Project no.:	**027657**
Project full title:	**Perception, Action & Cognition through learning of Object-Action Complexes**
Project Acronym:	**PACO-PLUS**
Deliverable no.:	**D1.1.2**
Title of the deliverable:	**Integration Report**

Contractual Date of Delivery to the CEC:	31. January 2009		
Actual Date of Delivery to the CEC:	04. February 2009		
Organisation name of lead contractor for this deliverable:	UniKarl		

Author(s): Tamim Asfour, Kai Welke, Alexander Bierbaum, Pedram Azad, Aleš Ude and Rüdiger Dillmann
Participant(s): UniKarl

Work package contributing to the deliverable:	WP1.1		
Nature:	R/D		
Version:	Draft		
Total number of pages:	10		
Start date of project:	1^{st} Feb. 2006	**Duration:**	48 month

Project co-funded by the European Commission within the Sixth Framework Programme (2002–2006)
Dissemination Level

PU	Public	X
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

Abstract:

In this report, we present the work on 1) the further development and improvements of the Karlsruhe Humanoid Head and the sensor and control system for the five-fingered hand including a new type of tactile sensors 2) the definition of software interfaces for integrating developed components within other work packages in the project and 3) the communication between the groups.

Keyword list: Integration of Software components in PACO-PLUS, hardware and software components, communication between the groups.

Table of Contents

1. Executive Summary

This deliverable reports on the ongoing activities related to the improvement of the Karlsruhe humanoid head, the sensor and control system of the pneumatic-driven five-fingered hand and gives a brief description of the developed and implemented software interfaces which are necessary for the integration of developed components in other work packages of the project.

The Karlsruhe humanoid head has been improved in terms of the extension of its sensorimotor capabilities to allow the implementation of several visual tasks on an active humanoid head. Accuracy tests have been performed, a kinematics calibration procedure for the robot eyes has been developed and both closed-loop and open-loop control strategies have been implemented and tested.

Due to the limitations of the FSR-based cursor navigation sensors used so far in the robot platforms at SDU and UniKarl, first prototypes of new tactile sensors based on capacitive sensing technology has been developed. First experimental results reveal good characteristics in terms of sensitivity. Furthermore, a hybrid position/force controller for the pneumatic-driven five-fingered hand has been developed and tested.

Section 3 gives a brief overview on the extension of the ARMAR software interfaces to allow the integration of various components that have been developed in other work packages and Section 4 summarizes the communication activities between the different partners.

2. Hardware Development

2.1 Further Development of the Karlsruhe Humanoid Head

We continued our work on the Karlsruhe Humanoid Head and the extension of its sensorimotor capabilities. Open-loop and closed-loop control strategies have been implemented and tested in the context of the implementation of several tasks such as foveation, object recognition and 3-D active vision methods. For more details the reader is referred to [A] and [C].

2.1.1 Head Accuracy

In [A] we evaluated the repeatability of joint movements on the Karlsruhe Humanoid Head. Therefore, a series of measurements on the basis of visual data was performed. A calibration rig was positioned at a fix location in front of the cameras. During the test procedure, all joints of the head were actuated successively. After moving, the head returned to its initial position and the pose of the calibration rig was determined visually.

Figure 1 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 0 degrees. 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 achieve an accuracy of about $\pm 0.13°$ and $\pm 0.075°$, respectively. The larger inaccuracy in these joints originates from dynamic effects in the gear belt driven joints caused by the weight of the head. The theoretically achievable accuracy can be derived from the number of encoder ticks which encode a rotation of one degree 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 $\pm 0.001°$.

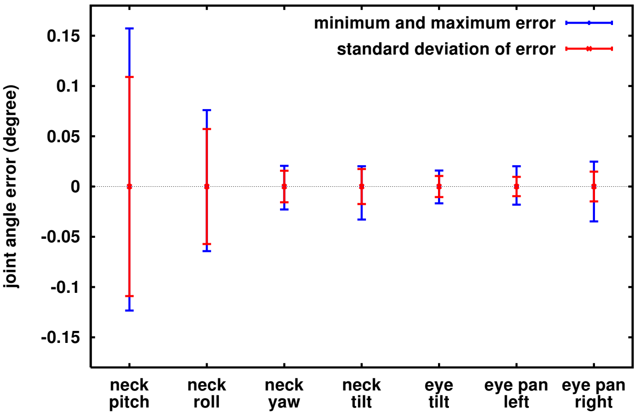
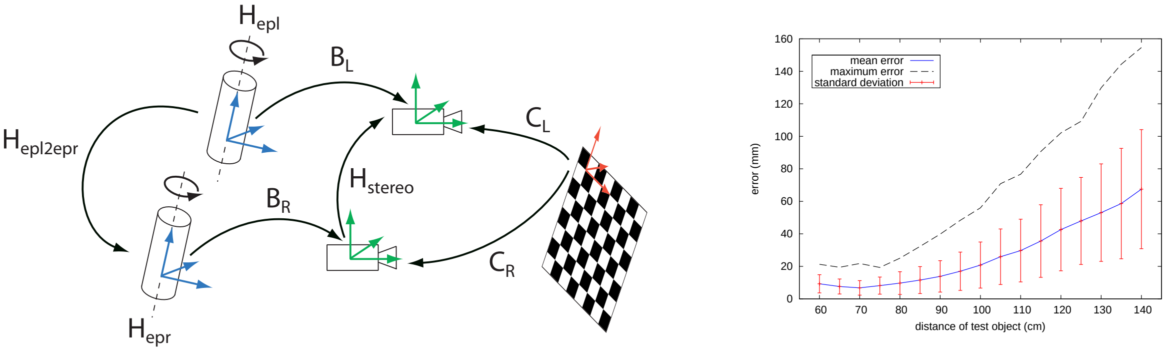


Figure 1: Accuracy of joint position during the repeatability tests. Each joint of the head was moved to the zero position starting from non-zero configurations 100 times. The pose of the head was measured visually using a calibration rig. The position error of the joints was measured using computer vision methods. The plot illustrates standard deviation, minimum and maximum of the joint angle errors in degrees.

2.1.2 Kinematic Calibration

We continued our work on kinematic calibration of the active camera system. Improvements were made concerning the accuracy of the approach. Furthermore, we evaluated the derived calibrated kinematic model in stereo triangulation and saccadic eye movement tasks. Figure 2(a) shows the first two DoF of the head-eye system that were used in the calibration procedure for stereo triangulation. The aim of the calibration was to determine the unknown transformations B_L and B_R between the optical center of the perspective cameras and the joint axes of left eye pan and right eye pan. A non-linear least squares approach was used to determine the unknown transformations from a set of extrinsic camera calibrations. Details of the calibration procedure are explained in [D]. Figure 2(b) shows the accuracy of the calibrated model in a stereo triangulation task with actuated eyes. Position errors of less then 1.5 cm could be achieved in manipulation distance.



(a) Transformations involved in the kinematic calibration process for active stereo calibration. The aim was to find the unknown transformation B_L, B_R in order to derive H_{stereo} for actuated eyes.

(b) Accuracy of stereo triangulation using the calibrated model. The translational part of the error is shown for different distances of the test object.

Figure 2: Kinematic calibration of the active camera system of the head.

2.1.3 Control Strategies

Two different control strategies were implemented on the Karlsruhe Humanoid Head: 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. Closed-loop control was implemented based on the approach proposed in [3] and integrated into the head control software. The approach uses a cascade of PD controllers, which are based on simplified mappings between visual coordinates and joint angles rather than on a full kinematic model. The open-loop control was implemented for the head-eye system on the basis of the kinematic calibration procedure described in the previous section. The open-loop control strategy can be divided into two problems. First an accurate kinematic model for the involved joints is established, and second the inverse kinematics problem is solved using the obtained kinematic model. Figure 3 shows the accuracy of open-loop control based on differential kinematics using a calibrated kinematic model.

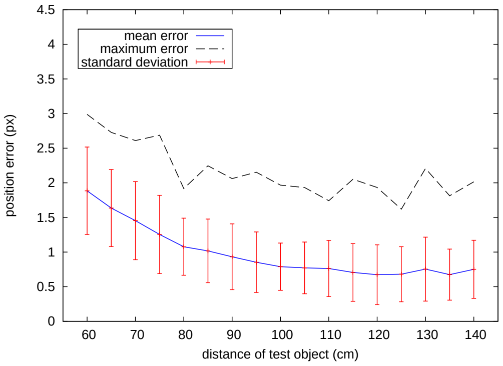


Figure 3: Accuracy of a saccadic eye movement tasks. The target object was positioned at arbitrary locations and different distances. The error is given in pixels.

2.2 Further Development of Tactile sensors for the five-fingered Hand

This work focusses on the investigation of new tactile sensor technologies needed for haptic exploration and advanced grasping on ARMAR as well as on the implementation of a hybrid position/force control for the pneumatic-driven five-fingered hand.

