Neural Networks 21 (2008) 1380–1391

Contents lists available at ScienceDirect

Neural Networks

journal homepage: www.elsevier.com/locate/neunet

A head–neck–eye system that learns fault-tolerant saccades to 3-D targets using a self-organizing neural model

Narayan Srinivasa [a,*], Stephen Grossberg [b]

[a] Department of Information and System Sciences, HRL Laboratories LLC 3011, Malibu Canyon Road, Malibu, CA – 90265, United States

[b] Department of Cognitive and Neural Systems, Center for Adaptive Systems and Center for Excellence for Learning in Education, Science and Technology, Boston University, 677 Beacon Street, Boston, MA – 02215, United States

ARTICLE INFO

Article history:
Received 1 February 2007
Revised and accepted 31 July 2008

ABSTRACT

This paper describes a head–neck–eye camera system that is capable of learning to saccade to 3-D targets in a self-organized fashion. The self-organized learning process is based on action perception cycles where the camera system performs micro saccades about a given head–neck–eye camera position and learns to map these micro saccades to changes in position of a 3-D target currently in view of the stereo camera. This motor babbling phase provides self-generated movement commands that activate correlated visual, spatial and motor information that are used to learn an internal coordinate transformation between vision and motor systems. The learned transform is used by resulting head–neck–eye camera system to accurately saccade to 3-D targets using many different combinations of head, neck, and eye positions. The interesting aspect of the learned transform is that it is robust to a wide variety of disturbances including reduced degrees of freedom of movement for the head, neck, one eye, or any combination of two of the three, movement of head and neck as a function of eye movements, changes in the stereo camera separation distance and changes in focal lengths of the cameras. These disturbances were not encountered during motor babbling phase. This feature points to general nature of the learned transform in its ability to control autonomous systems with redundant degrees of freedom in a very robust and fault-tolerant fashion.

1. Introduction

In humans, there are basically two types of eye movements. The most common eye movement is to keep the gaze affixed. These gaze holding movements are the result of compensating for head movements by moving the eyes in an equal and opposite amount to the direction of the head movements. These movements are either driven by the balance organs of the inner ear (called the vestibule-ocular reflexes or VOR), or alternatively they can be driven by the retinal image motion in a feedback loop (called the optokinetic responses or OKR). The other main class of eye movement comes about because the fovea (center high resolution portion of the retina), has a high concentration of color sensitive photoreceptor cells called cone cells. The rest of the retina is mainly made up of monochrome photoreceptor cell called rod cells, which are especially good for motion detection. By moving the eye so that small parts of a scene can be sensed with greater resolution, body resources can be used more efficiently. The eye movements disrupt vision and hence we have evolved to make these movements as fast, and therefore as short in duration, as they can possibly be; they are called *saccades*. In addition, since we have two eyes, they need to be coordinated so that images of an object fall on exactly the same parts of the two retinae.

A feature of the human eye system is that the total degrees of freedom available for use to perform coordinated eye movements is far greater than that required to fixate or saccade to 3-D targets. These redundant degrees of freedom are exploited by the human brain to derive flexible ways to saccade to 3-D targets. There have been several attempts to develop robotic camera systems that can saccade to 3-D targets (Aloimonos, 1990; Batista, Peixoto, & Ara'ujo, 1997; Batista, Dias, Araujo, & Almeida, 1995; Brown & Coombs, 1993; Dias et al., 1997; Murray, Bradshaw, MacLauchlan, Reid, & Sharkey, 1995; Sharma, 1994; Srinivasa & Ahuja, 1998; Srinivasa & Sharma, 1997, 1998; Wei & Ma, 1994). The novel aspect of this work is that we demonstrate a fully self-organized approach to learning how to perform saccadic control despite redundancies in the system. Furthermore, we also demonstrate that such a control system offers robustness to various disturbances that the system has not experienced *a priori*. This feature of our work has seldom been demonstrated in previous work for saccade control. In this paper, our goal is to develop a self-organized learning process that can enable a robotic head–neck–eye camera system

* Corresponding author. Fax: +1 310 317 5958.
E-mail address: nsrinivasa@hrl.com (N. Srinivasa).

0893-6080/$ – see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.neunet.2008.07.007

with 12 degrees of freedom to saccade to 3-D targets. Here the number of degrees of freedom available is more than the space in which the goal is specified (4-D coordinates — the desired location of the image coordinates of a 3-D target). This system is an example of a *motor equivalent* system (Bullock, Grossberg, & Guenther, 1993) because there are several possible alternatives to saccade to a given 3-D target. These alternatives are derived from various combinations of the redundant degrees of freedom of the head–neck–eye camera system. A redundant head–neck–eye camera system generates self-consistent signals between vision and motor systems via action perception cycles. These signals are then used by a self-organizing neural model to learn how to control the head–neck–eye movements to saccade to 3-D targets in a robust and fault-tolerant fashion.

The paper is organized as follows. The next section will introduce the notion of action perception cycles. This will be followed in Section 3 with an introduction to the head–neck–eye camera model. In Section 4, the self-organized learning process using the head–neck–eye camera model will be outlined. The next section will highlight the performance of the learned transform to saccade to new 3-D targets. In this section, the results on testing the saccade system for robustness to various disturbances and constraints will also be provided. In Section 6, the basis for model robustness will be provided. The biological plausibility of the model will also be discussed here. In Section 7, conclusions will be provided followed by details of the head–neck–eye model in the Appendix.

2. Action-perception cycles

Perceiving without acting is not common. For example, scrutinizing an object visually presupposes saccades at it and sometimes involves moving the head or even the whole body. Similarly, to accurately localize a sound source it becomes necessary to move head and ears towards the sound source. Acting without perceiving seldom makes sense; after all, actions defined as goal-directed behavior, aim at producing some perceivable event — the goal. Performing an appropriate action requires perceptual information about suitable starting and context conditions and, in the case of complex actions, about the current progress in the action sequence. Thus, perception and action are interlinked or interdependent.

There are several behavioral repertoires in which this interdependency is manifested in humans and other species. In the simplest form, a behavior is triggered by the present situation and reflects the animal's immediate environmental conditions. This type of behavior is often referred to as *stimulus-response reflexes*. A good example of this type of behavior in humans is provided by the orientation reflex, which we exhibit when encountering a novel and unexpected event. On the one hand, this reflex inhibits ongoing actions and tends to freeze the body — a stimulus triggered response. At the same time, it also draws attention towards the stimulus source by increasing arousal and facilitating stimulus-directed body movements. This interdependency between stimulus and response creates an *action perception cycle* (Piaget, 1963) wherein a novel stimulus triggers actions that lead to a better perception of itself or its immediate environmental condition and the cycle continues.

Human behavior is much more flexible than exclusive control by stimulus-response cycles. One of the hallmarks of human capabilities is the ability to *learn* new relations between environmental conditions and appropriate behavior during action perception cycles. This learning process provides an enormous gain in flexibility for an individual by creating the ability to adapt to environmental changes. Not only we learn to react to particular environmental conditions and situations in a certain way, we also can unlearn what we have acquired and learn new relationships between situations and actions. In fact, it is known (Barto & Sutton, 1982; Barto, 1995; Grossberg, 1972, 1982) that humans can condition their behaviors based on rewards/punishments that they may receive while interacting with its environment (also known as reinforcement learning). Furthermore, the ability to implement and switch between learned behaviors forms the basis of highest degrees of behavioral flexibility. It is the goal of this paper to study how action perception cycles could play a part in enabling a head–neck–eye camera system to *learn to saccade* to 3-D targets in a manner that is robust and tolerant to new disturbances and unexpected situations.

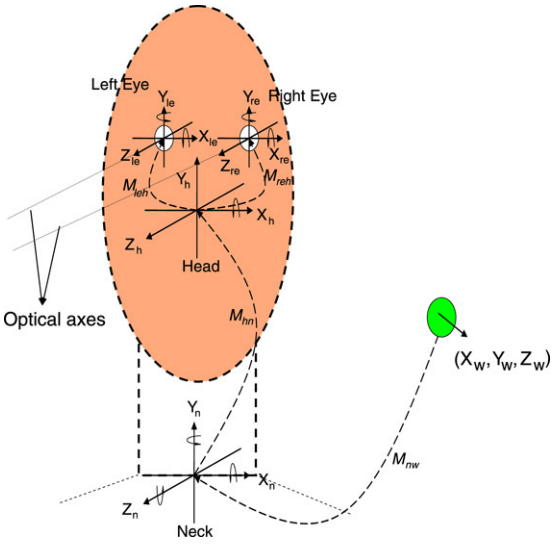

Fig. 1. A schematic of the head–neck–eye camera model is shown here. The details of the notations and transformations between the various coordinate frames are provided in the Appendix.

3. The head–neck–eye camera model

The human visual system is an active vision system that can be controlled by the brain in a deliberate fashion to extract useful information about the environment. The head–neck–eye camera model used in this work is an abstracted version of the human active vision system. It consists of a pair of cameras (eyes) mounted on a head and the whole system is supported by neck (refer to Fig. 1). The head–neck–eye system consists of 12 degrees of freedom: each eye has 8 degrees of freedom — a single tilt and two independent pan for rotation, two degrees of freedom for controlling the focal length of each camera, two degrees of freedom for the retinal image center for each camera and an additional degree of freedom in controlling the baseline distance between the two cameras; the neck has 3 degrees of freedom – tilt, pan and yaw (shoulder-to-shoulder) for rotation; and finally the head has one degree of freedom – tilt for rotation. During action perception cycles, only the rotational degrees of freedom (or extrinsic parameters) were exercised. The imaging parameters (or intrinsic parameters) were fixed. The kinematics that describes the imaging of a 3-D object in both the eyes as a function of the extrinsic parameters and intrinsic parameters is provided in the Appendix.

