2008 8th IEEE-RAS International Conference on Humanoid Robots
December 1 ~ 3, 2008 / Daejeon, Korea



# The Karlsruhe Humanoid Head

Tamim Asfour [1], Kai Welke [1], Pedram Azad [1], Ales Ude [2], Rüdiger Dillmann [1]

[1] *University of Karlsruhe, Germany* {asfour, welke, azad, dillmann}@ira.uka.de

[2] *Jožef Stefan Institute, Slovenia* ales.ude@ijs.si

***Abstract*— The design and construction of truly humanoid robots that can perceive and interact with the environment depends significantly on their perception capabilities. In this paper we present the Karlsruhe Humanoid Head, which has been designed to be used both as part of our humanoid robots ARMAR-IIIa and ARMAR-IIIb and as a stand-alone robot head for studying various visual perception tasks in the context of object recognition and human-robot interaction. The head has seven degrees of freedom (DoF). The eyes have a common tilt and can pan independently. Each eye is equipped with two digital color cameras, one with a wide-angle lens for peripheral vision and one with a narrow-angle lens for foveal vision to allow simple visuo-motor behaviors. Among these are tracking and saccadic motions towards salient regions, as well as more complex visual tasks such as hand-eye coordination. We present the mechatronic design concept, the motor control system, the sensor system and the computational system. To demonstrate the capabilities of the head, we present accuracy test results, and the implementation of both open-loop and closed-loop control on the head.**

## I. Introduction

The design and construction of cognitive humanoid robots that can perceive and interact with the environment is an extremely challenging task, which significantly depends on their perceptive capabilities and the ability of extracting meaning from sensor data flows. Therefore, the perception system of such robots should provide sensorial input necessary to implement various visuomotor behaviors, e.g. smooth pursuit and saccadic eye-movements targeting salient regions, and more complex sensorimotor tasks such as hand-eye coordination, gesture identification, human motion perception and imitation learning. Our goal is the design and construction of a humanoid head that allows the realization of such behaviors and to study higher level development of cognitive skills in humanoid robots.

Most current humanoid robots have simplified eye-head systems with a small number of degrees of freedom (DoF). The heads of ASIMO [1], HRP-3 [2] and HOAP-2 [3] have two DoF and fixed eyes. However, the design of humanoid systems able to execute manipulation and grasping tasks, interact with humans, and learn from human observation requires sophisticated perception systems, which are able to fulfill the therewith associated requirements. Humanoid robots with human-like heads have been developed for emotional human-robot interaction ([4], [5]) and for studying cognitive processes ([6], [7], [8]).

The design of artificial visual systems which mimic the foveated structure is of utmost importance for the realization of such behaviors. However, current sensor technology does not allow to exactly mimic the features of the human visual system because camera systems that provide both peripheral and foveal vision from a single camera are still experimental. Therefore, several humanoid vision systems have been realized using two cameras in each eye, i.e. a narrow-angle foveal camera and a wide-angle camera for peripheral vision ([9], [10], [11], [12], [7]).

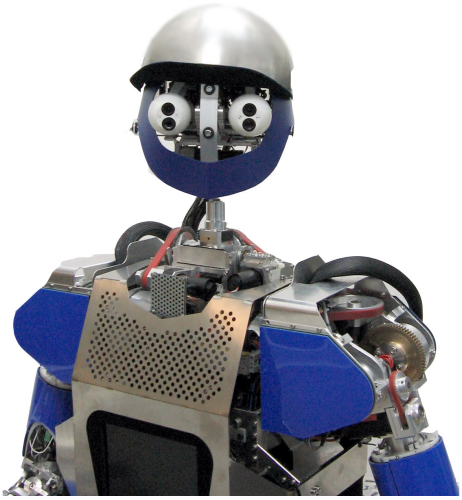

Fig. 1: The Karlsruhe humanoid head as part of the humanoid robot ARMAR-III. The head has two eyes and six microphones. Each eye has two camera

A retina-like vision sensor has been presented in [13] and a tendon driven robotic eye to emulate human saccadic and smooth pursuit movements has been presented in [14]. In [15], the biomimetic design of a humanoid head prototype with uncoupled eyes and vestibular sensors is presented.

In this paper, we present a new humanoid head with foveated vision (see Fig. 1), which has been developed as part of the humanoid robot ARMAR-III [16] and as a stand-alone vision system providing an experimental platform for the realization of interactive service tasks and cognitive vision research. In the next section we present the requirements for the development of the humanoid head. Section III provides the details about the motor, sensor and computation system of the head. The resulting accuracy tests and the realized head control strategies are presented in Section IV and V.

