Integration of Biological Neural Models for the Control of Eye Movements in a Robotic Head

Marcello Mulas(✉), Manxiu Zhan, and Jörg Conradt(✉)

Neuroscientific System Theory Group Department of Electric and Computer Engineering, Technische Universität München, Karlstrasse 45, 80333 Munich, Germany
{marcello.mulas,conradt}@tum.de
https://www.nst.ei.tum.de

Abstract. We developed a biologically plausible control algorithm to move the eyes of a six degrees of freedom robotic head in a human-like manner. Our neurocontroller, written with the neural simulator *Nengo*, integrates different biological neural models of eye movements, such as microsaccades, saccades, vestibular-ocular reflex, smooth pursuit and vergence. The coordination of the movements depends on the stream of sensory information acquired by two silicon retinas used as eyes and by an inertial measurement unit, which serves as a vestibular system. The eye movements generated by our neurocontroller resemble those of humans when exposed to the same visual input. This robotic platform can be used to investigate the most efficient exploration strategies used to extract salient features from either a static or dynamic visual scene. Future research should focus on technical enhancements and model refinements of the system.

Keywords: Neural control · Eye movements · Robotic head

1 Introduction

Traditional artificial vision algorithms analyze static visual scenes in order to extract the most salient features of an image. Unfortunately, state-of-the-art artificial vision is still far from reaching the performance of biological vision systems. A possible reason for such a gap is that biological systems might explicitly take advantage of information about the motion of the observer [1]. However, despite decades of research, it is not clear exactly how the human brain control eye movements in order to explore a visual scene.

Neuroscientists have developed several biologically plausible models of eye movement control. The models implemented for robotic control can reproduce human behavior but are usually limited to few specific eye movements (e.g., gaze fixation [2] and saccades [3]). A robotic control that integrates multiple

S.P. Wilson et al. (Eds.): Living Machines 2015, LNAI 9222, pp. 231–242, 2015.
DOI: 10.1007/978-3-319-22979-9_24

movement-specific models in a cohesive way is still lacking. This directly reflects the lack of neuroscientific knowledge about how top-down cognitive influences and bottom-up streams of sensory information interact in order to decide which movement should be executed at a given time.

It is plausible that the active exploration strategy adopted by the human brain might be tuned by evolution to speed up the process of information acquisition about the surrounding environment. An intelligent strategy that can quickly localize relevant features would also contribute to saving computational power. Aside from artificial vision, anthropomorphic robotics would also greatly benefit from a control algorithm that reproduces human-like behavior. Movement is one of the key factors taken into consideration by humans to judge the intelligence of an artificial system [4].

Several stereoscopic vision systems controlled by bio-inspired algorithms have been proposed for different purposes. However, to the best of our knowledge, a robotic head controlled by biologically realistic models that can reproduce human-like behavior is still lacking. For instance, Berthouze and Kuniyoshi [5] studied the emergence of sensorimotor coordination in a four DoF (degrees of freedom) robot head as it interacts with its surrounding environment. In order to provide both a wide field of view in the periphery and high resolution in the fovea, they used special camera lenses [6]. The learning of visual abilities such as gaze fixation and saccadic motion was based on Kohonen self-organizing maps (SOM) [7]. Bjorkman and Kragic proposed (2004) a robot head for object recognition, tracking and pose estimation [8]. They implemented foveal vision by using two sets of binocular cameras mounted on a four DoF robotic head. However, the authors did not aim at implementing neural plausible strategies, so the control program integrated different low level computer vision algorithms. Similarly to Berthouze et al., Asuni et al. used (2005) self-organizing maps to solve the inverse kinematics problem [2]. Once having learned sensory-motor relations, their seven DoF robotic head was able to direct its gaze towards a given target arbitrarily located in a 3D space. According to the authors, the robotic head was able to behave more similarly to humans thanks to additional constraints on the joint angles.

In this paper, we propose a biologically plausible control algorithm, implemented in *Nengo* [9], to move the eyes of a six DoF robotic head in a human-like manner. Our control integrates different biological neural models of specific eye movements, such as saccades, vestibular-ocular reflex, smooth pursuit and vergence. Coordination of these movements is based on the current internal and external context, dependent on the stream of sensory information continuously acquired by on-board sensors. The neural simulator *Nengo* was instrumental in easily defining information processing algorithms based on neural computation. The use of biological plausible neural circuitry can provide a more useful platform to investigate the basic mechanisms underlying human vision albeit at the expense of computational efficiency and movement control precision.

