This paper will appear in:
Proc. of the IEEE 2004 International Conference on Robotics and Automation (ICRA 2004), April 26-May 1, 2004, New Orleans, LA, USA,

Design of an Anthropomorphic Robot Head for Studying Autonomous Development and Learning

Hyundo Kim*, George York†, Greg Burton*, Erik Murphy-Chutorian*, and Jochen Triesch*,
* Natural Computation Group, Dept. of Cognitive Science, UC San Diego, La Jolla, California 92093-0515,
Emails: {h35kim,erikmc,triesch}@ucsd.edu
†YFX Studios P.O. BOX 292515, Sacramento, CA 95829, Email: contact@yfxstudio.com

***Abstract*— We describe the design of an anthropomorphic robot head intended as a research platform for studying autonomously learning active vision systems. The robot head closely mimics the major degrees of freedom of the human neck/eye apparatus and allows a number of facial expressions. We show that our robot head can shift its direction of gaze at speeds which come close to that of human saccades. Since our design only makes use of low cost consumer grade components, it paves the way for widespread use of anthropomorphic robot heads in science, education, health-care, and entertainment.**

I. Introduction

There are two major reasons why the study of biologically inspired or biomimetic robots has been receiving a lot of attention recently [5], [7], [21]. First, biological systems routinely manage to solve difficult problems whose solution is beyond the current state-of-the-art in artificial intelligence and robotics, and researchers try to build better robots by mimicking the solutions that nature found — to the extent that we understand them, e.g. [9]. Second, biologically inspired robots can serve as models for biological systems and help to better understand why and how biological systems function, i.e. it is of interest to biology and cognitive science, e.g. [14].

Within this general context, we are particularly interested in active vision [1], [4], autonomous learning and development [2], [13], [15], [22]. The human visual system is striking not only because it has solved the vision problem — especially compared to the current state of the art in computer vision — but also because it learns to do so with little or no explicit supervision. Several decades of experimental and theoretical work on infant vision and developmental neuroscience seem to suggest that infants *learn* how to see — and they learn to do so by themselves. This should be regarded as an existence proof for the hypothesis that vision can be learned autonomously. Thus, it should in principle be possible to construct a camera-equipped robotic learning system that, when switched on, will autonomously learn to see, i.e. to make sense of the patterns of light falling into its cameras. Of course, that is not to say that such a system can be a *tabula rasa* — the system will clearly need initial structure to facilitate learning, to focus learning on relevant things, and so forth. But what does this initial structure have to look like? What learning mechanisms will be needed? How will the system have to *interact* with the environment to uncover its statistical structure? In short, what exactly does it take to autonomously learn to see? Although rarely discussed, this question is arguably the most fundamental question in computer and human vision and its implications for autonomous robotics are profound. The long term goal of our lab is to advance our understanding of this question by building robotic vision systems that autonomously learn to see. These systems will be inspired by findings from developmental psychology and neuroscience and will serve as embodied computational models in these fields.

In the last two years we have been developing an anthropomorphic robot head with a stereo vision system as a platform for this line of research. The main goal of this paper is to describe our robotic head and its design trade-offs. The remainder of the paper is organized as follows. Section II describes our robot and the rationale for the design decisions made. Section III reports experiments to measure some performance metrics of our robot head. A first demonstration application is briefly described in Sec. IV. Section V discusses the work from a broader perspective and mentions some of our current research efforts on studying autonomous development of vision systems that utilize the robot head. Finally, Sec. VI concludes the paper.

II. System Description

A. Design Considerations

Our major design criteria were the following. First, the robot head should be of realistic human size and shape while modeling the major degrees of freedom (DoFs) found in the human neck/eye system, incorporating the redundancy between the neck and eye DoFs. Second, the "eyes" of the robot head were to be fitted with small CCD cameras with a IEEE1394 interface. Third, the robot head should have a small number of DoFs for facial expressions to be used in studies of learning in social settings. Fourth, the design should use low-cost off-the-shelf hardware components to keep costs to a minimum, while keeping the ability to approximate the dynamics of the human neck/eye system. In particular, we wanted the robot to be capable of *saccades*, very fast movements of the eyes to rapidly change the direction of gaze.