978-1-4244-2822-9/08/$25.00 ©2008 IEEE

## II. System Requirements

In designing the humanoid head, we paid special attention to the realization of foveation as several visual task e.g. object recognition, can benefit from foveated vision. Using two cameras in each eye, a humanoid robot will be able to bring the object into the center of the fovea based on information from the peripheral cameras. This is necessary because the area of interest, e. g. an object that is tracked by the robot, can easily be lost from the fovea due to its narrow field of view. It is much less likely that the object would be lost from the peripheral images, which have a wider field of view. On the other hand, operations such as grasping can benefit from high precision offered by foveal vision. The following design criteria were considered:

- 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 DoF.
- The robot head should feature human-like characteristics in motion and response, that is, the velocity of eye movements and the range of motion will be similar to the velocity and range of human eyes.
- The robot head must allow for saccadic motions, which are very fast eye movements allowing the robot to rapidly change the gaze direction, and smooth pursuit over a wide range of velocities.
- The optics should mimic the structure of the human eye, which has a higher resolution in the fovea.
- The vision system should mimic the human visual system while remaining easy to construct, easy to maintain and easy to control.
- The auditory system should allow acoustic localization in the 3D workspace.

With this set of requirements, we derive the mechatronical design of the humanoid head.

## III. Specification of the Head

### A. Head Kinematics

The neck-eye system in humans has a complex kinematics structure, which cannot be modeled as a simple kinematic chain due to the sliding characteristics of the articulations present in it. However, our goal is not to copy the anatomical and physiological details of the neck-eye system but rather to build a humanoid head that captures the essence and nature of human's head movements. The neck kinematics has been studied in human biomechanics and standard models of the human neck system have four DoF [17]. Each human eye is actuated by six muscles, which allows for movements around the three axis in space.

The kinematics of the developed head is shown in Fig. 2. The neck movements are realized by four DoF: Lower pitch, roll, yaw and upper pitch ($\theta_1, \ldots, \theta_4$), where the first three DoF intersect in one point. The vision system has three DoF $\theta_5, \theta_6$ and $\theta_7$, where both eyes share a common tilt axis ($\theta_5$) and each eye can independently rotate around a vertical

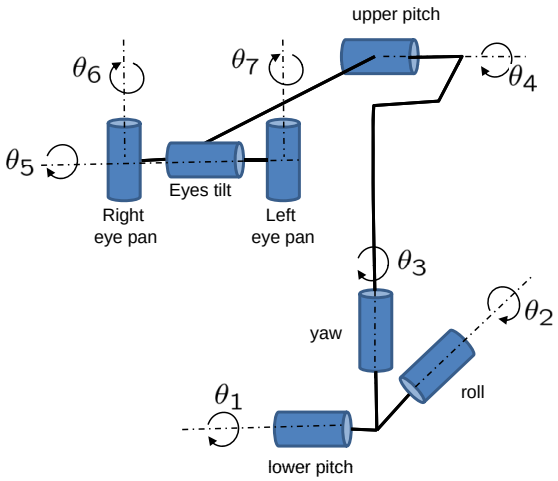


Fig. 2: The kinematics of the head with seven DoF arranged as lower pitch ($\theta_1$), roll ($\theta_2$), yaw ($\theta_3$), upper pitch ($\theta_4$), eyes tilt ($\theta_5$), right eye pan ($\theta_6$) and left eye pan ($\theta_7$).

axis ($\theta_6$ and $\theta_7$). These three DoF allow for human-like eye movements. Usually, human eyes can also rotate slightly about the direction of gaze. However, we decided to omit this DoF because the pan and tilt axes are sufficient to cover the visual space.

### B. Motor System

The head has seven DoF. Each eye can independently rotate around a vertical axis (pan DoF), and the two eyes share a horizontal axis (tilt DoF). All seven joints are driven by DC motors. For the pan joints we chose the brushless Faulhaber DC motor 1524-024 SR with backlash-free gear, IE2-512 encoder, 18/5 gear with 76:1 gear ratio, torque $2,5\,mNm$, and a weight of $70\,g$. For the tilt joint we chose the Harmonic Drive motor PMA-5A-50 with backlash-free gear, 50:1 gear ratio, and torque $0,47\,Nm$. For the four neck joints we chose brushless Faulhaber DC motors with IE2-512 encoders. The calculation of the actuators characteristics was based on the desired specifications and the moment of inertia, as well as the different weight of components, which were given by the CAD software.