2 Robotic Setup

The robotic head that we designed to develop and validate our neurocontroller incorporates six DoF (Figure 1 to the left). Two DoF are at the neck joint for the pan and tilt of the whole head; the others are two for each eye, one for horizontal and one for vertical rotation. All the components of the frame are made by a 3D printer using a rigid opaque white material (*Objet VeroWhitePlus* photopolymer) except for two T-shaped elements made out of metal to enhance robustness. Due to technical considerations we designed our robot head differently in comparison to the anatomy of a human head. A major difference is the interpupillary distance (IPD) that is equal to 230 mm in our robot head against the IPD of an average human head that is around 63 mm [10].

We chose Dynamixel servomotors as a good compromise between the high speed performance required to match biological performance and the need to keep the cost of the robotic system low. We used two Dynamixel MX-64 servomotors to actuate the joint of the neck and four Dynamixel MX-28 servomotors to actuate the eyes [11]. The servomotors are connected in chain formation, and they are controlled by a computer via serial communication. A power supply provides energy to all servomotors at a constant voltage of 12 V. The electric current necessary to actuate all the servomotors is approximately 300 mA in standby conditions.

We built the eyes of the robot head using two dynamic vision sensors (DVS) [12]. These cameras, also called silicon retinas, have a resolution of 128x128 pixels. Contrarily to standard cameras which record the intensity of light, silicon retinas are sensitive to light temporal contrast, i.e. each pixel signals a change in the intensity of light. Every time the incident light increases or decreases its intensity by certain amount an ON or OFF event is asynchronously generated. This working mechanism greatly reduces the redundancy of visual information providing a data rate that typically is orders of magnitude lower in comparison with the data rate of conventional frame-based cameras [12]. The large dynamic range (> 120 dB) and the asynchronous event-based output (similar to the spiking activity of retinal ganglion cells) make silicon retinas suitable to model human eyes. In addition, despite very different transduction mechanisms, the sensitivity to temporal contrast makes silicon retinas behave similarly to human retinas in presence of static scenes. In fact, during constant light conditions silicon retinas do not signal any changes in light intensity and likewise retinal cells become silent due to neural adaptation [13]. We exploited this property of DVS sensors for the control of eye movements as it seems that the functional role of microsaccades is to prevent the image from fading away [14].

Next, an inertial measurement unit (IMU) sensor is mounted in the center of the robot head to serve as an artificial vestibular system. The IMU provides a continuous stream of sensory information about rotational velocities and accelerations in all directions of the 3D space. However, in order to implement the vestibular-ocular reflex, we took into account only horizontal (i.e., yaw) and vertical (i.e., pitch) rotations of the robot head.

In order to validate our neurocontroller, we recorded the movements of the robotic head using a motion capture system [15] recording at a sampling rate of 120 Hz. To estimate the positions of reflective markers, the tracking software analyzes the recordings of eight cameras mounted close to the ceiling of the recording room and pointing downwards. The positions and orientations of both the robotic head and the two eyes can be accurately estimated from the positions of the reflective markers (1 cm in diameter) appropriately attached to the robotic frame (Figure 1, right panel). Then, based on simple geometric calculations, we computed for each eye the saccadic traces projected on a monitor screen placed in front of the robotic head. We then compared these traces with those generated by humans while watching the same video.

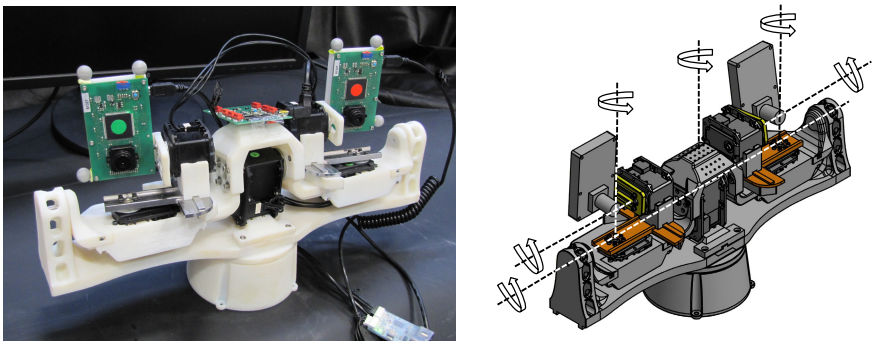

Fig. 1. Robotic head. Left: the robotic head is equipped with two silicon retinas, an inertial measurement unit and six servomotors. Right: the axes of rotation for each of the six DoF (two for each of the two eyes and two for the pan and tilt of the whole head).

