June 1998

Where To Look? Automating Some Visual Attending Behaviors of Human Characters

Sonu Chopra
University of Pennsylvania, schopra@gradient.cis.upenn.edu

University of Pennsylvania Institute for Research in Cognitive Science Technical Report No. IRCS-98-17.
(Dissertation Proposal)

Where To Look? Automating Some Visual Attending Behaviors of Human Characters

Abstract

We propose a method for automatically generating the appropriate attentional (eye gaze or looking) behavior for virtual characters existing or performing tasks in a dynamically changing environment. Such behavior is expected of human-like characters but is usually tedious to animate and often not specified at all as part of the character's explicit actions. In our system, referred to as the AVA (Automated Visual Attending), users enter a list of motor or cognitive actions as input in text format: (*walk to the lamp post, monitor the traffic light, reach for the box, etc*). The system generates the appropriate motions and automatically generates the corresponding attentional behavior. The resulting gaze behavior is produced not only by considering the explicit queue of required tasks, but also by factoring in involuntary visual functions known from human cognitive behavior (attentional capture by exogenous factors, spontaneous looking), the environment being viewed, task interactions, and task load. This method can be adapted to eye and head movement control for any facial model.

Comments

University of Pennsylvania Institute for Research in Cognitive Science Technical Report No. IRCS-98-17. (Dissertation Proposal)

Where To Look? Automating Some Visual Attending Behaviors of Human Characters

Dissertation Proposal

Sonu Chopra
Center for Human Modeling and Simulation
Computer and Information Science Department
University of Pennsylvania
email: schopra@gradient.cis.upenn.edu

June 29, 1998

This research is partially supported by U.S. Air Force through Delivery Order #8 on F41624-97-D-5002; Office of Naval Research (through Univ. of Houston) K-5-55043/3916-1552793; DARPA SB-MDA-97-2951001 (through the Franklin Institute); NSF IRI95-04372; NASA NRA NAG 5-3990; and JustSystem Japan.

Abstract

We propose a method for automatically generating the appropriate attentional (eye gaze or looking) behavior for virtual characters existing or performing tasks in a dynamically changing environment. Such behavior is expected of human-like characters but is usually tedious to animate and often not specified at all as part of the character's explicit actions. In our system, referred to as the *AVA* (Automated Visual Attending), users enter a list of motor or cognitive actions as input in text format: *(walk to the lamp post, monitor the traffic light, reach for the box, etc.)*. The system generates the appropriate motions and automatically generates the corresponding attentional behavior. The resulting gaze behavior is produced not only by considering the explicit queue of required tasks, but also by factoring in involuntary visual functions known from human cognitive behavior (attentional capture by exogenous factors, spontaneous looking), the environment being viewed, task interactions, and task load. This method can be adapted to eye and head movement control for any facial model.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Imagine a human character strolling through the park, noticing events, reaching for a paper, waiting for the light to change before crossing a road and generally avoiding objects (small children, pets, toys) that may be in his path. Where should the agent look? What if several events vie simultaneously for the person's attention? If he stops and just takes in the scene before him, how does the richness and complexity of the environment determine where he looks next?

This goal of this research is to automate the generation of visual attending (eye gaze or looking) behavior given a list of motor and cognitive activities a character should perform. Also, the resulting attending behavior reflects interactions or competition between deliberate (endogenous) tasks and involuntary (exogenous) attentional capture. Further, such behavior is generated in *real-time.* The system we implement takes as input, in text format, a set of actions such as *walk to a goal, monitor the traffic light, monitor oncoming traffic.* The system generates the corresponding figure animation of motion, using the *Jack* system's repertoire of motor skills, and also generates the appropriate eye gaze or looking behavior.

The basic premise of this work associates primitive motor activities (walk, reach, lift, manipulate, ...) and cognitive actions (monitor, visual search, ...) with *patterns* of pre-defined looking behavior. These patterns are estimated in our system based on empirical and qualitative data from the cognitive psychology and human factors literature [Moray, 1993; Abrams, Meyer, and Kornblum, 1990; Fisher, 1986; Yantis, 1993; Jonides, 1981; Yantis and Johnson, 1990; Yantis and Jonides, 1990; Kahneman, 1973; Kito, Haraguchi, Funatsu, Sato, and Kondo, 1989] as well as from computer vision research [Ballard, Hayhoe, Li, and Whitehead, 1992; Rao, Zelinsky, Hayhoe, and Ballard, 1996] and from simple observation. The inspiration for such a premise comes from experiments done in [Yarbus, 1967] that illustrate that people perform a characteristic *pattern and frequency* of eye movements for a given task. Secondary elements in the scene remain unexamined. When an agent is engaged in more than one task that requires the same sensory resource, as expected, performance degrades versus the single task condition [Hirst, 1986]. We also encode involuntary functions known to exist in the human visual system [Hillstrom and Yantis, 1994; Kahneman, 1973; Yantis, 1993]. The *AVA* will generate a character's pattern of attention based on competing voluntary and involuntary behaviors, anticipation, task load, and the environment being viewed.

Related experiments in psychology and vision provide insights regarding *specific* instances of attending behavior for given circumstances (e.g., visual search, eye targeting for reach motions). The contribution of this dissertation is to provide a framework in which eye movement patterns for specific actions can be *combined* and *interact* with each other (providing a simulation of cognitive load) and with exogenous factors (illustrating attentional capture by task unrelated events). This interaction is modeled as a three-level hierarchy in which behaviors related to intentional tasks have the highest precedence, exogenous behaviors the next highest precedence and spontaneous looking (or idling behavior) has the least.

1.1 Motivation and Applications

In order to simulate believable agents, gaze must be directed appropriately. Since gaze is significant in communication and behavioral representation, random or uncontrolled gaze is both misleading and disconcerting. Characters for which motion alone is animated while gaze remains fixed appear robot-like or mechanical. Also, *real-time* performance is necessary for interactive simulation and games.

Some potential applications of this research are:

- A tool for animators to generate the looking behavior appropriate to a set of tasks an agent should perform. Tasks may be non-manual cognitive activities (like visual search or monitoring) or motor activities like walk and reach. This behavior will also respond to task unrelated events (such as peripheral motion) and the dynamics of the agent's environment. The set of supported tasks and scope of scenarios for which this research is appropriate is presented in Section 1.3.

- Virtual reality immersive games. Human players anticipate that animated players move and behave appropriately to the circumstances of the game. Since game environments are typically changing, characters' responses cannot be scripted in advance.

- Determining the ergonomics of computer simulated environments. This research associates standard frequencies of eye movements for primitive cognitive and motor tasks. Frequencies are adjusted *automatically* in the implementation reflecting degradation in performance due to increasing cognitive load or interference from exogenous factors in the environment. Our model of eye behavior could indicate when critical events in the environment remain unattended.

1.2 Generating Attending Behavior Versus Planning or Learning to Attend

This research does not propose a theory of cognition or a production system approach to learning about and dealing with a simulated environment. A system such as SOAR [Lewis, Huffman, John, Laird, Lehman, Newell, Rosenbloom, Simon, and Tessler, 1993], alternately, provides a large, knowledge-based cognitive modeling implementation. Rather than an explicit representation of working memory, this research associates *memory uncertainty thresholds* with activities (uncertainty thresholds determined by empirical data or observation) that determine the frequency with which task related sites are glanced at. However, our technique will generate the appropriate *looking behavior* for an agent and can be used to determine if, due to the demands of simultaneous tasks or exogenous factors, critical events remain unattended. The goal of this approach is to automate visual attending behavior in *real-time* with minimal computational overhead.

Since the *AVA* is developed for the computer animation and graphics domain, all objects within an agent's *field of view* are assumed be known (with parameters such as object location, location of relevant landmarks or sites, and features such as color available to the agent). Computer vision approaches that are used to model spontaneous looking, visual search or visual targeting [Koch and Ullman, 1985; Rao, Zelinsky, Hayhoe, and Ballard, 1996; Ballard, Hayhoe, Li, and Whitehead, 1992; Marjanovic, Scassellati, and Williamson, 1996], on the other hand, must process camera images to determine the location of relevant features or objects in the environment. Since the graphics database (a complete listing of the environment's geometry including objects, edges, sites, color) is accessible to our implementation, we avoid the overhead of image processing except at the lowest level of gaze behavior (where we used a simplified image processing approach for spontaneous or idling gaze behavior in absence of task).

1.3 Scope of Potential Scenarios

The class of scenarios for which this research is appropriate are those which can be generated from combining the technique's supported set of motor and cognitive primitives. Motor primitives for which our technique supports eye behaviors are: walk, reach, grasp, lift, put down, pull and push. Cognitive primitives in the *AVA* with associated eye behaviors are: monitor, search (for a target), and limit monitor (monitor more frequently under a given circumstance). For example, a scenario where the agent searches for a target (e.g. "find the blue table"), walks to the target, reaches for and manipulates an object (e.g. "pick up the newspaper") and then walks to a destination ("walk to exit") is a simple case that combines several of this method's supported primitives. Recall that the *AVA* technique will generate behavior by considering the *amalgam* of simultaneously executing tasks (since some tasks such as monitoring and locomotion can proceed in parallel) *and* by factoring in involuntary looking behavior. In our simple example , if an object flies into the agent's field of view (and no other task demands are active), the agent will track it. Also, an agent will often lapse into idling behavior while a task is active. For example, when walking to a goal, the agent will not need to continuously look either at the goal or the ground in front of him (he will only do so when the memory uncertainty thresholds for those locations are reached).

Additional primitives may be added to the *AVA* if the frequency and general pattern of eye movements for the primitive are known or can be obtained from empirical data. For example, when climbing a ladder, a possible pattern and frequency input to our technique may be to look at the next rung and, when the next step is initiated, look at the following rung. Also, if a particular set of object-specific features are relevant, they may be added to our system. When glancing other characters in the simulation, for example, the eyes and mouth of the other agent may be scanned. When tracking a car, the headlights and driver may be looked at.

The focus of this research is to provide a psychologically plausible framework in which deliberate, involuntary and idling behaviors compete. Also, a set of *predefined* looking behaviors that correspond to common motor and cognitive primitives are provided in our implementation.

1.4 Requirements of the Human Model

This research is implemented using the *Jack* human modeling software. However, *any* virtual human model which supports a head and eye control mechanism can be integrated with the *AVA*. Essentially, our method provides either a site (a named 3D location) or 3D location in the environment which the head and eye controller must target. Also, since our method supports eye behaviors for various motor skills, any scenario requiring those skills (e.g., locomotion, reach) will require a human model that is capable of animating those motor capabilities (numerous models already exist that are capable of such basic skills).

1.5 Outline of Proposal

This proposal is organized in subsequent chapters as follows:

- We discuss alternate approaches that have been used in generating eye gaze behavior including: robotics and vision research, motion capture, facial animation and image processing techniques.

- We review related work in the psychology literature that provides the basis for our methodology.

- We outline the hierarchy of eye gaze behaviors which compete in our system.

- We relate cognitive and motor activity with patterns of looking and uncertainty levels.

- We provide a worked example illustrating how our technique's major data structures change and adapt over the course of a simulation.
- We summarize the goals of the *AVA* and provide an outline for future work and extensions.

Chapter 2

Related Work

2.1 Overview

This chapter discusses complementary or parallel research involving the determination of visual attending behavior.

Robotics and computer vision researchers are concerned with developing robots that exhibit human like behavior. Also, in computer vision applications, determining the focus of attention aids in reducing complexity of processing (attention acts as a a filter that selects which regions of interest to process in camera images) [Brooks, Breazeal, Irie, Kemp, Marjanovic, Scassellati, and Williamson, 1998; Marjanovic, Scassellati, and Williamson, 1996].

Image processing and vision techniques have been developed that attempt to model where humans look in the absence of task. Our method incorporates a simplified version of such approaches [Tsotsos, Culhane, Wai, Lai, and Nufflo, 1995; Koch and Ullman, 1985].

Research in animation has explored issues of eye engagement during social interactions or discourse between virtual agents [Cassell, Pelachaud, Badler, Steedman, Achorn, Becket, Douville, Prevost, and Stone, 1994]. Similarly, visual cues of attention between a robot and a human instructor are explored in [Scassellati, 1996]. The

AVA may be used to extend systems that deal with issues of facial animation and social interaction of virtual agents.

Motion capture methods, including eye tracking, are used to replay prerecorded motion or behaviors. While considerably more accurate then our method, such techniques are essentially scripted and unsuitable for dynamic simulations.

2.1.1 Cog

M.I.T.'s Cog Project [Brooks, Breazeal, Irie, Kemp, Marjanovic, Scassellati, and Williamson, 1998; Scassellati, 1996; Marjanovic, Scassellati, and Williamson, 1996] aims to develop a humanoid robot which learns or acquires skills during its interactions with its environment.