4. Self-organized learning of saccades via action perception cycles

During action perception cycles for learning to saccade, the head–neck–eye camera system is setup to look at 3-D targets in its visible space for various joint configurations of camera system.

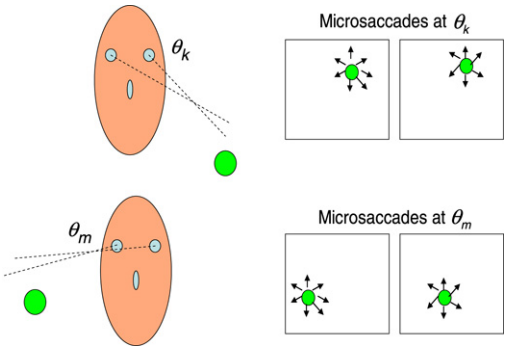

Fig. 2. A couple of steps during the action perception cycles are shown here for illustration. The camera joint configuration denoted by $\theta_k = \{\alpha_N, \beta_N, \gamma_N, \alpha_H, \alpha_e, \beta_{Le}, \beta_{Re}\}$ corresponds to the seven d.o.f of the camera. For each such camera configuration, the camera performs a set of actions — microsaccades and the resulting perception is a set of translations of the 3-D target (green sphere shown here) in various directions. Similar actions are performed at the second joint configuration θ_m. This process is repeated during the action perception cycles where the 3-D target is placed uniformly spanning the visible space of the head–neck–eye camera system.

Each head–neck–eye camera joint configuration $\boldsymbol{\theta}$ corresponds to a unique joint position of each of the seven degrees of freedom (three neck — $\boldsymbol{\alpha_N}, \boldsymbol{\beta_N}, \boldsymbol{\gamma_N}$, one head — $\boldsymbol{\alpha_H}$ and three eye — $\boldsymbol{\alpha_e}, \boldsymbol{\beta_{Le}}, \boldsymbol{\beta_{Re}}$ rotation angles). It should be noted that we will also use the notation θ_i (for $i = 1, \ldots, 7$) to refer to the seven rotational degrees of freedom interchangeably throughout the rest of the paper. These joint configurations represent the context for the learning system. At each joint configuration, the camera performed a set of microsaccades wherein the joints of the camera system was exercised to move in small increments. These actions resulted in translation of image of the 3-D target in various directions within the image plane of both the cameras (refer to Fig. 2).

The differential relationship between the spatial directions of target image in the retina to the joint rotations of the head–neck–eye camera system as a result of the microsaccades for a given context $\boldsymbol{\theta}$ during action perception cycles is a linear mapping. Our system learns this mapping in a self-organized fashion (as described below). For a redundant system like the head–neck–eye camera system used in this paper, this linear mapping is a one-to-many function. This implies that there exists several possible linear combinations of solutions from spatial directions of the target image in the stereo camera to head–neck–eye camera joint angle changes that can generate a single image space trajectory of the target that is continuous in joint space and correctly directed in the 4-D space (two direction vectors corresponding to the stereo pair directed towards the retinal center for each eye – i.e., to saccade). For example, to look at a 3-D target, it is possible to just move only the eyes with respect to $\boldsymbol{\theta}$ in order to fixate on the target provided the target is visible and the eye joints are within the physical limits of its joint rotation space or joint space. At the same time, it may also be possible to use some of the other joints including the head and neck in addition to the eye joints to fixate on the same 3-D target. Joint space continuity is ensured because all solutions are in the form of joint angle increments with respect to the present fixed joint configuration $\boldsymbol{\theta}$ of the head–neck–eye camera system.

This synchronous collection of increments to one or more joint angles of the head–neck–eye camera system is called a *joint synergy* (Bullock et al., 1993). During the self-organized learning process, the head–neck–eye motor system learns to associate a finite number of joint synergies to the spatial direction of image movements in the stereo camera when these synergies are activated for a given $\boldsymbol{\theta}$. During performance, a given desired movement direction of the target (in the case of saccades the desired direction is toward the retinal image center) can be achieved by activating in parallel any linear combination of the synergies that produces that image movement direction. This simple control strategy leads to motor equivalence when different linear combinations are used on different movement trials. The self-organized learning process utilizes a self-organizing neural model that will now be described.

The neural architecture for learning to saccade to 3-D targets is shown in Fig. 3. The network consists of four types of cells. The S cells encode the spatial directions of the target when the camera is either babbling or performing a learned saccadic movement. The V cells encode the difference between weighted inputs from the direction cells S and the R cells that encode the joint rotation directions or increments of the head–neck–eye system. The network adapts the weights Z between the S cells and the V cells based on the difference of activity between the spatial directions of the target motion in the cameras and joint rotations of the head–neck–eye camera system. We have adopted the VAM learning approach for this learning process (Gaudiano & Grossberg, 1991; Srinivasa & Sharma, 1998). During learning, the V cell activity drives the adjustment of weights. This process is akin to learning the pseudo-inverse of the Jacobian between spatial directions to joint rotations. During performance, the learned weights are used to drive the R cells to the desired increment in joint rotation.

In order to ensure that the correct linear mapping is learned and the motor equivalence can be addressed, the learning process has to account for the joint configuration $\boldsymbol{\theta}$ under which learning of the mapping takes place. We refer to each joint configuration as the context and a set of neurons or C cells that encode various contexts that span the joint space of the head–neck–eye camera system as the context field (Bullock et al., 1993; Fiala, 1994). The C cell neuron in the context field strongly inhibits the V cells allocated for that context (refer to Fig. 3). When a C cell neuron in the context field is excited due to the head–neck–eye system being in the appropriate configuration, it momentarily inhibits the V cells allocated for that context and this allows the learning process to adapt the weights in a manner that enables the computation of the correct linear mapping. An overall flowchart in Fig. 4 summarizes the neural network model and the various steps during both the training and performance phases. Table 1 summarizes the network equations of the model for these two phases.

5. Computer simulations

The head–neck–eye system used in this work is a seven-dof for angular position control (extrinsic parameters) and five degrees of freedom (change in focus and separation in baseline between the eyes and location(s) of the retinal center in the stereo images). The system has more degrees of freedom than required to saccade to a 3-D target thereby making the system redundant.

5.1. Learning

The system was trained using a single 3-D target (a sphere) during action–perception cycles to learn appropriate weights to saccade to the target (refer to Fig. 2). The process begins by motor-babbling where the head–neck–eye camera system performs random eye movements that exercise all the seven rotational degrees of freedom. These movements are two types. First, it performs a gross movement wherein the camera moves to distinctly different camera configurations. At each of these camera configurations, a set of microsaccades were performed and the direction-to-rotation transform was (as described earlier) learned at each camera configuration or context.

The weights Z were initialized to zero and subsequently adapted (refer to Table 1) on a context basis as and when the

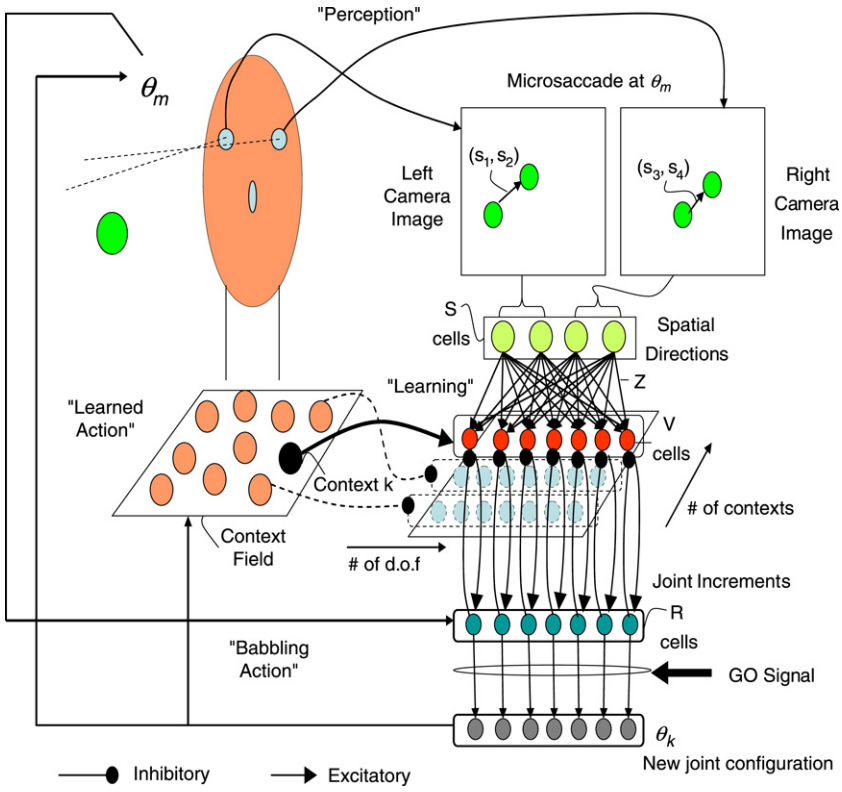

Fig. 3. The neural architecture for learning saccades to 3-D targets is shown here. The network shows the S, V and R cells and how their interactions during perception, learning and action cycles enable the network to adapt the weights z to learn to saccade. During performance, the network can integrate the R cell activity to produce new joint configurations that move the head–neck–eye camera system to saccade to a 3-D target.