3 Neurocontroller

We integrated different biological models of eye movement control using the neural simulator *Nengo* [9]. We chose to model neurons as Leaky Integrate and Fire (LIF) as this model represents a good compromise between biological plausibility and simulation performance. In *Nengo* LIF neurons are described by their tuning curves. Given a certain tuning curve, the spiking rate of a LIF neuron depends only on the input current. In order to cope with the massively parallel computation required by our neural simulation, we took advantage of *Nengo* support for GPU computing. Our neurocontroller runs on a Nvidia GeForce GTX Titan graphic card installed on a computer with an Intel Core i7-4770K processor and 16 Gb of RAM. The GPU we used has 2688 processing cores working at a base clock of 837 MHz.

Despite the available computational power, we had to reduce the spatial resolution of the silicon retinas in order to achieve acceptable real-time performance

for the control of the movements. First, accumulated the events coming from the two retinas with a decay time constant of 20 ms without distinguishing between their polarities (ON/OFF). Second, we summed up the accumulated events in contiguous regions of 14x14 pixels and we applied a binary threshold. The final input image to *Nengo* is a 9x9 binary matrix of visual activation.

In this work we integrated five models of eye movements in a single neural simulation. We considered three gaze-shifting movements (i.e., saccade, smooth pursuit and vergence), one gaze-stabilizing movement (i.e., the vestibular-ocular reflex) and one fixational movement (i.e., microsaccade). We implemented microsaccadic and saccadic motion with different models because of their different functional roles. However, it is still not clear if these models correspond to two distinct biological neural control modules [16].

3.1 Eye Movements Integration

A key feature of our neurocontroller is the coordination of several eye movements depending on external and internal contexts. The occurrence of a specific context is determined based on sensory streams from the silicon retinas and from the IMU that serves as a vestibular system. Figure 2A illustrates how information from these sensors flows through different neural modules to generate motor commands for the servomotors. Three main information processes run in parallel. One of them receives visual information from the two silicon retinas in order to control the vergence of the eyes. A second process receives information only from the IMU (*gyroscope*) to compensate for head movements. A third process is dedicated to the selection of one of three mutually exclusive eye movements (i.e., saccades, microsaccades and smooth pursuit).

The action selection working mechanism is further detailed in Figure 2B where excitatory connections are shown in red and inhibitory ones in black. Three neural modules (*saccades*, *microsaccades* and *smooth pursuit*) receive information about the level of retinal activity, the detection of moving objects in the scene and the motion of the eyes in order to determine how likely the activation of the corresponding eye movement is given the current context. More precisely, a saccadic movement is likely to occur if there is high activity from the retinas but no detection of motion and the eyes are not moving. Microsaccades are activated when the activity from the retinas is below a certain threshold and smooth pursuit can occur if a moving object is detected and the eyes are moving. From a biological perspective, the boundaries of a context are not precisely defined. As a consequence, it is necessary to use a neural mechanism that can decide at any given time which context is the most likely to occur. In the human brain a plausible candidate to perform action selection is the basal ganglia (BG) [17,18]. In our neurocontroller we used the *Nengo* implementation of BG developed by Eliasmith and collaborators [9], which is based on the work of Gurney et al. [19] in order to decide the current context. The BG model, which consists of five neural sub-modules receives excitatory inputs from the context modules. BG computes a *winner-takes-all* function determining the strongest input and selects one of three possible eye movements.