Two experiments relating attention and action in this research are visually guided pointing [Marjanovic, Scassellati, and Williamson, 1996] and estimating a human instructor's line of sight [Scassellati, 1996]. In the pointing task, Cog's head and eye controller mechanism learns the mapping between locations in the environment being viewed (a camera image) and the appropriate joint angles necessary to align the head and eyes with a target. Also, visual feedback is provided to the robot's arm control mechanism as a means of learning how to point to the visual target. A motion detection algorithm is used to determine the end point of the arm. Since the predicted position of the arm (within the center of field of view) and actual position may differ, the corresponding error term is used to tune weights in the arm control algorithm. In this experiment, attention is categorized as a neighborhood in the camera image where the arm end point is *expected* to be. Similarly, interactions between the Cog robot and a human instructor are examined in [Scassellati, 1996]. Such work explores issues such as responding to an instructor's attentional cues and pointing to request shared attention.

Unlike our technique, this research is concerned with acquiring or learning behavior from the ground up (i.e. as a child might learn how to fixate and point to targets).

Issues of competing events and interference from exogenous factors is not addressed and is not the intent of such work. Also, in our method, motor behavior may be modified due to increasing cognitive load. However, unlike Cog, it is understood that the agent has sufficient experience to perform a basic set of motor skills such as walk, reach, grasp, etc.

2.1.2 Conversational Agents and Social Interaction

Limited rules of eye engagement between animated participants *in conversation* are discussed in [Cassell, Pelachaud, Badler, Steedman, Achorn, Becket, Douville, Prevost, and Stone, 1994] based on psychological observations from [Argyle and Cook, 1976]. Looking behaviors such as head nods to signify turn taking in conversation, using gaze to determine how an utterance is being received and using gaze to accompany accent or emphasis are defined. The domain of this work , similar to research in modeling interactions between a human-like robot and instructor in [Scassellati, 1996], relates rules of social interaction, social cues and feedback with looking behaviors.

2.1.3 Facial Animation Systems

Facial animation systems [Parke and Waters, 1996; Kalra, Mangili, Magnenat-Thalmann, and Thalmann, 1991; Pearce, Wyvill, Wyvill, and Hill, 1986] relate expression and facial muscle movement to emotion. Eye expression rather than pattern of eye movement is addressed in such applications. A traditional hand animation reference [Thomas and Johnson, 1981] also discusses various known eye expressions (surprise, anger, happiness) but does not provide guidance for estimating *patterns* of looking. Our methodology, in contrast, is concerned with the *pattern and frequency* of eye movements in general settings.

2.1.4 Motion Capture Techniques

Motion capture and facial tracking systems are used to recreate the behavior of a human actor performing specified actions. Recovering line of sight from facial images is processing intensive [Scassellati, 1996; Marjanovic, Scassellati, and Williamson, 1996] while head mounted eye trackers are cumbersome or, at minimum, movement limiting [Crane, 1994]. In fact, technology for real-time body motion capture, whether optical or electromagnetic, virtually precludes the simultaneous capture of eye motions. Hence, when introducing characters to a changing environment as found in interactive multi-user games, pre-recorded behavior is not sufficient to animate the eyes. Human behavior in such systems should be reactive and even proactive: it cannot be scripted in advance.

2.1.5 Image Processing Approaches

Neural net [Tsotsos, Culhane, Wai, Lai, and Nufflo, 1995; Koch and Ullman, 1985] models of attention map task demands into feature primitives such as color, orientation, and luminance. In order to emulate voluntary task-driven control of attention, spatial areas in an image with relevant features are activated (combinations of important features will receive higher activation). Exogenous, or task unrelated stimulation of attention is modeled by activating areas with high local feature contrast. Such approaches are computationally intensive and are usually applied to a given *single* task [Koch and Ullman, 1985] or in the absence of any task motivation [Tsotsos, Culhane, Wai, Lai, and Nufflo, 1995; Koch and Ullman, 1985].

Since our goal is real-time animation, our technique operates at the level of object features or sites whenever feasible. In the absence of any deliberate task or exogenous capture, we model a type of idling behavior known as spontaneous looking. We incorporate a simplified image processing technique, explained in Section 4.2.5, to generate such behavior.

2.1.6 Other Related Techniques

Parallel distributed models in the cognitive science literature [Cohen and Huston, 1994; Cohen, Dunbar, and McClelland, 1990] map task features such as color or words into network units. Such models are applied in the context of a single given task. Activation and network weights determine task response times. Noser, Renault, and Thalmann [Noser, Renault, and Thalmann, 1995] used visual-guided agent locomotion in their work. Approaches in the visual display design literature examine which preattentive visual features should be used and combined to convey information in a manner that requires the least processing overhead [Lohse, Biolsi, Walker, and Rueter, 1994; Healey, Booth, and Enns, 1996]. All these techniques are usually difficult to generalize and, other than Noser's locomotion work, not applied in the context of combining motor activity and attention.

Chapter 3

Related Psychology Experiments - Parameters for Our Method

A character's attention is directed by volitional, goal-directed aims known as endogenous factors that correspond to the current task(s) being performed. Involuntary attentional capture by irrelevant stimuli such as peripheral motion or local feature contrast are said to be exogenous factors [Yantis, 1993].

The demands of a particular task generate a characteristic *pattern* of eye movements. Depending on an observer's intentions or goals, eye fixations will vary even when directed at the same image. In [Yarbus, 1967], observers were shown a picture and asked to estimate the ages of figures in the picture. Patterns of fixations were directed at the face of each figure. When asked to estimate the "material circumstances" of participants, fixations were directed at the clothes of each figure. Accordingly, in the *AVA* we associate patterns of eye behavior for broad categories of motor and cognitive activity.

The transitioning between simultaneous tasks is characterized in [Allport, Styles, and Hsieh, 1994] as "shifting intentional set." When engaged in more than one task that requires the same sensory modality, performance degrades versus the single task condition (a review of divided attention experiments is found in [Hirst, 1986]). We account for this phenomenon in our method by increasing response time to task targets as the number of events vying for an agent's attention increase.

Attention may be directed *covertly* without explicit shifts of gaze or overtly. Covert shifts of attention are measured by line-motion illusion [Hikosaka, Miyauchi, and Shimojo, 1996], brain activity increase in the V5 area using functional MRI [G. Rees, 1997] or facilitated response times to targets in attended regions [Posner and Cohen, 1980]. Overt shifts in gaze are preceded by shifts in attention [Klein and Pontefract, 1994]. The *AVA* seeks to characterize the *observable* effects of attention shifts relevant to character animation. Hence, covert shifts are relevant in so much as they *interfere* with or increase response time to targets [Jonides, 1981] in unattended locations.

Once attention has shifted, perception of targets in the attended location is facilitated as long as targets appear within 100ms of the shift [Posner and Cohen, 1980]. If targets appear after 300ms, an increase in target detection time occurs [Posner, Rafal, Choate, and Vaughan, 1985]. This phenomenon is known as *inhibition of return* and accounts for attention shifting through space.

When attention is not engaged, eye saccades to targets are within the order of 100ms and are known as express saccades [Fisher, 1986]. When a character is attending to a task, however, eye saccade time between relevant sites will increase to 200ms [Fisher, 1986]. Voluntary engagement of attention acts as a "hold mechanism" [Allport, 1993] and suppresses express saccades to irrelevant stimuli. The tendency to orient gaze toward irrelevant distractors is found in patients with frontal-parietal brain lesions [Ladavas, Zeloni, Zaccara, and Gangeni, 1997] (reflecting impairment of oculomotor control) and in early infancy [Johnson, 1994] (reflecting the underdevelopment of selective attention). This range of behavior could be characterized in our method by a *distractability* parameter that allows a probabilistic sampling of irrelevant stimuli.

What sorts of exogenous factors capture attention and with what frequency? A review of the literature suggests that peripheral events [Jonides, 1981] and *abrupt on-*

sets, the introduction of new perceptual objects into a scene, capture attention [Yantis, 1993] when attention is in a diffuse or divided mode (i.e. the target may appear anywhere). However, when attention is fully engaged in a particular location, capture by onset does not occur [Yantis and Jonides, 1990]. Similarly, functional imaging of brain activity indicates that perception of motion, even in the periphery, is reduced or eliminated when attention is fully consumed by current task demands [G. Rees, 1997]. Since onsets appear to be rare phenomenon in general settings, the *AVA* attempts instead to detect and predict interference from peripheral events.

Moving objects within the center of view do not necessarily capture attention unless motion detection is a necessary feature of the given task [Hillstrom and Yantis, 1994]. Motion generates an onset when it segregates an object from a surrounding perceptual grouping [Hillstrom and Yantis, 1994]. Generally, feature singletons, perceptual features that differ from their backgrounds by color, motion or orientation, interfere with goal directed attention only when the task itself requires "singleton detection mode" [Egeth and Yantis, 1997; Folk, Remington, and Wright, 1994].

In the absence of any given task, attention follows patterns of spontaneous looking [Kahneman, 1973] where areas of high local feature contrast capture interest.

In summary, we see that tasks impose a voluntary pattern of eye movements. As several tasks are simultaneously attempted, performance degrades. Peripheral events capture attention when the agent is engaged in a task which requires diffuse attentiveness (i.e. visual search or divided attention). In the absence of tasks or peripheral stimuli, attention follows patterns of spontaneous looking.

3.1 Patterns of Looking Associated with Cognitive and Motor Activities

Avionics engineering studies have constructed memory uncertainty models that predict the allocation of attention when monitoring cockpit instruments [Moray, 1993].

We generalize such activity by incorporating uncertainty thresholds in our monitoring eye behaviors. For example, we consider locomotion as a generalized class of monitoring task.

We model eye behavior that corresponds to reach and grasp motions by looking at the relevant grasp sites. Once the hand is in close proximity to the goal site, we initiate attention behavior for the next motor action in the task queue. The notion of when to initiate the next eye movement is examined in Section 4.2.6.

We model two classes of exogenous eye behavior: capture by motion in the periphery of view and spontaneous looking. Spontaneous looking, a term coined in [Kahneman, 1973], is a pattern of eye movement in the absence of any explicit task. We propose a technique adapted from image processing approaches in [Koch and Ullman, 1985; Tsotsos, Culhane, Wai, Lai, and Nufflo, 1995].

We implement visual search behavior by having the agent scan the environment until the object of interest is acquired [Rao, Zelinsky, Hayhoe, and Ballard, 1996; Rabbit, 1983]. No computer vision routines are assumed; the graphics database is presumed directly accessible.

Chapter 4

Methods for Attention Control

Our methodology characterizes volitional control and effects of simultaneous task performance. Also, interference between voluntary attention, peripheral events and local feature contrast is examined.

In the following sections, we present the major components of the *AVA*. First, we examine the motor control mechanisms that must be available in any human model which employs our method. We discuss our particular human model's technique of coordinating head and eye position parameters with attended locations. Next, we define the classes of eye behavior used in our system. Behaviors are organized in a three-level hierarchy corresponding to voluntary, exogenous, and spontaneous and are described in Section 4.2. Behaviors, modeled as finite state machines, compete and their interaction generates the *AVA's* animation of eye movements. We then expand several example action requests. Finally, we summarize our goals and suggest a research plan for future work.

4.1 Head-Eye Movement and Motor Control

The *AVA* can be integrated with any virtual human model that supports a mechanism for head and eye movement control. Additionally, we associate attentional behavior for locomotion, arm reach and hand grasps as well as cognitive actions such as visual search and monitoring.

The animated human model used in our method's implementation employs inverse kinematics to create arm motion and collision detection between fingers and grasp object segments to animate hand grasps.

Our human model employs eyes with two degrees of freedom: vertical rotation corresponding to eye tilt and horizontal rotation corresponding to eye pan. Head motion is controlled by a three degree of freedom joint (which controls head tilt, pan and roll). Currently, we set only the head's pan and tilt parameters.

Target object coordinates are converted into joint angles that manipulate our human model's head, neck and eyes. The mechanism which controls our model's head and eye movement is based on a study of eye-head coordination in [Sparks, 1989]. Small gaze shifts produce only eye movement while larger shifts (between 20 to 90 degrees) generate combined head and eye movement. We plan to expand in future work the mechanism that distributes motion between head and eyes. Experiments in [Freedman and Sparks, 1997] indicate that head contribution increases linearly with shift amplitude. The current algorithm which controls our model's eye-head coordination is given in Appendix A.

Given a final position and orientation, our human model's locomotion system generates the corresponding walking motion. Attention behaviors are therefore appropriately and reactively generated during whatever collision-free path the agent actually chooses during locomotion. And, of course, the perception and detection of obstacles and imminent collisions may themselves modify the path taken. Other locomotion models that use sensed information, such as Reynolds' flocks and herds, use sensing to control path [Reynolds, 1987]. What we add are the observations that (1) sensing is a resource to be allocated and directed and that (2) sensing takes time.

4.2 Modules for Attention Behavior

Eye behaviors in our system, referred to as *nets*, are implemented as finite state machines that execute in parallel. Nets embody three types of activities: deliberate or intentional tasks, capture by peripheral motion, and spontaneous looking. We maintain two lists that eye behaviors access and modify: an **IntentionList** which identifies sites or objects (corresponding to deliberate actions) that are currently vying for attention and a **PList** that identifies moving objects within an agent's peripheral field of view. Table 4.1 summarizes the nets used in our system. Additionally, users enter action requests in text format which are stored on a task queue. Sample requests and task queue processing are examined in Section 4.2.7.