### C. Sensor System

*1) Vision System:* To perform various visuo-motor behaviours it is useful to first identify regions that potentially contain objects of interest and secondly analyze these regions to build higher-level representations. While the first task is closely related to visual search and can benefit from a wide field of view, a narrower field of view resulting in higher-resolution images of objects is better suited for the second task. While the current technology does not allow us to exactly mimic the features of the human visual system and because camera systems that provide both peripheral and foveal vision from a single camera are still experimental, we decided for an

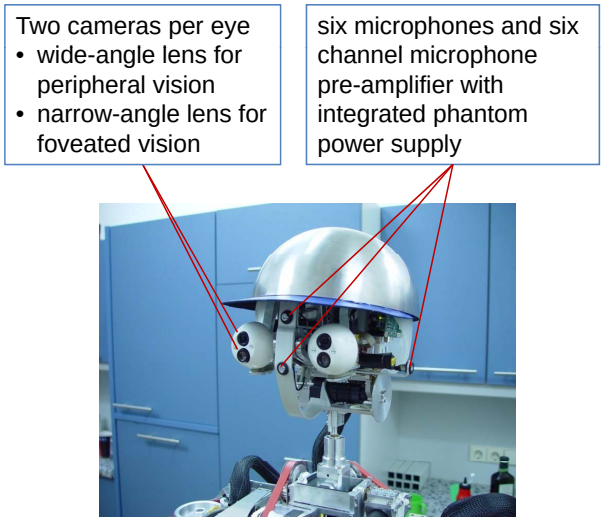


Fig. 3: The humanoid head with seven DoF arranged as lower pitch, roll, yaw, upper pitch, eyes tilt, left eye pan and right eye pan.

alternative that allows to use commercially available camera systems, which are less expensive and more reliable.

Therefore, foveated vision in our head is realized using two cameras per eye, one with a wide-angle lens for peripheral vision and one with a narrow-angle lens for foveal vision. We use the Point Grey Research Dragonfly2 IEEE-1394 camera in the extended version (`www.ptgrey.com`). The extended version allows the CCD to be located up to 6 inches away from the camera interface board. This arrangement helps with accessing hard to reach places, and with placing the lens into a small volume. Since the cameras are very light and are extended from the interface board by a flexible extension cable, they can be moved with small and low-torque servos.

The cameras can capture color images at a frame rate of up to $60\,Hz$. They implement the DCAM standard, and transmit a raw 8 bit Bayer Pattern with a resolution of $640\times480$, which is then converted on the PC to a 24 bit RGB image. The cameras have a FireWire interface, which is capable of delivering data rates of up to 400 Mbps. The benefit of transmitting the Bayer Pattern is that only a third of the bandwidth is needed for transmitting the color image without loosing any information. Thus, it is possible to run one camera pair at a frame rate of 30 Hz and the other at a frame rate of 15 Hz, all being synchronized over the same FireWire bus, without any additional hardware or software effort. Running the foveal cameras, which have a smaller focal length and thus a narrower view angle, at a lower frame rate is not a drawback because these cameras are not crucial for time critical applications such as tracking, but are utilized for detailed scene analysis, which does not need to be performed at full frame rate in most cases anyway.

The camera is delivered as a development kit with three micro lenses with the focal lengths $4, 6$, and $8\,mm$. In addition, one can use micro lenses with other focal lengths as well. We have chosen a $4\,mm$ micro lens for the peripheral cameras and a $12\,mm$ micro lens for the narrow angle cameras.

*2) Audio System:* The head is equipped with a six channel microphone system for 3D localization of acoustic events. As acoustic sensors, off-the-shelf miniature condensor microphones were selected. One microphone pair is placed at the ear locations in the frontal plane of the head. Another microphone pair is placed on the median plane of the head at the vertical level of the nose, one microphone on the face side and one microphone at the back of the head. The third microphone pair is mounted on the median plane but at the level of the forehead.

For each microphone a pre-amplifier with phantom power supply is required. These units are commercially not available in the required dimensions. Therefore, a miniature six channel condenser microphone preamplifier with integrated phantom power supply was developed as a single printed circuit board (PCB) with dimensions of only $70 \times 40\,mm$. The amplified microphone signal is conducted to a multi-channel sound card on the PC side. The acoustic sensor system proved high sensitivity for detecting acoustic events while providing a good signal to noise ratio. In preliminary experiments we successfully performed audio tracking of acoustic events.

*3) Inertial System:* Though the drives of the head kinematics are equipped with incremental encoders, we decided to add a gyroscope-based orientation and heading reference sensor. The sensor is an integrated attitude and heading reference system manufactured by the XSense company (`www.xsens.com`). It provides drift-free 6D orientation and acceleration measurement data and interfaces to a host PC (head control PC) via USB. The sensor will serve as a robot-equivalent sense of balance. It is especially useful for calibration and referencing of the head attitude and the detection of the body posture. In conjunction with the kinematics model and incremental encoder readings, partly redundant information about heading and orientation of the head is determined, which may further be used for diagnostics purposes. This is superior to the exclusive deployment of encoder readings as the kinematic model exposes uncertainty due to mechanical tolerances. Currently, the support of the attitude reference system in the head positioning control software is being implemented.

