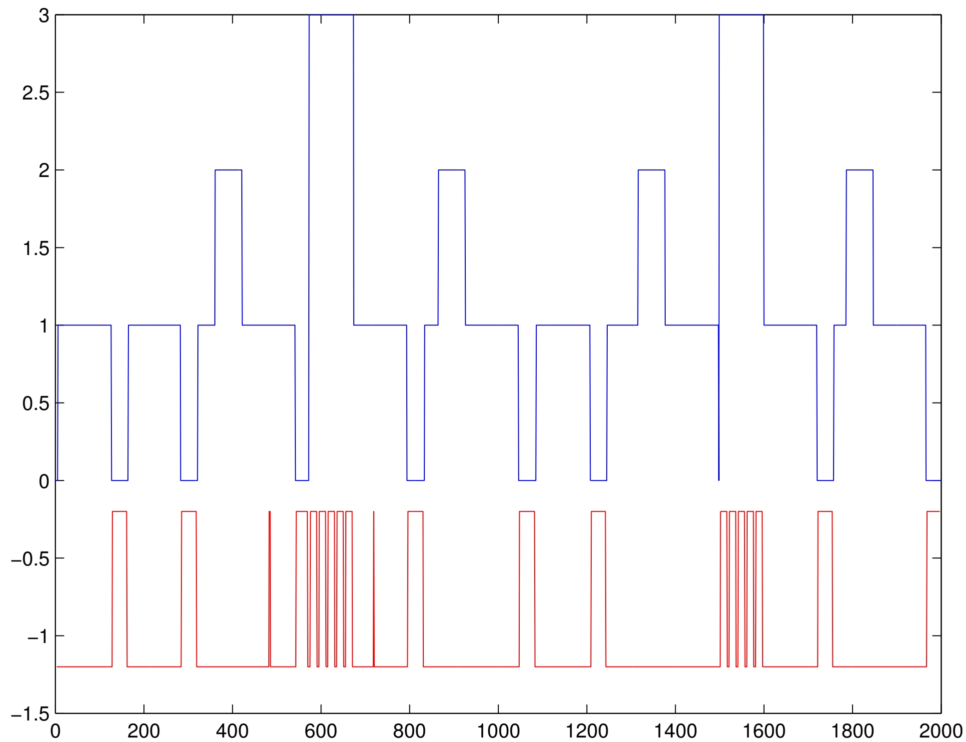


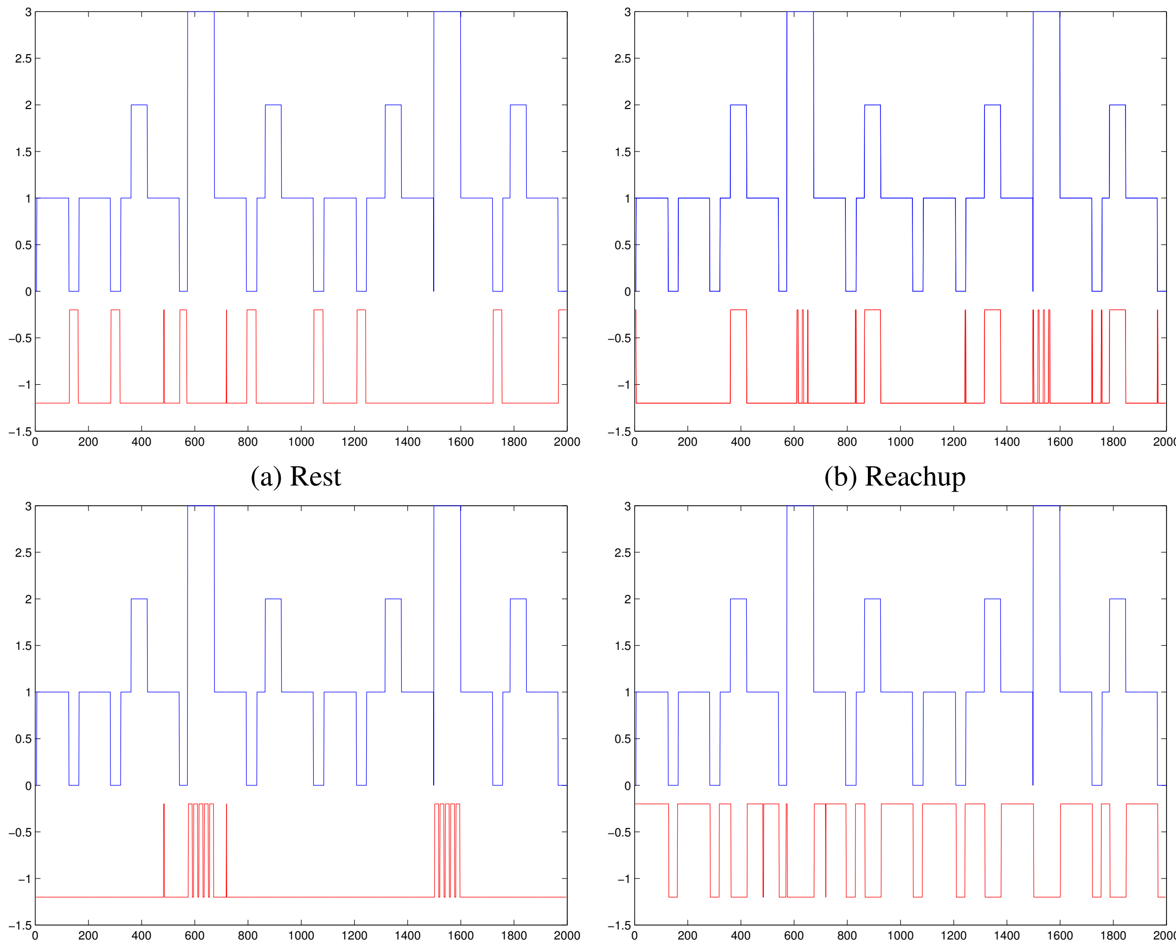
Figure 19: BBU Discovery for Static Frames (*rest* and *groom*). The blue line is the ground truth, and the red line is the detected result.