Figure 4.1 illustrates our method's architecture. Users enter task requests as text input. A task queue manager consumes such requests and generates the appropriate eye gaze or looking behaviors for given motor or cognitive tasks. The motions which correspond to motor tasks are also generated. Behaviors may:

- sense motion and tag moving objects in the environment,
- identify sites or objects which must be attended due to uncertainty thresholds,
- identify sites relevant to the current motor or cognitive activity,
- remove objects or sites from contention when a motor activity completes,
- generate a series of eye movements.

Behaviors of the same *type* compete equally for an agent's attention. Task related eye behaviors have the highest precedence. As the number of concurrent task eye behaviors increase, response time to targets increases. A probability factor is used to determine overt orienting toward peripheral stimuli. If the agent is engaged in visual search or in a series of parallel tasks (requiring divided attention), the presence and number of peripheral events will increase response time to task-related targets.

Net	**Summary**
Motion Sensor	If an object in the agent's periphery view moves from the previous frame to the current and is not already on **PList**, add the object to **PList**.
Monitor Net	When the memory uncertainty threshold is reached for a relevant object site, add the site to **IntentionList**.
Reach Eye Behavior	Add the relevant reach/grasp sites to **Intentionlist**. When the end effector is close to the target, remove site from **IntentionList**.
Visual Search	Generate intermediate eye movements to target. Add to **IntentionList**.
Limit Nets	If the limit condition (usually when the agent is in close proximity to an object) is satisfied, add object or site to **IntentionList**.
Spontaneous Looking Net	Find locally conspicuous pixels in the agent's field of view. Convert their locations back into pan and tilt coordinates for the agent's head and eyes.
Task Queue Manager	Read the head of TaskQ. If not empty, do the motor or cognitive behavior for the action. Spawn corresponding attention behavior nets. Wait for motor behavior to complete if actions done in sequence.
GazeNet	Arbitrate between three levels of eye behaviors: intentional, exogenous and spontaneous.

Table 4.1: Eye Behavior Nets

Spontaneous looking has the lowest precedence and can be interrupted by any other type of behavior.

A **GazeNet**, illustrated in Figure 4.2, coordinates flow of control between levels of eye behaviors. The algorithm used in **GazeNet** processing is described in Figure 4.3.

We assign behaviors to broad classes of motor and cognitive activities: monitoring and locomotion, reach and grasp, and visual search. Such activities are entered as tasks on a queue. If multiple actions are put on the queue, eye behavior that corresponds to a subsequent action may be initiated before a previous motor activity completes. A task queue manager process, explained in Section 4.2.7, coordinates requested motor actions and spawns the appropriate attentional behavior for each action.

Spontaneous looking behavior has the lowest priority and is invoked in the absence of any sites or objects on the **IntentionList** or peripheral stimuli in the agent's view.

4.2.1 Monitoring and Locomotion

Monitoring tasks (locomotion being a general case) use uncertainty thresholds [Moray, 1993] that relate how often a signal, event, or goal should be glanced at in order to maintain an accurate view of the signal's state in memory. When the uncertainty threshold for a given monitoring task is reached in our system, the relevant site is added to **IntentionList**.

While walking, for example, an agent in our system looks toward the horizon or destination and occasionally glances at the ground [Swain and Stricker, 1993] . This is an example of a monitoring task with high uncertainty thresholds. If the state of the terrain changes, becoming slippery or uneven, for example, the uncertainty threshold associated with the ground plane is reduced, causing the agent to glance more frequently at the ground in front of his feet.

Limit Monitoring

Monitoring may also be associated with limit conditions [Moray, 1993]. As a signal's state approaches a critical or cautionary level, it will occasion more frequent eye fixations. For example, when crossing the road, an agent will more frequently glance at the light or crossing signal if it is yellow rather than green.

Obstacles, or objects in an agent's path, may be considered limit signals in our system. Such objects will not occasion eye fixations until the agent is in very close proximity.

LimitNets are associated with object sites and a boolean function that indicates when a limit condition has been reached. If such a condition becomes true, the site(s) in the net (i.e. the center of the red light panel in the road crossing example) are added to **IntentionList**.

4.2.2 Reaching and Grasp

Traditional experiments indicate that eye movements precede hand movements and since eye saccades are extremely fast [Abrams, Meyer, and Kornblum, 1990], the eye arrives before the hand motion is started. However, as the target size increases beyond extremely small (0.5 degree visual angle), eye and hand movement are initiated almost concurrently [Bekkering, Adam, van den Aarssen, Kingman, and Whiting, 1995].

When initiating a reach and grasp motion, we generate eye movement toward the relevant grasp site by adding it to **IntentionList**. We continue to add the grasp site to **IntentionList** at periodic intervals as long as the reach motion is active. If an agent is picking up a cup, we look at the cup handle. If the agent is lifting a box, we generate a sequence of eye motions to the box grips [Ballard, Hayhoe, Li, and Whitehead, 1992]. Clearly, the eye is supposed to establish targeting for the hand [Abrams, Meyer, and Kornblum, 1990]. Also, note that if the two eyes are independently positionable, this 3D target will properly set the convergence angle for each eye (up to physical rotation limits). We animate the reach motion by calling our human model's inverse kinematics routines. The alignment constraint on the eye is established through the algorithm previously described in Figure A.1.

When the hand is in close proximity to the grasp site, we begin eye movement to the next fixation site (either as a result of the current reach or due to the next motor action on the queue). We associate with the motor activity a confidence or repeat factor. If the agent has done the same reach before or the action is *a priori* marked with a high confidence factor, abandonment of the current eye behavior happens earlier than it would otherwise.

4.2.3 Motion in the Periphery

The human eye perceives a moving object if it appears in the visual field for more than 0.15 seconds and travels at a speed greater than 1 minute/second [Yarbus, 1967]. The eyes are directed toward the moving target by saccadic eye movement. Once the eye fixates on a moving object, however, it follows a pattern of smooth pursuit [Kahneman, 1973].

We model visual pursuit by fixating and subsequently tracking any moving target that enters an agent's peripheral field of view. A motion sensor net continuously checks the environment for objects that change position between frames. Such objects are added to **PList** if they fall within an agent's periphery.

4.2.4 Visual Search

We model visual search by first determining the angle between the center of fixation and the target. We generate a sequence of intermediate positions that move the eye from its current position to the target location. Each position is placed, in order, on **IntentionList**. This eye behavior corresponds to experiments and a computational model proposed in [Rao, Zelinsky, Hayhoe, and Ballard, 1996]. Visual processing proceeds in a low to high accuracy manner. When asked to locate a specific object in a scene, subjects in [Rao, Zelinsky, Hayhoe, and Ballard, 1996] performed a *series* of (progressively more accurate) eye saccades toward the object rather than immediately fixating it.

4.2.5 Spontaneous Looking

In the absence of a specific goal or task, attention follows patterns of spontaneous looking [Kahneman, 1973]. Attention is drawn to items in the environment that are likely to be informative or significant. Psychologists argue this is due to a need to reduce uncertainty about our surroundings.

Novel or complex items are considered significant. Novelty may be measured by motion, color, isolation, or complexity of shape. Image processing approaches in [Tsotsos, Culhane, Wai, Lai, and Nufflo, 1995; Koch and Ullman, 1985] look for areas in the field of view that are locally conspicuous. Luminance is considered salient in [Tsotsos, Culhane, Wai, Lai, and Nufflo, 1995] while color and orientation of edges are the measure of conspicuousness in [Koch and Ullman, 1985].

Since we wish to generate real-time eye behavior, we use a simplified novelty measure. The system copies a 1000x400 snapshot of the agent's field of view into a pixel buffer. We select those pixels whose color values are the furthest from their neighbors in RGB space. In future work, we will consider pixel color values in the CIE-LUV* color space. The human eye is more sensitive to certain frequencies or colors of light than others. We will consider color distance in the CIE-LUV* space since it is more perceptually uniform than traditional RGB space. We convert the most conspicuous pixel locations into their corresponding pan and tilt parameters in eye and head coordinates. Figure 4.4 draws rays between the agent's eyes and those pixels of the image with the highest color contrast (the location that each ray first intersects a figure in our environment indicates a site that will be looked at).

4.2.6 Interleaving and Confidence Levels

The interleaving of an agent's attention will happen as a natural consequence of competing behaviors in our system. In contrast, given a set of sequential motor activities, our system must determine when to abandon the current eye behavior and initiate a subsequent one. A boolean variable is maintained in each net that implements eye behavior based on a reach or locomotion. This variable indicates an expectation that the current activity will complete successfully. Normally, such a variable is set when the hand is in close proximity to the relevant grasp site or the agent is close to his destination. If an agent is confident or expert, however, this variable will be set earlier in the execution of the reach motion reflecting greater confidence in the agent's skill. Setting this boolean variable thus allows attention to be directed to the next activity while the motor system completes the motor task. Notice that if this variable is set at the beginning of the task, the interpretation is consistent with human behavior: it means the agent knows where to reach or walk even without looking at the object or goal.

4.2.7 Task Queue Manager and Examples

A task queue net consumes action requests posed for the agent and invokes the appropriate eye behaviors. Figures 4.5(a)-(e) illustrate the structure and content of sample requests.

If 4.5(a) were entered on the queue, the task manager would spawn a monitoring locomotion eye behavior, and initiate the walk motor activity for our human model. The monitoring behavior in our system associates an uncertainty threshold of 100 with the goal and 200 with the ground (implying that every 100 frames the agent should glance at the destination and every 200 frames, he should glance at the ground). Since toy objects are to be avoided during the activity, a **LimitNet** is spawned with a limit function that returns true whenever the agent walks in close proximity to a toy (i.e. such objects will be attended to only if they are too close). The distance limit in our implementation is 1 meter.

If 4.5(c) were entered on the queue, the task manager would spawn a reach eye behavior (with the relevant sites on the box passed as arguments) and invoke our human model's reach and grasp mechanism. The reach eye behavior would indicate, by polling end effector position, when the motor activity was close to completion. This value, when set to true, would allow the task manager to initiate eye behavior for the next task on the queue.

If 4.5(d) were entered on the queue, a visual search behavior would be spawned. This behavior would generate a sequence of eye movements to the target. These intermediate positions would be added to **IntentionList**.

Figure 4.6 illustrates processing performed by the task queue manager.

4.3 Worked Example

Consider a scenario where an agent is asked to walk to a destination: in order to reach the destination, he must cross a road, watch out for oncoming traffic and monitor the appropriate traffic signal. We ask our system to handle such a scenario with the input illustrated in Figure 4.7.

A task queue manager net for agent “Stanley” will consume these actions requests. A walking eye behavior net will be spawned that periodically adds sites to **IntentionList** (the sites will be the destination and, infrequently, the ground in front of Stanley's feet). Also, the walking motor activity will be spawned (the corresponding eye behavior will remain active as long as the motor activity is not complete).

A monitoring eye behavior will be spawned that periodically adds the traffic light as a figure to be monitored on **IntentionList**. If the light turns yellow, the frequency with which the monitoring behavior adds the traffic light to **IntentionList** will increase. The monitoring behavior will only remain active while the agent is crossing the street.

A monitoring eye behavior will also be spawned to check for oncoming traffic on the right side of the road. If a car approaches within a certain distance of Stanley (5 meters), the frequency of monitoring will increase. This behavior will also remain active until Stanley crosses the street.

Behaviors modeling exogenous factors (involuntary attention capture by task unrelated events) will be a a peripheral motion sensor and spontaneous looking. A peripheral motion sensor will sample for moving objects in agent Stanley's field of view. If a moving object is detected, it will be added to **PList**.

Whenever both **IntentionList** and **PList** are empty, spontaneous looking behavior will be activated for Stanley. This behavior will determine the pixel in Stanley's field of view with the greatest local feature contrast, convert the pixel location back into the corresponding 3D environment coordinates and cause Stanley to glance at the appropriate target.

4.3.1 Details of Simulation

Figures 4.8 – 4.15 are some snapshots from the animated version of this scenario. While the complete animation is approximately 360 frames, we examine some representative frames that illustrate how **IntentionList** (the queue of task related figures or sites that need to be attended) and **PList** (the queue of moving objects in the agent's periphery) are modified and adapt over the course of the simulation. Also, we indicate when spontaneous looking behavior becomes active. A line is drawn indicating Stanley's line of sight when looking at task related or exogenous targets.

By frame 6, a monitoring eye behavior has placed the site "traffic_light.yellow" on **IntentionList** (indicating that the agent should look at the traffic light). Also, Stanley's peripheral motion sensor has noticed that a car object is moving and placed it on **PList**. Figure 4.8(a) shows the close up view of Stanley looking at the traffic light and Figure 4.8(b) shows a small window into Stanley's field of view.

By frame 17, a monitoring eye behavior has placed the site "road.right" on **IntentionList** indicating that Stanley needs to look at the right side of the road (to ascertain if other cars are coming). Also, the walking eye behavior has added the the figure "table" to **IntentionList** indicating that Stanley needs to glance at his destination. Figure 4.9 shows Stanley looking right and Figure 4.10 shows Stanley subsequently tracking the car.

By frame 75 (figure 4.11), Stanley looks back at his destination (the table). At frame 96, the traffic light monitoring behavior places a site on **IntentionList** again (figure 4.12) (indicating that Stanley needs to look at the light to ascertain its color).