The initial design of the head was done in collaboration with YFX studios (http://www.yfxstudios.com), a Sacramento based special effects company, which specializes in animatronics. Since then, we have made a number of modifications to

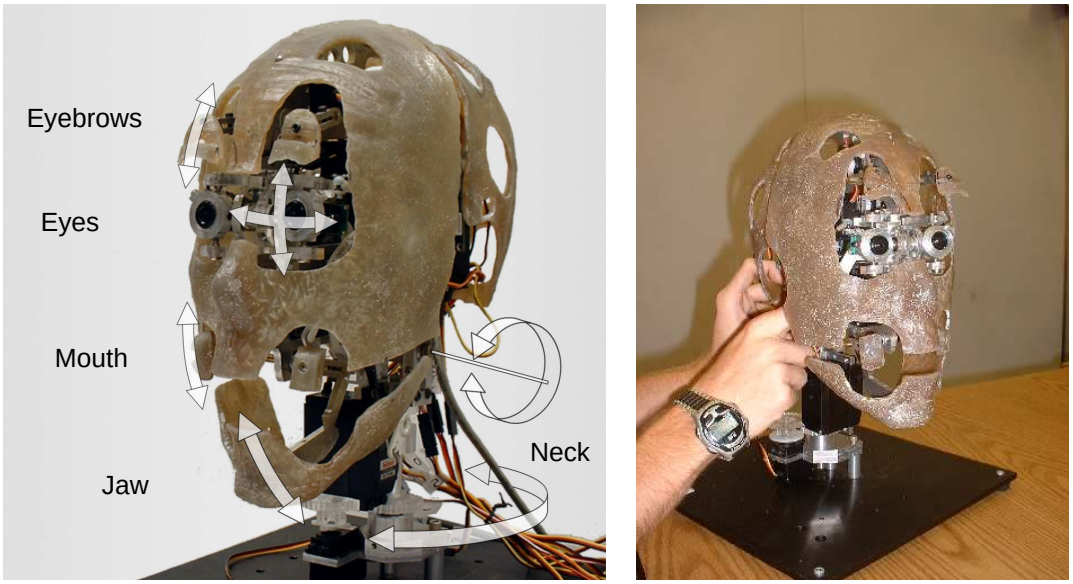

Fig. 1. Illustration of the degrees of freedom of the head and its overall appearance. There is a total of nine degrees of freedom, six for the neck/eye system and three for facial expressions. See text for details.

TABLE I
DIMENSIONS OF THE HEAD.

Height	24.1 cm
Width	13.2 cm
Depth	20.6 cm
Circumference	62.0 cm
Weight w/ base plate	2.72 kg
Weight w/o base plate	1.81 kg

increase the range and stability of several DoFs, but the basic design is essentially the same. Figure 1 gives an overview of the degrees of freedom of the head and its appearance. The backbone of the system is formed by an aluminum base plate in saggital orientation, on which all other mechansism are mounted. Lightweight materials including aluminum, fiber glass, and acrylics are used throughout. Measures of its size and weight are given in Table I.

B. Motor System

There are a total number of nine DoFs. The neck can pan left-right and tilt up-down. Similarly, both eyes can independently pan and tilt, introducing redundancy as in the human system. In humans, each eye can also rotate around the line of sight, which is used to compensate for the head tilting left-right. Since our head does not have the ability to tilt the head left-right, we chose to omit this DoF in the eyes. In addition, an in-plane rotation of the images provided by the robot's cameras could be easily done in software.

The remaining three DoFs are for facial expressions. The robot can open and close the mouth, raise and lower the corners of the mouth to crudely mimic expressions of smiling of frowning, and it can raise and lower its "eyebrows", to express surprise. We plan to use these DoFs for facial expressions in future work studying learning in social settings.