Meanwhile, from our experiments with the synthetic data, we have noticed that in applying the affinity method in BBU discovery, optimal feature (spatio-temporal features with strong BBU-distinguishing power) and parameter (size of temporal subsequence, and number of frames to skip) selection are critical to successfully cluster the BBUs. We plan to apply the proposed statistical method in this task.

Furthermore, conducting complex video animal behavior analysis and uncovering underlying behavior models is another area for our future research effort. Here the complex behaviors involve the spatial-temporal constraints (between object BBU and environment, between BBUs of the same object, and between different objects) from the contextual level and the goal-directed rules from the conceptual level. We plan to explore the HMM model or the hierarchical approach for the complex behavior recognition.

Finally, we are going to explore the application of the four-level behavior models (i.e., the temporal state models for physical and physiological levels, the contextual relations, and the goal-directed or motivational model as shown in Figure 8) to feature extraction, BBU and complex behavior recognition. Our hope is that the behavior model will help increase the accuracy of behavior recognition.

Once these models are successfully applied to the real mouse video, we will then extend this frame-



(a) Rest (b) Reachup (c) Groom (d) Explore

Figure 20: BBU Discovery Result: a) *rest* b) *reachup* c) *groom* d) *explore*. The blue line is the ground truth. The red line is the detected result.

work and methodology to other video-based object behavior analysis applications, e.g., human activity analysis, sport analysis, etc.

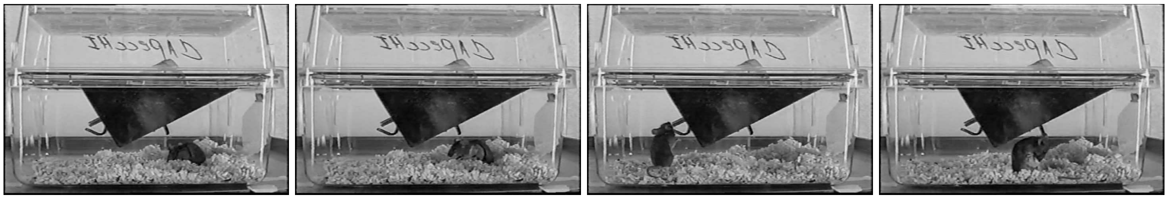
References

[1] *http://www.sandiegoinstruments.com/prod_smart.htm*.

[2] *http://www.med-associates.com/new_prod/video.htm*.