By frame 127, the road monitoring behavior has added a site to **IntentionList** indicating that Stanley needs to look toward the road again (figure 4.13).

By frame 145, both **IntentionList** and **PList** are empty, so Stanley lapses into spontaneous looking behavior(Figure 4.14).

By frame 211, a moving ball has arrived in Stanley's periphery (his peripheral motion sensor has placed the ball on **Plist**). Figures 4.15(a) and 4.15(b) illustrate Stanley tracking the ball.

4.3.2 Analysis of Simulation

There are several important points that should be noted from the animation generated from our scenario.

First, certain predefined data is associated with the tasks entered into the system (the monitoring frequencies, or memory uncertainty thresholds, associated with walking activity and with watching the traffic light and oncoming traffic). However, the *AVA* generates eye movements that consider the combination of simultaneously executing tasks (and hence produce *timings* of eye movements that differ from the standard uncertainty thresholds). The state of the **IntentionList** at frame 17, for example, shows two sites that need to be attended and a car object on **PList** that needs to be tracked. Essentially, the *AVA* is generating behavior as a result of increasing cognitive *load*.

Second, the *AVA* generates eye movements that are the result of changes in the environment and not explicitly the result of a task on the task queue. Whenever tasks demands do not require attention, for example, the agent lapses into idling behavior. Also, when a ball flies into the agent's field of view (see frames 211 and 226), the agent tracks it in the absence of other task demands. Also, when the traffic light turns yellow, the frequency of monitoring increases so that the agent will glance at the light more often than previously.

The plausibility of our method is determined not so much by how it accurately reproduces the empirical data on which it is based but rather in how it *adapts* and in how it *fails*. If too many deliberate tasks had vied for the agent's attention in our simulation, he would have ignored a potentially critical event (a car bearing down on him!) and been run over. If several moving objects appeared in his field of view as he was walking to the table, one of them may not have been observed and he could have bumped into it (just as would happen during a real life walk in the park).

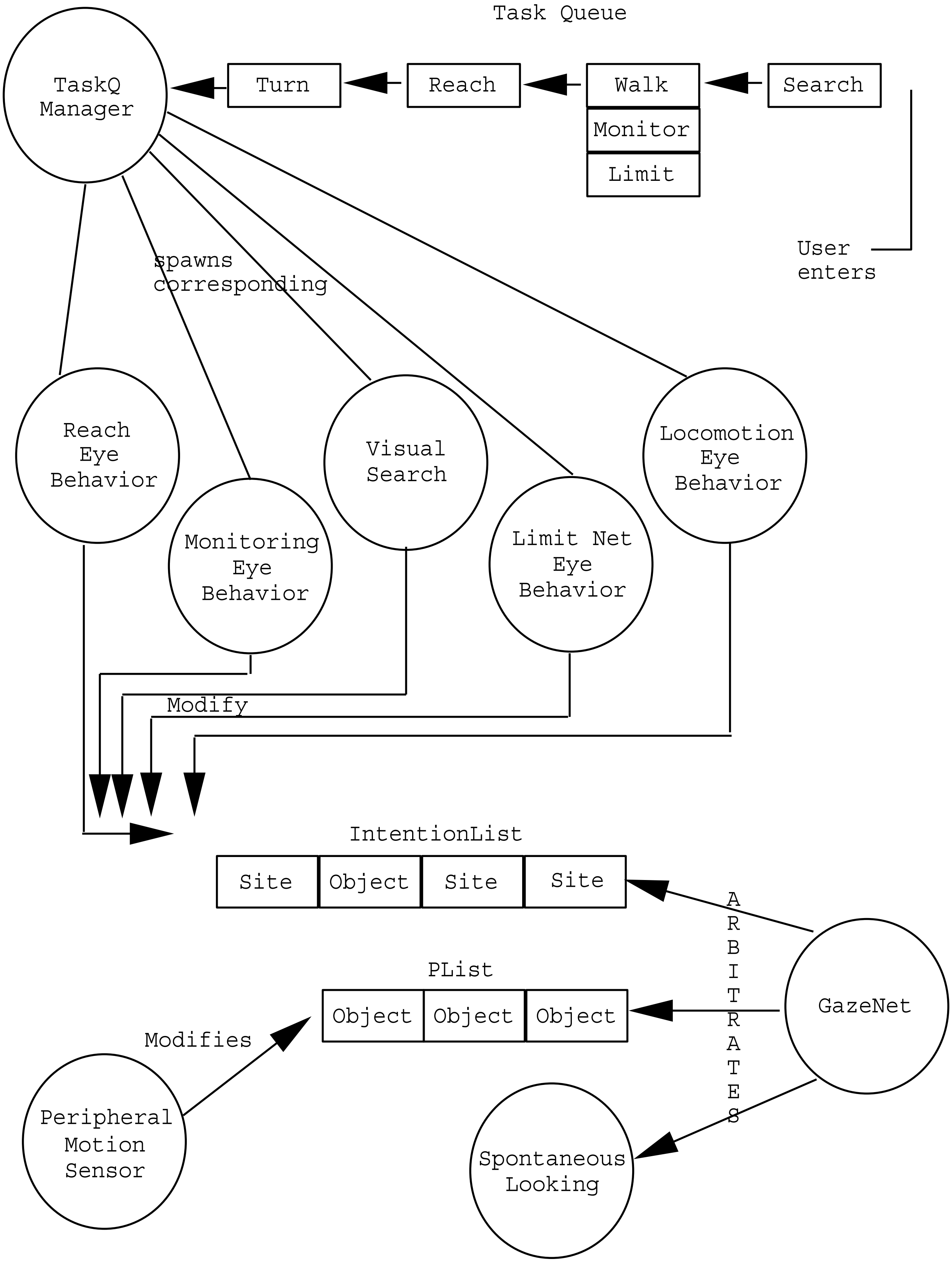


Figure 4.1: Method Architecture

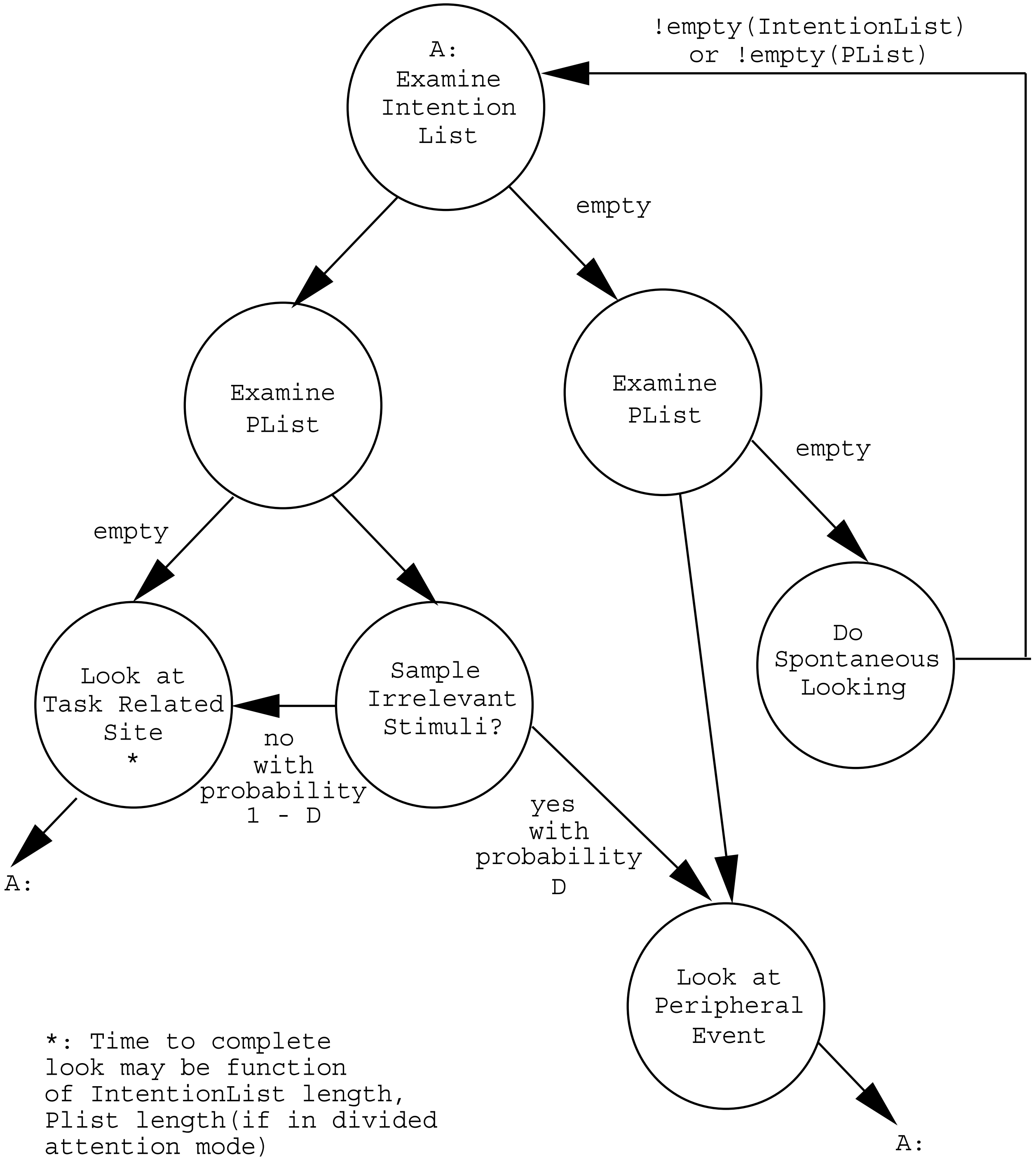


Figure 4.2: GazeNet

```
While not empty(IntentionList)
    Object ⇐ Removehead(IntentionList)
    If not empty(PList)
        look at head of PList with probability D
    else
        If the Object is a figure, look at its
        center of mass.
        If the object is a site,
        look at its location.

If not empty(PList)
    Object ⇐ Removehead(PList)
    Look at Object's center of mass.

While empty(PList) and empty(IntentionList)
    Do spontaneous looking. Take snapshot of field
    of view. Determine locally conspicuous pixels.
    Aim head and eyes at most conspicuous locations
    in succession.
```

Figure 4.3: GazeNet Algorithm

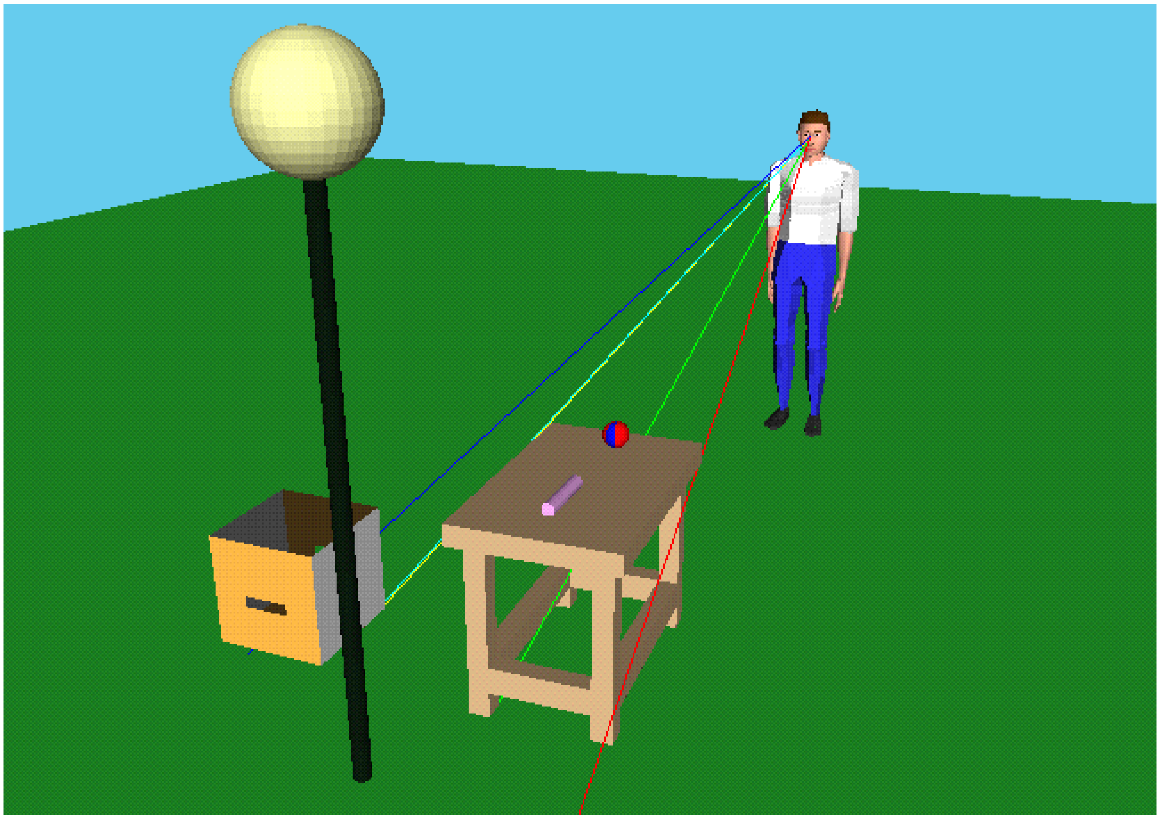


Figure 4.4: Spontaneous Looking - Rays Intersect Features with Local Contrast

(a)