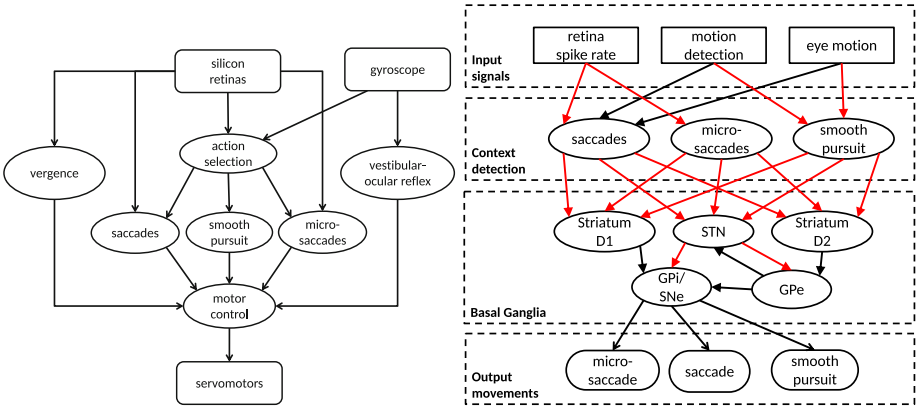

Fig. 2. Neural processing diagrams. Left: the arrows show the information flow between hardware components (rectangular boxes) and functional neural modules (rounded boxes). Right: the BG model receives information about the likelihood of different contexts and select the action associated with the strongest activation.

3.2 Saccades

Saccades are movements of the eye that direct the fovea to a new point of interest in the visual field. Saccades are ballistic movements, that is the final target position cannot change during the execution of the motion. In case the target is missed, smaller corrective saccadic movements are possible, but only after the current saccade is concluded. The velocity of a saccade increases with the amplitude of the movement and can reach up to 900 degrees per second. In our controller we determine the next fixation target on the basis of a bottom-up saliency map model [20], computed at the lowest level on retinal windows of 3x3 pixels. The level of saliency of such a window depends on the detection of basic geometric primitives like points, edges or corners. The neural mechanism that implements this function is based on a standard model of V1 simple cells [21,22]. A generic simple cell receives a strong excitatory connection from the photoreceptor in the center of its receptive field and weaker inhibitory connections from the photoreceptors in the surrounding periphery. The outputs of contiguous simple cells are then combined to estimate the saliency of larger areas of the visual field. In addition, we introduced a global saliency map with a memory effect. This prevents the saliency map from generating fixations lasting too long and also the return of the gaze to previous locations (i.e., inhibition of return). Memory is implemented with recurrent inhibitory connections with a time constant of 50 ms. The saliency map of the current visual field is inhibited by the corresponding area of the global saliency map.

3.3 Smooth Pursuit

Smooth pursuit allows an observer to track a moving object by constantly keeping its image centered in the fovea. This movement is voluntary but the presence of a moving stimulus in the visual field is necessary for untrained people in order to accomplish this motion. Our implementation of smooth pursuit is based on the detection of motion due to direction selective (DS) ganglion cells. These cells are modeled as Elaborated Reichardt Detectors [23]. The presence of the delay in only one of the two input connections allows for the coincident arrival of two excitatory stimuli to the DS ganglion cell only if the upstream photoreceptors are activated in a specific temporal order. In order to solve the motion ambiguity inherent in a local motion detector (known as the aperture problem), the outputs of multiple DS ganglion cells sensitive to different directions of motion at different speeds are combined. The motion detected in the fovea of the silicon retinas is then used to estimate the speed of the eyes required to keep tracking the moving object.

3.4 Vergence

Vergence is a coordinated movement of both eyes in opposite directions in order to converge on a common fixation point. A biologically-inspired technical solution to allow vergence is to compute matching landmarks in stereo images [24]; a biologically plausible way to determine the right velocities of the eyes is to compute a disparity map [25]. Following the model described in Patel et al. [26] we implemented a population of neurons for each specific degree of disparity. Each population receives excitatory connections from the right retinal photoreceptors and inhibitory connections from the left retinal photoreceptors. A *winner-takes-all* mechanism selects the population corresponding to the most likely disparity. Depending on the estimated disparity d, the eyes converge (if $d > 0$) or diverge (if $d < 0$).

3.5 Vestibular-Ocular Reflex (VOR)

VOR is a very important mechanism to stabilize vision based only on the sensory information provided by the vestibular system. During either the translation or rotation of the head, this reflex induces the eyes to move in the opposite direction in order to compensate for the motion of the head. In our neural simulation four vestibular nuclei act as integrators of acceleration signals coming from the vestibular system. Two of the nuclei control the angular velocity of the eyes on the horizontal plane and the other two in the vertical plane.