2.2.1 New tactile sensors for the five-fingered Hand

The characteristics of the FSR-based cursor navigation sensors presented in D1.1.1 from the last reporting period have been further investigated to identify their limitations as well as methods for technical improvement. The sensor system is used in robot platforms at SDU and UniKarl and works in a reliable way on both robot platforms. However, experiments revealed that the level of sensitivity still should be increased to enable haptic exploration of unknown objects. A reliable improvement of the sensitivity by modifying the sensor actuation layer as proposed earlier could not be achieved. Furthermore, full coverage of the robot hand with such kind of sensors or skin segments was impossible due to the fact that the sensing area of each sensor element is embedded in a small but insensitive carrier frame. Therefore, our work focusses

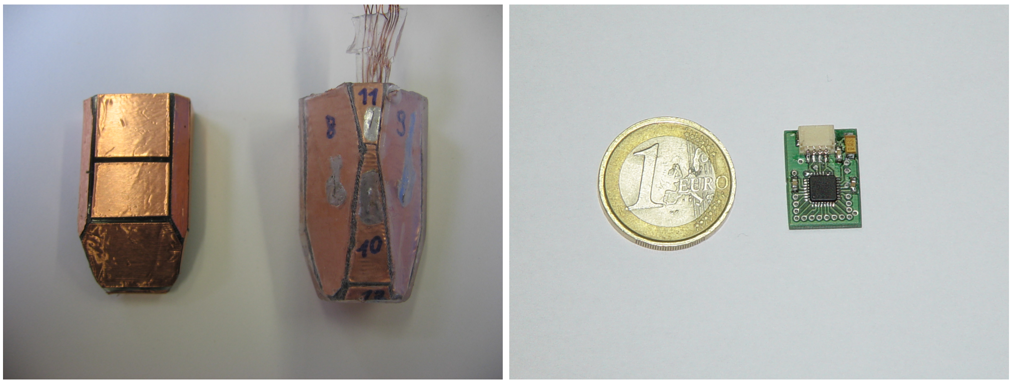


Figure 4: Finger tip with capacitive tactile sensor (left) and capacitive sensor electronics (right).

on the investigation of alternative tactile sensing technologies that provide better sensitivity and coverage features. A promising approach is the deployment of highly integrated capacitive touch controllers as they have become available recently with the spread of hardware devices offering human machine interfaces via touch, e.g. in cell phones. Based on this capacitive sensing technology, we have developed a prototype of a finger tip sensor for the five-fingered robot hand. The finger tip is fully covered with a touch sensing layer subdivided into twelve separate sensing regions (see Figure 4).

Such capacitive sensors may be designed in a great variety of shapes by using flexible electrode materials. In our case it was possible to create a sensor covering the complete surface of a robot finger tip, which allows the detection of contacts in all directions as it may occur during operation in unstructured environments.

An elastic silicone layer and a conductive electrode layer encompass the fingertip sensor. The sensor electronics PCB is situated inside the fingertip. As in the case of the FSR-based cursor navigation sensors, the capacitive sensor offers an I^2C-bus based communication interface to connect to the micro-controller of the robot hand.

The sensor is currently being evaluated. It offers continuous measurement characteristics and does not exhibit a lower force limit during measurement as it is the case with FSRs. This results in significantly better sensitivity. First results indicate very sensitive force characteristics. Subject to investigation are the capacitive interferences between sensor regions, long term stability and methods of calibration of the sensor.

2.2.2 Hybrid position/force hand control

Significant progress was made in the development of a hybrid position/force controller for the five-fingered robot hand. The complete sensor system comprising nine miniature pressure and eight joint position sensors has been integrated into the left hand of the humanoid robot ARMAR-IIIb at UniKarl (see Figure 5, for detailed information the reader is referred to [B].). The robot hand offers eight active controllable degrees of freedom. A low-level controller has been developed, which allows the specification of both pressure and position targets for the joint actuators. The balance between the targets is adjustable according to the desired task the hand is executing, thus allowing active compliance control. Furthermore, diagnostic information is generated by the control software comprising the estimated leakage rate and external force for each joint actuator. The external force estimate serves as additional proprioceptive sensory feedback and complements the application of tactile sensors for detecting contact between the robot's fingers and an object. As the actuators exhibit individual force transmission characteristics, the controller parameters currently are adjusted for optimal control performance in a final step. In Figure 6, the typical time responses of the actual joint angle, actuator pressure and the external force estimates are shown.

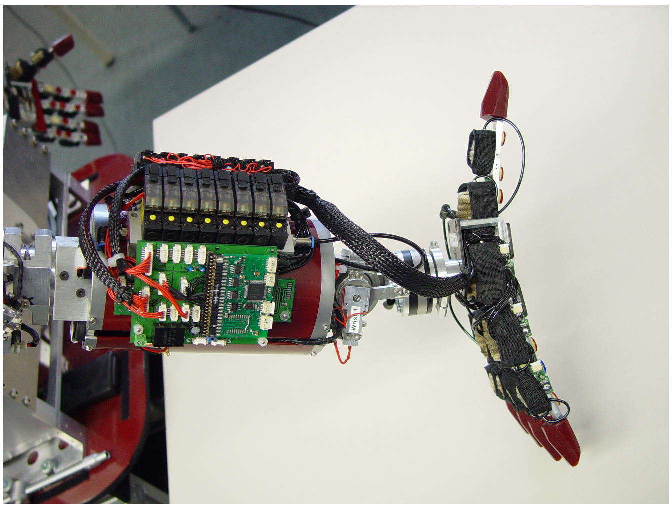


Figure 5: Pneumatic robot hand with sensors, valves and control system.

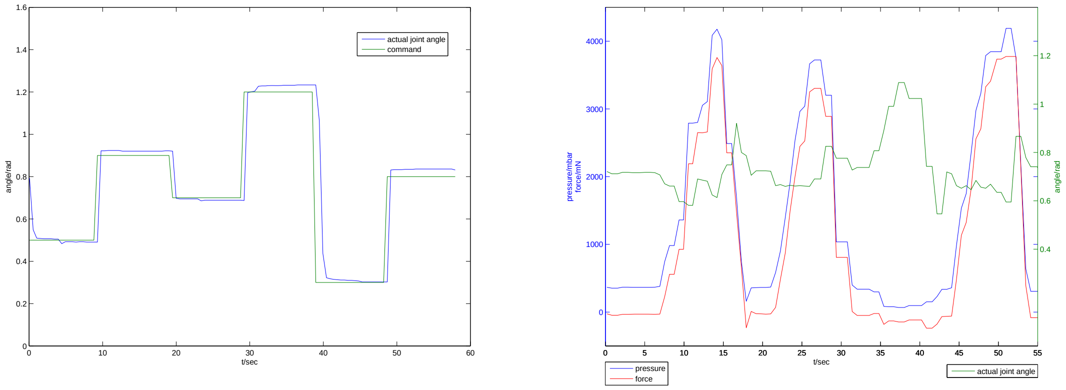


Figure 6: Response characteristics of a joint angle step command sequence for the index proximal joint (left) and typical response to external disturbances at a joint angle command of 0.7 rad and maximum force command (right). The three peaks in the pressure signal correspond to disturbances applied via external force.

3. Software Interfaces on ARMAR

Robot interface: In D1.1.1 of the last report period, the so-called *robot interface* was presented, which offers an API for convenient access to the robot's sensors and actors. The provided abstraction levels are *skills*, *tasks*, and *scenarios*. The robot interface is being constantly extended. One important skill that has been added is a *general purpose* visual servoing skill, which allows for accurate vision-based operational space control of the end-effector. This skill is complementary to conventional open-loop inverse kinematics and is required for the integration of the grasp reflex on ARMAR-III (UniKarl, SDU), grasping based on box decomposition (UniKarl, KTH), and pushing for grasping (UniKarl, JSI).

Network interface: In order to communicate with ARMAR-III from external applications, the robot network interface has been extended to provide access to the sensor and target values and the robot interface API of ARMAR-III via TCP and UDP communication. Furthermore, the exchange of application specific data is supported on scenario level. Several components, which have been integrated, make use of the network interface, e.g. Q-Learning (UniKarl, BCCN), grasp reflex on ARMAR-III

(UniKarl, SDU), grasping based on box decomposition (UniKarl, KTH) and grasp pushing (UniKarl, JSI).

Tools for Computer Vision: In D1.2 of the first report period, the *Integrating Vision Toolkit* (IVT[1]) was introduced, which serves as the framework for the integration of computer vision modules. For the integration of the results from other partners, images must be provided that satisfy the assumptions made by the modules to be integrated. Depending on the module, the necessary image transformations are rectification and undistortion. For this purpose a highly optimized image mapping routine has been implemented that outperforms the OpenCV implementation by a factor of 4. Furthermore, calibration data and projection matrices are provided also for the *transformed* images, as it is required by the CoVis software for the integration of the grasp reflex on ARMAR-III (UniKarl, SDU).

Image Transmission: The integration of vision components that are implemented as external applications by the partners requires transmission of the camera images. For this purpose, the robot interface has been extended by a module that allows for image transmission by using the above-mentioned network interface. Depending on the application, the raw images or compressed images are transmitted. This interface is used for the integration of the grasp reflex on ARMAR-III (UniKarl, SDU).

Master Motor Map: In D8.2.2 of the last report period, the *Master Motor Map* (MMM) [1] was introduced, which serves as an exchange data format for human joint angle trajectories by defining a reference kinematics model. In [1], segment lengths of the body parts are not included. In [2] (see attached paper from D8.2.3 from this report period), the MMM has been extended to include segment lengths as well as a designated target object position. This extension allows for representation of human joint angle trajectories of goal-directed actions. The MMM is used for the integration of contributions related to action synthesis (UniKarl, JSI) as well as grasp recognition and mapping to ARMAR (UniKarl, KTH).

4. Communication between the groups

The members held a number of meetings and phone conferences, which will be listed in the activity report. In addition, there were numerous one-to-one discussions between the subgroup members as well as bilateral visits of project members on the sites of the partners. The meetings related to integration efforts took place as follows:

20-21 November 2008 General meeting in Ljubljana attended by all work package leaders.

8 January - 25 March 2008 Visit of Dennis Herzog (AAU) in Ljubljana (JSI). Work on the implementation of parametric hidden Markov models on HOAP-3.

14-22 April 2008 Work on integration of the object representation (”Feature Hierarchies”) of ULg with the visual primitives (CoViS) developed in Odense. Focus on applying the resulting system on object pose estimation.

5-23 May 2008 Summer school with participations of members from all partners (see D9.2.2).

13-23 May 2008 Andrej Kos and Ales Ude (JSI) visited Karlsruhe (UniKarl) to work on the transfer of goal-directed action synthesis to ARMAR.

1 July - 31 August 2008 Ales Ude (JSI) visited ATR (Japan) to experiment with visuomotor processes needed to implement goal-directed action synthesis.

[1] `http://ivt.sourceforge.net`

20 August - 7 September 2008 Nils Adermann (UniKarl) visited Edinburgh to work on the integration of high-level planning (PKS) to ARMAR.

26 September 2008 Visit of Norbert Krüger (SDU) at BCCN. Discussion on semantic scene graphs

1-2 October 2008 Meeting in Karlsruhe attended by Florentin Wörgötter (BCCN), Norbert Krüger (SDU), Christopher Geib (UEDIN), Rüdiger Dillmann (UniKarl) and Tamim Asfour (UniKarl). Discussion on learning and planning issues.