All DoFs are actuated by standard hobby grade servo motors as used in radio controlled cars or air planes. Since some of the DoFs are actuated via four bar linkages, in particular those of the eyes, there is a non-linear relation between the rotation angle of the servo motor, and the rotation of the respective eye. While from an engineering stand point, this makes control somewhat more difficult, it provides an interesting test case for methods aiming to learn visuo-motor control completely autonomously, i.e. without a prior model of the kinematics, in which we are interested.

The servo motors are controlled through two Mini SSC II interface boards (Scott Edwards Electronics Inc.) via a 9600 bit/s serial connection from a PC. On the positive side, the Mini SSC II is a cheap, convenient, and robust solution. On the negative side it only provides 8 bit resolution over either 90° or 180°. For instance, the smallest horizontal eye movement we can make corresponds to a shift of about 3 pixels at the highest image resolution. This resolution is sufficient for our purposes. Since each movement command requires the transmission of 3 Bytes across the serial connection, up to 400 commands can be sent each second. Thus, we can send a motor command to each servo every 22.5 ms. An overview of the entire system is given in Fig. 3.

C. Vision System

The robot head has two digital color CCD cameras as eyes. We use the Point Grey Research Firefly cameras (http://www.ptgrey.com). Each camera can capture 24 bit color images through a 1/3" progressive scan CCD at up to 30 fps. Three modes of image resolution (160x120, 320x240, and 640x480) with different color formats (RGB, YUV444,

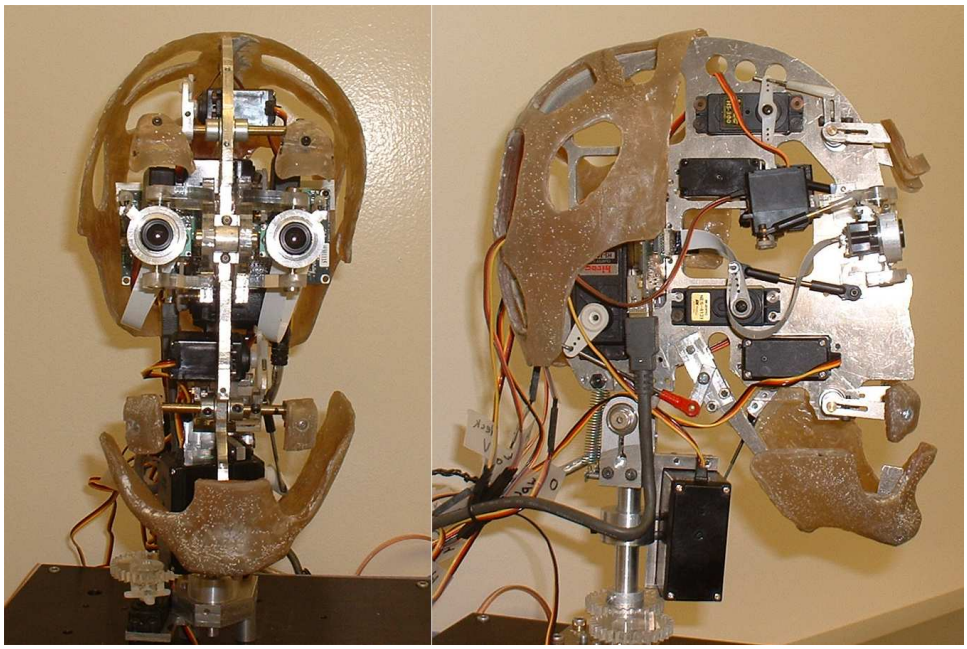

Fig. 2. Frontal and side view of the robot head. As in Fig 1 the face plate has been removed to allow better view of the internal structure. See text for details.

TABLE II
RANGE OF MOVEMENT AND RESOLUTION OF THE NECK AND EYE DEGREES OF FREEDOM.