3.6 Microsaccades

Microsaccades are small involuntary fixational eye movements. Their functional role is still a matter of debate. However, a common hypothesis assumes that they prevent the retinal image from fading in presence of a perfectly static visual

scene [27]. Because silicon retinas behave similarly to human retinas in this regard, our neurocontroller triggers microsaccadic movements every time the stream of retinal events decreases below a certain threshold as occurs in the absence of movement. Moreover, it is unclear how microsaccades are generated [28]. A possibility is that they are generated from the noise in the Superior Colliculus (SC) [29]. In our implementation we took advantage of spiking rate fluctuations in the LIF neurons with constant input in order to generate noise sources. We then connected two of these sources with the neural populations that control the horizontal and vertical eye movements.

4 Validation

In order to validate the similarity between our neurocontroller-generated behavior and human behavior, we compared the movements of the robotic eyes to those generated by humans when exposed to the same visual input. We placed a monitor in front of the robotic head and tracked the saccadic movements while showing a video ($2'29''$ of duration) selected from the *DIEM* database [30]. The video is part of a documentary about monkeys living in the tropical rainforest jungle and was selected amongst others because of the presence of distinct moving elements in the scene such as monkeys jumping from one tree to another.

For a very preliminary assessment of the accuracy of our neurocontroller in reproducing human behavior we directly compared the eye traces on the screen recorded from the eyes of the robot and from those of a randomly chosen human subject as shown in Figure 3. On the one hand, similar to the human counterpart, each of the eyes of the robotic head seems to follow a random trajectory that explores a large area of the screen close to the center. In addition, microsaccadic and saccadic movements alternate in a similar way to human eye motion. On the other hand, the trajectories of the robotic eyes do not converge on close fixation points as observed in human behavior. This is likely due to the blur of the image

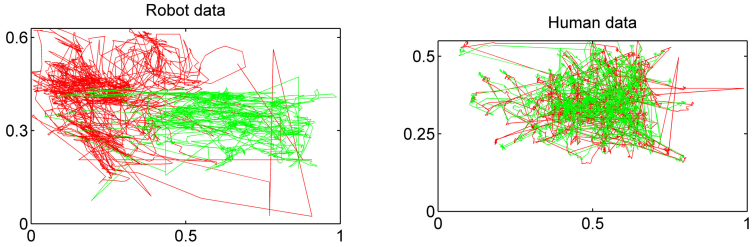

Fig. 3. Saccadic traces on the monitor screen for the left (red) and right (green) eye. The monitor screen size is normalized in reference to its width. The trajectory patterns for each robotic eye resemble the corresponding human one, but the traces do not overlap due to an insufficient convergence of the eyes.

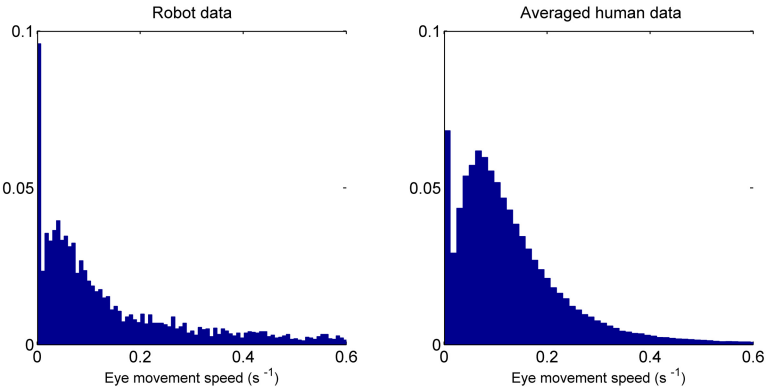

Fig. 4. Velocity distribution of eye motion. The distribution of velocities of robotic (left) and human (right) saccades and microsaccades is similar except for a scale factor.