Agent:	bill
Action:	walkto
Object:	
Goal:	lamp post
Sites:	
Avoid Figures:	
Limit Condition:	

(b)

Agent:	bill
Action:	monitor
Object:	traffic light
Goal:	
Sites:	red lamp, yellow lamp, green lamp
Avoid Figures:	
Limit Condition:	function pointer: returns true when traffic light is yellow

(c)

Agent:	bill
Action:	reach
Object:	box1
Goal:	
Sites:	box1's left handle, box1's right handle
Avoid Figures:	
Limit Condition:	

(d)

Agent:	monica
Action:	search
Object:	ice cream truck
Goal:	
Sites:	
Avoid Figures:	
Limit Condition:	

(e)

Agent:	bill
Action:	reach
Object:	box1
Goal:	table
Sites:	table center
Avoid Figures:	
Limit Condition:	

Figure 4.5: Sample Action Requests Processed by Task Queue Manager

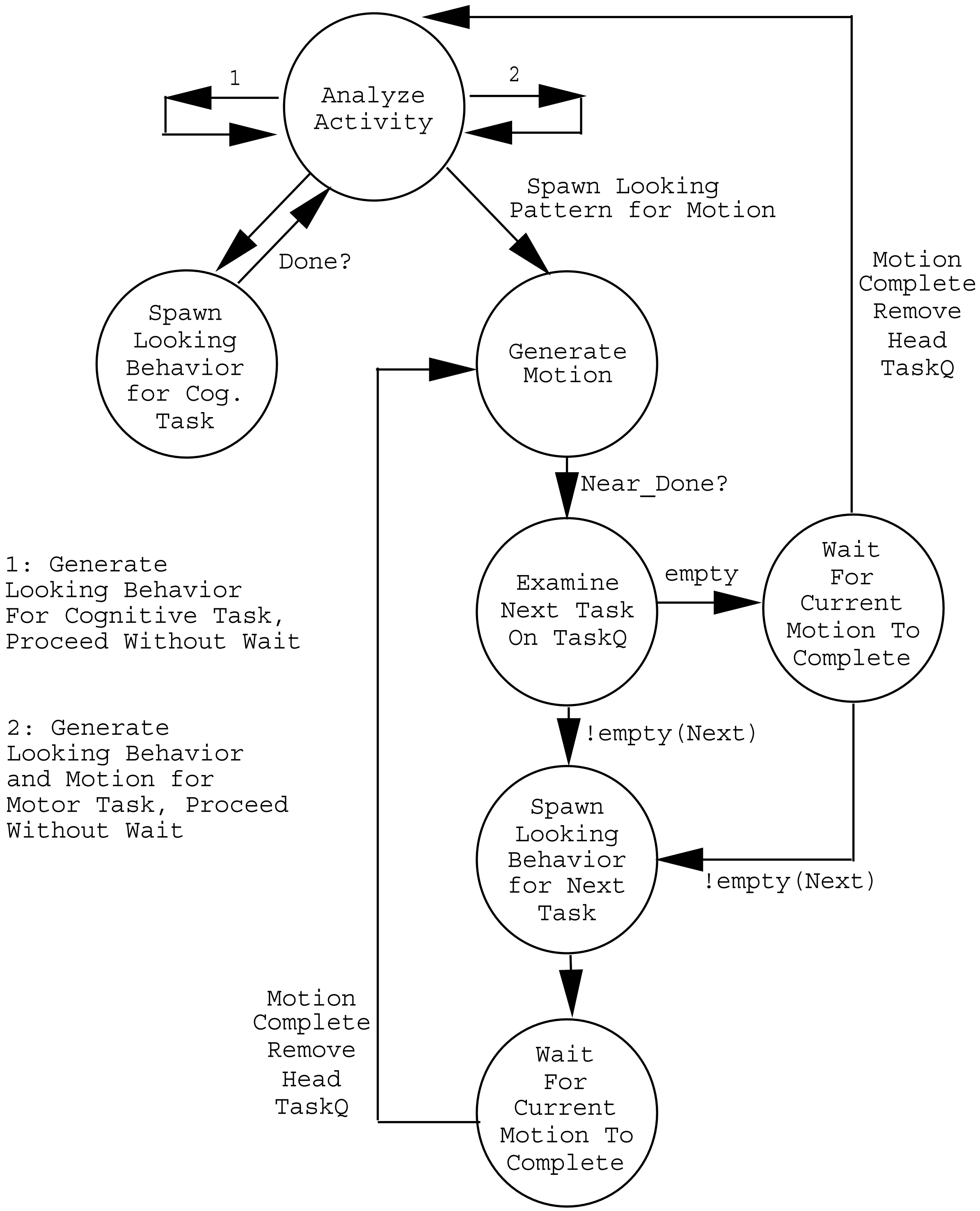

Analyze Activity
1
2
Spawn Looking Pattern for Motion
Done?
Spawn Looking Behavior for Cog. Task
Generate Motion
Motion Complete Remove Head TaskQ
Near_Done?
1: Generate Looking Behavior For Cognitive Task, Proceed Without Wait
Examine Next Task On TaskQ
empty
Wait For Current Motion To Complete
!empty(Next)
2: Generate Looking Behavior and Motion for Motor Task, Proceed Without Wait
Spawn Looking Behavior for Next Task
!empty(Next)
Motion Complete Remove Head TaskQ
Wait For Current Motion To Complete

Figure 4.6: Task Queue Manager

(a)

Agent:	stanley
Action:	walk
Object:	
Goal:	table
Sites:	
Avoid Figures:	
Limit Condition:	

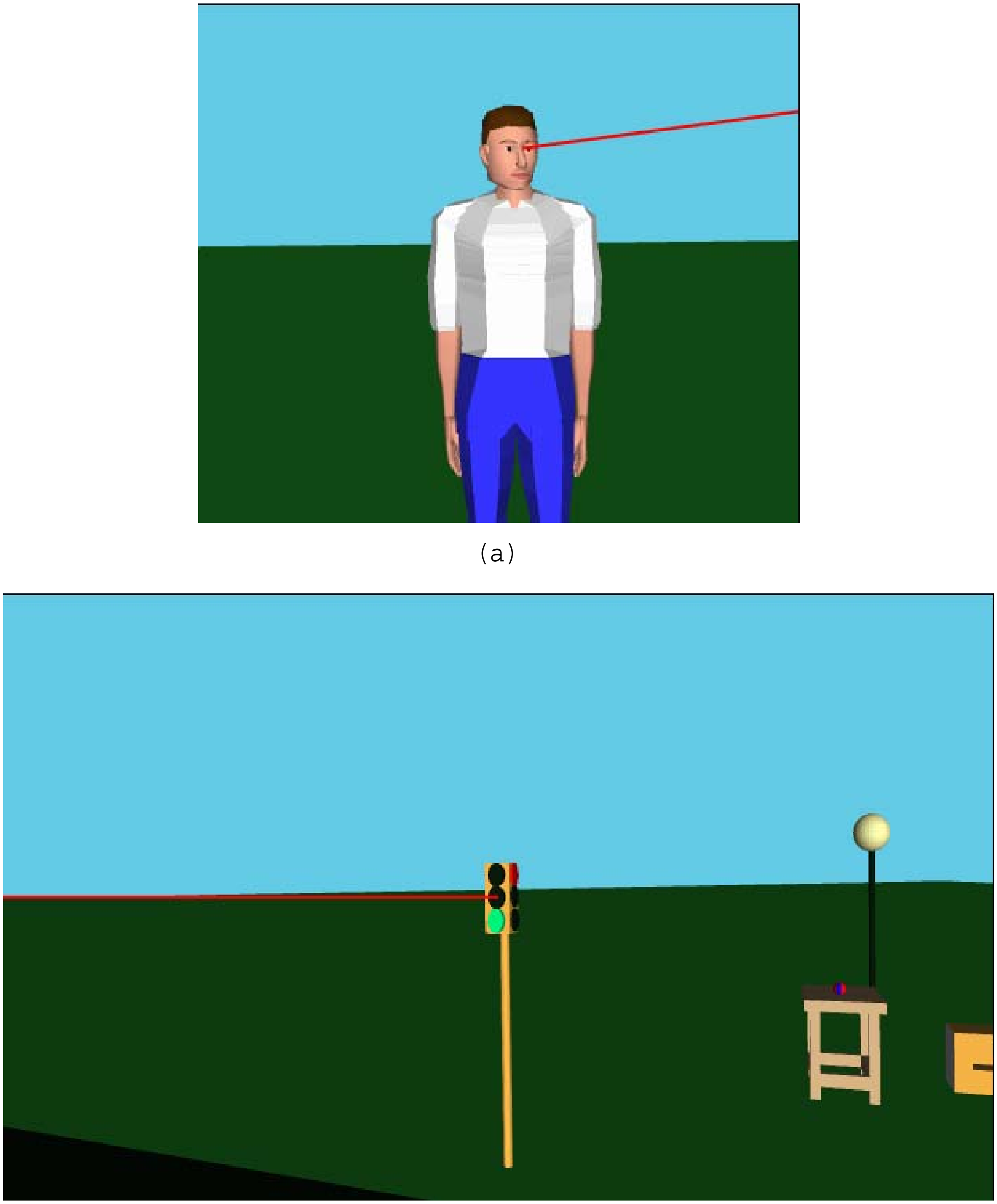
(b)

Agent:	stanley
Action:	monitor
Object:	traffic light
Goal:	
Sites:	yellow lamp
Avoid Figures:	
Limit Condition:	function pointer: returns true when traffic light is yellow
Duration Condition:	function pointer: returns true while agent is crossing road

(c)

Agent:	stanley
Action:	monitor
Object:	road
Goal:	
Sites:	right side of road(direction from which cars appear)
Avoid Figures:	
Limit Condition:	function pointer: returns true when car is in close proximity
Duration Condition:	function pointer: returns true while agent is crossing road

Figure 4.7: Example Action Requests on Taskq

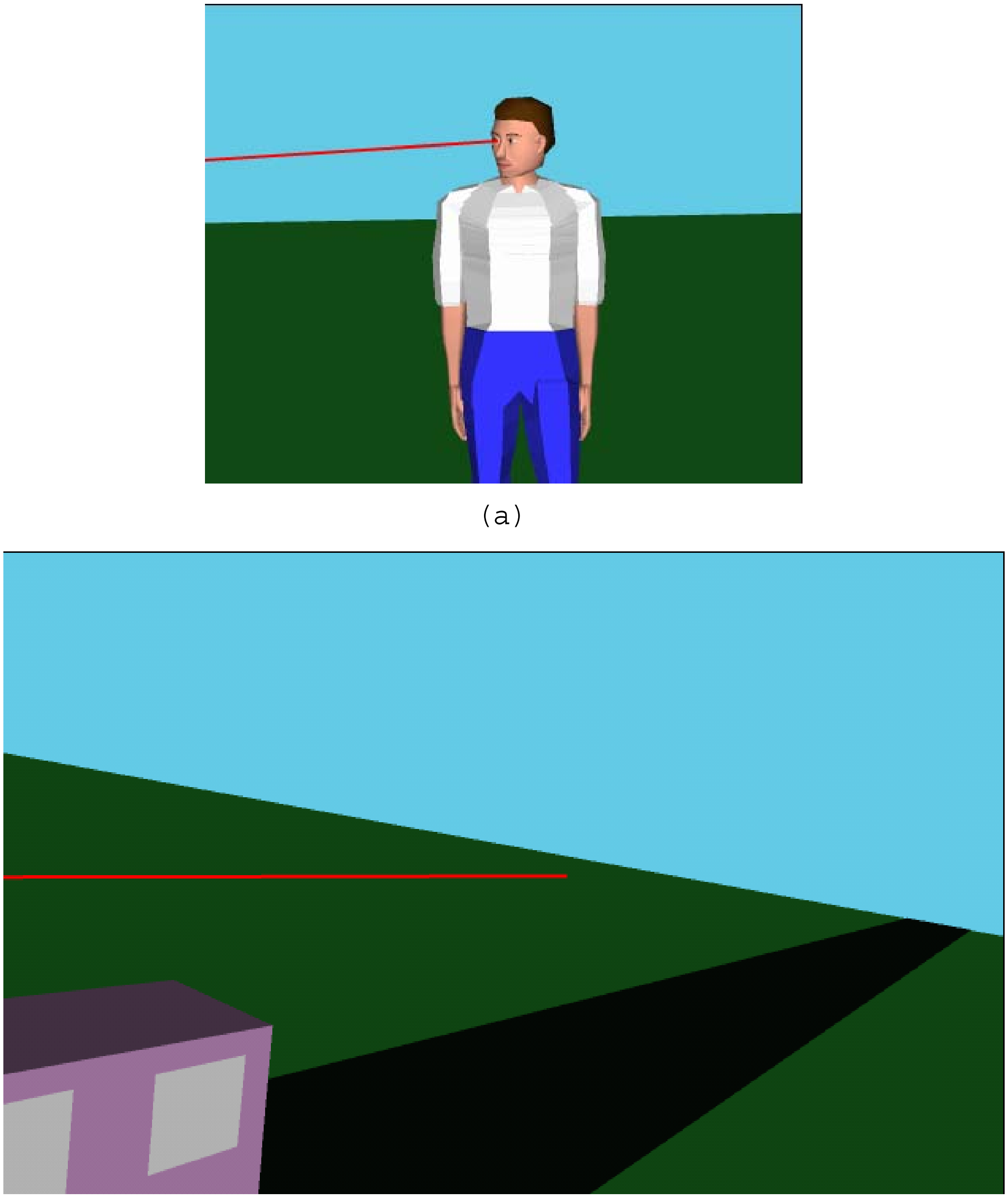


(a)