	Range		Resolution	
	Hor.	Ver.	Hor.	Ver.
Left Eye	42.8°	35.3°	0.37°	0.28°
Right Eye	37.9°	33.4°	0.33°	0.27°
Neck	180°	40.2°	0.71°	0.17°

YUV422, and YUV411) are supported. The cameras use a IEEE1394 interface which is capable of delivering data rates up to 400 Mbps. Since the cameras are very light and are extended from the interface board by a flexible extension cable, they can be easily moved inside the head with fairly small, low-torque servos. The lenses are changeable with focal length of 2, 4, 6, and 8 mm. We can either use both cameras with the same focal length in order to perform stereo image processing, or we can use each camera with a different focal length lens such that one camera can model high resolution foveal vision with a narrow field of view while the other models peripheral vision by having a wide field of view with low resolution. Currently, we are using the former setup using 4 mm focal length lenses. These have a diagonal field of view of 100 degrees, which corresponds to approximately 80 degrees horizontally and 60 degrees vertically. This is a good compromise between high resolution and big field of view.

D. Computational System

Visual and motor systems are controlled by a desktop computer with Pentium 4, 3.06 GHz processor and 2 GB of RAM running Red Hat Linux 9. The interfaces (IEEE 1394 standard for the cameras and a standard serial interface for the servo motor control) ensure easy maintenance and upgrading. We have developed C++ class libraries for low level motor control and image acquisition. We use the Intel IPP and OpenCV libraries as a basis for developing computer vision algorithms. We are currently working to parallelize some of the computationally expensive visual processing. To this end, we are installing a Gigabit ethernet network between our lab's PCs.

E. Planned Hardware Extensions

Over the next year or so, we are planning to incorporate an auditory system to study the role of auditory cues in the development of attentional mechanisms and shared attention skills. In addition, we are interested in developing models of the learning of multi-modal integration strategies in infants.

Next to that, we are currently working on a simple low-cost 4 DoF arm. It will enable us to study the autonomous development of manipulation skills such as reaching and grasping. In addition, we are interested in imitation learning of grasping movements [19]. We also plan to study the rôle of active object manipulation for learning visual representations of objects.

III. PERFORMANCE EVALUATION

In order to measure the performance of the robot head, we conveniently used its own visual system as a measurement

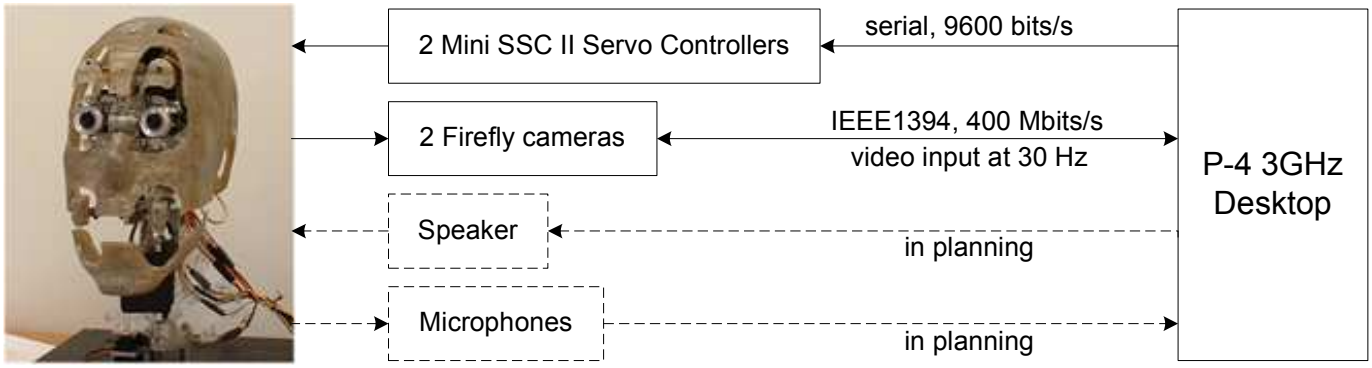

Fig. 3. System overview, dashed components are still in planning. See text for details.

device. The range of movement for the DoFs in the eyes was measured by moving the eyes across the entire range between their mechanical limits while observing how many pixels an object in the field of view would shift during the movement. The range of motion of the neck was measured directly with a goniometer. The results are given in Table II. Also given are the angular resolution of these DoFs. Since the four bar linkage actuation leads to a non-linear relation between the rotation of the servo shafts and the rotation of the cameras, these values are average resolutions obtained by dividing the angle of rotation of the cameras between their minimum and maximum position by the number of position increments (steps) between these positions.