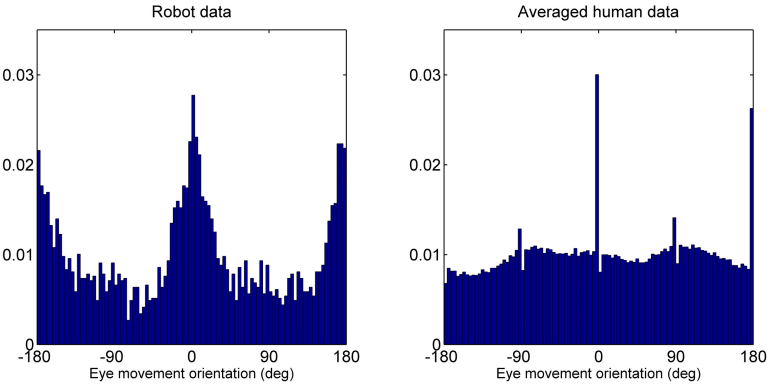

Fig. 5. Orientation distribution of eye motion. The eye movements of the robot head are biased towards horizontal directions in contrast to the behavior of human eyes.

captured by the silicon retinas that is responsible for the lack of a clear common point of fixation.

Figure 4 shows a comparison between the distribution of velocities of robotic and human microsaccades and saccades. To more accurately estimate the velocity distribution, we analyzed data from the DIEM database recorded from 50 human subjects watching the same selected video. As it can be seen, the profiles of velocities are similar except for a scale factor. However, if we consider the whole range of velocities, that is including saccadic motion as well, the average and standard deviation of the distribution match the human data particularly well (0.32 ± 0.80 for the robot head, 0.36 ± 0.99 for humans).

Figure 5 shows a comparison between the distribution of orientations of robotic and human eye movements (averaged across 50 subjects). Contrarily

to velocity distributions, orientation distributions show a clear difference in the preferred directions of motions. In fact, the control of the eyes in our robotic setup seems to be biased towards horizontal movements.

5 Conclusions

We hypothesized it is possible to generate high level human-like behavior by coordinating several low level models of eye movements. Based on this assumption, we implemented a neurocontroller that integrates different biologically plausible models of eye movements. The integration crucially relies on a model of basal ganglia, which selects the most appropriate action depending on internal and external contexts. Because it implements a *winner-takes-all* strategy, this model of BG can resolve situations where different competing contexts occur at the same time.

Unfortunately, the performance of our neurocontroller is still far from accurately matching human behavior. One of the major limitations of our current implementation is the severe down-sampling of the silicon retina resolution necessary to make the system operate in real-time. The exploitation of the full resolution would require a number of neurons in *Nengo* that is far beyond the capabilities of our computing hardware. The most important limiting factor for our neural simulations was memory rather than the number of processing cores. A possible solution to this technical limitation is to further preprocess visual information to input in Nengo.

Overall, our neurocontroller still lacks the ability to efficiently explore a visual scene like a human. Besides technical enhancements, which would allow for a richer stream of visual information and more powerful computational neural modules, performance would likely benefit from a more careful tuning of neural model parameters. In order to achieve this, more comprehensive and accurate behavioral tests need to be done. Once our robotic system will reproduce in a more accurate way human-like behavior it could be used to study the importance of active motion for object recognition tasks. In particular, the possibility to disable specific neural modules in our neurocontroller might be helpful for validating current models of visual perception based on lesion studies.

References

1. Rothkopf, C.A., Ballard, D.H., Hayhoe, M.M.: Task and context determine where you look. J Vis. Dec. 19 **7**(14), 16.1–20 (2007). doi:10.1167/7.14.16
2. Asuni, G., Teti, G., Laschi, C., Guglielmelli, E., Dario, P.: A robotic head neurocontroller based on biologically-inspired neural models. In: Proceedings of the 2005 IEEE International Conference on Robotics and Automation, ICRA 2005, pp. 2362–2367, 18–22 (2005)
3. He, H., Ge, S.S., Zhang, Z.: A saliency-driven robotic head with bio-inspired saccadic behaviors for social robotics. Autonomous Robots **36**(3), 225–240 (2013). doi:10.1007/s10514-013-9346-z