### D. Computational System

The head (visual and motor system) are controlled by three Universal Controller Module (UCoM) units for low-level motor control and sensory data acquisition: The UCoM is a DSP-FPGA-based device which communicates with the embedded PCs via CAN-Bus [18]. By using a combination of a DSP and a FPGA, a high flexibility is achieved. The DSP is dedicated for calculations and data processing, whereas the FPGA offers the flexibility and the hardware acceleration for special functionalities. One off-the-shelf PC104 with a Pentium 4 with 2 GHz processor and 2 GB of RAM running under Debian Linux, kernel 2.6.8 with the Real Time Application Interface RTAI/LXRT-Linux is used for motor

control. The PC is equipped with a dual FireWire card and a CAN bus card. The communication between the UCoMs and the PC104 system takes place via CAN bus. The basic control software is implemented in the Modular Controller Architecture framework MCA2 (`www.mca2.org`). Table I summarizes the motor, sensor and computational system of the humanoid head.

## IV. Head Accuracy

In order to prove the accuracy of the head, we evaluated the repeatability of joint positioning, which gives a good hint on the feasibility of the design and construction of the head. In contrast to tests on absolute accuracy, the knowledge of an approximated kinematic model of the head is sufficient for repeatability test.

In the following, the positioning accuracy of the left camera of the head was measured visually. Therefore, a calibration pattern was mounted in front of the calibrated camera. With the extrinsic camera parameters, the position of the calibration pattern was determined. Using an approximated model of the kinematics, the position could be transformed to each rotation axis and projected to the plane perpendicular to the axis. The angle between two projected positions describes the relative movement of the corresponding joint.

In the course of one test cycle, one joint was rotated to the positions 10° and -10° relative to the zero position. After each actuation, the joint returned to the zero position and the angle was measured as described above. For each joint this procedure was repeated 100 times.

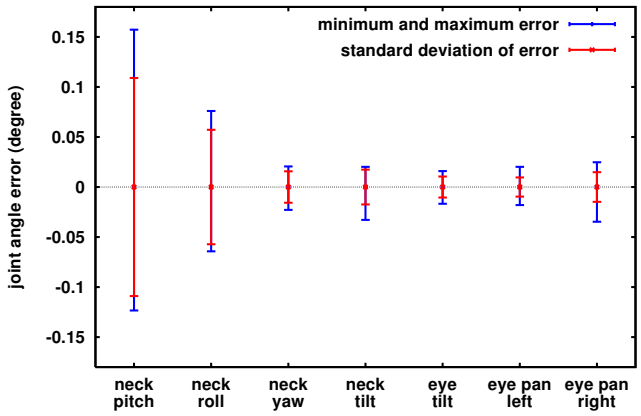


Fig. 4: Results of the head accuracy test for all seven joints of the head (from bottom to top). Each joint of the head was moved to the zero position starting from non-zero configurations 100 times. The position error of the joints was measured using computer vision methods. The plot illustrates standard deviation, minimum and maximum of the joint angle errors in degrees.

Fig. 4 illustrates the standard deviation of all angles as well as the minimum and maximum angle errors. The mean of all measured angles per joint was assigned to zero degree. The results show that the last five joints in the kinematic chain achieve an accuracy of about $\pm 0.025°$. The neck pitch and neck roll joints ($\theta_1$ and $\theta_2$ in Fig. 2) achieve an accuracy of about $\pm 0.13°$ and $\pm 0.075°$ respectively. The larger inaccuracy in these joints originates from dynamic effects caused in the gear belt driven joints by the weight of the head. The theoretical achievable accuracy can be derived from the number of encoder ticks which encode one degree of rotation for a joint. Using these values, the maximum accuracy lies between $0.0027°$ and $0.0066°$. The accuracy of the measurement process was measured with about $\pm 0.001°$.

## V. Head Control Strategies

Movements of the head and eyes are usually initiated by perceptual stimuli from the somatosensory, auditory or visual system. The goal of such movements consists of focusing the source of a stimuli with the camera system for further visual inspection. There are essentially two possible strategies to execute the required movements: closed-loop control and open-loop control. In closed-loop control, usually visual feedback is used in order to derive the position error of the eyes iteratively. In contrast, open-loop control does not depend on visual feedback but uses the kinematic model of the system to determine the desired posture. While closed-loop control can be applied to a wide range of problems concerning with foveation, there are cases where the necessary visual feedback cannot be provided, e.g. during the acquisition of unknown objects where the object model required for generation of visual feedback is unknown.

In the following sections we will present the implementations of both open-loop and closed-loop control strategies on the developed head.