Since our explicit design goal was to allow for rapid, saccade-like eye movements, we measured durations of gaze shifts and compared them to human saccades. Human saccade speeds are well approximated by a linear relation between saccade amplitude A and saccade duration $D = D_0 + d \cdot A$, where D_0 is an offset typically between 20–30 ms, and the slope d is of the order of 2–3 ms/deg [6]. This relation is fairly accurate for horizontal saccades across the midline with amplitdues of up to about 50–60 deg. Peak velocities are reported to exceed 500 deg/s for saccades of that amplitude.

In order to measure the durations of the robot's gaze shifts as a function of their size we again used the robot's visual system as a measurement tool. We recorded continuous, time stamped video while the robot made gaze shifts of varying amplitudes. By subsequently analyzing the recorded images from the moving camera we could determine the duration of the gaze shift with an accuracy of up to the inverse frame rate, which is 30 Hz. We varied the movement amplitudes in steps of either 20 or 40 ticks and repeated the measurements 5 times for each horizontal or vertical movement amplitude. Measurements for the left eye and the neck are plotted in Figs. 4 and 5. Errors bars indicate the standard deviation of the measurements and we fitted a line through the given data points using Matlab. Slope and y-axis intersects for the fitted line are given in Table III. As a comparison, the relation for human saccades is plotted as well, with lower dotted line denoting $d = 2$ ms/deg and $D_0 = 20$ ms and upper, $d = 3$ ms/deg and $D_0 = 30$ ms. The area between the two dotted lines represents the range of typical human saccade durations. We can see that although our robot's gaze shifts are still slower than human saccades, they are within a factor of two of typical saccade durations reported for humans.

TABLE III
LINEAR FIT OF SACCADE METRICS.

	Left Eye		Neck	
	Hor.	Ver.	Hor.	Ver.
Slope [ms/deg]	3.86	5.21	5.43	11.67
y-axis intersect [ms]	68.86	52.42	110.40	60.53

The jaw and eyebrows can move about one inch vertically, and the corners of the mouth can move about half an inch vertically. In most of the DoFs, we are not able to exploit the full 8 bits of resolution that the servo controller provides since the range of motion is limited by the internal structure of the robot head.

IV. A Simple Demonstration Application

As a first demonstration application for the robot head, we implemented a simple face finding and tracking application. In this application, the robot only uses a single camera. On each frame the robot runs a modified version of the real time face detection algorithm by Viola and Jones [20]. We use the open source implementation of this method in the OpenCV computer vision library (see http://sourceforge.net/projects/opencvlibrary/). The largest face detected in the scene is the target of the next gaze shift. Gaze shifts are achieved by a simple closed-loop feedback controller that utilizes both eye and neck DoFs. Figure 6 gives an example of successful face detection. A demonstration video can be viewed on our web pages at http://csclab.ucsd.edu.

V. Discussion

We have demonstrated an anthropomorphic robot head as a research platform for studying autonomous learning of visual skills. It mimics the major degrees of freedom of the redundant human neck/eye system and provides additional degrees of freedom for simple facial expressions that can be