Table 1
The network equations during learning and performance phases are listed for various cells and computations

Network cells/computations	Learning phase	Performance phase
Direction cells (S)	$\frac{dS_j}{dt} = -\lambda S_j + (1 - S_j)s_j - S_j \sum_{l \neq j} s_l$	$\frac{dS_j}{dt} = -\lambda S_j + (1 - S_j)d_j - S_j \sum_{l \neq j} d_l$
Difference cells (V)	$\frac{dV_{ik}}{dt} = -\alpha \left[V_{ik} + \sum_j z_{ijk} S_j - R_i\right]$	$\frac{dV_{ik}}{dt} = -\alpha \left[V_{ik} + \sum_j z_{ijk} S_j - R_i\right]$
Joint rotation cells (R)	$\frac{dR_i}{dt} = \delta(r_i - R_i)$	$\frac{dR_i}{dt} = \delta(-R_i + \sum_k V_{ik})$
Weights (Z)	$\frac{dz_{ijk}}{dt} = \gamma V_{ik} S_j$	No learning
Joint configuration (θ)	$\theta_i = \theta_i^{\min} + rand^*(\theta_i^{\max} - \theta_i^{\min})\ i = 1, \ldots, 7$ $\theta_1 = \alpha_N; \theta_2 = \beta_N; \theta_3 = \gamma_N; \theta_4 = \alpha_H$ $\theta_5 = \alpha_e; \theta_6 = \beta_{Le}; \theta_7 = \beta_{Re};$	$\frac{d\theta_i}{dt} = -\eta\theta_i + G[R_i] + \theta_{old};\ i = 1, \ldots, 7$
Context cell selection	$C_i = \|\theta_c - \theta_i\|$ $k = \max_k(C_k)$	$C_i = \|\theta_c - \theta_i\|$ $k = \max_k(C_k)$

The direction cells S for the learning phase obey a center-surround type of computation (Grossberg, 1988) where the direction inputs s are normalized by the S field. For the performance phase, the S cells are normalized in a similar fashion. However, the desired direction to image center d is used in the computation. The V cell computations are based on the difference between the weighted direction inputs and joint rotation increments. The R cell computations during learning phase is based on the motor babbled increment r whereas during performance is derived from the population of V cell activity whose contexts k are relevant to the head–neck–eye configuration. The learning updates of z are based on the activity of the V and S cells. During learning phase the joint configurations are computed randomly as part of the motor babbling phase while they are updated during performance phase by the product of the GO signal which controls the speed of the camera movement. The context cell selection during learning and performance is based on selecting the cell k that best matches (L_2 norm or Euclidean distance) the current head–neck–eye camera configuration.

appropriate context node became active. In our simulations, the range of the joints used for the seven rotational degrees of freedom and the number of discretized zones for each angle is listed in Table 2. This discretization process yielded a total of 77175 contexts cells ($5 \times 5 \times 3 \times 3 \times 7 \times 7 \times 7$). At each camera configuration a total of 100 randomly generated microsaccades were performed to compute the direction-to-rotation transform for that context. The sequence of steps during the learning phase is summarized in Fig. 4(i). The various parameters used during the learning phase for the equations listed in Table 1 are: $\lambda = 0.01, \alpha = 10.0, \delta = 32.0$ and $\gamma = 8.0$. All simulations were performed using the 4th order Runge–Kutta ODE solver with a time step of 0.001. The total duration of learning was 2.5 h on Dell XPS computer with a 2 GB RAM.

5.2. Performance

The performance phase begins when the learning phase is completed. It should be noted that this does not have to be the case in general. The nature of the learning algorithm (i.e., VAM learning — Gaudiano and Grossberg (1991)) allows training and performance phases to be interleaved. In order to saccade to a visible target, the desired spatial direction from the current location of target in the stereo images to the image centers is computed. The context or camera configuration of head–neck–eye system is used to access the appropriate direction-to-rotation transform. This transform is used to compute the desired joint angle increments of the head–neck–eye camera system. These increments are finally integrated over time to saccade to the 3-D

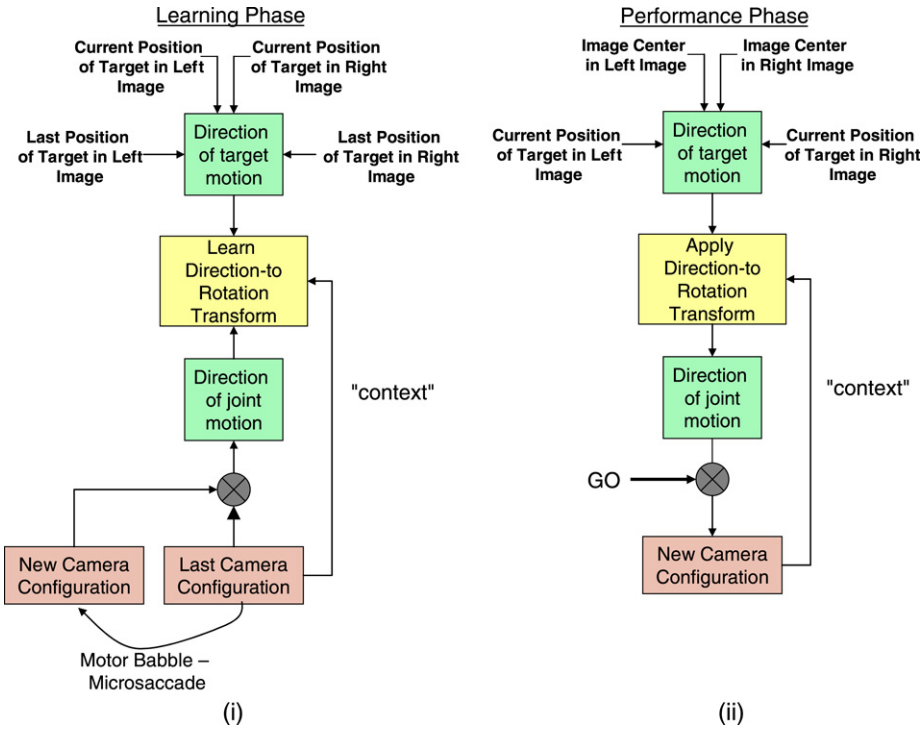

Fig. 4. An overall summary flowchart for sequence of steps during (i) Learning phase — the process begins at the bottom of the flowchart by motor-babbling step or the action step that induces the joint directions cells to activate. At the same time, the corresponding perception step computes the spatial direction of target movement in the stereo camera. Using the context of head–neck–eye system, the system learns the correct spatial direction to joint rotations during action–perception cycles. (ii) Performance phase — the process begins in the opposite fashion from the learning phase. In order to saccade to a visible target, the desired spatial direction to the retinal image centers is computed. This is then combined with the context to compute the desired joint angle increments of the head–neck–eye camera system. These increments are finally integrated over time to saccade to the 3-D target.

Table 2
The joint ranges of the seven degrees of freedom as well as the number of angular zones used for discretization of each angle are listed here.

Joint	Min (deg)	Max (deg)	Angular zones
α_N	−60	30	5
β_N	−90	90	5
γ_N	−60	60	3
α_H	−45	45	3
α_e	−60	60	7
β_{le}	−60	60	7
β_{re}	−60	60	7

target. This integration step involves a GO signal $G(t) = G_0{}^*t$ (Bullock et al., 1993) that defines the speed at which saccadic movement is performed. The joint angle increment is multiplied with $G(t)$ to obtain joint angle velocity vector which is integrated to get new joint angles. In our simulations, $G_0 = 25.0$. The sequence of steps during performance phase is summarized in Fig. 4(ii). The various parameters used during performance phase for equations listed in Table 1 are: $\lambda = 0.01, \alpha = 40.0, \delta = 32.0, \gamma = 8.0$ and $\eta = 0.001$.

The system was tested for its ability to saccade to new 3-D targets after learning phase was completed. The system was able to accurately saccade to 3-D targets within its view and within its visible space (i.e., target is in front of the camera and within the controllable joint space of camera). The system used all the seven rotational degrees of freedom to generate motor synergies during its saccadic movements. An example saccade using its seven degrees of freedom is shown in Fig. 5. In this figure, the 3-D target is shown in the form of a sphere. The camera images of sphere were processed as a binary image to extract the centroid of the images and the controller then moved the camera joints to bring this centroid to center of the camera during a saccade. The seven degree of freedom joint positions can be seen in Fig. 5(i). The trajectory of centroid during the saccade is traced for the stereo images in Fig. 5(ii). Example snapshots from the saccade sequence including the initial, intermediate and final configurations are shown in Fig. 5(iii). The optical axis of both the camera intersects on the sphere in the final configuration indicating the completion of the saccade. In all our simulations, the system began its saccade with the same initial configuration of the camera without any loss of generality. Also, in all our simulations the system was expected to converge to within 4 pixel square width of the true camera image center (shown as the cross hair location in Fig. 5(ii)).

5.3. Performance: Loss of one degree of freedom

The system was tested for its ability to handle various constraints and disturbances. The first case was to reduce the degrees of freedom of the system to six by preventing the neck of the camera from shoulder-to-shoulder movement ($\gamma_N = 0°$). This situation could be viewed as a change in the extrinsic parameters of the system and is new compared to how the system was trained during the learning phase. The camera was still able to accurately saccade to the 3-D target as shown in Fig. 6. The plots of the joint positions show the γ_N to be flat indicating that it was locked during the saccade.