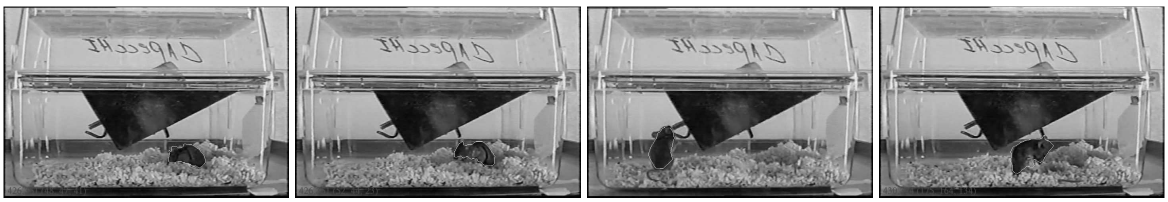
[3] *http://www.qubitsystems.com/track.html*.

[4] *http://www.biobserve.com/products/trackit/index.html*.



(a) Rest (b) Explore (c) Reachup (d) Groom

Figure 21: Mouse-in-Cage Scenario Real Video Clips.



(a) Rest (b) Explore (c) Reachup (d) Groom

Figure 22: Mouse-in-Cage Scenario Real Video Tracking Result.

[5] *http://www.vicon.com/products/peakmotussoftware.html*.

[6] *http://www.actimetrics.com/bigbrother/BigBrother.html*.

[7] J. Aggarval and Q. Cai. Human motion analysis: A review. *Computer Vision and Image Understanding*, 73(3), 1999.

[8] J. Aggarval, Q. Cai, W. Liao, and B. Sabata. Articulated and elastic non-rigid motion: A review. In *In Workshop on Motion of Non-Rigid and Articulated Objects*, pages 2–14, Austin, Texas, USA, 1994.

[9] J. Albus. *Brain, Behavior, and Robotics*. Byte Books, Peterborough, NH, 1981.

[10] J. Albus and A. J. Barbera. Rcs: a cognitive architecture for intelligent multi-agent systems. In *Plenary Talk, PerMISÓ4*, 2004.

[11] J. F. Allen and G. Ferguson. Actions and events in interval temporal logic. *Journal of Logic and Computation*, 4(5):531–579, 1994.

[12] M. Arens and H.-H. Nagel. Behavioral knowledge representation for the understanding and creation of video sequences. In A. Gnther, R. Kruse, and B. Neumann, editors, *Proceedings of the 26th German Conference on Artificial Intelligence (KI-2003)*, volume LNAI 2821. Springer-Verlag, 2003.

[13] R. C. Arkin. *Behavior-Based Robotics*. The MIT Press, Cambridge, Massachusetts, 1998.

[14] D. Ayers and M. Shah. Monitoring human behavior from video taken in an office environment. *Journal of Image and Viion Computing*, 19(12):833–846, 2001.

[15] R. Bakeman and J. M. Gottman. *Observing Interaction: An Introduction to Sequential Analysis*. Cambridge University Press, Cambridge, UK, second edition, 1997.

[16] T. Balch, Z. Khan, and M. Veloso. Automatically tracking and analyzing the behavior of live insect colonies. In *AGENTS*, 2001.

[17] J. Barbic, A. Safonova, J.-Y. Pan, C. Faloutsos, J. K. Hodgins, and N. S. Polland. Segmenting motion capture data into distinct behaviors. In *Proceedings of Graphics Interface 2004 (GI'04)*, Canada, May 2004.

[18] J. Bart, M. A. Fligner, and W. I. Notz. *Sampling and Statistical Methods for Behavioral Ecologists*. Cambridge University Press, New York, 1998.

[19] C. BenAbdelkader, R. G. Cutler, and L. S. Davis. Gait recognition using image self-similarity. *EURASIP Journal on Applied Signal Processing*, 4:572–585, 2004.

[20] M. Brand and V. Kettnaker. Discovery and segmentation of activities in video. *IEEE Transactions on PAMI*, 22(8):844–851, 2000.

[21] T. Burghardt, J. Calic, and B. Thomas. Tracking animals in wildlife videos using face detection. In *European Workshop on the Integration of Knowledge, Semantics and Digital Media Technology*, 2004.

[22] C. Cedras and M. Shah. Motion-based recognition: A survey. *Image and Vision Computing*, 13(2), 1995.