2-17 October 2008 Visit of Renaud Detry (ULg) in Odense (SDU). Integration of the grasp reflex software of SDU with the grasp densities of ULg. Continued integration of the object representation ("Feature Hierarchies") with the visual primitives (CoViS).

5-9 October 2008 Minija Tamosiunaite (BCCN) visited JSI to prepare an experiment for learning of filling actions using reinforcement learning.

9-15 October 2008 Nicolas Pugeault in Odense (SDU). Work on pose estimation.

20-30 October Mila Popovic (SDU) visited Karlsruhe to work on the integration of CoVis and grasp reflex on ARMAR.

11-22 November 2008 Minija Tamosiunaite and Florentin Wörgötter (BCCN) visited JSI to work on the implementation of reinforcement learning on HOAP-3.

Skype conferences Multiple Skype calls between SDU and UEDIN to discuss ongoing integration of high-level plan execution monitoring on the SDU robot/vision platform.

Skype conferences Multiple Skype calls between UniKarl and UEDIN to discuss the design of the high-level planning domain for the UniKarl kitchen environment, and PKS software integration on the ARMAR robot platform.

Planned visits

2-13 March 2009 Visit of Alejandro Agositin (CISC) in Karlsruhe. Work on the integration of the rule learning system on ARMAR.

2-6 March 2009 Visit of Aleš Ude (JSI) in Karlsruhe. Work on the generalization of example movements with dynamic systems.

14-30 April 2009 Visit of Javier Romero (KTH) and Dennis Herzog (AAU) in Karlsruhe. Work on the integration of grasp recognition and action recognition and synthesis on ARMAR.

19-24 April 2009 Visit of Damir Omrcen (JSI) in Karlsruhe. Working on integration of pushing for grasping on ARMAR.

Attached Papers

[A] T. Asfour, K. Welke, P. Azad, Ales Ude, and R. Dillmann. The Karlsruhe Humanoid Head. In *Proc. IEEE/RAS International Conference on Humanoid Robots (HUMANOIDS)*, 2008.

[B] I. Gaiser, S. Schulz, A. Kargov, H. Klosek, A. Bierbaum, C. Pylatiuk, R. Oberle, T. Werner, T. Asfour, G. Bretthauer, and R. Dillmann. A New Anthropomorphic Robotic Hand. In *IEEE/RAS International Conference on Humanoid Robots*, Daejeon, Korea, 2008.

[C] Aleš Ude and Tamim Asfour. Control and recognition on a humanoid head with cameras having different field of view. In *Proc. IAPR Conf. Pattern Recognition (ICPR)*, Tampa, Florida, December 2008.

[D] K. Welke, M. Przybylski, T. Asfour, and R. Dillmann. Kinematic calibration for saccadic eye movements. In *Robotics: Science and Systems (submitted to)*, 2009.

References

[1] P. Azad, T. Asfour, and R. Dillmann. Toward an Unified Representation for Imitation of Human Motion on Humanoids. In *IEEE International Conference on Robotics and Automation*, pages 2558–2563, Roma, Italy, 2007.

[2] M. Do, P. Azad, T. Asfour, and R. Dillmann. Imitation of Human Motion on a Humanoid Robot using Nonlinear Optimization. In *IEEE/RAS International Conference on Humanoid Robots*, Daejeon, Korea, 2008.

[3] A. Ude, C. Gaskett, and G. Cheng. Foveated vision systems with two cameras per eye. In *Robotics and Automation, 2006. ICRA 2006. Proceedings 2006 IEEE International Conference on*, pages 3457–3462, 2006.

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]).

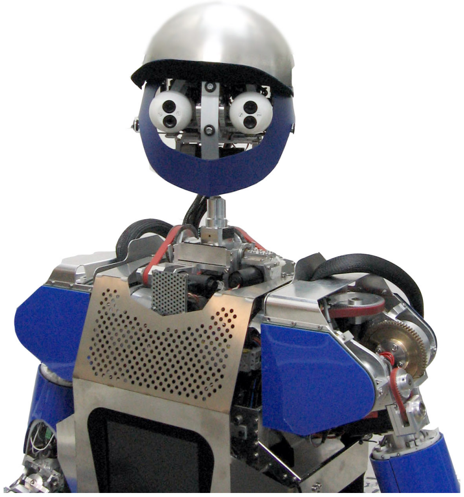


Fig. 1: The Karlsruhe humanoid head as part of the humanoid robot ARMAR-III. The head has two eyes and six microphones. Each eye has two camera

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.

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 θ_5, θ_6 and θ_7, where both eyes share a common tilt axis (θ_5) and each eye can independently rotate around a vertical axis (θ_6 and θ_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.

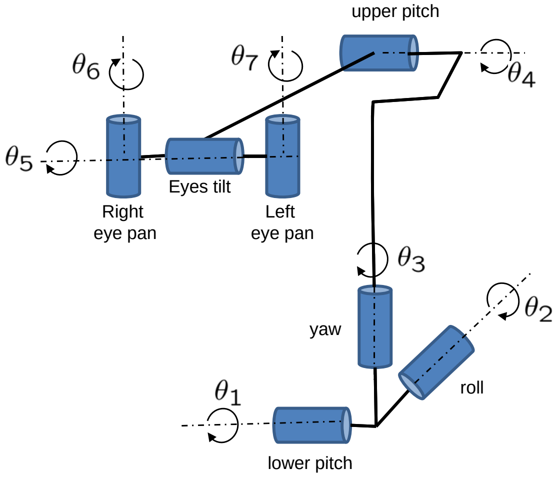


Fig. 2: The kinematics of the head with seven DoF arranged as lower pitch (θ_1), roll (θ_2), yaw (θ_3), upper pitch (θ_4), eyes tilt (θ_5), right eye pan (θ_6) and left eye pan (θ_7).

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

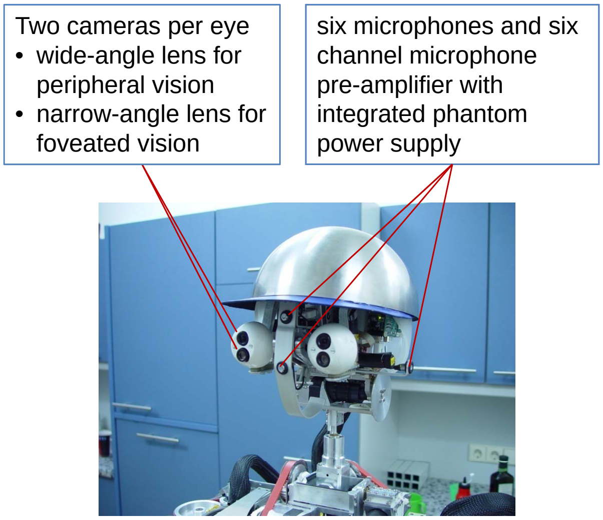

Two cameras per eye
• wide-angle lens for peripheral vision
• narrow-angle lens for foveated vision
six microphones and six channel microphone pre-amplifier with integrated phantom power supply

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 640×480, 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.

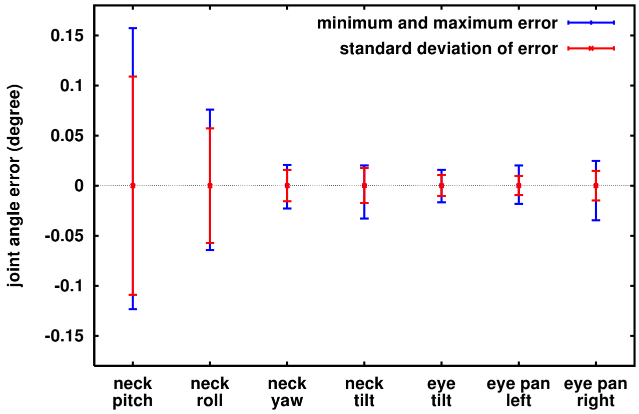


Fig. 4: Results of the head accuracy test for all seven joints of the head (from bottom to top). Each joint of the head was moved to the zero position starting from non-zero configurations 100 times. The position error of the joints was measured using computer vision methods. The plot illustrates standard 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 ±0.025°. The neck pitch and neck roll joints (θ_1 and θ_2 in Fig. 2) achieve an accuracy of about ±0.13° and ±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 ±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 × 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 α_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 α_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:

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

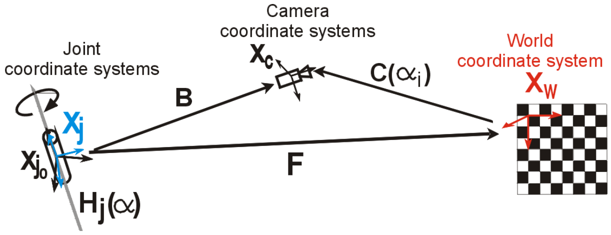


Fig. 5: Coordinate systems and transformations required in the kinematic calibration process.

[1] `ivt.sourceforge.com`

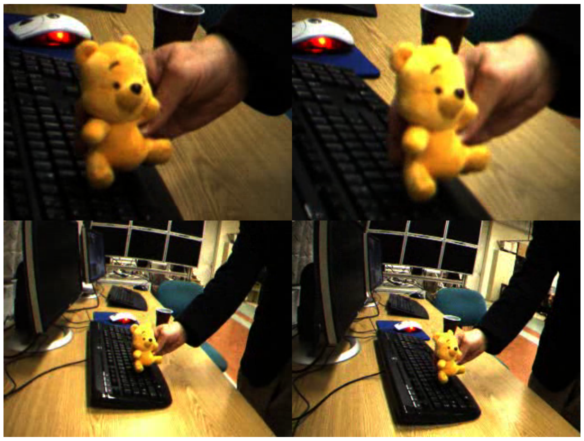


Fig. 6: Simultaneous stereo view from peripheral (below) and foveal cameras (above)

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 $\mathbb{R}^{3\times 3}$, 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 (θ_3) and the eyes pan axes (θ_6 and θ_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 (θ_1 and θ_4) and the eyes tilt axis (θ_6) result in movement of the object along the vertical axis in the image. On the other hand, a movement around the roll axis (θ_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; θ 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; 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:

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

where $\widehat{\dot{\theta}}_{\mathsf{L}EP}$ is the new target velocity for the joint; L and 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 $K_{\text{cross-target}\to EP} < K_{\text{target}\to EP}$.