(b)

```
IntentionList: traffic_light.yellow
Plist:   car
Spontaneous Looking Active?: No
```

Figure 4.8: Frame 6 - Stanley monitors the traffic light

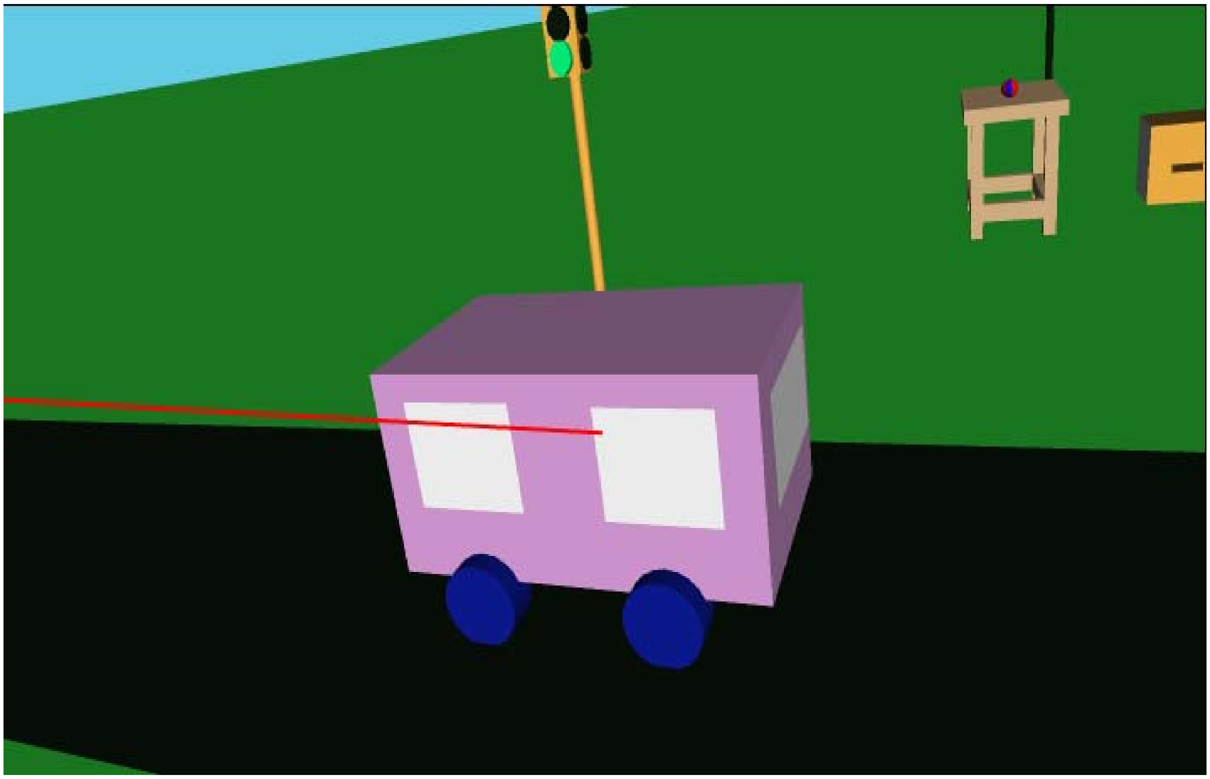


(a)

(b)

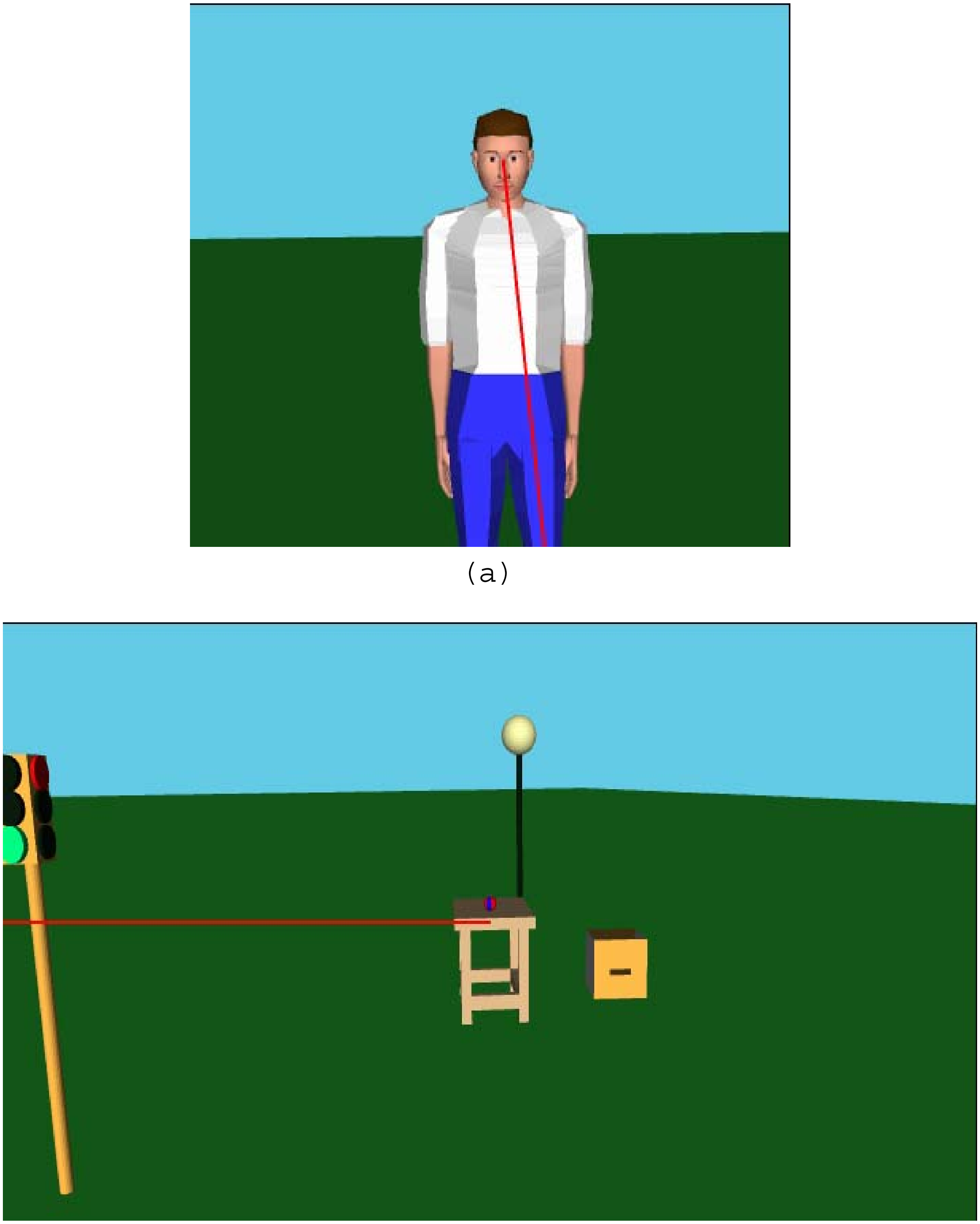
```
IntentionList: road.right, table
Plist: car
Spontaneous Looking Active?: No
```

Figure 4.9: Frame 17 – Stanley glances right



```
IntentionList: table
Plist: car
Spontaneous Looking Active?: No
```

Figure 4.10: Frame 38 – Stanley tracks the car

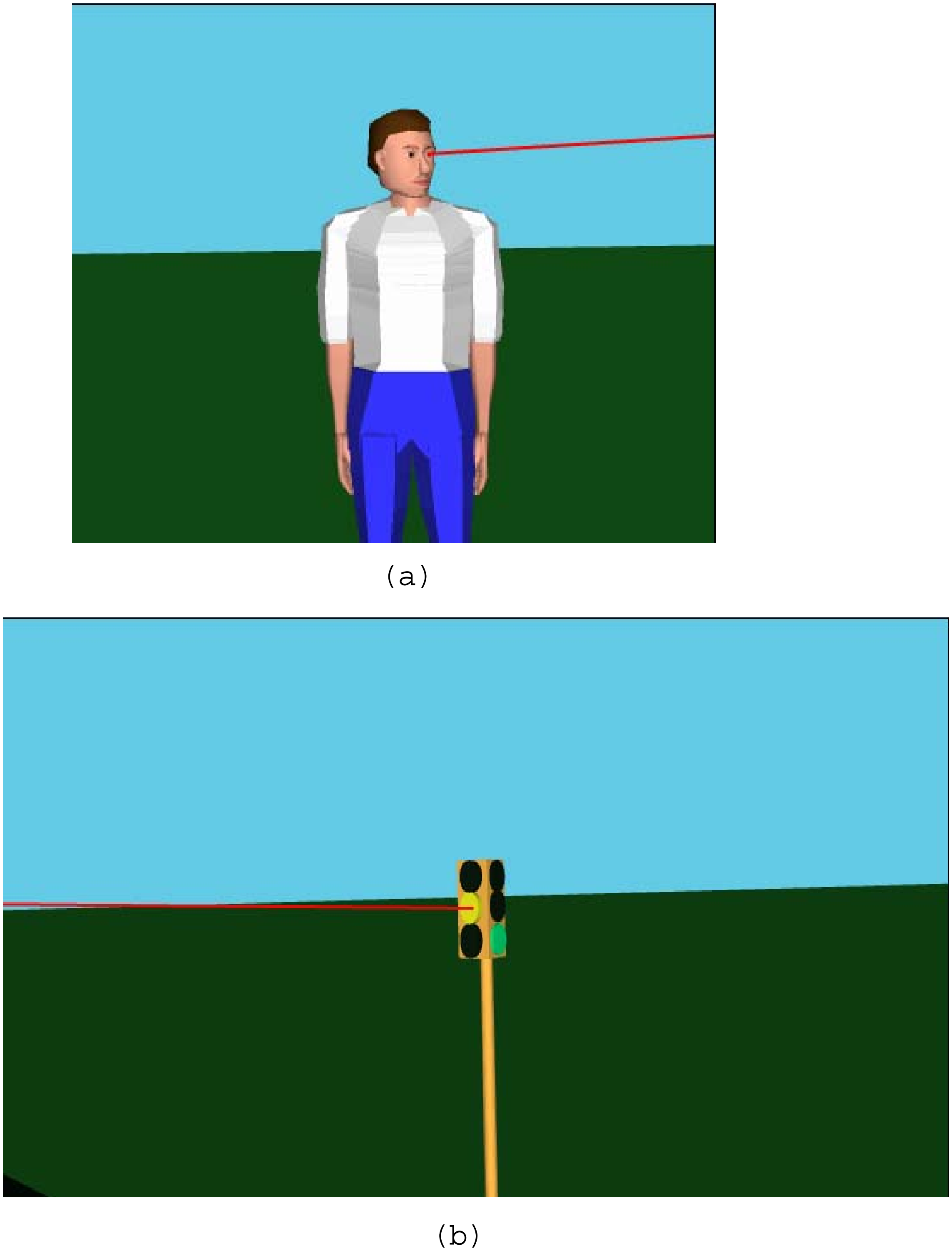


(a)