[23] C. W. Clark and M. Mangel. *Dynamic State Variable Models in Ecology: Methods and Applications*. Oxford University Press, New York, 2000.

[24] P. W. Colgan. *Quantitative Ethology*. John Wiley & Sons, New York, 1978.

[25] J. W. Davis and A. F. Bobick. The representation and recognition of action using temporal templates. In *Proc. IEEE CVPR9́7*, 1997.

[26] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern classification*. Wiley, New York, 2nd edition, 2001.

[27] L. A. Dugatkin and H. K. Reeve. *Game Theory and Animal Behavior*. Oxford University Press, Cambridge, UK, 1998.

[28] A. A. Efros, A. C. Berg, G. Mori, and J. Malik. Recognizing action at a distance. In *IEEE International Conference on Computer Vision*, pages 726–733, Nice, France, 2003.

[29] I. Eibl-Eibesfeldt. *Ethology: The Biology of Behavior*. Holt, Rinehart and Winston, second edition, 1975.

[30] H.-L. Eng, K.-A. Toh, W.-Y. Yau, and T.-K. Chiew. Recognition of complex human behaviors in pool environment using foreground silhouette. In *Proceedings of International Symposium on Visual Computing (LNCS 3804)*, pages 371–379, 2005.

[31] G. Figueiredo, T. Dickerson, E. Benson, G. V. Wicklen, and N. Gedamu. Development of machine vision based poultry behavior analysis system. In *ASAE*, Las Vegas, Nevada, 2003.

[32] D. Forsyth and J. Ponce. *Computer Vision: A Modern Approach*. Prentice Hall, Upper Saddle River, NJ, 2003.

[33] J. M. Gottman and A. K. Roy. *Sequential Analysis: A Guide For Behavioral Researchers*. Cambridge University Press, Cambridge, 1990.

[34] B. A. Hazlett. *Quantitative Methods in the Study of Animal Behavior*. Academic Press, New York, second edition, 1977.

[35] T. C. Henderson and X. Xue. Complex behavior analysis: A simulation study. In *ISCA 18th International Conference on Computer Applications in Industry and Engineering (CAINEÓ5)*, Hawaii, 2005.

[36] O. Holland and D. McFarland. *Artificial Ethology*. Oxford University Press, New York, 2001.

[37] W. Hu, T. Tan, L. Wang, and S. Maybank. A survey on visual surveillance of object motion and behaviors. *IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS*, 34(3):334–351, 2004.

[38] Y. Ivanov and A. Bobick. Recognition of visual activities and interactions by stochastic parsing. *IEEE Transactions on PAMI*, 22(8):852–872, 2000.

[39] O. C. Jenkins and M. J. Mataric. Deriving action and behavior primitives from human motion data. In *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems(IROS)*, pages 2551–2556, 2002.

[40] S. Kumar, F. Ramos, B. Upcroft, and H. Durrant-Whyte. A statitical framework for natural feature representation. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2005.

[41] I. L. MacDonald and W. Zucchini. *Hidden Markov and Other Models for Discrete-valued Time Series*. Chapman & Hall, New York, 1997.

[42] M. Magnusson. Discovering hidden time patterns in behavior: T-patterns and their detection. *Behavior Research Methods, Instruments & Computers*, 32:93–110, 2000.

[43] M. Mangel and C. W. Clark. *Dynamic Modeling in Behavioral Ecology*. Princeton University Press, Princeton, NJ, 1988.

[44] P. Martin and F. P. Bateson. *Measuring Behavior: An Introductory Guide*. Cambridge University Press, Cambridge,UK, second edition, 1993.

[45] D. McFarland and A. Houston. *Quantitative Ethology: The State Space Approach*. Pitman Advanced Publishing Program, Boston,MA, 1981.

[46] G. G. Medioni, I. Cohen, F. Bremond, S. Hongeng, and R. Nevatia. Event detection and analysis from video streams. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(8):873–889, 2001.

[47] T. Moeslund and E. Granum. A survey of computer vision-based human motion capture. *Computer Vision and Image Understanding*, 81(3):231–268, 2001.

[48] R. Nevatia, S. Hongeng, and F. Bremond. Video-based event recognition: activity representation and probabilistic recognition methods. *Journal of Computer Vision and Image Understanding*, 96(2):129–162, 2004.