5.4. Performance: Six degrees of freedom with shift in retinal image center in both cameras

The system was tested with same loss of degrees of freedom (i. e., $\gamma_N = 0°$) and an additional new constraint of generating saccades to a new shifted retinal image center. In all our simulations, (160, 120) were coordinates of the original retinal image center. The new shifted retinal image center coordinates was (130, 140). This set of new constraints can be seen as a

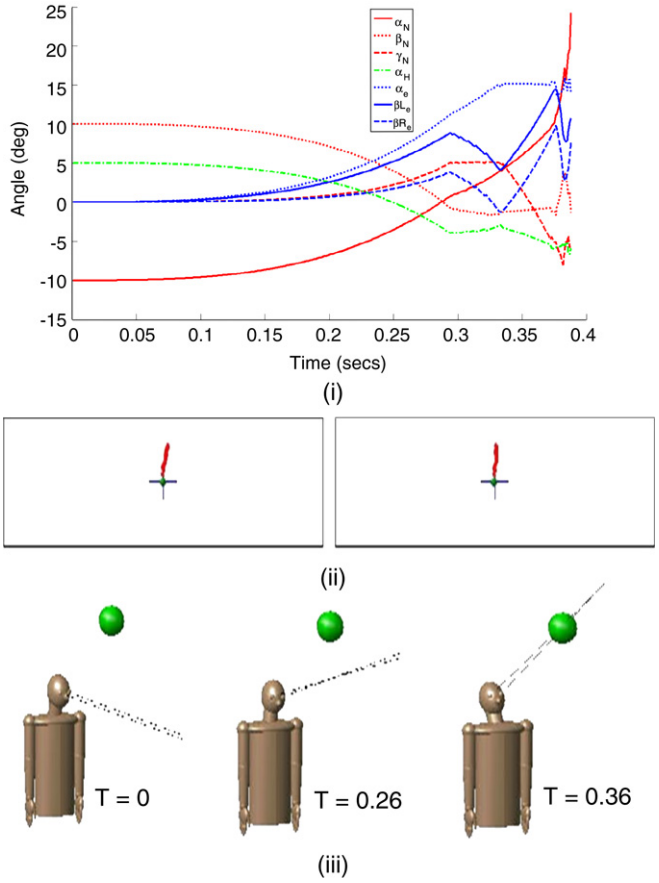

Fig. 5. This shows an example saccade sequence during normal performance. (i) Joint position trajectories during the saccade. (ii) Spatial trajectory of the target in the stereo camera during the saccade. (iii) Initial, intermediate and final images of system during saccade.

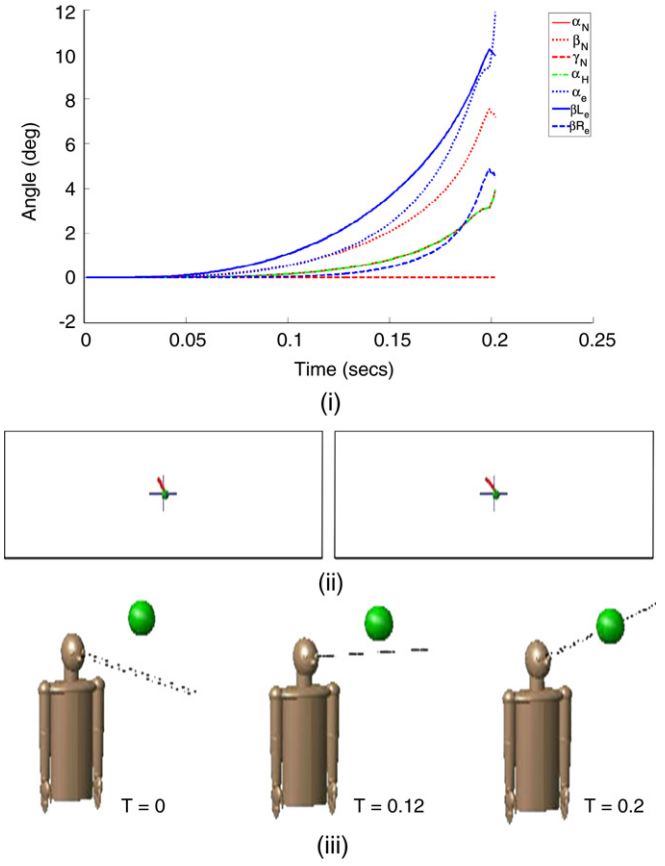

Fig. 6. This shows an example saccade sequence with loss of γ_N is fixed throughout the saccade. (i) Joint position trajectories during the saccade. (ii) Spatial trajectory of the target in the stereo camera during the saccade. (iii) Initial, intermediate and final images of system during saccade.

combination of extrinsic and intrinsic parameter disturbances. The results in Fig. 7 show that the system is able to cope with both these conditions not encountered during learning phase. It should be noted that the optical axis of the cameras are not intersecting on the sphere due to new shift in retinal image center. The system demonstrates that it is able to compensate its control for this new constraint by fixating on some other point in 3-D space such that the 3-D target is imaged at the new center of the image.

5.5. Performance: Three degrees of freedom

The system was tested by preventing the entire head and neck from moving (i.e., $\alpha_N = -10°$; $\beta_N = 10°$; $\gamma_N = 0°$; $\alpha_H = 5°$). This reduced the degrees of freedom for the extrinsic parameters of system from seven to three. Since the system was expected to saccade to 3-D targets, this constraint still provided sufficient degrees of freedom to saccade to 3-D targets. The results in Fig. 8 show an example case of how the system adapted to the new constraints not seen during learning phase. Here the system is able to remarkably just move its eyes to saccade to the target much like humans can.

5.6. Performance: Three degrees of freedom with change in baseline distance

The system was then tested by adding further constraints to the system in Section 5.5. Here the baseline distance between the cameras was increased from 0.17 units during learning phase to 0.27 units (refer to parameter B in the model equations in Appendix). This corresponds to an intrinsic parameter change and will affect the retinal images. However, the learned transform is immune to this change as demonstrated in Fig. 9. The system is able to cope with both these constraints and perform accurate saccades to 3-D targets.

5.7. Performance: Three degrees of freedom with change in focal length

The system was tested with same three degrees of freedom for the eyes as above but the focal length of both its cameras was now changed from 0.15 units in the original system to 0.25 units in the new system. Since the focal length of the cameras were both increased, the size of the sphere and hence the center of the sphere were affected by the change. This resulted in shifts in the image registration for both cameras. This situation corresponds to an extreme change in extrinsic parameters and also changes in intrinsic parameters. The learned transform was immune to this change as shown in Fig. 10. The change in focal length manifests as a change in eye size in the model as can be seen in Fig. 10(iii).

5.8. Performance: Three degrees of freedom with change in focal length and baseline distance

The system above was now further constrained by changing the baseline in distance between the eyes from 0.17 units to 0.27 units. This corresponds to further changes/disturbances to intrinsic parameters of the system. The system is able to perform remarkably well and can perform accurate saccades even with these extreme changes in both the extrinsic and intrinsic

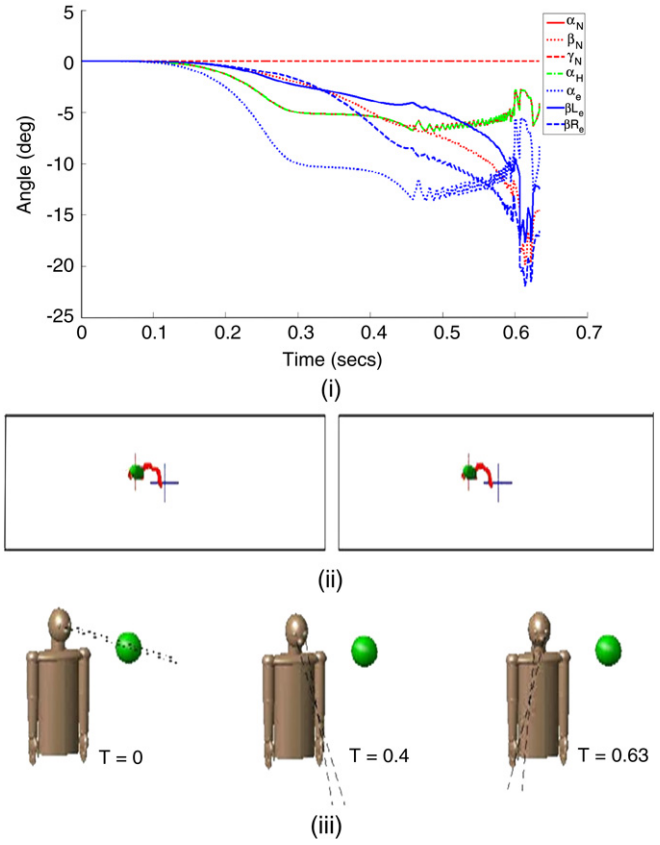

Fig. 7. This shows an example saccade sequence with loss of γ_N and a change in the image center (the red cross hair indicates the new center). (i) Joint position trajectories during the saccade. (ii) Spatial trajectory of the target in the camera during the saccade. (iii) Initial, intermediate and final images of system during saccade.

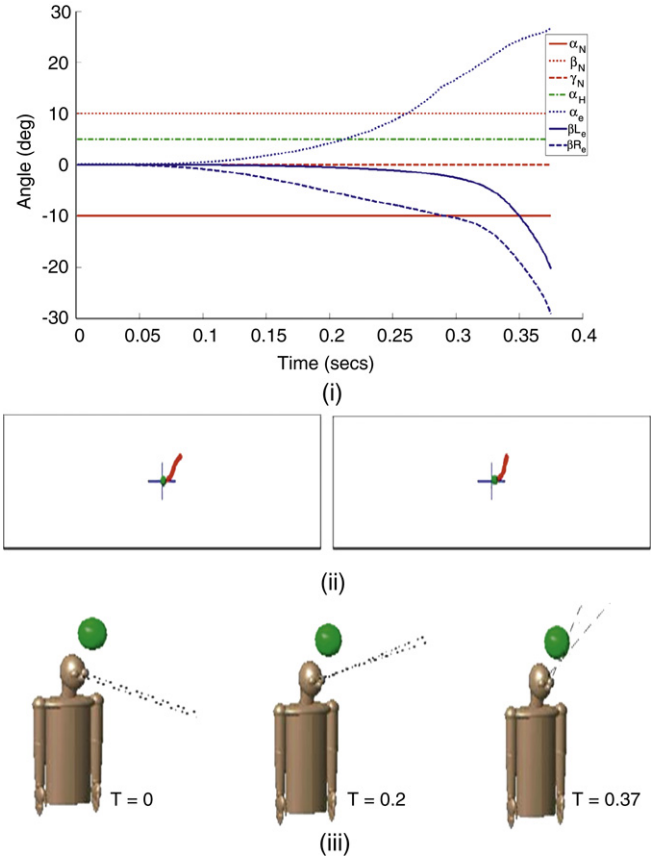

Fig. 8. This shows an example saccade sequence with loss of the entire head and neck movement. (i) Joint position trajectories during the saccade. (ii) Spatial trajectory of the target in the stereo camera during the saccade. (iii) Initial, intermediate and final images of system during saccade.

parameters. An example result for this experiment is shown in Fig. 11.