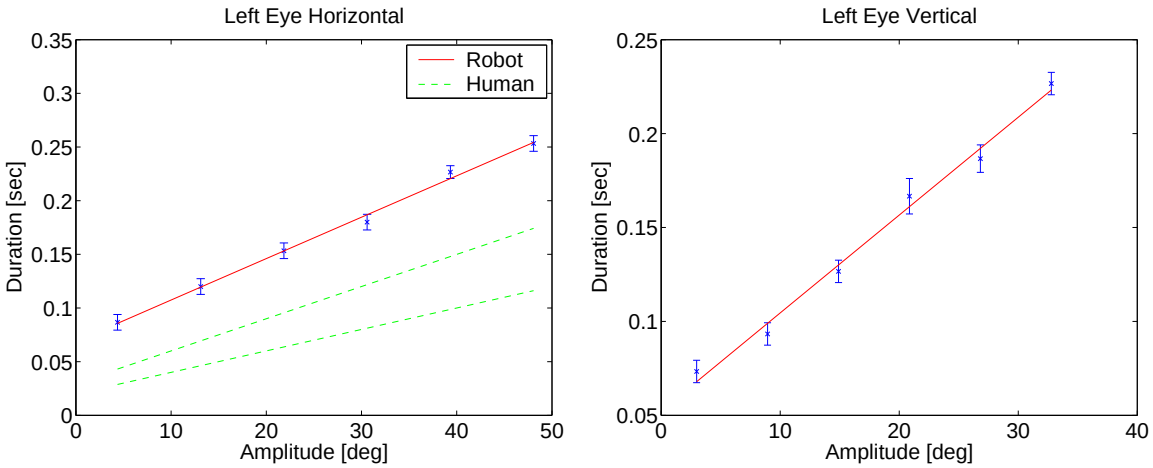

Fig. 4. Average time elapsed vs. amplitude of motion for horizontal (pan, left graph) and vertical (tilt, right graph) movements of the left camera. The relation is approximately linear and comes close to the metrics of human saccades where the dashed lines indicate the range of saccade durations typically observed in humans. Error bars represent the standard error of the mean.

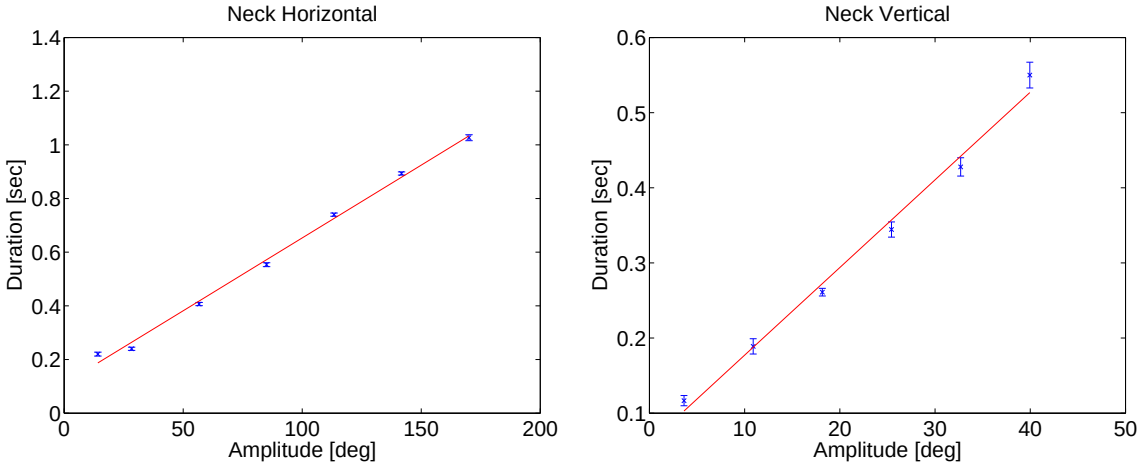

Fig. 5. Average time elapsed vs. amplitude of motion for horizontal (pan, left graph) and vertical (tilt, right graph) movements of the neck. Error bars represent the standard error of the mean.

used for learning in social settings. Despite using consumer grade, low-cost actuators and control hardware, we managed to demonstrate fast, saccade like eye movements with good resolution. While our main intent is to use the robot to study models of autonomous learning of visual skills in infancy, the design may also have potential applications in the development of more natural and intuitive computer interfaces.