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

$$\widehat{\dot{\theta}}_{HP} = K_p \times \Big[K_{\text{relaxation}} D_{HP} - K_{ET\to HP} D_{ET}\Big]. \tag{7}$$

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.

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

A New Anthropomorphic Robotic Hand

I. Gaiser[1], S. Schulz[1], A. Kargov[1], H. Klosek[1], A. Bierbaum[2], C. Pylatiuk[1], R. Oberle[1], T. Werner[1], T. Asfour[2], *Member IEEE*, G. Bretthauer[1], *Member IEEE*, R. Dillmann[2], *Member IEEE*

[1] Research Center Karlsruhe (KIT), Germany, Institute for Applied Computer Science
immanuel.gaiser@iai.fzk.de, schulz@iai.fzk.de

[2]University of Karlsruhe (KIT), Germany, Institute of Computer Science and Engineering bierbaum@ira.uka.de

Abstract—This paper presents the new robotic FRH-4 hand. The FRH-4 hand constitutes a new hybrid concept of an anthropomorphic five fingered hand and a three jaw robotic gripper. The hand has a humanoid appearance while maintaining the precision of a robotic gripper. Since it is actuated with flexible fluidic actuators, it exhibits an excellent power to weight ratio. These elastic actuators also ensure that the hand is safe for interacting with humans. In order to fully control the joints, it is equipped with position sensors on all of the 11 joints. The hand is also fitted with tactile sensors based on cursor navigation sensor elements, which allows it to have grasping feedback and the ability for exploration.

I. INTRODUCTION

THE number of scenarios predicting the future use of robots in everyday life increases constantly [1-4]. One of the most important parts of a service robot is to manage safe interaction with humans and to manipulate objects. For social acceptance reasons a humanoid appearance is needed. Service robots are meant to operate in environments that are designed to be operated by the human hand. This is another reason why a service robot needs an anthropomorphic end-effector.

During the past years several robotic hands were developed with fascinating manipulation abilities [5-15]. These hands have either built in actuators or have actuators mounted in the forearm region. In the case of the latter, the actuators are mostly connected to the joints by using tendon cables. However, to create a compliant system, complex mechanical designs and/or controlling systems are necessary. One way to achieve compliance is by using the series elastic actuators as shown in [12]. Another way is to use pneumatic actuators, which are compliant due to the compressibility of air [9-15].

In order to create a compliant, lightweight system that is easy to control and can be made available as a series, the new hand was developed. This includes providing a hand that is modularly and does not need any space proximal to the wrist so that it can basically be attached to any kind of robotic arm. For operation only air supply and a five wire cable is necessary. A complete evolution of the FRH hand series was developed at the Research Center Karlsruhe. The most recent hand, a further development of [20], the FRH-4 is topic of this paper. A prototype is shown in Figure 1. The design and characteristics of this hand are discussed in this paper.

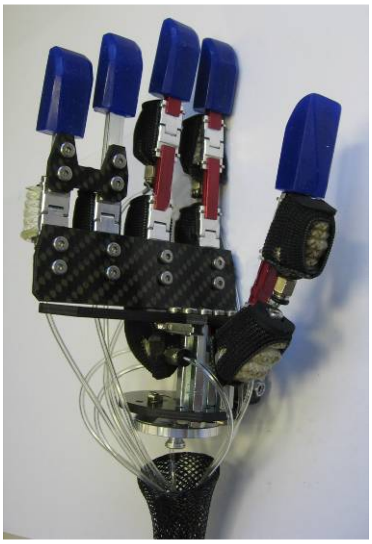


Fig. 1: The new FRH-4 hand

II. CONCEPT OF THE NEW ANTROPOMORPHIC FLUID HAND FRH-4

A. Kinematics

The most significant property of the developed hand is that it represents a hybrid concept of an anthropomorphic humanoid hand and a robotic gripper. This characteristic turns the new hand into an ideal end-effector for humanoid robotics applications because it allows for precise and stable grasping over a wide range of grasping force. The arrangement of artificial joints and bones is shown in Figure 2f. The circles mark the joints and the lines mark the bones. The filled circles mark the position of joints that occupy an outstanding function. These joints are located in the palm of hand and are mainly responsible for the humanoid appearance of the hand. While the hand has a robotic body structure as shown in Figure 2f, it has a humanoid look with inflated palm actuators, as shown in Figure 1, and 8. This also describes the main differences to former FRH models. The passive flexible abduction joints were eliminated since they do not allow precise grasping. Figure 1, and 8 show fully functional prototypes of the FRH-4 hand. The joints of the distal phalanxes of the ring finger and the little finger have not been implemented yet.

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

The position with bent palm joints is the starting position for most of the grasping types needed. Figure 2 a)-e) shows the achievable grasping patterns. The significant difference to the formerly developed fluid hand FRH 1 [14, 15] is that here all the fingers are arranged parallel, so that precise grasping and controlling can be accomplished.

The new hand has 8 independent degrees of freedom, two in each, thumb, index and middle finger, one coupled DOF for ring finger and small finger, and one, double actuated, in the palm of the hand. Another significance of the FRH-4 is the totally symmetric arrangement of all components which is shown Figure 2. The distance between two joints is 40 mm the width of the hand is 93 mm, which corresponds to a large human hand. The thumb is mounted opposite of the fingers exactly in the middle between index finger and middle finger. The symmetry allows precise controlling and freely programmable grasping patterns. The precise three jaw grasp is shown in Figure 3.

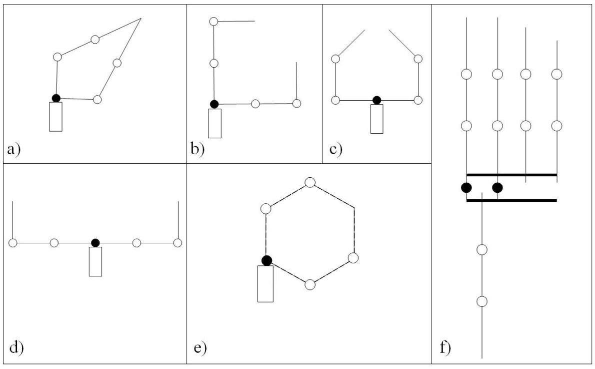


Fig. 2: Schematic sketches of the realizable grip patterns (a-e) and basic outline of the hand structure (f).

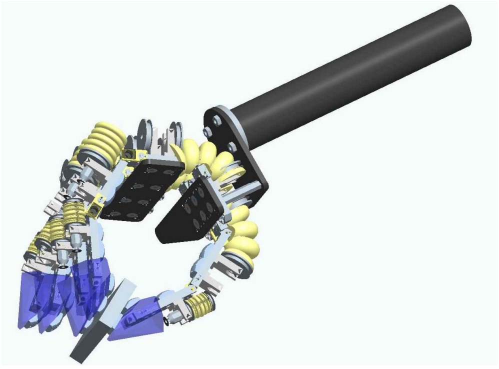


Fig. 3: Precise Three jaw grasp of the robotic hand

This setup allows for high functionality and a wide range of movement. The maximum opening span of the hand is 120 mm. In addition to adaptive grasping of varying objects the hand is able to use one of its fingers as an index finger in order to operate any type of switch or button.

B. Actuation

The new FRH-4 hand is actuated with flexible fluidic actuators as described in [20]. This actuation system consists of reinforced flexible bellows that are attached to a joint in a way so that they apply a torque to the joint by inflating the bellows. A principal setup of the system is demonstrated in Figure 4. All proximal finger joints are actuated with actuators 20 mm in diameter. The distal phalanxes are moved with actuators 12 mm in diameter. The characteristic of the 12mm actuators is shown in the torque over angle plot in Figure 5.

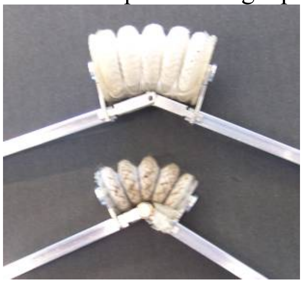


Fig. 4: Flexible fluidic actuators 20 mm (top), 12 mm (bottom)

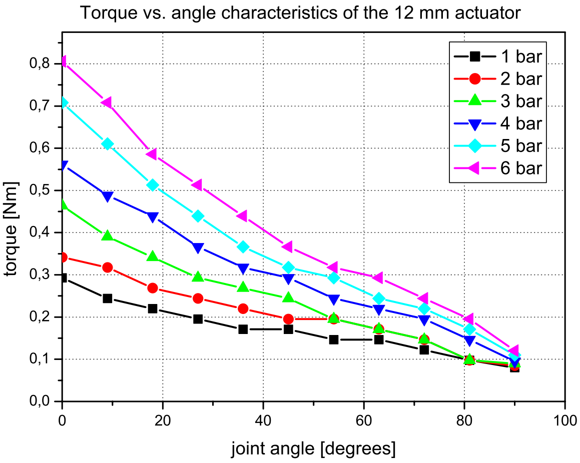


Fig 5: Torque vs. Angle plot for a 12 mm actuator

Another innovation of the new FRH-4 hand is that enhanced actuators are used with an effective diameter of 20 mm. The characteristic torque curve of the 20 mm actuator can be seen in Figure 6. By comparing the two plots it becomes obvious that the new hand will have a higher grasping force than the FRH 2-3 hands built with the 12 mm actuator or will need lower actuation pressure respectively. The 20 mm actuator shows a three times higher torque at an applied pressure of 6 bar.