5.9. Performance: Three degrees of freedom with different focal length for each camera

The system with same degrees of freedom for the eyes was tested with different focal lengths for each eye. The focal length of left eye was set to 0.13 while that for the right eye was set to 0.27. This is another example of changes the system has not been exposed during learning phase with constraints on extrinsic parameters and changes in the intrinsic parameters. The system is able to saccade to 3-D targets even with these constraints and changes as shown in Fig. 12. The difference in focal lengths of the two cameras is manifested in the form of smaller (larger) eye size for the left (right) camera.

5.10. Performance: Three degrees of freedom, different focal lengths and shift in image center in one of the camera

The system above was now further subjected to an additional change in that the image center of the left camera was changed from the original location (160, 120) to (130, 120). The image center of the right camera was unchanged at (160, 120). The focal lengths of the left and right eyes were set to 0.13 and 0.27. The system was able to saccade to 3-D targets and an example is shown in Fig. 13. Since the eyes had a common vertical tilt degree of freedom, any unequal offset in the vertical direction cannot be accomplished. This could however, be possible if the system had learned to control the head–neck–eye system with an independent vertical degree of freedom for each eye (much like the human eye) which has an independent set of muscles that control the vertical movements of each eye (Grossberg & Kuperstein, 1989; Sparks & Mays, 1983).

5.11. Performance: Original system with change in focal length, baseline distance and shift in image center in both cameras

The original system was also tested to see if it were robust to all the three changes in the intrinsic parameters of camera: change in focal length from 0.15 units to 0.25 units, change in baseline distance from 0.17 units to 0.27 units and also a shift in the center of the image. The system performed accurate saccades and was found to be robust to these changes despite not being trained on it during the learning phase. An example result of this experiment is shown in Fig. 14.

5.12. Performance: Original system with one camera lost

The original system was tested to see if it could tolerate the complete loss of a camera. In order to simulate this, our model essentially performs direction computation for only the functional camera. The joint angles for the camera being lost is not used during the control process. The system performed saccades to the 3-D target but used the extra degrees of freedom available in the form of head and neck movements to accomplish the saccade. This was achieved without having being exposed to this unforeseen perturbation in the system. An example result of this experiment is shown for the case of the left camera being lost in Fig. 15. The system performed equally well if the other camera was lost as well.

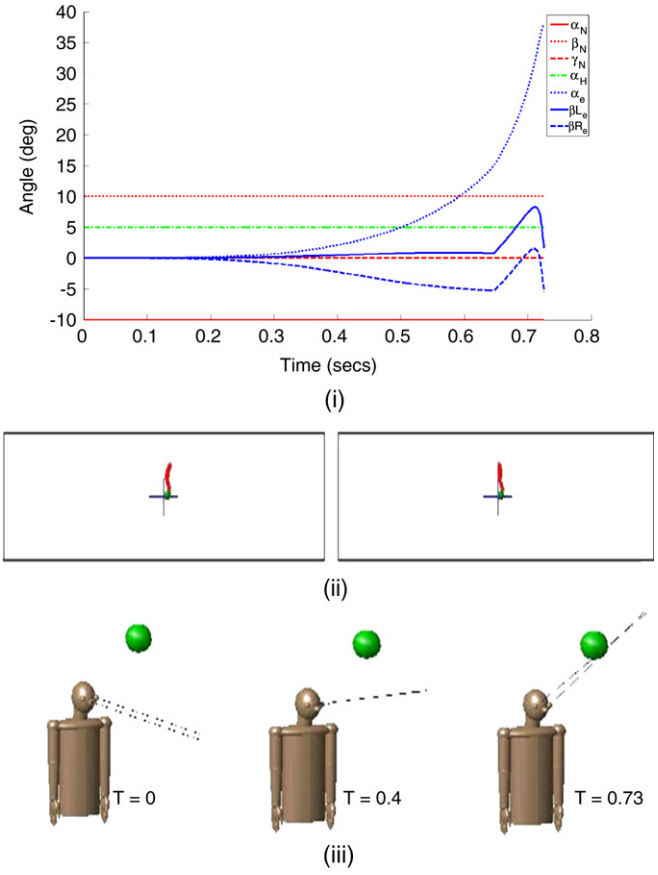

Fig. 9. This shows an example saccade sequence with loss of the entire head and neck movement and increase in baseline distance between eyes. (i) Joint position trajectories during the saccade. (ii) Spatial trajectory of the target in the stereo camera during the saccade. (iii) Initial, intermediate and final images of system during saccade.

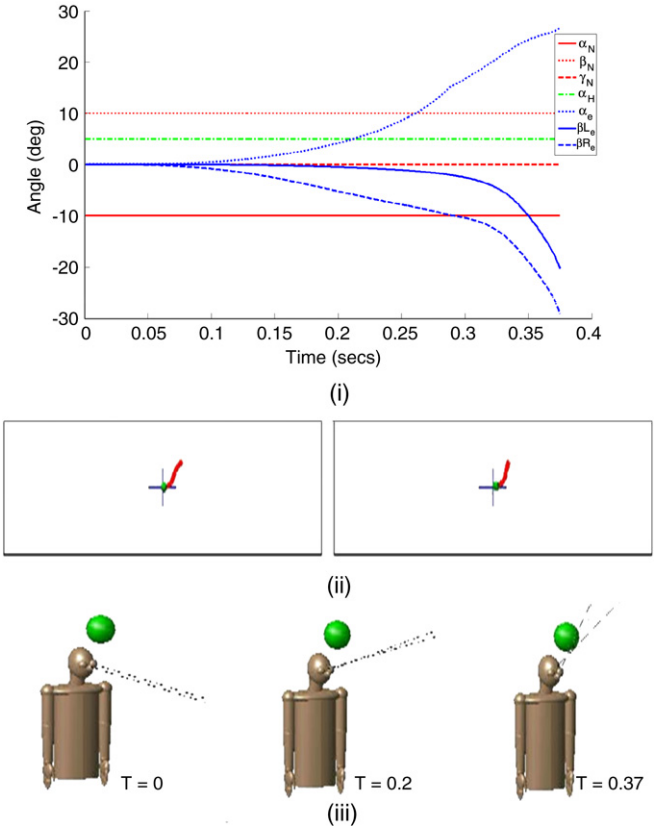

Fig. 10. This shows an example saccade sequence with loss of the entire head and neck movement and increase in focal length. (i) Joint position trajectories during the saccade. (ii) Spatial trajectory of the target in the stereo camera during the saccade. (iii) Initial, intermediate and final images of system during saccade.

5.13. Performance: Original system with stereo image expansion/compression

The original system was tested for the case of stereo image expansion and compression. This feature was manifested in the model by performing a translation of the image pixels with respect to the center: either away from image centers in a radial direction (expansion) or towards the image centers in a radial direction (compression). The system was able to successfully saccade despite image expansion or compression. An example result is shown in Fig. 16 for an expansion factor of 2.5. This result extends the capability DIRECT model much like the new results (Grosse-Wentrup & Contreras-Vidal, 2007) for handling image expansion and compression during reaching tasks. Our results shows that the saccades can be performed using a stereo camera despite these perturbations.

5.14. Performance: Original system with stereo image rotation

The original system was tested for the case of stereo image rotations. Image rotations are manifested in our system using pure rotations of image pixels about the image centers. The system was found to be able to perform saccades to targets in a manner robust to image rotations of up to 90° rotations. An example result is shown in Fig. 17 for an image rotation of 30°. This result also extends the capability of the original DIRECT model much like the new results (Grosse-Wentrup & Contreras-Vidal, 2007) for handling image rotations during reaching tasks. Our results shows that the saccades can be performed using a stereo camera despite these perturbations.

6. Discussion

An important feature to point out in all these performance experiments (Section 5.3 through 5.14) is that none of these new disturbances or changes was experienced during learning via action–perception cycles. The main reason for this robustness is the nature of linear mapping learned during the self-organized learning process. For example, loss of motion of one or more degrees of freedom during a saccade will cause the actual saccadic movement produced by the head–neck–eye system to mismatch desired movement.

If the actual movement differs by less than 90° from desired movement direction, which is the case even with loss four rotational degrees of freedom, our system is able to accurately finish saccade, provided geometry of the head–neck–eye system with reduced degrees of freedom allows the joint configuration that is required to bring the image of the 3-D target to image centers. This is because the desired movement continuously reflects the effects of errant actual saccadic movements caused by changes in both extrinsic and intrinsic parameters. This is ensured as long as the accurate information about both target position (i.e., the image centers of the two cameras) and current location of target in the stereo images is available. This enables our system to continuously update the desired saccadic movement direction thus ensuring that the error due to actual errant saccadic movement does not accumulate. Using this desired saccadic movement direction, the redundant direction to rotation linear mapping always moves the head–neck–eye camera system towards the image centers of the two cameras (as simulations have shown) albeit in a sub-optimal fashion.