### A. Open-loop control

Open-loop control only depends on the current state and the kinematic model of the system. In order to direct the gaze of the head-eye system to a specific position in Cartesian space, the joint positions for all involved joints can be derived by solving the inverse kinematics problem. With this in mind, the open-loop control strategy can be divided into two problems. First an accurate kinematic model for the involved joints has to be established, second the inverse kinematic problem has to be solved on base of the kinematic model. In the following we will describe solutions to both problems as implemented for the eye system of the Karlsruhe Humanoid Head.

The exact kinematic model of the head-eye system is not known because of inaccuracies in the construction process and because of the unknown pose of the optical sensors of the cameras in relation to the kinematic chain. In order to derive a more accurate kinematic model, a kinematic calibration process is performed. The classical formulation of the head-eye calibration problem (see [19], [20]) is extended with a model that prevents the introduction of methodical errors into the calibration process. For more details, the reader is referred to [21]. The procedure does not assume that the rotation axes of two joints intersect. Extrinsic camera calibration matrices

TABLE I: Overview on the motor, sensor and computational systems of the humanoid head.

| | |
|---|---|
| Kinematics | 3 DoF in the eyes arranged as common eyes tilt and independent eye pan. |
| | 4 DoF in the neck arranged as lower pitch, roll, yaw and upper pitch. |
| Actuator | DC motors and Harmonic Drives. |
| Vision system | Each eye is realized by two Point Grey Dragonfly2 color cameras in the extended version with a resolution of $640 \times 480$ at $60\,Hz$. (See `www.ptgrey.com`). |
| Auditory system | Six microphones (SONY ECMC115.CE7): two in the ears, tow in the front and two in the rear of the head. |
| Inertial system | Xsens MTIx gyroscope-based orientation sensor, which provides drift-free 3D orientation as well as 3D acceleration. (See `www.xsens.com`). |
| Universal Controller Module (UCoM) | Three UCoM units for motor control: The UCoM is a DSP-FPGA-based device, which communicates with the embedded PCs via CAN-Bus. By using a combination of a DSP and a FPGA, a high flexibility is achieved. The DSP is dedicated to calculations and data processing, whereas the FPGA offers the flexibility and hardware acceleration for special functionalities. |
| Control PC | Embedded PC with a dual FireWire card and a CAN card. Communication between the UCoMs and the PC104 system takes place via CAN bus. |
| Operation System | The embedded system is running under Linux, kernel 2.6.8 with Real Time Application Interface RTAI/LXRT-Linux (Debian distribution). |
| Control Software | The basic control software is implemented within the Modular Controller Architecture framework MCA (`www.mca2.org`). The control parts can be executed under Linux, RTAI/LXRT-Linux, Windows or Mac OS, and communicate beyond operating system borders. |
| | Graphical debugging tool (mcabrowser), which can be connected via TCP/IP to the MCA processes to visualize the connection structure of the control parts. |
| | Graphical User Interface (mcagui) with various input and output entities. |
| | Both tools (mcabrowser and mcagui) provide access to the interfaces and control parameters at runtime. |
| Integrating Vision Toolkit (IVT) [1] | Computer vision library, which allows to start the development of vision components within minimum time and provides support for the operating systems Windows, Linux, and Mac OS. The library contains a considerable amount of functions and features like the integration of various cameras, generic and integrated camera models and stereo camera systems, distortion correction and rectification, various filters and segmentation methods, efficient mathematical routines, especially for 3-D computations, stereo reconstruction, particle filter framework, platform-independent multithreading, convenient visualization of images and the integration of the library Qt for the development of Graphical User Interfaces. |

$C(\alpha_i)$ relative to a static calibration pattern are collected while the joint $j$ to be calibrated is moved to different positions $\alpha_i$. Fig. 5 illustrates the involved transformation matrices in the calibration. The transformation $F$ from the static joint coordinate system $X_{j0}$ to the world coordinate system $X_w$ remains constant over different actuations of the joint $\alpha_i$. The matrix $F$ can be rewritten using the extrinsic camera calibration $C(\alpha_i)$, the rotation of the joint to be calibrated $H_j(\alpha_i)$ and the desired calibration matrix $B$ in the following way:

$$F_i = C(\alpha_i)^{-1} B H_j(\alpha_i), \tag{1}$$

where $i$ denotes the index of the extrinsic calibration data. The calibration matrix $B$ is calculated using a non-linear least squares optimization approach using the difference of two matrices $F_i$ and $F_k$ which belong to two different sets of extrinsic calibrations as the target function for optimization:

$$min \sum_{k=1}^{N-1} ||F_i - F_k|| \tag{2}$$

The calibration procedure has been applied to the eye tilt, left eye pan and right eye pan joints.