Over the last couple of years, a variety of anthropomorphic robot heads have been built whose designs were driven by very different goals. At one end of the spectrum, roboticists have been trying to build robotic heads that look as similar to the human head as possible. Such designs typically include a more ore less realistically looking skin and the explicit goal is often to explore the boundaries of Masahiro Mori's famous uncanny valley [16] or to overcome it altogether. Maybe the most advanced example to date is David Hanson's K-Bot. Other researchers have focused on trying to make robots more sociable by endowing them with the ability to produce and respond to facial cues, without necessarily striving for a high degree of visual realism [8]. Another popular approach is to build robot heads that will function as embodied models of human active vision and learning, i.e. the goal is to use the robot as a tool for testing the feasability of theories about brain function [2], [3], [13], [14]. In this case, a high degree of realism of the underlying kinematics and dynamics is often desired. Realistic appearance may not be necessary, but can be important if models of learning in social settings are to be developed. The design of our robot head was driven by this last set of goals.

We plan to use our robot head as a research platform for several projects. Current work focuses on the completely autonomous learning of visuo-motor control strategies, i.e. learning controllers for accurate saccades [23], smooth pursuit, and vergence movements, in a redundant neck/eye system [17]. Beyond, this we will develop an active tracking and segmentation system based on the work described in [11],

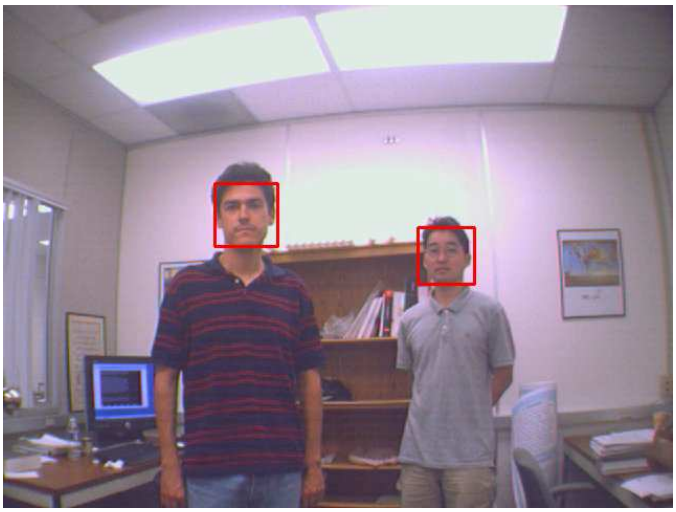

Fig. 6. Example of successful face detection with the Viola and Jones method. The 2 boxes near the center of the image are the face candidates found by the face detector. The robot makes gaze shifts to the largest face detected.

[12], [18]. Our vision is the construction of a system that seeks out and tracks moving objects and gradually learns and refines representations of these objects in a completely unsupervised fashion. Beyond this, we will also use the robot head to model the development of early social interactions skills. In particular, we are working on a project aiming to better understand the emergence of gaze following behaviors in late infancy [10] by building embodied models of the hypothesized learning processes underlying the development.

VI. Conclusion

Over the last few years a number of sophisticated anthropomorphic robots have been developed — for a variety of different purposes. Our goal was the creation of a flexible and easy to maintain humanoid robot head that can serve as a research platform for studying autonomous learning in active vision systems. In this paper we have demonstrated an anthropomorphic robot head with a redundant neck/eye system that closely mimics the major degrees of freedom of the human neck/eye apparatus. It is capable of fast saccade-like eye movements. Importantly, the robot head is built from consumer grade components, paving the way for widespread application of such technology in research, education, entertainment, and other domains.

Acknowledgments

The authors would like to thank the National Science Foundation for support under grant NSF IIS-0208451. We would like to thank Nathan Delson for continuing support of the project. We also thank Boris Lau for providing the picture of the robot head in Figure 1 (left).

References

[1] J. Aloimonos, I. Weiss, and A. Bandyopadhyay. Active vision. *Int. J. of Computer Vision*, 2:333–356, 1988.

[2] M. Asada, K. F. MacDorman, H. Ishiguro, and Y. Kuniyoshi. Cognitive developmental robotics as a new paradigm for the desgin of humanoid robots. *Robotics and Autonomous Systems*, 37:185–193, 2001.