The robustness and fault tolerance exhibited by the head–neck–eye system trained using the self-organizing neural model is akin

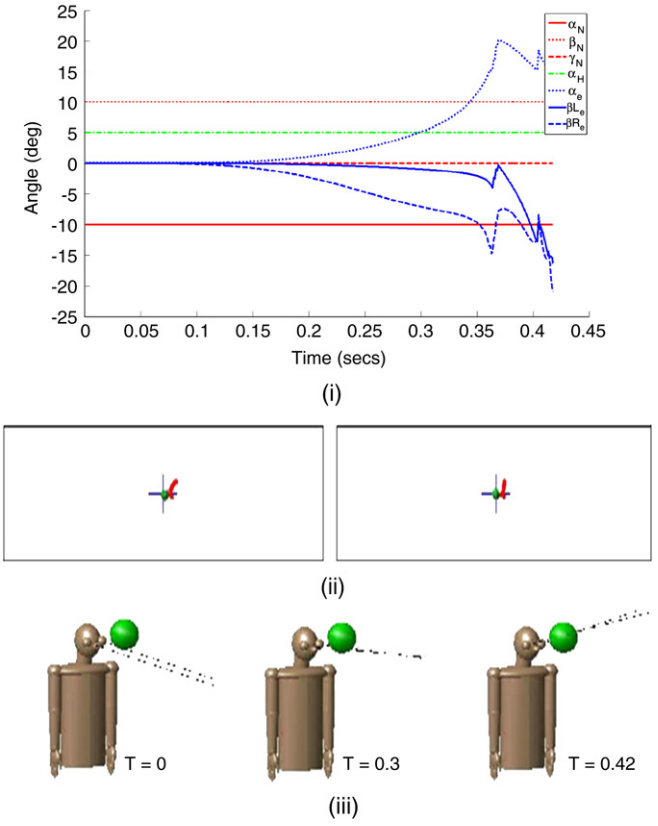

Fig. 11. This shows an example saccade sequence with loss of the entire head and neck movement, increase in focal length and increase in baseline distance between eyes. (i) Joint position trajectories during the saccade. (ii) Spatial trajectory of the target in the stereo camera during the saccade. (iii) Initial, intermediate and final images of system during saccade.

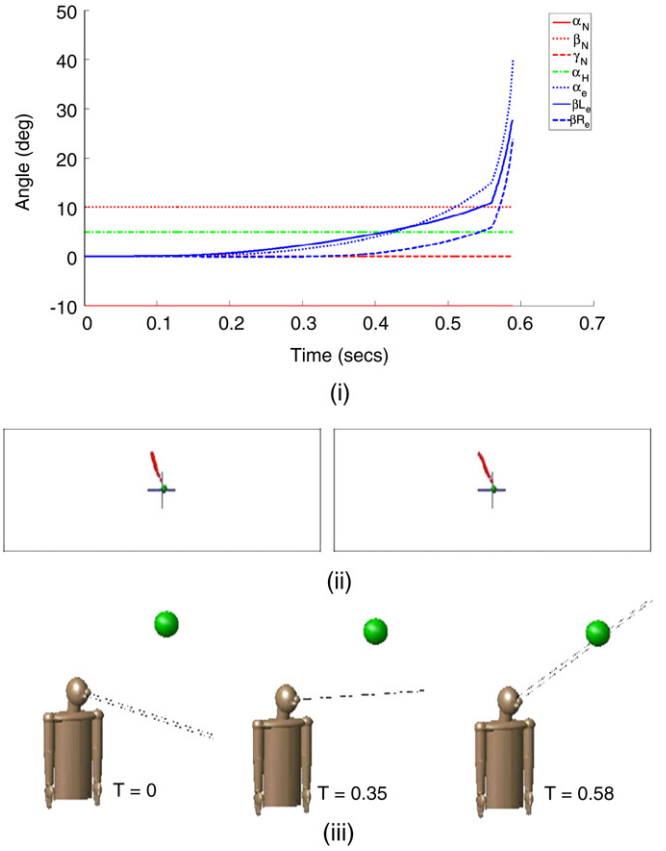

Fig. 12. This shows an example saccade sequence with loss of the entire head and neck movement, decrease in focal length for left eye and increase in focal length for the right eye. (i) Joint position trajectories during the saccade. (ii) Spatial trajectory of the target in the stereo camera during the saccade. (iii) Initial, intermediate and final images of system during saccade.

to behaviors by biological systems wherein reliability and fault-tolerance are cornerstones of performance. These key features ensure the survivability of biological systems when faced with previously unforeseen environments and disturbances. It seems that by learning the appropriate transform between various sensory modalities, biological systems are able to exhibit fault –tolerance and robustness to unexpected changes. A similar transform was learned in Bullock et al. (1993) where they demonstrated motor-equivalent reaching and tool use of a 3-dof planar arm. There the arm was able to adapt to new disturbances such as restrictions to joint rotations, tool-use and prism rotations of images. They also point out that their model matches human data from neuroscience and psychophysical point of view. The results from this paper further point to the utility of this learned transform to control multimodal interacting systems with redundant degrees of freedom in 3-D.

The head–neck–eye camera system presented in this paper is a model that utilizes a basic mechanism for learning from action–perception cycles. These cycles provides self-consistent movement commands that activate correlated visual, spatial and motor information that are used to learn an internal coordinate transformation between vision and motor systems. While this strategy is adopted by biological systems, we do not claim that the underlying neural architecture replicates the human brain. The time taken for some eye movements during testing are not in the physiological range and one could argue that the movements are not *saccadic* in nature. However, it should be noted that these movements are not performed under normal conditions where the system has all the redundancy available. Our model predicts that for simulations under altered conditions, not experienced during the training, can cause the system to be slow in performing eye movements. For example, the eye movements during the saccade sequence with loss of the entire head and neck movement and increase in baseline distance between eyes (Fig. 9) shows that the system needed nearly 700 ms to perform the eye movement to the target. Increasing the GO parameter under these circumstances does not provide much faster performance. On the contrary, increasing the GO parameter under extremely limited joint mobility situations causes the system to become unstable. The system seems to require making the movements more slowly in order to be able to correct for the mistakes made after each control movement (or microsaccades). The speed of response may however be improved by learning with adaptive timing as performed by the sub-cortical cerebellum of the brain which is implicated for learning rapid saccade control in conjunction with other brain regions including the superior colliculus, frontal eye fields, etc. The reader is referred to a complete set of models that address how these brain regions interact to enable accurate and rapid saccade control (Brown, Bullock, & Grossberg, 2004; Carpenter, 1988; Gancarz & Grossberg, 1999, 1998; Grossberg & Kuperstein, 1989; Grossberg, Roberts, Aguilar, & Bullock, 1997; Sparks & Mays, 1983).

7. Conclusions

We developed a self-organizing neural model that is capable of learning to generate saccades to 3-D targets by self-generated vision and motor signals during action–perception cycles. The learned knowledge is in the form of local linear mappings at each camera joint configuration between direction of motion

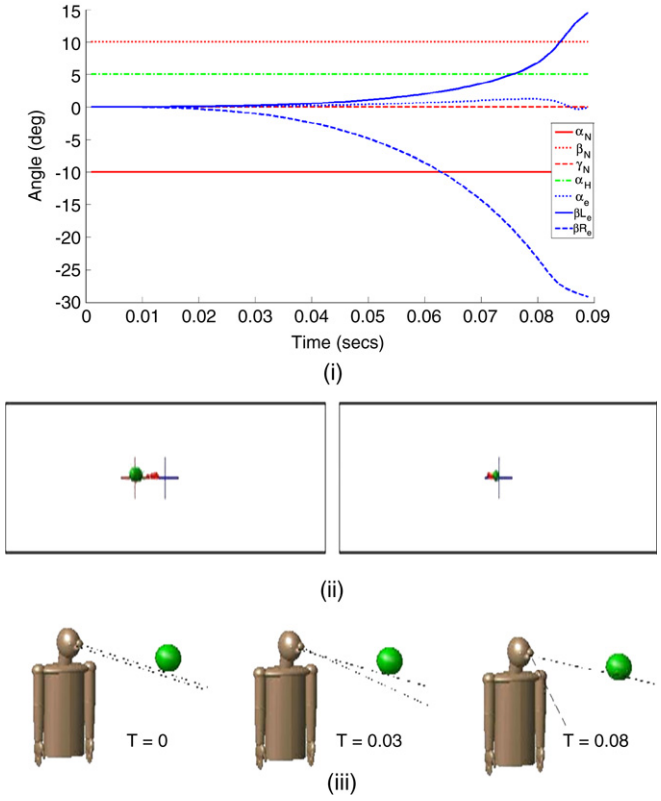

Fig. 13. This shows an example saccade sequence with loss of the entire head and neck movement, decrease in focal length for left eye and increase in focal length for the right eye and shift in retinal image center for the left eye. (i) Joint position trajectories during the saccade. (ii) Spatial trajectory of the target in the stereo camera during the saccade. (iii) Initial, intermediate and final images of system during saccade.