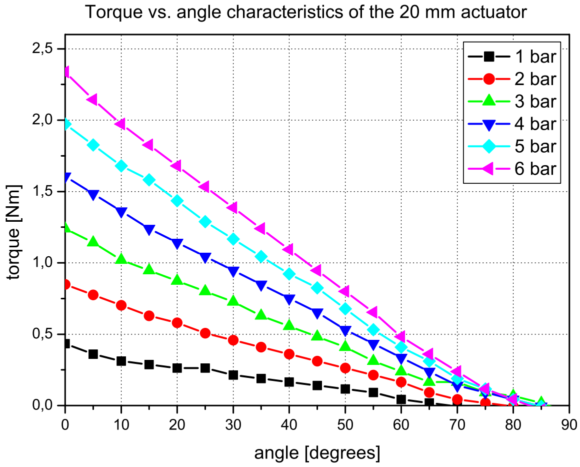


Fig. 6: Torque vs. Angle plot for a 20 mm actuator

Since there is no antagonist actuator in the joints, a retraction force is necessary to move the joints back to their starting position at 0°. In order to accomplish this requirement, tension springs or elastic rubber bands can be attached to each joint. The setup of one finger including actuators and retraction units is shown in Figure 7.

The hand can be driven with any type of pressured air source. The FRH-4 hand is designed for an operating pressure of 6 bar.

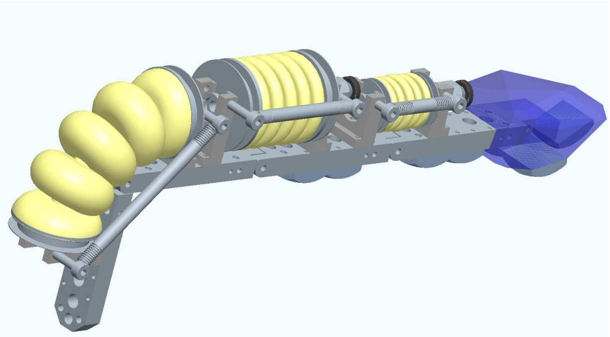


Fig.7: Side view of a finger including actuation and retraction unit

In order to be able to operate and control the hand miniaturized valves were developed [21]. These valves are operated at a frequency of 200 Hz.

The technical specifications of the new robotic hand can be summarized as shown in Table 1.

TABLE I

TECHNICAL SPECIFICATIONS OF THE ROBOT HAND

Parameters	Dimensions
Total weight	216 g
Number of actuators	12
Independent DOF	11
Hand dynamics (grasping speed)	0 – 2 rad/sec
Holding force (hook grasp)	up to 110 N
Average phalange contact force (stable holding with a power grasp)	from 1N
Torque joint 12 mm actuator (6 bar)	up to 0.7 Nm
Torque joint 20 mm actuator (6 bar)	up to 2.4 Nm
Power supply prototype / optional	8.5 – 14 VDC
Pressure supply	air at 6 bar
Interface external/internal	CAN bus/I^2C, SPI
Frequency of valves	200 Hz
Noise level (valves / 1 meter distance)	55 dB
Length (wrist to fingertip)	149 mm
Width	93 mm

C. *Position Sensors*

Position measurement is carried out using low cost contact-free 12-bit programmable magnetic rotary encoders of the type AS5045 provided by Austriamicrosystems. These encoders work in a contact-free manner and over a complete rotation of 360°. Figure 9 shows the integration of the two encoder parts, magnet and Hall sensor, in a joint. When magnet and sensor are precisely positioned the accuracy as described in the product specification provided by Austriamicrosystems could be proofed.

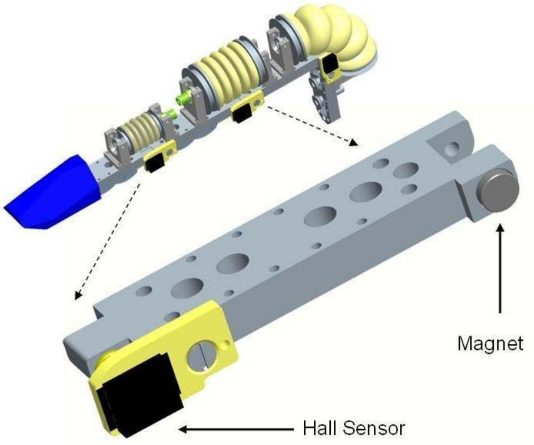
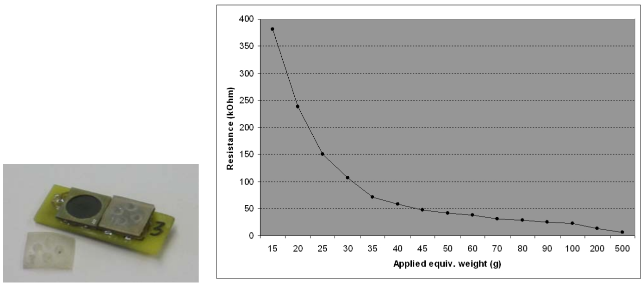


Fig. 9: Integration of the parts of the position encoder in a joint

D. *Tactile Sensors*

The tactile sensor system for the robot hand uses Force Sensing Resistor (FSR) Technology [16] as a basis. This technology makes use of a resistive effect in the sensing material, which leads to a non-linear change in the electrical resistance, when a mechanical load is applied. The sensor system consists of integrated modules as PCBs with dimensions 30x12x4.5 mm, each carrying two FSR cursor navigation sensor elements from Interlink Electronics [17], as shown in Figure 10a. Similar sensors have been used for intrinsic force measurement in combination with a tactile sensor array [18]. This type of sensor has originally been developed as cursor navigation input device for hand held devices. It is therefore low cost and available off-the-shelf. The sensor elements provide a circular sensitive area with four distinct sectors as sub-sensors. A rubber actuation layer with small buckles at the centre of each sector was designed as a force concentrator that is attached to the sensor topside. The resistance characteristic of a sub-sensor using this actuation layer is shown in Figure 10b. It exhibits the typical wide dynamic range with high sensitivity in the lower force range and decreasing resolution towards higher forces until saturation. The characteristics were acquired by applying a mechanical load to individual sub-sensors and measuring the equivalent weight with a digital scale while the response in electrical resistance was measured.



(a) Sensor array PCB with rubber actuation layer.

(b) Characteristics of a subsensor

Fig. 10: Tactile sensor system

The resistance of the FSR elements is measured using the combination of a voltage divider circuit and an 8- channel Analog-Digital-Converter (ADC) with I²C interface [19] which allows for interconnection of up to 5 modules on a single I²C bus instance. Due to a limited set of available bus addresses for the selected ADC, this number may be increased by using a standard bus expander circuit to a maximum of 128 devices in a two-level bus hierarchy. The I²C bus connection scheme was chosen for the availability and low costs of compatible devices and furthermore for its easy software interfacing and the low wire count. It is planned to attach a module to the proximal and distal phalanxes of index finger and middle finger, one to the distal phalanx of the thumb and two modules to the palm. The tactile sensor modules may be connected to either the hand controller unit or directly to a control PC for data processing. The described setup provides the structure for tactile exploration as described in [22]

E. Controlling Unit

The controlling concept is based on a decentralized concept. Hence each finger has a controller which has to collect all the data from the angle, pressure and tactile sensors. Then it has to interpret and transmit it via CAN-bus to the central controlling unit. The other important function the local controller has to accomplish is to interpret the received data transmitted trough the CAN-bus and process the data for the driver of the valves.

The microcontroller was chosen in order to have several interfaces which can operate simultaneously to complete all the controlling tasks in a short time.

For these reasons the microcontroller PIC24H256 provided by MICROCHIP was chosen.

The PIC24H256 microcontroller has an 8 channel hardware DMA with a 2Kb dual ported DMA buffer area implemented. This allows parallel data transfer between the RAM and the communication modules without any interruption of the main program. The detailed specifications are as follows:

- 2 SPI modules which support 8 Bit as well as 16 Bit data formats
- 2 I²C™ modules which support 7 and 10 Bit addressing
- 2 UART modules which support LIN and IrDA®
- 2 CAN (ECAN™) 2.0B modules with FIFO options and DMA support

Beyond that the PIC24H256 has a A/D module, which allows a 10–bit with 1.1Msps or a 12-bit with 500kscp conversion. That allows for sampling of 4 Samples parallel and even scanning in sleep mode is possible.

All components are compatible to a 3 V logic level. The circuit boards are equipped with ICSP (In-Circuit Serial Programming) interfaces. Thus it is possible to load programs directly to the circuit board. The interface can also be used for comfortable program development in connection with a MPLAB In-Circuit Debugger.

The communication between the local controllers and the superior controlling computer is realized via CAN-bus. A picture of the controller unit can be seen in Figure 11.

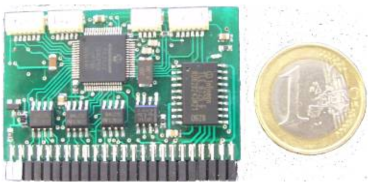


Fig. 11: Controller Unit

III. Application

The development of a mobile assistive robot at the University of Karlsruhe is the main objective of the Collaborative Research Centre 588 project “Humanoid Robots – Learning and Cooperating Multimodal Robots” [13]. The new mobile robot is a service robot with a humanoid design. According to the main aim of the project, this service robot should be equipped with two artificial hands and work directly in cooperation with the user in a kitchen environment to assist elderly and disabled humans.

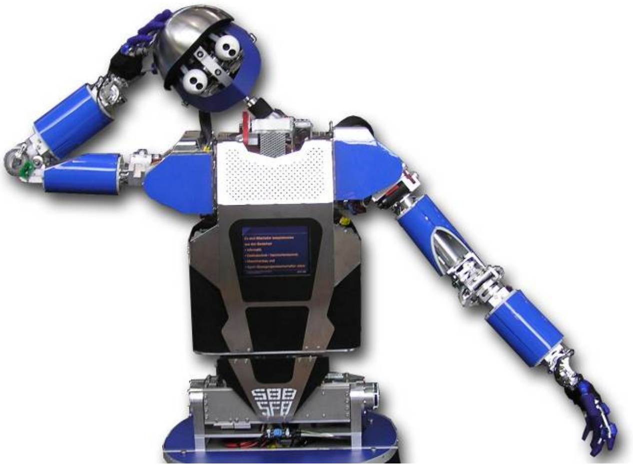


Fig. 12: Humanoid Robot ARMAR

Features of this robot system are a human-like motion system and intelligent control. In order to use the FRH-4 hand, manipulation tasks were performed in a kitchen environment. The experiments demonstrated that the hand is able to grasp and hold objects. Objects like a bottle, cups, drawer handles, and dishwasher handles were common objects for test grasping. The manipulation and grasping results as well as the controlling approach are described in [23, 24]. A picture of the newest ARMAR version is showed in Figure 12.