In order to direct the gaze of the eye system, the optical axes of the respective cameras have to be aimed at a given point $\vec{x}$ in Cartesian space. For this purpose, the kinematic model resulting from the calibration process is extended with a virtual prismatic joint which is attached to the optical center of the cameras and which slides along the optical axis. Therefore, the movement of each camera can be described with the three-dimensional joint space vector $\vec{\theta} = (\theta_{tilt}, \theta_{pan}, \theta_{virt})^T$, which corresponds to a rotation around the eye tilt and around the eye pan axes and a translation along the optical axis. For each camera, the joint velocities that move the gaze toward the point $\vec{x}$ are calculated using the inverse reduced Jacobian:

$$\begin{bmatrix} \dot{\theta}_{tilt} \\ \dot{\theta}_{pan} \\ \dot{\theta}_{virt} \end{bmatrix} = J_r^{-1}(\vec{\theta}) \begin{bmatrix} \dot{x} \\ \dot{y} \\ \dot{z} \end{bmatrix} \tag{3}$$

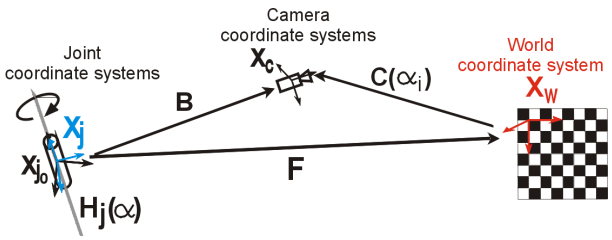


Fig. 5: Coordinate systems and transformations required in the kinematic calibration process.

[1] `ivt.sourceforge.com`

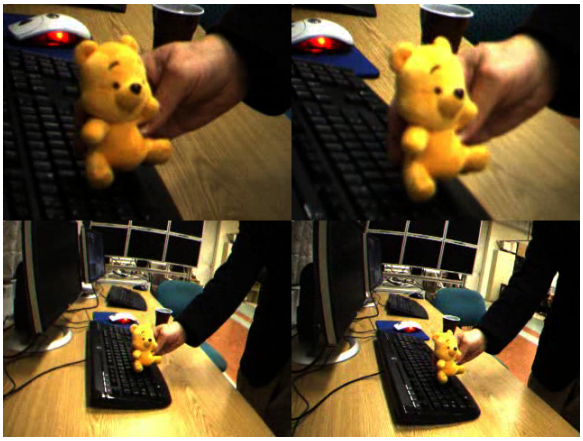

Fig. 6: Simultaneous stereo view from peripheral (below) and foveal cameras (above)

The reduced Jacobian $J_r$ is derived from the kinematic model using the geometrical method proposed by Orin et al. [22]. Since the Jacobian is a regular matrix in $\mathbb{R}^{3\times3}$, the inverse $J_r^{-1}$ always exists. The joint error $\Delta\vec{\theta}$ is calculated iteratively by evaluating the product of the inverse Jacobian at the current joint position $\vec{\theta}$ and the Cartesian position error $\Delta\vec{x}$. In order to prevent solutions that are not reachable due to joint limits, the joint positions $\vec{\theta}$ are initialized with values close to the correct positions using a simplified kinematic model from the construction process of the head.

In order to bring an object at position $\vec{x}$ to the center of one stereo camera pair, the inverse kinematic problem is solved for both cameras and the common tilt joint is actuated with the mean of both eye tilt target values.

### B. Closed-loop control (Foveation Control)

Humanoid vision systems that realize foveation using two cameras in each eye should be able to bring the object into the center of the fovea based on information from the peripheral cameras. This is necessary because the area of interest, e. g. an object that is tracked by the robot, can easily be lost from the fovea due to its narrow field of view. It is much less likely that the object would be lost from peripheral images that have a wider field of view. On the other hand, operations such as grasping can benefit from high precision offered by foveal vision. It is therefore advantageous to simultaneously use both peripheral and foveal vision (see Fig. 6). Since the foveal cameras are vertically displaced from the peripheral cameras, bringing the object into the center of peripheral images will not result in an object being projected onto the center of the foveal images. It is, however, possible to show that by directing the gaze so that the object center is projected onto the peripheral image at position $(x_p^*, y_p^*)$, which is displaced from the center of the peripheral image in the vertical direction, we achieve that the object is approximately projected onto the center of the foveal image provided that the head is not too close to the object.

The head has seven degrees of freedom: lower pitch, roll, yaw, upper pitch, eyes tilt, right eye pan and left eye pan (see Fig. 2). Instead of accurately modeling the kinematics of the head for foveation, we rather realized a simplified control system that exploits a rough knowledge about how the object moves in the image when the head (neck and eyes) moves. Obviously, moving the yaw axis ($\theta_3$) and the eyes pan axes ($\theta_6$ and $\theta_7$) results in movement of an object located in front of the head along the horizontal axis in the images, whereas moving the lower and upper pitch axes ($\theta_1$ and $\theta_4$) and the eyes tilt axis ($\theta_6$) result in movement of the object along the vertical axis in the image. On the other hand, a movement around the roll axis ($\theta_2$) results in object movements along both axes. Head roll follows the head lower pitch, therefore the above relationship is not completely true for the head lower pitch when the head roll is equal to zero. However, the approximation is good enough because the system is closed-loop and can make corrective movements to converge towards the desired configuration. Compared to classic kinematic control, our approach has the advantage that it does not change over the life time of the robot and we do not need to recalibrate the system due to factors such as wear and tear.