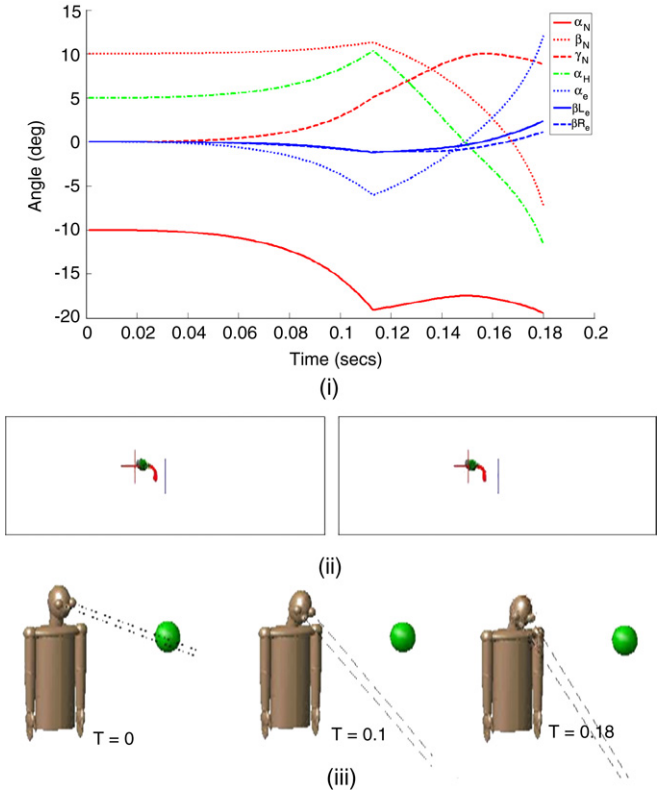

Fig. 14. This shows an example saccade sequence with original head–neck–eye system with an increase in focal length, increase in baseline distance between eyes and shift in retinal image center. (i) Joint position trajectories during the saccade. (ii) Spatial trajectory of the target in the stereo camera during the saccade. (iii) Initial, intermediate and final images of system during saccade.

of a target in images of the stereo camera to corresponding change in joint angles (in the form of microsaccades) that caused the change in direction. These local linear mappings can be construed as an internally calibrated model at each joint configuration that the system can utilize to perform accurate saccades to 3-D targets. The interesting aspect of this model is its ability to exploit redundancies in the head–neck–eye system to overcome new disturbances and constraints not seen during learning. This fault tolerant nature of the transform and its applications to motor equivalent reaching (Bullock et al., 1993) and motor equivalent saccades as shown in this work points to a biologically plausible control strategy that may be adopted by humans and other animals. This strategy seems to offer a trade-off between optimality in functioning within a limited scope or rigid environments for flexible and robust performance in a wide variety of environments.

Acknowledgements

The authors would like to thank Youngkwan Cho for his help in developing the simulation model of the robot and the head–neck–eye system. We would like to thank an internal R&D grant at HRL under SR060029 that supported this effort.

Appendix

The series of coordinate transformations that transform the world coordinates of a target $P_w = (X_w, Y_w, Z_w, 1)^T$ to target image coordinates $P_{le} = [x_{le}, y_{le}]$ and $P_{re} = [x_{re}, y_{re}]$ for our head–neck–eye model is now provided.

World to neck transformation

The matrix M_{wn} transforms the coordinates of the target P_w in world frame to P_n in the neck coordinate frame (refer to Fig. 1). Transformations between coordinate systems are represented by 4×4 matrices (Walker & Orin, 1982), and the points in each coordinate system are represented by 1×4 column vectors in homogeneous coordinate systems. The matrix M_{wn} is a function of the three degrees of freedom of the neck: α_N, β_N, γ_N and can be derived as follows. The local transform matrix M_n of the neck can be expressed as:

$$M_n = R^z(\gamma_N) * R^y(\beta_N) * R^x(\alpha_N) \quad \text{where}$$

$$R^z(\circ) = \begin{bmatrix} \cos(\circ) & -\sin(\circ) & 0 & 0 \\ \sin(\circ) & \cos(\circ) & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

$$R^y(\circ) = \begin{bmatrix} \cos(\circ) & 0 & \sin(\circ) & 0 \\ 0 & 1 & 0 & 0 \\ -\sin(\circ) & 0 & \cos(\circ) & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

$$R^x(\circ) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos(\circ) & -\sin(\circ) & 0 \\ 0 & \sin(\circ) & \cos(\circ) & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

The transformation of neck coordinates to world coordinates is then expressed as:

$$M_{wn} = R^z(-\pi/2) * R^x(\pi/2) * M_n * R^x(-\pi/2) * S * T_{nw} \quad \text{where}$$

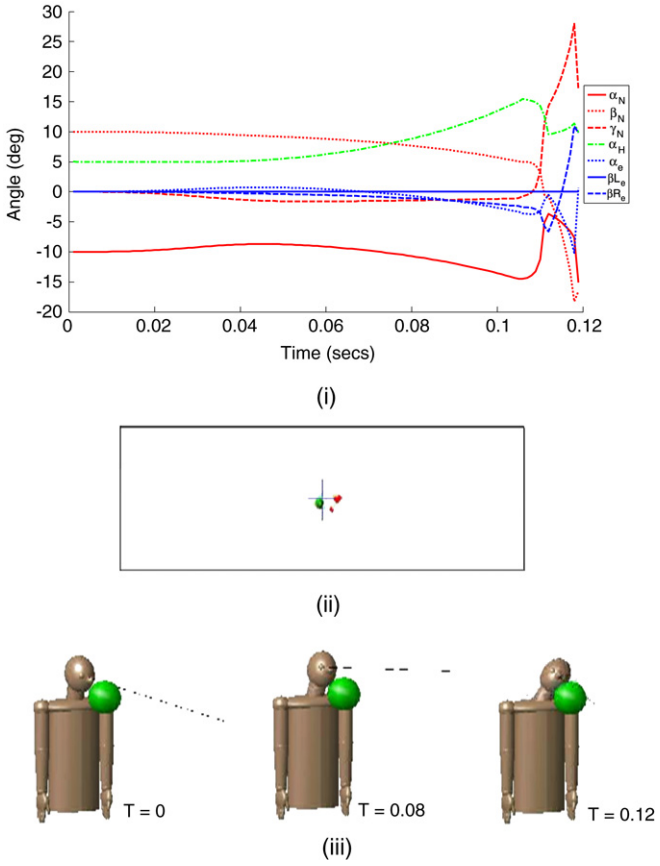

Fig. 15. This shows an example saccade sequence with original head–neck–eye system with loss of the left camera. (i) Joint position trajectories during the saccade. Notice that the joint angle for the left camera remains at zero throughout the saccade. (ii) Spatial trajectory of the target in the stereo camera during the saccade for the right camera. (iii) Initial, intermediate and final images of system during saccade.

$$S = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \frac{3}{4}K & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \qquad T_{nw} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & \frac{K}{2} \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

and the rotations can be computed as above.

Neck to head transformation

The matrix M_{nh} transforms the coordinates of the target in neck coordinates to targets in the head coordinate frame. The matrix M_{nh} is a function of one degree of freedom of the head α_H and is given by:

$$M_{nh} = R^x(\alpha_h) * T_h \quad \text{where}$$

$$T_h = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0.3K \\ 0 & 0 & 1 & -0.1K \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

and K is constant.

Head to left-eye transformation

The matrix M_{hl} transforms the coordinates of the target in head coordinate frame to targets in the left eye coordinate frame. This matrix is a function of the two degrees of freedom of the left eye α_e and β_{le} and is given by

$$M_{hl} = T_{le} * R^y(\beta_{le}) * R^x(\alpha_e) * T_{\text{lens}} \quad \text{where}$$

$$T_{le} = \begin{bmatrix} 1 & 0 & 0 & BK \\ 0 & 1 & 0 & -0.03K \\ 0 & 0 & 1 & -0.29K \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

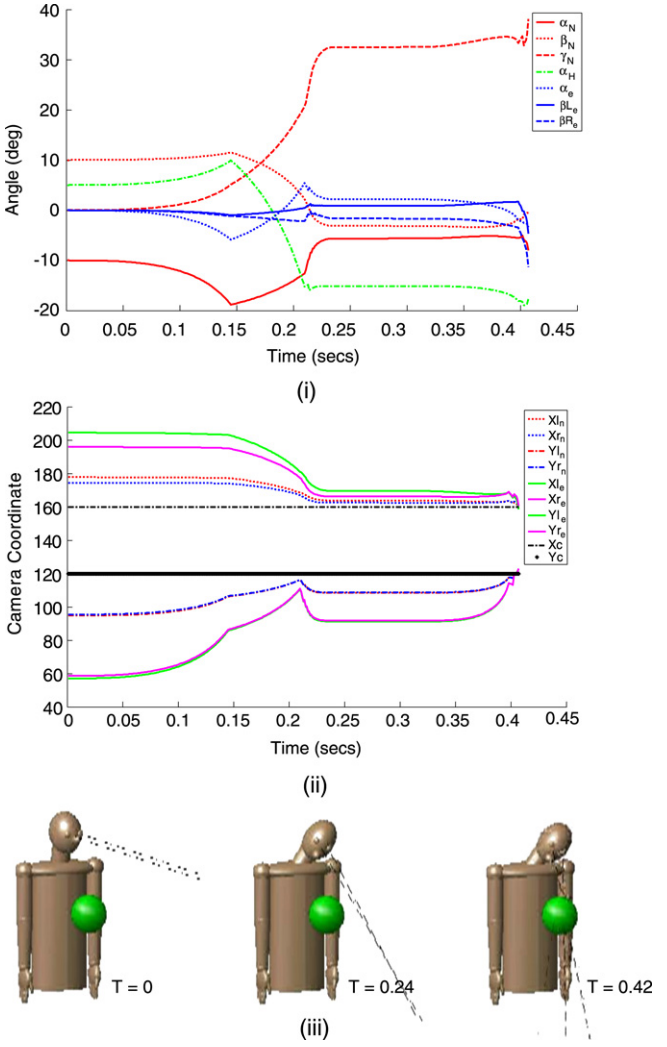

Fig. 16. This shows an example saccade sequence with original head–neck–eye system with stereo image expansion. (i) Joint position trajectories during the saccade. (ii) Spatial trajectory of the target in the stereo camera during the saccade is shown for with (subscript e) and without image expansion (subscript n). The convergence of the saccade to the image center (160, 120) is seen despite stereo image expansion. (iii) Initial, intermediate and final images of system during saccade.

and B is half the baseline distance between the cameras.

$$T_{\text{lens}} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & \frac{-fK}{2} \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

and f is the focal length of the camera lens.

Head to right-eye transformation

The matrix M_{hr} transforms the coordinates of the target in head coordinate frame to targets in the right eye coordinate frame. This matrix is a function of the two degrees of freedom of the right eye α_e and β_{re} and is given by

$$M_{hr} = T_{re} * R^y(\beta_{re}) * R^x(\alpha_e) * T_{\text{lens}} \quad \text{where}$$

$$T_{re} = \begin{bmatrix} 1 & 0 & 0 & -BK \\ 0 & 1 & 0 & 0.03K \\ 0 & 0 & 1 & 0.29K \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

and B is half the baseline distance between the cameras and T_{lens} is the same as the equation above.