[49] R. Nevatia, T. Zhao, and S. Hongeng. Hierarchical language-based representation of events in video streams. In *The 2nd IEEE Second IEEE Workshop on Event Mining in conjunction with IEEE CVPR03*, 2003.

[50] L. Noldus, A. J. Spink, and R. A. Tegelenbosch. Ethovision: A versatile video tracking system for automation of behavioral experiments. *Behavior Research Methods, Instruments, & Computers*, 33(3):398–414, 2001.

[51] N. M. Oliver, B. Rosario, and A. P. Pentland. A bayesian computer vision system for modeling human interactions. *IEEE Transactions on PAMI*, 22(8), 2000.

[52] N. Paragios and R. Deriche. Geodecic active contours and level sets for detection and tracking of moving objects. *IEEE Transactions on PAMI*, 22(3):266–280, 2000.

[53] P. Perner. Motion tracking of animal for behavior analysis. In *Visual Forum, LNAI*, volume 2059, pages 779–787. Springer-Verlag, 2001.

[54] F. Porikli and T. Haga. Event detection by eigenvector decomposition using object and frame features. In *Workshop on Event Mining, IEEE ICCV*, 2004.

[55] H. Prendinger. *Life-Like Characters: Tools, Affective Functions and Applications*. Springer, New York, 2003.

[56] L. R. Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, 1989.

[57] F. Ramos, B. Upcroft, S. Kumar, and H. Durrant-Whyte. A bayesian approach for place recognition. In *IJCAI Workshop on Reasoning with Uncertainty in Robotics*, Edinburgh, Scotland, 2005.

[58] N. A. Rota and M. Thonnat. Activity recognition from video sequences using declarative models. In *Proc. European Conf. on A.I.*, 2002.

[59] D. Sergeant, R. Boyle, and M. Forbes. Computer visual tracking of poultry. *Computers and Electronics in Agriculture*, 21:1–18, 1998.

[60] J. Sethian. *Level Set Methods and Fast Marching Methods*. Cambridge Univ. Press, 1999.

[61] Y. Shi and W. C. Karl. Real-time tracking using level sets. In *Proceedings of IEEE CVPR\'05*, 2005.

[62] J. B. Tenenbaum, V. de Silva, and J. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science Magazine*, 290, 2000.

[63] F. M. Toates. *Animal Behavior: A Systems Approach*. John Wiley & Sons, New York, 1980.

[64] K. Torkkola. Feature extraction by non-parametric mutual information maximization. *Journal of Machine Learning Research*, 3:1415–1438, 2003.

[65] C. Twining, C. Taylor, and P.Courtney. Robust tracking and posture description for laboratory rodents using active shape models. *Behavior Research Methods, Instruments, & Computers*, 33(3):381–391, 2001.

[66] B. Vemuri, J. Ye, and C. Leonard. Image registration via level-set motion: applications to altas-based segmentation. *Medical Image Analysis*, 20:1–20, 2003.

[67] L. Wang, W. Hu, and T. Tan. Recent developments in human motion analysis. *Chinese Journal of Computers*, 25(3):225–237, 2002.

[68] Y. Weiss. Segmentation using eigenvectors: a unifying view. In *Proc. IEEE International Conference on Computer Vision*, pages 975–982, 1999.

[69] R. T. Whitaker. A level-set approach to 3d reconstruction from range data. *International Journal of Computer Vision*, 29(3):203–231, 1998.

[70] R. T. Whitaker and X. Xue. Variable-conductance, level-set curvature for image denoising. In *Proceedings of IEEE ICIP\'01*, 2001.

[71] A. D. Wilson and A. F. Bobick. Parametric hidden markov models for gesture recognition. *IEEE Transactions on PAMI*, 21(9), 1999.

[72] L. Zelnik-Manor and M. Irani. Event-based analysis of video. In *Proceedings of IEEE CVPR*, 2001.

[73] S. C. Zhu and A. Yuille. Region competition: unifying snakes, region growing, and bayes/mdl for multiband image segmentation. *IEEE Transactions on PAMI*, 22(3):266–280, 2000.

[74] J. B. Zurn, D. Hohmann, S. Dworkin, , and Y. Motai. A real-time rodent tracking system for both light and dark cycle behavior analysis. In *Proceedings of IEEE Workshop on Applications of Computer Vision*, 2005.