Computationally, to aid in coordinating the joints, we assign a relaxation position to each joint and 2-D object position. The relaxation position for the object is at $(x_p^*, y_p^*)$ and the eyes' task is to bring the object to that position. The relaxation position for the 3 eye joints is to face forward, and the head's task is to bring the eyes to that position. Further, the head tilt and the 3 neck joints have a relaxation position, and the control system attempts not too deviate too much from this position. For example, if the object of interest is up and to the left, the eyes would tilt up and pan left, causing the head would tilt up and turn left.

The complete control system is implemented as a network of PD controllers expressing the assistive relationships. As mentioned above, the PD controllers are based on simplified mappings between visual coordinates and joint angles rather than on a full kinematic model. They fully exploit the redundancy of the head. Below we illustrate the implementation of the controller network by describing how the left eye pan and head nod motion is generated. Other degrees of freedom are treated in a similar way.

We define the *desired change* for self-relaxation, $D$, for each joint,

$$D_{joint} = \left(\theta^*_{joint} - \theta_{joint}\right) - K_d\dot{\theta}_{joint}, \tag{4}$$

where $K_d$ is the derivative gain for joints; $\theta$ is the current joint angle; $\dot{\theta}$ is the current joint angular velocity, and the asterisk indicates the relaxation position. The derivative components help to compensate for the speed of the object and assisted joints.

The desired change for the object position is:

$$D_{\mathsf{X}object} = \left(x^*_p - x_{object}\right) - K_{dv}\dot{x}_{object}, \tag{5}$$

where $K_{dv}$ is the derivative gain for 2-D object position; $\mathsf{X}$ represents the $x$ pixels axis; and $x_{object}$ is 2-D object position

in pixels.

The purpose of the *left eye pan* ($\mathsf{L}EP$) joint is to move the target into the center of the left camera's field of view:

$$\begin{aligned}\widehat{\dot{\theta}}_{\mathsf{L}EP} = K_p \times \Big[ & K_{\text{relaxation}} D_{\mathsf{L}EP} \\ & - K_{\text{target}\rightarrow EP} K_v C_{\mathsf{L}object} D_{\mathsf{LX}object} \\ & + K_{\text{cross-target}\rightarrow EP} K_v C_{\mathsf{R}object} D_{\mathsf{RX}object} \Big], \quad (6)\end{aligned}$$

where $\widehat{\dot{\theta}}_{\mathsf{L}EP}$ is the new target velocity for the joint; $\mathsf{L}$ and $\mathsf{R}$ represent left and right; $K_p$ is the proportional gain; $K_v$ is the proportional gain for 2-D object position; $C_{object}$ is the tracking confidence for the object; and the gain $K_{\text{cross-target}\rightarrow EP} < K_{\text{target}\rightarrow EP}$.

*Head pitch joint* ($HP$) assists the eye tilt joint:

$$\widehat{\dot{\theta}}_{HP} = K_p \times \Big[ K_{\text{relaxation}} D_{HP} - K_{ET\rightarrow HP} D_{ET} \Big]. \quad (7)$$

Other joints are controlled in a similar way. The controller gains need to be set experimentally.

## VI. Conclusions

In this paper, we presented the Karlsruhe Humanoid Head as an active foveated vision system with two cameras per eye. The head has a sophisticated sensor system, which allows the realization of simple visuo-motor behaviors such as tracking and saccadic motions towards salient regions, as well as more complex visual tasks such as hand-eye coordination. The head is used as part of our humanoid robots ARMAR-IIIa [16] and ARMAR-IIIb, an exact copy of ARMAR-IIIa. Using the active head, several manipulation and grasping tasks in a kitchen environment have been implemented and successfully demonstrated [23], where all perception tasks were performed using the active head. In addition, seven copies of the head are used as a stand-alone system in different laboratories in Europe in the context of oculomotor control, object recognition, visual attention, human-robot interaction and vestibulo-ocular control.

We also presented accuracy results of the head and the implementation of both open-loop and closed-loop control strategies on the head.

## Acknowledgment

The work described in this paper was partially conducted within the EU Cognitive Systems project PACO-PLUS (FP6-2004-IST-4-027657) funded by the European Commission and the German Humanoid Research project SFB588 funded by the German Research Foundation (DFG: Deutsche Forschungsgemeinschaft).