4. Shibata T.: An overview of human interactive robots for psychological enrichment. In: Proceedings of the IEEE, vol. 91, no. 11 (2004)
5. Berthouze, L., Kuniyoshi, Y.: Emergence and Categorization of Coordinated Visual Behavior Through Embodied Interaction. Machine Learning **31**, 187–200 (1998)
6. Kuniyoshi, Y., Kita, N., Sugimoto, K., Nakamura, S., Suehiro, T.: A foveated wide angle lens for active vision. In: Proceedings 1995 IEEE International Conference on Robotics and Automation, vol. 3, pp. 2982–2988, 21–27 (1995)
7. Kohonen, T.: Self-Organized Formation of Topologically Correct Feature Maps. Biological Cybernetics **49**(1), 59–69 (1982)
8. Bjorkman, M., Kragic, D.: Combination of foveal and peripheral vision for object recognition and pose estimation. In: Proceedings 2004 IEEE International Conference on Robotics and Automation, ICRA 2004, vol. 5, pp. 5135–5140 (2004)
9. Eliasmith, C., Stewart, T.C., Choo, X., Bekolay, T., DeWolf, T., Tang, Y., Rasmussen, D.: A Large-Scale Model of the Functioning Brain. Science **338**(6111), 1202–1205 (2012)
10. Dodgson N.A.: Variation and extrema of human interpupillary distance. Electronic Imaging 36–46 (2004)
11. Robotis Inc.: http://www.robotis.com
12. Lichtsteiner, P., Posch, C., Delbruck, T.: A 128×128 120 dB 15 μs Latency Asynchronous Temporal Contrast Vision Sensor. IEEE Journal of Solid-State Circuits **43**(2), 566–576 (2008)
13. Riggs, L.A., Ratliff, F.: The effects of counteracting the normal movements of the eye. Journal of the Optical Society of America **42**, 872–873 (1952)
14. Martinez-Conde, S., Macknik, S.L., Troncoso, X.G., Dyar, T.A.: Microsaccades counteract visual fading during fixation. Neuron. **49**(2), 297–305 (2006)
15. OptiTrak Tracking System: https://www.optitrack.com
16. Martinez-Conde, S., Otero-Millan, J., Macknik, S.L.: The impact of microsaccades on vision: towards a unified theory of saccadic function. Nature Reviews Neuroscience **14**, 83–96 (2013)
17. Mink, J.: The basal ganglia: focused selection and inhibition of competing motor programs. Progress in Neurobiology **50**(4), 381–425 (1996)
18. Redgrave, P., Prescott, T., Gurney, K.N.: The basal ganglia: A vertebrate solution to the selection problem? Neuroscience **89**, 1009–1023 (1999)
19. Gurney, K., Prescott, T.J., Redgrave, P.: A computational model of action selection in the basal ganglia. II. Analysis and simulation of behaviour. Biological Cybernetics **84**(6), 411–423 (2001)
20. Itti, L., Koch, C.: Computational modelling of visual attention. Nature Reviews Neuroscience **2**, 194–203 (2001). doi:10.1038/35058500
21. Hubel D.H., Wiesel T.N.: Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. J. Physiol. (1962)
22. Li, Z.: A saliency map in primary visual cortex. Trends in cognitive sciences **6**(1), 9–16 (2002)
23. Van Santen, J.P.H., Sperling, G.: Elaborated Reichardt detectors. JOSA A **2**(2), 300 (1985)
24. Conradt, J., Simon, P., Pescatore, M., Verschure, P.F.M.J.: Saliency maps operating on stereo images detect landmarks and their distance. In: Dorronsoro, J.R. (ed.) ICANN 2002. LNCS, vol. 2415, pp. 795–800. Springer, Heidelberg (2002)
25. Franz A., Triesch J.: Emergence of disparity tuning during the development of vergence eye movements. In: IEEE 6th International Conference on Development and Learning, pp. 31–36, pp. 11–13 (2007) doi:10.1109/DEVLRN.2007.4354029

26. Patel, S.S., Ogmen, H., White, J.M., Jiang, B.C.: Neural network model of short-term horizontal disparity vergence dynamics. Vision Research **37**(10), 1383–1399 (1997)
27. Pritchard, R.M.: Stabilized images on the retina. Sci. Am. **204**, 72–78 (1961)
28. Rolfs, M., Kliegl, R., Engbert, R.: Toward a model of microsaccade generation: the case of microsaccadic inhibition. J. Vis. **8**(11), 5.1–23 (2008)
29. Krekelberg, B.: Microsaccades. Current Biology **21**(11), 416 (2011)
30. Mital, P.K., Smith, T.J., Hill, R., Henderson, J.M.: Clustering of Gaze during Dynamic Scene Viewing is Predicted by Motion. Cognitive Computation **3**(1), 5–24 (2011)