IV. Conclusion

The developed hand broadens the field of application compared to formerly developed hands [14, 15]. Especially the ability for fulfilling feedback and exploration tasks has been extended [22]. The described hand has position sensors in every joint and several touch sensors on every finger. To achieve higher flexibility as well as a higher degree of anthropomorphic appearance, actuators for abduction will be integrated, which will increase the number of DOF's to 16. This will also extend the amount of grasps that can be accomplished. Future work will include optimized constructive integration of the current sensors as well as the integration of pressure sensors in the hand. This extended sensory infrastructure will allow a more dynamical and precise control, as well as better exploration capabilities.

References

[1] United Nations Economic Commission for Europe. Press release ECE/STAT/05/P03, Geneva, 11 October 2005

[2] Specific targeted research or innovation project. "*Physical Human-Robot Interaction: Dependability and Safety"*. Information sheet. Available: http://www.phriends.eu

[3] IEEE and Robotics & Automation Society; Technical Committee on Service Robotics, http://www.service-robots.org, 2005

[4] R. D. Schraft, G. Schmierer: "Service Robots – Products, Scenarios, Visions". PETERS, NATICK Verlag, 2000. ISBN 1568811098

[5] G. Hirzinger: "A new Roboter-Generation for Space , Service and Surgery", *it – Information Technology,* Oldenbourg Wissenschaftsverlag, Vol. 49, No. 04/2007, pp. 247-259

[6] N. Furihata, S. Hirose: "Development of Mine Hands: Extended Prodder for Protected Demining Operation", *Autonomous Robot,* Springer Netherlands, Vol. 18, Number 3 / May, 2005, pp. 337-350

[7] T. Mouri, H. Kawasaki, K. Yoshikawa, J. Takai, S. Ito: "Anthropomorphic Robot Hand: Gifu Hand III", *ICCAS (2002),* Oct. 16th – 19th. Muiu Resort. Jeonbuk. Korea

[8] S. C. Jacobsen, E. K. Iversen, D. F. Knutti, R.T. Johnson and K. B. Biggers: "Design of the Utah/MIT Dextrous Hand", *ICRA (1986),* pp. 1520-1532

[9] http://www.shadowrobot.com

[10] http://www.festo.com/INetDomino/coorp_sites/en/ffeed49f2394ea43c12572b9006f7032.htm

[11] Lovchik, C., Diftler, M. "The Robonaut Hand: A Dexterous Robot Hand For Space," *Proceedings of the 1999 IEEE International Conference on Automation and Robotics*, Detroit, Michigan, May, 1999, pp 907-912

[12] I. Yamano, T. Maeno: "Five-fingered Robot Hand using Ultrasonic Motors and Elastic Elements", *Proceedings of the 2005 IEEE International Conference on Automation and Robotic,* Barcelona, Spain, April 2005, pp. 2684-2689

[13] R. Becher, P. Steinhaus, R. Dillmann: „The Collaborative Research Center: Humanoid Robots – Learning and Cooperating Multimodal Robots.", *Proceedings of the 2003 IEEE International Conference on Humanoid Robots,*Karlsruhe and Munich, Germany, 2003

[14] S. Schulz and G. Bretthauer: "A Fluidic Humanoid Robothand". *Proceedings of the IEEE-RAS International Conference on Humanoid Robots*, 2001

[15] S. Schulz, C. Pylatiuk and G. Bretthauer: "A New Ultralight Antropomorphic Hand." Proceedings of the 2001 IEEE International Conference on Robotics & Automation, Seoul, Corea, May, 2001

[16] S. I. Yaniger, "Force sensing resistors: A review of the technology," in *Electro International, 1991*, April 16-18, pp. 666-668.

[17] Interlink Electronics, *MicroNav Integration Guide,* v3.0, http://www.interlinkelectronics.com

[18] G. Cannata and M. Maggaiali, " An embedded tactile and force sensor for robotic manipulation and grasping," *2005 5th IEEE-RAS International Conference on Humanoid Robots*, Dec. 5, 2005, pp. 80-85

[19] NXP (former Phillips Semiconductors), I²C-bus specification and user manual, Rev. 03, 19th June 2007, http://www.nxp.com

[20] A. Kargov, C. Pylatiuk, S. Schulz: "Study of fluidic actuators in prosthetic hands." *10th International Conference on New Actuators & 4th International Exhibition on Smart Actuators and Drive Systems,* June 14-16, 2006, Bremen, Germany

[21] A. Kargov, H. Breitwieser, H. Klosek, C. Pylatiuk, S.. Schulz, and G. Bretthauer: „Design of a modular Arm Robot System based on Flexible Fluidic Drive Elements", *10th IEEE conference on Rehabilitation Robotics,* June 13-15, 2007, Netherlands

[22] Bierbaum, A.; Rambow, M.; Asfour, T. & Dillmann, R. A Potential Field Approach to Dexterous Tactile Exploration, Inproceedings of the *IEEE-RAS International Conference on Humanoid Robots 2008*, Daejeon, Korea, 2008

[23] T. Asfour, P. Azad, N. Vahrenkamp, K. Regenstein, A. Bierbaum, K. Welke, J. Schröder & R. Dillmann: „Toward humanoid manipulation in human-centred environments" *Robotics and Autonomous Systems*, North-Holland Publishing Co., Vol. 56, No. 01/2008, pp. 54-65

[24] N. Vahrenkamp, S. Wieland, P. Azad, D. Gonzalez, T. Asfour & R. Dillmann: "Visual Servoing for Humanoid Grasping and Manipulation Tasks", *Inproceedings of the IEEE-RAS International Conference on Humanoid Robots 2008*, Daejeon, Korea, 2008

Control and recognition on a humanoid head with cameras having different field of view

Aleš Ude
Jožef Stefan Institute, ABR
Jamova 39, Ljubljana, Slovenia
ales.ude@ijs.si

Tamim Asfour
University of Karlsruhe, ITEC
Haid-und-Neu-Strasse 7, Karlsruhe, Germany
asfour@ira.uka.de

Abstract

In this paper we study object recognition on a humanoid robotic head. The head is equipped with a stereo vision system with two cameras in each eye, where the cameras have lenses with different view angles. Such a system models the foveated structure of a human eye. To facilitate the pursuit of moving objects, we provide mathematical analysis that enables the robot to guide the narrow-view cameras toward the object of interest based on information extracted from the wider views. Images acquired by narrow-view cameras, which produce object images at higher resolutions, are used for recognition. The proposed recognition approach is view-based and is built around a classifier using nonlinear multi-class support vector machines with a special kernel function. We show experimentally that the increased resolution leads to higher recognition rates.

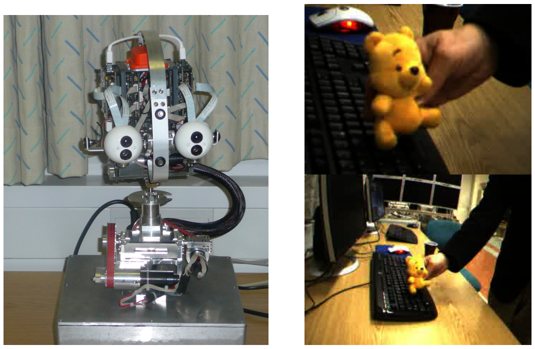


Figure 1. An example humanoid head (left). The narrow-angle cameras are positioned above the wide-angle ones. On the right are the images simultaneously taken from the wide- and narrow-angle view.

1 Introduction

Designers of a number of humanoid robots attempted to replicate human oculomotor system. For the optical part, this means that the optics should model the foveated structure of the human eye and allow *simultaneuos* processing of images of varying resolution. For the motor part, this means that the head must have suffficent mobility to perform typical eye movements such as smooth pursuit and saccades. Such an arrangement is useful because, firstly, it enables the robot to monitor and explore its surroundings in wide-angle views that contain most of the environment at low resolution, thereby increasing the efficiency of the search process. Secondly, it makes it possible to simultaneously extract additional information – once the area of interest is determined – from narrow-angle camera images that contain more detail. This kind of system is especially useful for object recognition on a humanoid robot. General object recognition is difficult because it requires the robot to detect objects in dynamic environments and to control the eye gaze to get the objects into the fovea and to keep them there. These tasks can be accomplished using information from wide-angle views, which enables the robot to determine the identity of the object by processing narrow-angle views.

There are various ways to construct humanoid vision system in hardware. The approach we followed is is to use two cameras in each eye equipped with lenses with different focal lengths [1, 4, 5]. This has the advantage of allowing us to use small-form cameras for the construction of the head.

2 Wide- and Narrow-Angle Views

The humanoid head of Fig. 1 has narrow-angle cameras rigidly connected to the wide-angle cameras and

978-1-4244-2175-6/08/$25.00 ©2008 IEEE

placed above them with roughly aligned optical axes. In the following we show that objects can be placed in the central field of view of narrow-angle cameras by bringing them to a certain position in the wide views. This position is displaced from the center of wide-angle camera images. The necessary displacement depends on the distance of the object from the cameras.

For theoretical analysis, we model both cameras by a standard pinhole camera model. The relationship between a 3-D point $\boldsymbol{M} = [X, Y, Z]^T$ and its projection $\boldsymbol{m} = [x, y]^T$ is given by

$$s\tilde{\boldsymbol{m}} = \boldsymbol{A}\tilde{\boldsymbol{M}},\ \boldsymbol{A} = \begin{bmatrix} \alpha & \gamma & x_0 \\ 0 & \beta & y_0 \\ 0 & 0 & 1 \end{bmatrix}. \qquad (1)$$

where $\tilde{\boldsymbol{M}} = [\boldsymbol{M}^T, 1]^T$, $\tilde{\boldsymbol{m}} = [\boldsymbol{m}^T, 1]^T$ are both points in homogeneous coordinates, s is an arbitrary scale factor, and $(\alpha, \beta, \gamma, x_0, y_0)$ are the intrinsic parameters of the camera. The world coordinate system is assumed to coincide with the camera coordinate system. In the following we assume without loss of generality that the origin of the image coordinate system coincides with the principal point (x_0, y_0), thus $x_0 = y_0 = 0$.