(b)

```
IntentionList: table
Plist:
Spontaneous Looking Active?:  No
```

Figure 4.11: Frame 75 – Stanley glances at his destination

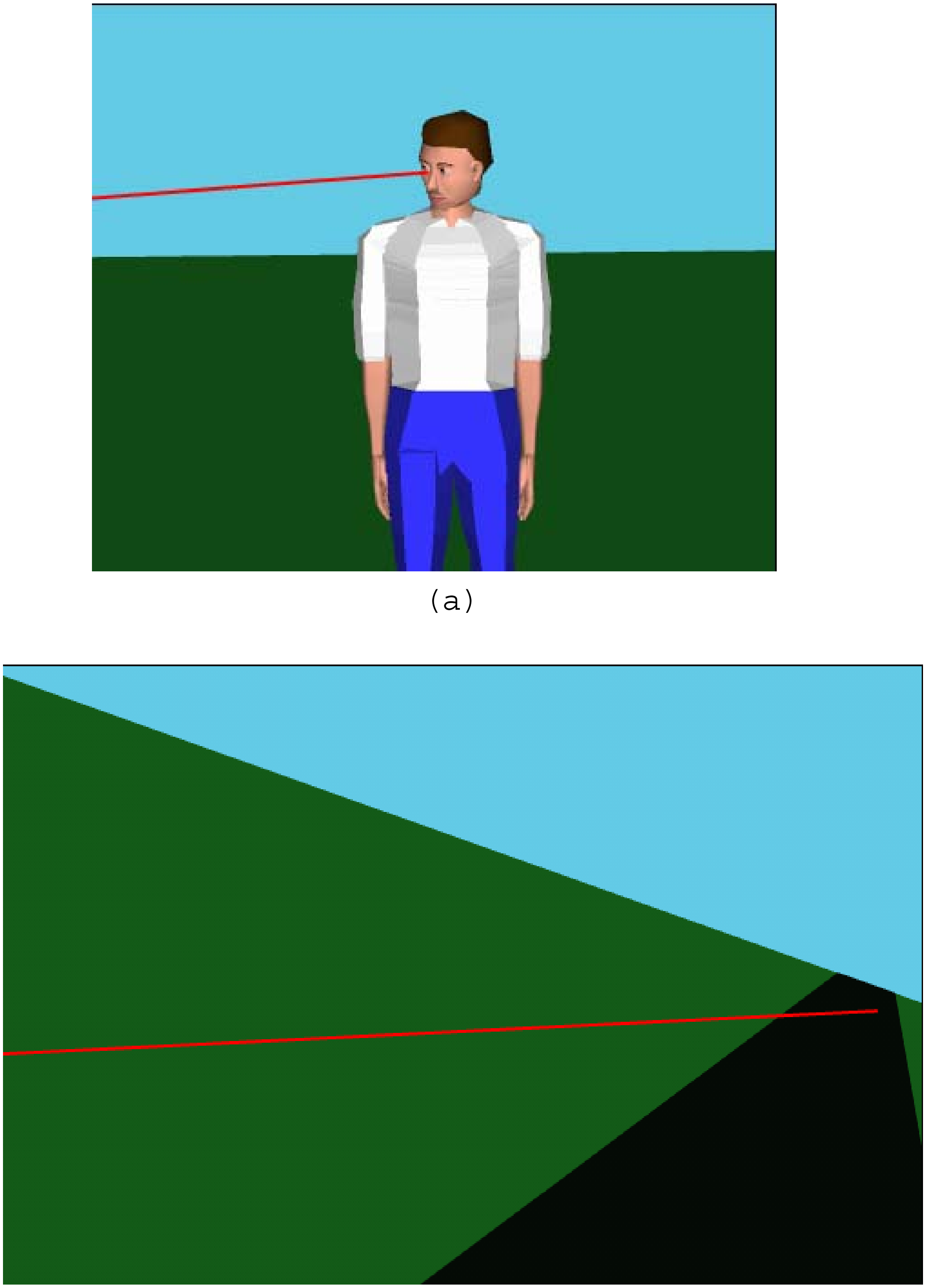


(a)

(b)

```
IntentionList: traffic_light.yellow
Plist:
Spontaneous Looking Active?: No
```

Figure 4.12: Frame 96 – Stanley glances back at traffic light

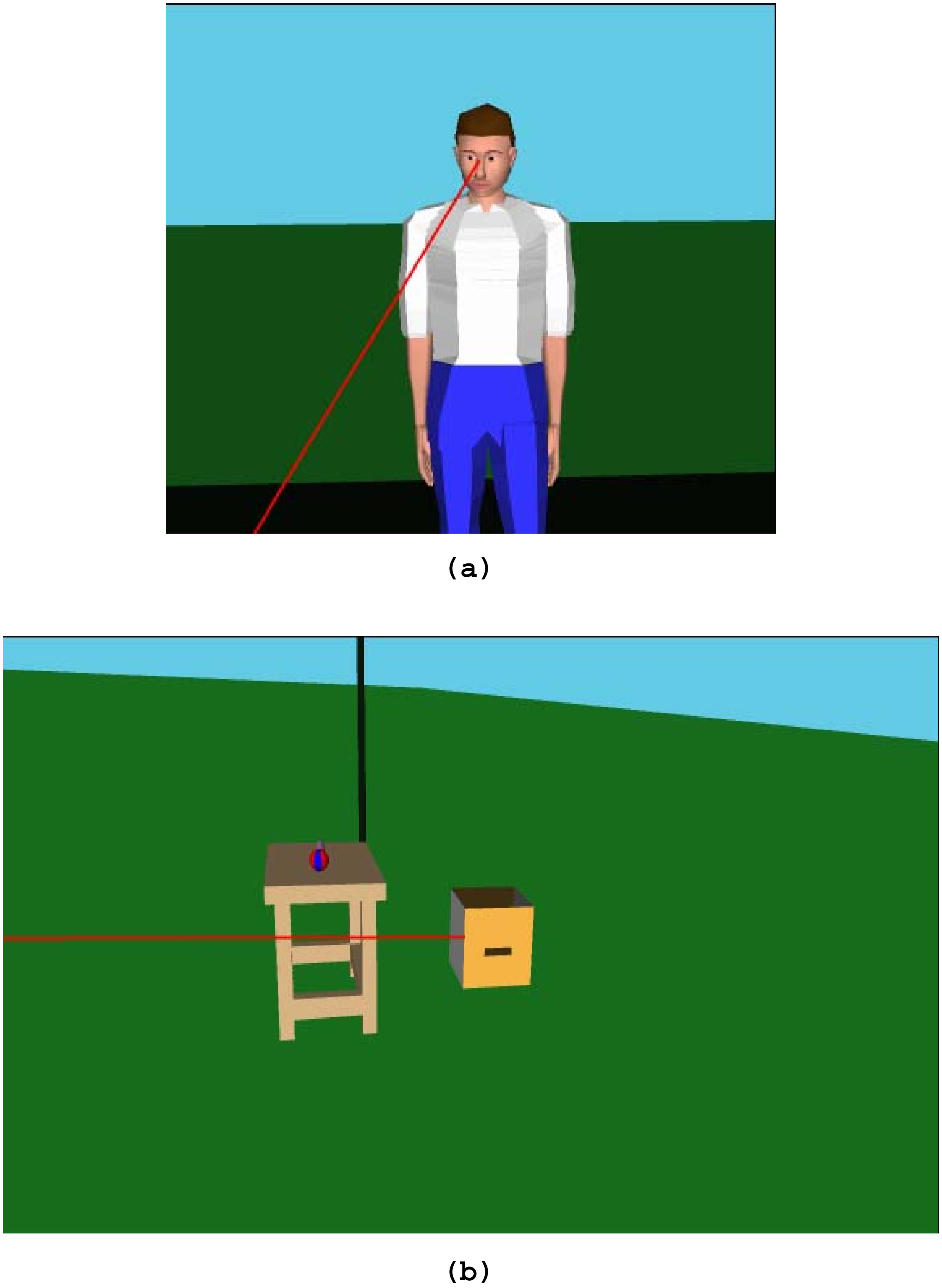


(a)

(b)

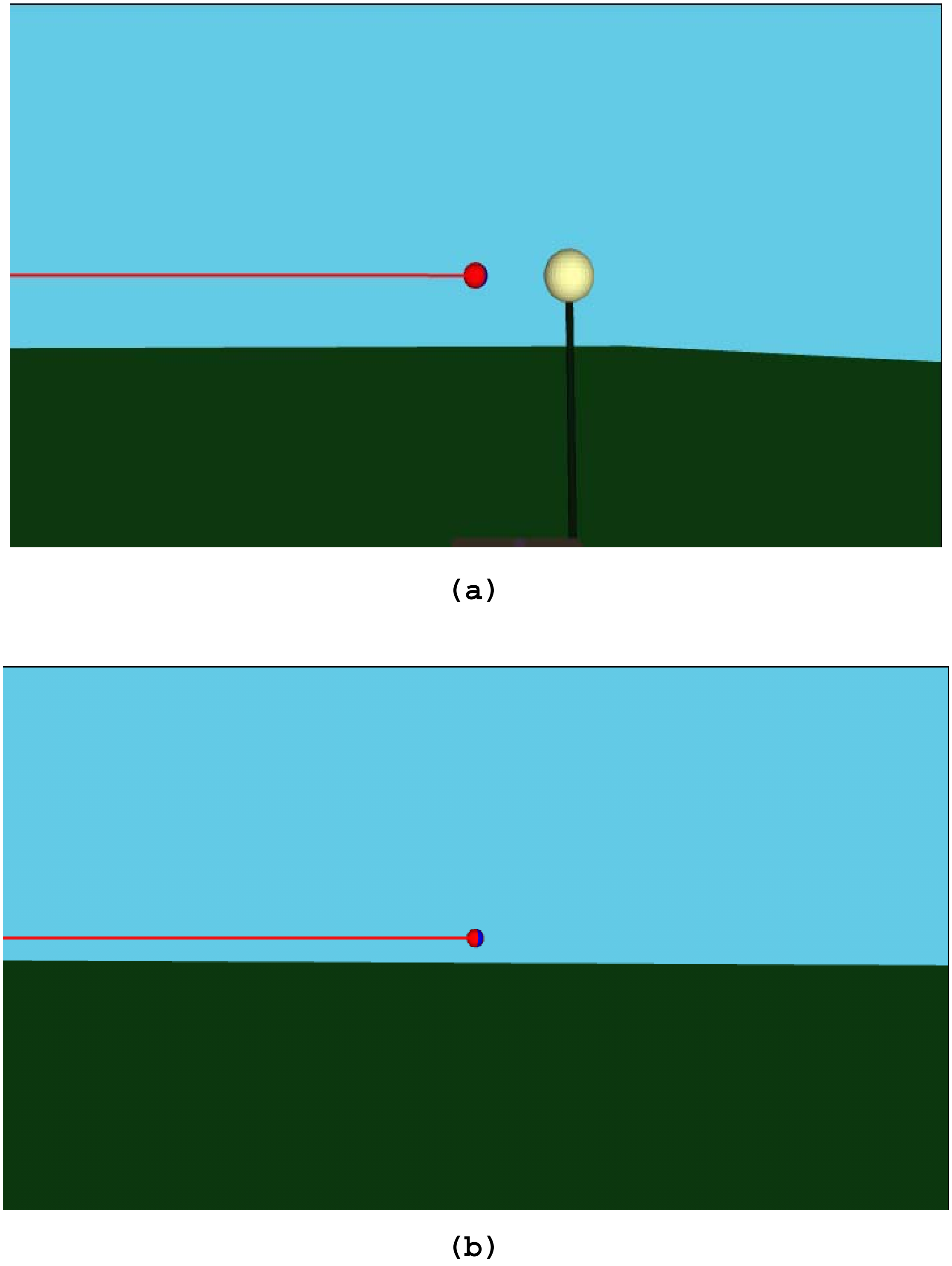
```
IntentionList: road.right
Plist:
Spontaneous Looking Active?: No
```

Figure 4.13: Frame 127 – Stanley glances back at road



```
IntentionList:
Plist:
Spontaneous Looking Active?: Yes
```

Figure 4.14: Frame 145 – Stanley spontaneously looking



(a)

(b)

```
IntentionList:
Plist: ball
Spontaneous Looking Active?: No
```

Figure 4.15: Frame 211-226 – Stanley tracking a moving ball

Chapter 5

Summary and Outline of Research Plan

We have proposed a computational framework for generating attending behavior, referred to as the *AVA*, using emprical and qualitative observations from the psychology, human factors and computer vision literature. This method drives simulated attention in novel, unscripted, and resource-bounded ways. The goal of this research is to generate more natural looking animated human characters by directing line of sight appropriately.

Our work associates a pattern and frequency of eye movements with cognitive (visual search and monitoring) and motor (reach, walk, lift, pull, put down) activities. Such behaviors are *reactive* and may be modified by external changes in the virtual environment and situation. Further, the interactions between voluntary, task-related eye movements and attention capture by exogenous factors (peripheral events and feature singletons) as well as lapses into idling behavior are modeled in our technique.

Psychology experiments reveal that tasks impose a voluntary pattern of eye movements [Yarbus, 1967; Moray, 1993]. As several tasks are simultaneously attempted, performance degrades versus the single task condition [Hirst, 1986]. Peripheral events capture attention when the agent is engaged in a task which requires diffuse attentiveness (e.g., visual search or divided attention) [Yantis and Jonides, 1990; Egeth and Yantis, 1997]. When a task target can be isolated by a a feature such as shape or color, then other *feature singletons* (objects which differ from their surroundings by unique color, shape, orientation or motion) act as *distractors* and increase response time to a target [Folk, Remington, and Wright, 1994; Hillstrom and Yantis, 1994]. In the absence of tasks or peripheral stimuli, attention follows patterns of spontaneous looking [Kahneman, 1973].

Complex tasks may be modeled as combinations or instances of modules we provide. Driving, for example, can be considered a monitoring task where several signals must be simultaneously attended (other cars, traffic lights, pedestrians).

The validity of our method may be judged by how behavior generated *adapts* in situations of increasing cognitive load or as a consequence of a changing environment. Various experiments in progress have evolved the design of our proposed technique. We intend to fully implement the *AVA* in a variety of scenarios combining reach, grasp, locomotion, and interference from exogenous effects. Also, we will add a *feedback* mechanism in future work where human motion may be modified as a consequence of increasing cognitive load.