[3] C.G. Atkeson, J. Hale, M. Kawato, S. Kotosaka, F. Pollick, M. Riley, S. Schaal, S. Shibata, G. Tevatia, and A. Ude. Using humanoid robots to study human behaviour. *IEEE Intelligent Systems*, 15(4):46–56, 2000.

[4] D. H. Ballard. Animate vision. *Artificial Intelligence*, 48:57–86, 1991.

[5] Y. Bar-Cohen and C. Breazeal. *Biologically Inspired Intelligent Robots*. SPIE, 2003.

[6] W. Becker. Saccades. In R. H. S. Carpenter, editor, *Vision & Visual Dysfunction Vol 8: Eye Movements*, pages 95–137. CRC Press, 1991.

[7] R. D. Beer, H. J. Chiel, R. D. Quinn, and R. E. Ritzmann. Biorobotic approaches to the study of motor systems. *Current Opinion in Neurobiology*, 8:777–782, 1998.

[8] C. Breazeal and B. Scassellati. Infant-like social interactions between a robot and a human caretaker. *Adaptive Behvior*, 8(1), 2000.

[9] R. Brooks, C. Breazeal, I. Robert, C. C. Kemp, M. Marjanovic, B. Scassellati, and M. Williamson. Alternate essences of intelligence. In *AAAI-98*, 1998.

[10] E. Carlson and J. Triesch. A computational model of the emergence of gaze following. 8th Neural Compuation and Psycholgy Workshop (NCPW8), Canterbury, UK, in press, 2003.

[11] E. Hayman and J.-O. Eklundh. Probabilistic and voting approaches to cue integration for fi gure-ground segmentation. In A. Heyden, G. Sparr, M. Nielsen, and P. Johansen, editors, *Proc. of ECCV 2002, 7th European Conference on Computer Vision*. Springer, 2002.

[12] H. Kim, B. Lau, and J. Triesch. Adaptive object tracking with an anthropomorphic robot head. In *Proc. of the 8th Int. Conf. on the Simulation of Adaptive Behaviors (SAB'04)*, 2004.

[13] H. Kozima. Infanoid: An experimental tool for developmental psychorobotics. In *Proc. Intl. Workshop on Developmental Study*, 2000.

[14] J. L. Krichmar and G. M. Edelman. Machine psychology: Autonomous behavior, perceptual categorization and conditioning in a brain-based device. *Cerebral Cortex*, 12:818–830, 2002.

[15] G. Metta, F. Panerai, R. Manzotti, and G. Sandini. Babybot: an artifi cial developing robotic agent. In *SAB 2000 Paris, France. Sep. 11-16*, 2000.

[16] Masahiro Mori. *The Buddha in the Robot*. Charles E. Tuttle Co., 1982. ISBN: 4333010020.

[17] T. Shibata, S. Vijayakumar, J. Conradt, and S. Schaal. Biomimetic oculomotor control. *Adaptive Behavior*, 9:189–208, 2001.

[18] J. Triesch and C. v.d. Malsburg. Democratic integration: Self-organized integration of adaptive cues. *Neural Computation*, 13(9):2049–2074, 2001.

[19] J. Triesch, J. Wieghardt, C. v.d. Malsburg, and E. Maël. Towards imitation learning of grasping movements by an autonomous robot. *Lecture Notes in Artificial Intelligence*, 1739:73–84, 1999.

[20] P. Viola and M. Jones. Robust real-time object detection. *International Journal of Computer Vision*, 2001.

[21] B. Webb. Can robots make good models of biological behaviour? *Behavioral and Brain Sciences*, 24(6), 2001.

[22] J. Weng, J. McClelland, A. Pentland, O. Sporns, I. Stockman, M. Sur, and E. Thelen. Autonomous mental development in robots and animals. *Science*, 291(5504):599–600, 2001.

[23] E. Wiewiora, D. Berg, J. Triesch, and T. Hashiyama. Learning optimal gaze decomposition. *Journal of Vision*, 3(9):437a, 2003. Abstract published at the Vision Science Society Meeting.