Let $\hat{\boldsymbol{t}}$ be the position of the origin of the wide-angle camera coordinate system expressed in the narrow-angle camera coordinate system and let $\hat{\boldsymbol{R}}$ be the rotation matrix that rotates the basis vectors of the wide-angle camera coordinate system into the basis vectors of the narrow-angle camera coordinate system. We denote by $\boldsymbol{M}_n$ and $\boldsymbol{M}_w$ the position of a 3-D point expressed in the narrow- and wide-angle camera system, respectively. We then have

$$\boldsymbol{M}_w = \hat{\boldsymbol{R}}(\boldsymbol{M}_n - \hat{\boldsymbol{t}}). \qquad (2)$$

The projections of a 3-D point $\boldsymbol{M}_n = (X, Y, Z)$ onto the planes of both cameras are given by

$$x_n = \frac{\alpha_n X + \gamma_n Y}{Z}, \qquad (3)$$

$$y_n = \frac{\beta_n Y}{Z}, \qquad (4)$$

and

$$x_w = \frac{\alpha_w \boldsymbol{r}_1 \cdot (\boldsymbol{M}_n - \hat{\boldsymbol{t}}) + \gamma_w \boldsymbol{r}_2 \cdot (\boldsymbol{M}_n - \hat{\boldsymbol{t}})}{\boldsymbol{r}_3 \cdot (\boldsymbol{M}_n - \hat{\boldsymbol{t}})}, \qquad (5)$$

$$y_w = \frac{\beta_w \boldsymbol{r}_2 \cdot (\boldsymbol{M}_n - \hat{\boldsymbol{t}})}{\boldsymbol{r}_3 \cdot (\boldsymbol{M}_n - \hat{\boldsymbol{t}})}, \qquad (6)$$

where $\boldsymbol{r}_1$, $\boldsymbol{r}_2$, and $\boldsymbol{r}_3$ are the rows of the rotation matrix $\hat{\boldsymbol{R}} = \begin{bmatrix} \boldsymbol{r}_1^T & \boldsymbol{r}_2^T & \boldsymbol{r}_3^T \end{bmatrix}^T$. $\boldsymbol{M}_n$ projects onto the principal point in the narrow-angle camera if $x_n = y_n = 0$. Assuming that the point is in front of the camera, hence $Z > 0$, we obtain from Eq. (3) and (4) that $X = Y = 0$, which means that the point must lie on the optical axis of the narrow-angle camera. Inserting this into Eq. (5) and (6), we obtain the following expression for the ideal position $(\hat{x}_w, \hat{y}_w)$ in the wide-angle camera image that results in the projection onto the principal point in the narrow-angle camera image

$$\hat{x}_w = \frac{\alpha_w \boldsymbol{r}_1 \cdot \hat{\boldsymbol{t}} + \gamma_w \boldsymbol{r}_2 \cdot \hat{\boldsymbol{t}} - (\alpha_w r_{13} + \gamma_w r_{23})Z}{\boldsymbol{r}_3 \cdot \hat{\boldsymbol{t}} - r_{33}Z}, \qquad (7)$$

$$\hat{y}_w = \frac{\beta_w \boldsymbol{r}_2 \cdot \hat{\boldsymbol{t}} - \beta_w r_{23} Z}{\boldsymbol{r}_3 \cdot \hat{\boldsymbol{t}} - r_{33}Z}, \qquad (8)$$

where $\begin{bmatrix} r_{13} & r_{23} & r_{33} \end{bmatrix}^T$ is the third column of $\hat{\boldsymbol{R}}$. Note that the ideal position in the periphery is independent from the intrinsic parameters of the foveal camera. It depends, however, on the distance of the point of interest from the cameras.

Utilizing these formulas we can turn the eye gaze towards the object and keep the object in the center of narrow-angle cameras based on information from wide-angle views. This is important because it is difficult to move the cameras quick enough to keep the object in the center of narrow-angle views. For this reason the object can easily be lost from narrow-angle views. Therefore it is advantageous to control the cameras using information from wide-angle views.

3 Learning Object Representations

We developed an object tracking system [6] that allows the robot to find objects of interest and locate them in the images. Using the formulas described in Section 2 and stereo vision, the robot can apply the results of the tracking process to center the object of interest in the narrow-angle view, where the object image has relatively high resolution. Since our tracker can estimate both the location and scale of the object in the image, we can warp, i. e. translate, rotate and scale along the principal axes, the object images to a window of constant size.

Our goal is to learn a view-based representation for all available objects. To achieve this, it is necessary to show the objects to the humanoid from all relevant viewing directions. In computer vision this is normally achieved by accurate turntables that enable the collection of images from regularly distributed viewpoints. However, this solution is not practical for humanoid robotics, where on-line interaction is often paramount. We therefore explored whether it is possible to reliably learn models from images collected while a human teacher randomly moves the object in front of the robot. In this case the training process is started by a teacher who moves the object to be learnt in front of the robot. Snapshots from various viewpoints are collected and processed. Warping the snapshots onto a window of constant size ensures invariance against scaling and planar rotations.

To ensure maximum classification performance, the data is further processed before training a general classifier. Most modern view-based approaches characterize the views by ensembles of local features. We use complex Gabor kernels to identify local structure in the images. A Gabor jet at pixel $\boldsymbol{x}$ is defined as a set of complex coefficients $\{J_j^{\boldsymbol{x}}\}$ obtained by convolving the image with a number of Gabor kernels at this pixel. The kernels are normally selected so that they sample a number of different wavelengths k_ν and orientations ϕ_μ. Wiskott et al. [7] proposed to use $k_\nu = 2^{-\frac{\nu+2}{2}}$, $\nu = 0, \ldots, 4$, and $\phi_\mu = \mu\frac{\pi}{8}$, $\mu = 0, \ldots, 7$, but this depends both on the size of the incoming images and the image structure. They showed that the similarity between the jets can be measured by

$$\mathrm{S}\left(\{J_i^{\boldsymbol{x}}\}, \{J_i^{\boldsymbol{y}}\}\right) = \frac{\boldsymbol{a}_{\boldsymbol{x}}^T * \boldsymbol{a}_{\boldsymbol{y}}}{\|\boldsymbol{a}_{\boldsymbol{x}}\| \|\boldsymbol{a}_{\boldsymbol{y}}\|}, \qquad (9)$$

where $\boldsymbol{a}_{\boldsymbol{x}} = [|J_1^{\boldsymbol{x}}|, \ldots, |J_s^{\boldsymbol{x}}|]^T$ and s is the number of complex Gabor kernels. This is based on the fact that the magnitudes of complex coefficients vary slowly with the position of the jet in the image.

Our system builds feature vectors by sampling Gabor jets on a regular grid of pixels $\boldsymbol{X}_G$. At each grid point we calculate the Gabor jet and add it to the feature vector. The grid points need to be parsed in the same order in every image. The grid size used in our experiments was 6×6, the warped image size was 160×120 with pixels outside the enclosing ellipse excluded, and the dimension of each Gabor jet was 40, which resulted in feature vectors of dimension 16080. These feature vectors were supplied to the SVM for training.

4 Nonlinear Multi-Class SVMs

Utilizing the similarity measure (9), we developed a classifier for object recognition based on nonlinear multi-class support vector machines. Nonlinear multi-class support vector machines (SVMs) [2] use the following decision function

$$\boldsymbol{H}(\boldsymbol{x}) = \arg\max_{r \in \boldsymbol{\Omega}} \left\{ \sum_{i=1}^{m} \tau_{i,r} \mathrm{K}(\boldsymbol{x}_i, \boldsymbol{x}) + b_r \right\}. \qquad (10)$$

Here $\boldsymbol{x}$ is the input feature vector to be classified (in our case a collection of Gabor jets), $\boldsymbol{x}_i$ are the feature vectors supplied to the SVM training, $\tau_{i,r},\ b_r$ are the values estimated by SVM training, and $\Omega = \{1, \ldots, N\}$ are the class identities (objects in our case). The feature vectors $\boldsymbol{x}_i$ with $\tau_{i,r} \neq 0$ are called the support vectors. The SVM training consists of solving a quadratic optimization problem whose convergence is guaranteed for all kernel functions K that fulfill the Mercer's theorem.

The similarity measure for Gabor jets (9) provides a good motivation for the design of a kernel function for the classification of feature vectors consisting of Gabor jets. Let $\boldsymbol{X}_G$ be the set of all grid points within two normalized images on which Gabor jets are calculated and let $J_{\boldsymbol{X}_G}$ and $L_{\boldsymbol{X}_G}$ be the Gabor jets calculated in two different images, but on the same grid points. A suitable kernel function can be defined as follows

$$\begin{aligned}&\mathrm{K}_G(J_{\boldsymbol{X}_G}, L_{\boldsymbol{X}_G}) = \\ &\exp\left(-\rho\frac{1}{M}\sum_{\boldsymbol{x}\in\boldsymbol{X}_G}\left(1 - \frac{\boldsymbol{a}_{\boldsymbol{x}}^T * \boldsymbol{b}_{\boldsymbol{x}}^T}{\|\boldsymbol{a}_{\boldsymbol{x}}\|\|\boldsymbol{b}_{\boldsymbol{x}}\|}\right)\right),\end{aligned} \qquad (11)$$

where M is the number of grid points in $\boldsymbol{X}_G$. This function satisfies the Mercer's condition [2] and can thus be used for support vector learning. Parameter ρ needs to be supplied experimentally.