The following is a list of prioritized items that we plan to address in future work:

- High Priority
  - Tag object-specific features in the environment for particular attending purposes. For example, the headlights and windshield are relevant sites when monitoring cars. Characters' eyes may be relevant sites when glancing at other agents.
  - Add motion prediction to the motion sensor behavior (e.g., attend to those objects that deviate from their predicted path or that are in danger of collision).
  - Add a motion tracking behavior that allows an agent to visually pursue ob-

jects, using predicted motion, even when those objects may be temporarily occluded.

  - Regulate motion under situations of heavy attentional load (e.g., If events begin to lengthen the attention queue, our agent should slow down to accommodate the required attention and subsequent decision-making).

- Medium Priority

  - If the target of a task can be distinguished by isolation of color, shape or orientation, a strategy of singleton detection may be profitable (e.g., if the target has a *unique* shape or color in the environment). In such situations, even though the task itself may involve searching for a given shape, objects which differ from their surroundings by color, orientation and motion act as *distractors* and interfere with performance. When generating eye behaviors for such tasks, develop sensors that determine the presence of motion or color singletons. Incorporate the presence of these singletons when determining response time to task targets (e.g., if the task requires finding a moving target, multiple moving objects in the agent's field of view will act as distractors and increase response time to the specified target).

- If Time Allows:

  - Expand the mechanism which distributes motion between head and eyes. Allow for feedback and correction [Freedman and Sparks, 1997].

5.1 Conclusion

Attention is a process that utilizes a allocatable, steerable resource (the eyes) and as such requires a control algorithm and a time budget for movement and sensing. Competing behaviors require prioritizing and arbitration. Visual perception is a significant component of the human behavior repertoire. Through our methodology we have shown that automatic attention control is both feasible and useful for animated human-like characters.

Appendix A

Head Eye Coordination Algorithm

This algorithm is executed every *frame* while eye tracking for a given target is active. Hence, head and eyes can be continuously aligned with a moving target. The pan and tilt displacement needed to align the head with a target is calculated independently of the displacements necessary to align the eyes. Horizontal gaze shifts greater than 20 degrees or vertical shifts greater than 10 degrees produce combined head and eye movement. In the combined motion scenario, the eyes move as far as oculomotor limits (joint limits set in our human model) allow. Once the head is aligned with the target, correct realignment of the eyes to the target occurs in the next frame.

In the human visual system, the vestibulo-ocular reflex (VOR) stabilizes images on the retina by generating compensatory eye motion of equal magnitude and in the opposite direction to head motion. For large gaze shifts, VOR activity is turned off until line of sight is aligned with the target. Then, VOR resumes and eyes may counter rotate to compensate for head motion. [Sparks, 1989]

The mechanism of eye head coordination for our human model was implemented by [Xhao and Achorn, 1994] and is given in Figure A.1.

```
For each eye,
    determine view vector between center of eyeball and target
    transform this vector into the eyeball coordinate frame
    solve for:
            $\theta_{left\ or\ right,v} = vertical\ displacement$
            and
            $\theta_{left\ or\ right,h} = horizontal\ displacement$
    needed to move eye coordinate frame to the target

Similarly, for the head,
    determine view vector between base of head and target
    transform this vector into the base of head coordinate frame
    solve for:
            $\beta_v = vertical\ displacement$
            and
            $\beta_h = horizontal\ displacement$
    needed to move the head coordinate frame to the target.

If ($\theta_{left,v} > 10\ degrees$) or ($\theta_{right,v} > 10\ degrees$) or
   ($\theta_{left,h} > 20\ degrees$) or ($\theta_{right,h} > 20\ degrees$) then
            $\theta_{left,v} = \theta_{left,v} + \beta_v$ and
            $\theta_{right,v} = \theta_{right,v} + \beta_v$ and
            $\theta_{left,h} = \theta_{left,h} + \beta_h$ and
            $\theta_{right,h} = \theta_{right,h} + \beta_h$
            threshold $\theta_{i,j}$ so that displacement is within
            joint limits
Else
            $\beta_v = 0$ and $\beta_h = 0$
            (only the eyes move for small shifts)

Let:
    $new\ head\ position =$
    $(current\ head\ pan + \beta h, current\ head\ tilt\ +\ \beta v)$
    $left\ eye\ position =$
    $(current\ l.\ eye\ pan + \theta_{left,h}, current\ l.\ eye\ tilt + \theta_{left,v})$
    $right\ eye\ position =$
    $(current\ r.\ eye\ pan + \theta_{right,h}, current\ r.\ eye\ tilt + \theta_{right,v})$
```

Figure A.1: Head-Eye Coordination Algorithm

Bibliography

R.A. Abrams, D.E.M. Meyer, and S. Kornblum. Eye-hand coordination: Oculomotor control in rapid aimed limb movements. *Journal of Experimental Psychology: Human Perception and Performance*, 16:248–267, 1990.

A. Allport. Attention and control: Have we been asking the wrong questions? a critical review of 25 years. *Attention and Performance*, 14:183–218, 1993.

A. Allport, E. Styles, and S. Hsieh. Shifting intentional set: Exploring the dynamic control of tasks. *Attention and Performance*, 15:421–452, 1994.

M. Argyle and M. Cook. *Gaze and Mutual Gaze*. Cambridge University Press, 1976.

D. Ballard, M. Hayhoe, F. Li, and S. Whitehead. Hand-eye coordination during complex tasks. *Investigative Ophthalmology and Visual Science*, 33(4):1355, 1992.

H. Bekkering, J. Adam, A. van den Aarssen, H. Kingman, and J. Whiting. Interference between saccadic eye and goal-directed hand movements. *Experimental Brain Research*, 106:475–484, 1995.

R. Brooks, C. Breazeal, R. Irie, C. Kemp, M. Marjanovic, B. Scassellati, and M. Williamson. Alternative essences of intelligence. In *AAAI98*, 1998.

J. Cassell, C. Pelachaud, N. Badler, M. Steedman, B. Achorn, T. Becket, B. Douville, S. Prevost, and M. Stone. Animated conversation: Rule-based generation of facial expression, gesture and spoken intonation for multiple conversational agents. In *ACM SIGGRAPH Annual Conference Series*, pages 413–420, 1994.

J. D. Cohen and T. A. Huston. Progress in the use of interactive models for understanding attention and performance. *Attention and Performance*, 15:453–476, 1994.

J.D. Cohen, K. Dunbar, and J.L. McClelland. On the control of automatic processes: A parallel distributed processing account of the stroop effect. *Psychological Review*, 97(3):332–361, 1990.

H. Crane. The purkinjee image eyetracker. In D. Kelly, editor, *Visual Science and Engineering*, chapter 2. Dekker, 1994.

H. Egeth and S. Yantis. Visual attention: Control, representation, and time course. *Annual Review of Psychology*, 48:269–297, 1997.

B. Fisher. The role of attention in visually guided eye movements in monkey and man. *Psychological Research*, 48:251–257, 1986.

C. Folk, R. Remington, and J. Wright. The structure of attentional control: Contingent attentional capture by apparent motion, abrupt onset and color. *Journal of Experimental Psychology: Human Perception and Performance*, 20(2):317–329, 1994.

E.G. Freedman and D.L. Sparks. Eye-head coordination during head-unrestrained gaze shifts in rhesus monkeys. *Journal of Neurophysiology*, 77(5):2328–2348, 1997.

N. Lavie G. Rees, C. Frith. Modulating irrelevant motion perception by varying attentional load in an unrelated task. *Science*, 278:1616–1619, 1997.

C. Healey, K.S. Booth, and J.T. Enns. High-speed visual estimation using preattentive processing. *ACM Transactions on Computer-Human Interaction*, 3(2):107–135, 1996.

O. Hikosaka, S. Miyauchi, and S. Shimojo. Orienting of spatial attention: its reflexive, compensatory and voluntary mechanisms. *Cognitive Brain Research*, 5:1–9, 1996.

A. Hillstrom and S. Yantis. Visual motion and attentional capture. *Perception and Psychophysics*, 55(4):399–411, 1994.

W. Hirst. The psychology of attention. In *Mind and Brain: Dialogues in Cognitive Neuroscience*, pages 105–141, 1986.

M. Johnson. Visual attention and the control of eye movements in early infancy. *Attention and Performance*, 15:291–310, 1994.

J. Jonides. Voluntary versus automatic control over the mind's eye movement. *Attention and Performance*, 9:187–203, 1981.

D. Kahneman. *Attention and Effort.* Prentice-Hall, 1973.

P. Kalra, A. Mangili, N. Magnenat-Thalmann, and D. Thalmann. Smile: A multilayered facial animation system. In T.L. Kunii, editor, *Modeling in Computer Graphics.* Springer-Verlag, 1991.

T. Kito, M. Haraguchi, M. Funatsu, M. Sato, and M. Kondo. Measurements of gaze movements while driving. *Perceptual and Motor Skills*, 68:19–25, 1989.

R. M. Klein and A. Pontefract. Does oculomotor readiness mediate cognitive control of visual attention? revisited! *Attention and Performance*, 15:333–350, 1994.

C. Koch and S. Ullman. Shifts in selective visual attention: Toward the underlying neural circuitry. *Human Neurobiology*, 4:219–227, 1985.

E. Ladavas, G. Zeloni, G. Zaccara, and P. Gangeni. Eye movements and orienting of attention in patients with visual neglect. *Journal of Cognitive Neuroscience*, 9(1): 67–75, 1997.

R.L. Lewis, S.B. Huffman, B.E. John, J. E. Laird, J.F. Lehman, A. Newell, P.S. Rosenbloom, T. Simon, and S.G. Tessler. SOAR as a unified theory of cognition. In P. Rosenbloom, J. Laird, and A. Newell, editors, *The SOAR Papers: Readings on Integrated Intelligence*, chapter 51. MIT Press, 1993.

G. Lohse, K. Biolsi, N. Walker, and H. Rueter. A classification of visual representations. *Communications of the ACM*, 37(12):36–50, 1994.

M. Marjanovic, B. Scassellati, and M. Williamson. Self-taught visually-guided pointing for a humanoid robot. In *Fourth International Conference on Simulation of Adaptive Behavior*, Cape Cod, Massachusetts, 1996.

N. Moray. Designing for attention. In *Attention, Selection, Awareness, and Control: A Tribute to Donald Broadbent*, pages 53–72. Clarendon Press, Oxford, 1993.

H. Noser, O. Renault, and D. Thalmann. Navigation for digital actors based on synthetic vision, memory, and learning. *Computers & Graphics*, 19(1):7–19, 1995.

Fred Parke and Keith Waters. *Computer Facial Animation.* A K Peters, 1996.

A. Pearce, B. Wyvill, G. Wyvill, and D. Hill. Speech and expression: A computer solution to face animation. In M. Green, editor, *Proceedings of Graphics Interface '86*, pages 136–140, 1986.

M. I. Posner and Y. Cohen. Attention and the control of movements. In *Tutorials in Motor Behavior*, pages 243–258, 1980.

M. I. Posner, R. D. Rafal, L.S. Choate, and J. Vaughan. Inhibition of return: Neural basis and function. *Cognitive Neuropsychology*, 2:211–228, 1985.

P. Rabbit. Control of attention in visual search. In *Varieties of Attention*, Series in Cognition and Perception, pages 273–288. Academic Press, 1983.

R. Rao, G. Zelinsky, M. Hayhoe, and D. Ballard. Modeling saccadic targeting in visual search. In D. Touretzky, M. Mozer, and M. Hasselmo, editors, *Advances in Neural Information Processing Systems*. MIT Press, 1996.

C. Reynolds. Flocks, herds, and schools: A distributed behavioral model. *Computer Graphics*, 21(4):25–34, 1987.

B. Scassellati. Mechanisms of shared attention for a humanoid robot. In *AAAI Fall Symposium on Embodied Cognition and Action*, 1996.

D. L. Sparks. The neural control of orienting eye and head movements. In *Proceedings Dahlen Conference on Motor Control*, 1989.

M. Swain and M. Stricker. Promising directions in active vision. *International Journal of Computer Vision*, 11:109–126, 1993.

F. Thomas and O. Johnson. *Disney Animation: The Illusion of Life*. Abbeville Press, New York, NY, 1981.

J.K. Tsotsos, S.M. Culhane, W.Y.K. Wai, Y. Lai, and F. Nufflo. Modeling visual attention via selective tuning. *Artificial Intelligence*, 78:507–545, 1995.

X. Xhao and B. Achorn. *Implemented as part of doctoral research.* Center for Human Modeling and Simulation, University of Pennsylvania, 1994.

S. Yantis. Stimulus-driven attentional capture and attentional control settings. *Journal of Experimental Psychology: Human Perception and Performance*, 19(3):676–681, 1993.

S. Yantis and D. Johnson. Mechanisms of attentional priority. *Journal of Experimental Psychology: Human Perception and Performance*, 16(4):812–825, 1990.

S. Yantis and J. Jonides. Abrupt visual onsets and selective attention: Voluntary versus automatic allocation. *Journal of Experimental Psychology: Human Perception and Performance*, 16(1):121–134, 1990.

A. L. Yarbus. *Eye Movements and Vision.* Plenum Press, 1967.