## References

[1] S. Sakagami, T. Watanabe, C. Aoyama, S. Matsunage, N. Higaki, and K. Fujimura, "The Intelligent ASIMO: System Overview and Integration," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2002, pp. 2478–2483.

[2] K. Akachi, K. Kaneko, N. Kanehira, S. Ota, G. Miyamori, M. Hirata, S. Kajita, and F. Kanehiro, "Development of Humanoid Robot HRP-3," in *IEEE/RAS International Conference on Humanoid Robots*, 2005.

[3] "Fujitsu, humanoid robot hoap-2, www.automation.fujitsu.com," 2003.

[4] H. Miwa, T. Okuchi, H. Takanobu, and A. Takanishi, "Development of a new human-like head robot WE-4," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2002.

[5] L. Aryananda and J. Weber, "MERTZ: a quest for a robust and scalable active vision humanoid head robot," in *IEEE/RAS International Conference on Humanoid Robots*, 2004.

[6] G. Cheng, S.-H. Hyon, J. Morimoto, A. Ude, J. G. Hale, G. Colvin, W. Scroggin, and S. C. Jacobsen, "CB: a humanoid research platform for exploring neuroscience," *Advanced Robotics*, vol. 21, no. 10, pp. 1097–1114, 2007.

[7] H. Kozima and H. Yano, "A robot that learns to communicate with human caregivers," in *International Workshop on Epigenetic Robotics*, Lund, Sweden, 2001.

[8] R. Beira, M. Lopes, M. Praca, J. Santos-Victor, A. Bernardino, G. Metta, F. Becchi, and R. Saltaren, "Design of the robot-cub (iCub) head," in *IEEE International Conference on Robotics and Automation*, 2006.

[9] A. Ude, C. Gaskett, and G. Cheng, "Foveated vision systems with two cameras per eye," in *IEEE International Conference on Robotics and Automation*, Orlando, Florida, USA, 2006.

[10] B. Scasselatti, "A binocular, foveated active vision system," MIT, Artificial Intelligence Laboratory, Tech. rep. A.I. Memo No. 1628, Tech. Rep., 1999.

[11] 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 and Their Applications*, vol. 15, no. 4, pp. 46–56, 2000.

[12] T. Shibata, S. Vijayakumar, J. Conradt, and S. Schaal, "Biomimetic oculomotor control," *Adaptive Behavior*, vol. 9, no. 3/4, pp. 189–207, 2001.

[13] G. Sandini and G. Metta, "Retina-like sensors: motivations, technology and applications." in *Sensors and Sensing in Biology and Engineering*, T. Secomb, F. Barth, and P. Humphrey, Eds. Wien, New York: Springer-Verlag, 2002.

[14] D. Biamino, G. Cannata, M. Maggiali, and A. Piazza, "MAC-EYE: a tendon driven fully embedded robot eye," in *IEEE/RAS International Conference on Humanoid Robots*, 2005.

[15] F. Ouezdou, S. Alfayad, P. Pirim, and S. Barthelemy, "Humanoid head prototype with uncoupled eyes and vestibular sensors," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2006.

[16] T. Asfour, K. Regenstein, P. Azad, J. Schröder, N. Vahrenkamp, and R. Dillmann, "ARMAR-III: An integrated humanoid platform for sensory-motor control," in *IEEE/RAS International Conference on Humanoid Robots*, 2006.

[17] V. Zatsiorsky, *Kinematics of Human Motion*. Champaign, Illinois: Human Kinetics Publishers, 1998.

[18] K. Regenstein, T. Kerscher, C. Birkenhofer, T. Asfour, M. Zöllner, and R. Dillmann, "Universal Controller Module (UCoM) - component of a modular concept in robotic systems," in *IEEE International Symposium on Industrial Electronics*, 2007.

[19] Y. Shiu and S. Ahmad, "Calibration of wrist-mounted robotic sensors by solving homogeneous transform equations of the form AX=XB," in *IEEE Transactions on Robotics and Automation*, vol. 5, 1989, pp. 16–29.

[20] M. Li, "Kinematic calibration of an active head-eye system," in *IEEE Transactions on Robotics and Automation*, vol. 14, 1998, pp. 153–158.

[21] M. Przybylski, "Kinematic calibration of an active camera system," Master's thesis, University of Karlsruhe, 2008.

[22] D. Orin and W. Schrader, "Efficient computation of the jacobian for robot manipulators," in *International Journal of Robotics Research*, 1984, pp. 66–75.

[23] T. Asfour, P. Azad, N. Vahrenkamp, K. Regenstein, A. Bierbaum, K. Welke, J. Schröder, and R. Dillmann, "Toward humanoid manipulation in human-centred environments," *Robot. Auton. Syst.*, vol. 56, no. 1, pp. 54–65, 2008.