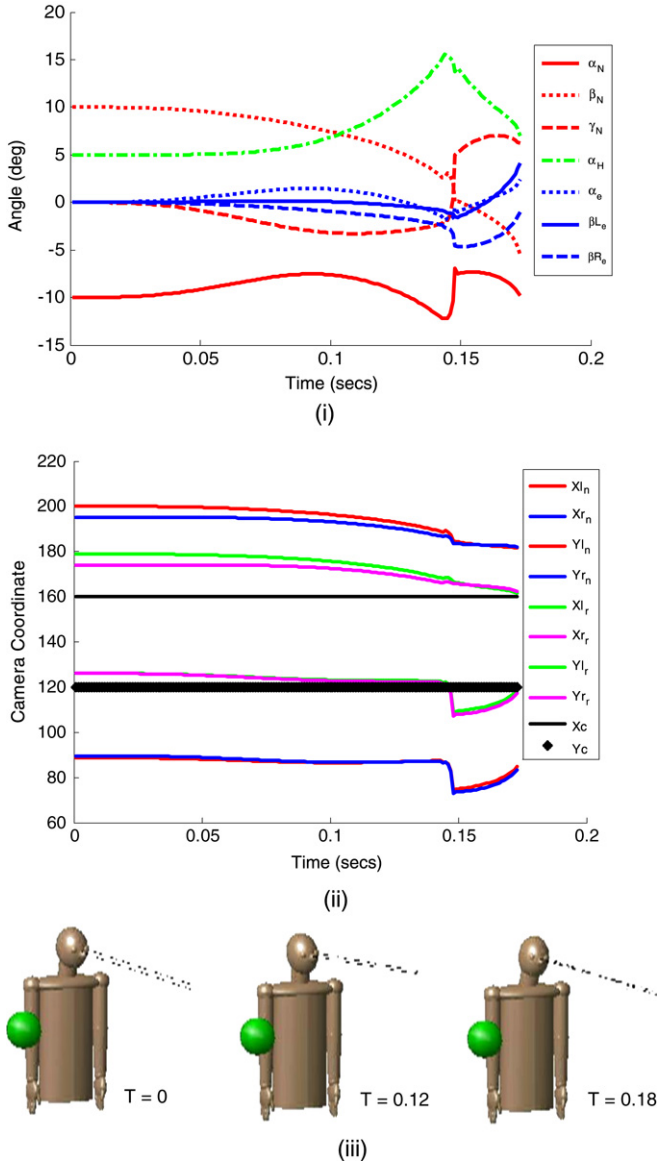

Fig. 17. This shows an example saccade sequence with original head–neck–eye system with stereo image rotation. (i) Joint position trajectories during the saccade. (ii) Spatial trajectory of the target in the stereo camera during the saccade is shown for with (subscript r) and without image rotation (subscript n). The convergence of the saccade to the image center (160, 120) is seen despite stereo image rotation. (iii) Initial, intermediate and final images of system during saccade.

World to eye transformations

The transformation from world coordinates to the camera coordinates is given by

$$\text{Left-Eye: } P_{le} = [M_{wn} * M_{nh} * M_{hl}]^{-1} * P_w$$
$$\text{Right-Eye: } P_{re} = [M_{wn} * M_{nh} * M_{hr}]^{-1} * P_w.$$

Camera image coordinates

The camera image coordinates for the left image can be computed as:

$$\begin{bmatrix} x_{le} \\ y_{le} \end{bmatrix} = \begin{bmatrix} \dfrac{(P_{le}(1,\ 1) + \text{FOV}_X) * W_X}{2\text{FOV}_X} \\ \dfrac{(P_{re}(2,\ 1) + \text{FOV}_Y) * W_Y}{2\text{FOV}_Y} \end{bmatrix}$$

where FOV_x is the field of view of the camera in the x direction and FOV_y is the field of view of camera in the y direction. W_x and W_y are the x and y pixel resolution of each camera. The camera image coordinates for the right image can similarly be computed as:

$$\begin{bmatrix} x_{re} \\ y_{re} \end{bmatrix} = \begin{bmatrix} \dfrac{(P_{re}(1,\ 1) + \text{FOV}_X) * W_X}{2\text{FOV}_X} \\ \dfrac{(P_{re}(2,\ 1) + \text{FOV}_Y) * W_Y}{2\text{FOV}_Y} \end{bmatrix}.$$

In all our simulations, $K = 1.0$, $W_x = 320$, $W_y = 240$, $\text{FOV}_x = 5.67$ and $\text{FOV}_y = 1.732$. The intrinsic parameters $B = 0.15$ and $f = 0.17$ during the learning phase. During performance phase, these two parameters were changed to test the robustness of the system to unforeseen changes.

References

Aloimonos, Y. (1990). Purposive and qualitative active vision. In *Proc. image understanding workshop* (pp. 816–828).

Barto, A. G., & Sutton, R. S. (1982). Simulation of anticipatory responses in classical conditioning by a neuron-like adaptive element. *Behavioral Brain Research*, *4*, 221–235.

Barto, A. G. (1995). Reinforcement learning. In M. A. Arbib (Ed.), *Handbook of brain theory and neural networks* (pp. 804–809). Cambridge, MA: MIT Press.

Batista, J., Peixoto, P., & Ara'ujo, H. (1997). Real-time vergence and binocular gaze control. In *Int. conf. on intelligent robots and systems*.

Batista, J., Dias, J., Araujo, H., & Almeida, A. (1995). The ISR multi-degree of freedom active vision robot head: design and calibration, M2VIP-95- In *Second international conference on mechatronics and machine vision in practice*.

Brown, J. W., Bullock, D., & Grossberg, S. (2004). How laminar frontal cortex and basal ganglia circuits interact to control planned and reactive saccades. *Neural Networks*, *17*, 471–510.

Brown, C., & Coombs, D. (1993). Real-time binocular smooth pursuit. *International Journal of Computer Vision*, *11*(2), 147–165.

Bullock, D., Grossberg, S., & Guenther, F. H. (1993). A self-organizing neural model of motor equivalent reaching and tool use by a multijoint arm. *Journal of Cognitive Neuroscience*, *5*, 408–435.

Carpenter, R. H. S. (1988). *Movements of the eye*. Pion Publishers.

Dias, J., Paredes, C., Fonseca, I., Batista, J., Araujo, H., & Almeida, A. (1997). Simulating pursuit with machines: experiments with robots and artificial vision. In *Proc. international conf. on robotics and automation*.

Fiala, J. C. (1994). A network of learning kinematics with application to human reaching models. In *IEEE international conference on neural networks*.

Gancarz, G., & Grossberg, S. (1999). A neural model of the saccadic eye movement control explains task-specific adaptation. *Vision Research*, *39*, 3123–3143.

Gancarz, G., & Grossberg, S. (1998). A neural model of the saccade generator of reticular formation. *Neural Networks*, *11*, 1159–1174.

Gaudiano, P., & Grossberg, S. (1991). Vector associative maps: Unsupervised real-time error-based learning and control of movement trajectories. *Neural Networks*, *4*(2), 147–183.

Grossberg, S. (1972). A neural theory of punishment and avoidance, II: Quantitative theory. *Mathematical Biosciences*, *15*, 253–285.

Grossberg, S. (1982). A psychophysiological theory of reinforcement, drive, motivation, and attention. *Journal of Theoretical Neurobiology*, *1*, 283–369.

Grossberg, S. (1988). Nonlinear neural networks: Principles, mechanisms, and architectures. *Neural Networks*, *1*, 17–61.

Grossberg, S., & Kuperstein, M. (1989). *Neural dynamics of adaptive sensory-motor control: Expanded edition*. Elmsford, NY: Pergamon Press.

Grossberg, S., Roberts, K., Aguilar, M., & Bullock, D. (1997). A neural model of multimodal adaptive saccadic eye movement control by superior colliculus. *Journal of Neuroscience*, *17*(24), 9706–9725.

Grosse-Wentrup, M., & Contreras-Vidal, J. L. (2007). The role of the striatum in adaptation learning: A computational model. *Biological Cybernetics*, *96*, 377–388.

Murray, D., Bradshaw, K., MacLauchlan, P., Reid, I., & Sharkey, P. (1995). Driving saccade to pursuit using image motion. *International Journal of Computer Vision*, *16*(3), 205–228.

Piaget, J. (1963). *The origins of intelligence in children*. New York: Norton.

Sharma, R. (1994). Active vision for visual servoing: A review. In *IEEE workshop on visual servoing: Achievements, applications and open problems*.

Sparks, D., & Mays, L. E. (1983). Spatial localization of saccade targets I: Compensation for stimulation induced perturbations in eye position. *Journal of Neurophysiology*, *49*, 45–63.

Srinivasa, N., & Ahuja, N. (1998). A learning approach to fixate on 3D targets with active cameras. In *Lecture notes in computer science*: *Vol. 1351* (pp. 623–631). Springer-Verlag.

Srinivasa, N., & Sharma, R. (1997). Execution of saccades for active vision using a neuro-controller. *IEEE Control Systems*, 18–29 (Special issue on intelligent control).

Srinivasa, N., & Sharma, R. (1998). Efficient learning of VAM-based representation of 3D targets and its active vision applications. *Neural Networks*, *11*(1), 153–172.

Walker, M. W., & Orin, D. E. (1982). Efficient dynamic computer simulation of robotic mechanisms. *Journal of Dynamic Systems, Measurement and Control*, *104*, 205–211.

Wei, G. Q., & Ma, S. D. (1994). Implicit and explicit camera calibration: Theory and experiments. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, *16*, 469–480.