5 Experimental Results

We used a set of ten objects to test the performance of the developed recognition system on a humanoid robot. For each object we recorded two or more movies using a video stream coming from the narrow-angle cameras, which were controlled by information acquired from wide-angle views. In each of the recording sessions the teacher attempted to show one of the objects to the robot from all relevant viewing directions. One movie per object was used to construct the SVM classifier, while one of the other movies was used to test the classifiers. Each movie was one minute long and we used at most 4 images per second for training. Since slightly more than first ten seconds of the movies were needed to initialize the tracker, we had at most 208 training images per object. For testing we used 10 images per second, which resulted in 487 test images per object. All the percentages presented here were calculated using the classification results obtained from 4870 test images. Gabor jets were calculated as proposed by Wiskott et al. [7] and the grid size was 6 pixels in both directions. The filters were scaled appropriately when using lower resolution images. To show the usefulness of foveated vision for recognition, we tested the performance of the system on images of varying resolution. We also compared the developed SVM-based classifier with the nearest neighbor classifier (NNC) that uses the similarity measure (9) – summed over all grid points – to deterrmine the class of the nearest neighbor by comparing Gabor jets directly.

Table 1. Correct classification rate (image resolution 120×160 pixels)

Training views per object	SVM	NNC
208	97.6 %	95.9 %
104	96.7 %	93.7 %
52	95.1 %	91.5 %
26	91.9 %	86.7 %

Table 2. Correct classification rate (image resolution 60×80 pixels)

Training views per object	SVM	NNC
208	94.2 %	89.3 %
104	92.4 %	87.3 %
52	90.7 %	84.4 %
26	86.7 %	79.2 %

Table 3. Correct classification rate (image resolution 30×40 pixels)

Training views per object	SVM	NNC
208	91.0 %	84.7 %
104	87.2 %	81.5 %
52	82.4 %	77.8 %
26	77.1 %	72.1 %

Results in Tables 1 - 3 prove that foveation is very useful for recognition. The classification results clearly become worse with the decreasing resolution. Our results also show that we can collect enough training data even without using accurate turntables to systematically collect the views. As expected the recognition rate decreases with the number of images, but we can conclude that collecting the training views statistically is sufficient to build models for 3-D object recognition.

The presented results cannot be directly compared to the results on standard databases for benchmarking object recognition algorithms because here the training sets are much less complete. Some of the classification errors are caused by the lack of training data rather than by a deficient classification approach. Unlike many approaches from the computer vision literature that avoid the problem of finding objects, we tested the system on images obtained through a realistic object tracking and segmentation procedure. Only such data is relevant for foveated object recognition because without some kind of segmentation it is not possible to direct the fovea towards the objects of interest.

6 Conclusions

Our experiments demonstrate that by exploiting the properties of a humanoid vision we can construct an effective object recognition system. Wide-angle views are necessary to search for objects, direct the gaze towards them and keep them in the center of narrow-angle views. Narrow-angle views provide object images at a higher resolution, which significantly improves the recognition rate. Having both views at the same time is essential. Most of previous approaches that employed support vector machines for object recognition used binary SVMs combined with decision trees [3]. Our system makes use of nonlinear multi-class SVMs to solve the multi-class recognition problem. By normalizing the views with respect to scale and planar rotations based on the results of the tracker, we were able to reduce the amount of data needed to train the SVMs. Object representations can be learnt just by collecting the data statistically while the demonstrator attempts to show the objects from all relevant viewing directions. Experimental results show high recognition rates in realistic test environments.

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.

References

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

[2] K. Crammer and Y. Singer. On the algorithmic implementation of multiclass kernel-based vector machines. *J. Machine Learning Research*, 2:265–292, 2001.

[3] G. Guo, S. Z. Li, and K. L. Chan. Support vector machines for face recognition. *Image and Vision Computing*, 19(9-10):631–638, 2001.

[4] H. Kozima and H. Yano. A robot that learns to communicate with human caregivers. In *Proc. Int. Workshop on Epigenetic Robotics*, Lund, Sweden, 2001.

[5] B. Scassellati. A binocular, foveated active vision system. Technical Report A.I. Memo No. 1628, MIT, Artificial Intelligence Laboratory, 1999.

[6] A. Ude and C. G. Atkeson. Probabilistic detection and tracking at high frame rates using affine warping. In *Proc. 16th Int. Conf. Pattern Recognition, Vol. II*, pages 6–9, Quebec City, Canada, 2002.

[7] L. Wiskott, J.-M. Fellous, N. Krüger, and C. von der Malsburg. Face recognition by elastic bunch graph matching. *IEEE Trans. Pattern Anal. Machine Intell.*, 19(7):775–779, 1997.

Kinematic Calibration for Saccadic Eye Movements

Kai Welke, Markus Przybylski, Tamim Asfour and Rüdiger Dillmann
University of Karlsruhe (TH), IAIM, Institute of Computer Science and Engineering (CSE)
P.O. Box 6980, 76128 Karlsruhe, Germany
{welke,przybyls,asfour,dillmann}@ira.uka.de

Paper-ID 70

***Abstract*— The visual perception system of humanoid robots should provide sensorial information that fulfills the requirements imposed by perceptual tasks in natural environments. One prerequisite for such systems is the ability to observe the environment by actively moving its visual sensors. This ability allows to implement two essential behaviors for a cognitive visual system: smooth pursuit and saccadic eye movement.**

In the work presented in this paper we propose a kinematic calibration approach for the active camera system of the Karlsruhe Humanoid Head. The proposed method solves two fundamental problems when performing saccadic eye movements: the required kinematic model for open-loop control and the ability of stereo reconstruction with active cameras. We present experiments on the accuracy of the kinematic model, the stereo triangulation and the saccadic eye movement.

I. Introduction

Most current humanoid robots have simplified head-eye 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, humanoid systems that are able to execute manipulation and grasping tasks, that interact with humans and learn from human observation require sophisticated perception systems, which are able to fulfill the therewith associated requirements. The Karlsruhe Humanoid Head [4] (see fig.1) used in this work offers an active vision system with two cameras, which can be moved independently. This ability is usually exploited to implement behaviors that are essential to common visual perception tasks, namely smooth pursuit and saccadic eye movements.

The smooth pursuit behavior (also called fixation or tracking) consists in focusing on a known combination of visual stimuli, e.g. the face of a person during interaction. The common control strategy deployed in such behaviors is closed-loop control, where the prior knowledge of the visual stimuli is exploited (see e.g. [5]).

Saccadic eye movements play an important role in the serialisation of visual information processing within a perceived environment. Saccadic eye movements are usually initiated by attention mechanisms, in order to focus on salient parts of the scene. In such mechanisms, the target of the saccade is determined by the spike of a single neuron ([6],[7]). The information necessary for closed-loop control is not available. Consequently, open-loop control strategies have to be provided in order to execute saccadic eye movements.

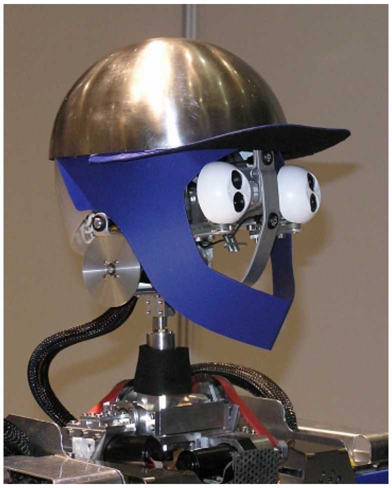


Fig. 1. The Karlsruhe Humanoid Head offers an active camera system.

Another problem arising from the application of active stereo camera systems is the ability of performing stereo reconstruction. When working with fixed eyes, the stereo calibration required for stereo triangulation is fixed and only has to be calculated once. In contrast the calibration for active stereo camera systems varies with each actuation of the eyes thus making static stereo calibration impossible.

In this work we present a new method for kinematic calibration of an active camera system which solves the problem of open-loop control for saccadic eye movements as well as the stereo calibration problem with actuated cameras. The calibration procedure results in a kinematic model which allows to solve the inverse kinematics problem for the eye system required for saccadic eye movements as well as the calculation of the stereo calibration required for stereo vision.

The paper is organized as follows. In Section II an overview of the different approaches for the problem classes of head-eye and hand-eye calibration in the literature is given and the approaches are analyzed according to their feasibility for our problem. In Section III a brief description of the system configuration is provided. Section IV describes the proposed approach for kinematic calibration. In Section V, the proposed method is evaluated on the Karlsruhe Humanoid Head. We provide experiments on the accuracy of the kinematic model itself, the accuracy of the stereo calibration and the accuracy of saccadic eye movements.

II. Related Work

The literature offers a variety of methods to solve the two related problems of head-eye and hand-eye calibration. Most of them are based on the traditional $AX = XB$ and $AX = ZB$ formulations of correspondences between coordinate frames in the system to be calibrated.

The $AX = XB$ formulation arises from the problem of head-eye calibration. In this case A denotes the coordinate transformation between two distinct camera positions when moving the camera by a transformation B of a joint. X denotes the unknown relationship from the actuated joint to the camera which is to be determined.

The $AX = ZB$ formulation has been proposed for the case of hand-eye calibration problems. Here A describes the transformation from the camera to the world frame. B denotes the transformation between the robot's hand coordinate frame and the base coordinate frame. The unknowns to be determined are the hand-to-camera transformation X and the base-to-world transformation Z.

Considering different methods which make use of the $AX = XB$ and $AX = ZB$ formulations, there are two essential decisions to be made when developing a method based on them. First, there are different possibilities to model the rotational parts of the involved coordinate transformations. The proposed models comprise Euler angles and quaternions as well as representations using a rotation axis and an angle or such based on Lie theory. The second decision concerns the mathematical method, which is used to actually find a solution for the unknown coordinate transformation, given its representation. Most solutions are based on linear or non-linear least squares optimization methods. Other approaches suggest the use of Lagrange multipliers or avoid any kind of optimization.

Some of the most important contributions in the field can be categorized in the following way. Tsai and Lenz [8] use an axis-and-angle representation for the rotation matrices. They solve separately for rotation and translation and present a linear least squares solution for both. Shiu and Ahmad [9] use a similar representation as Tsai and Lenz but develop a different linear solution. Li's method ([10],[11]) uses rotation matrices to model the problem. In his experiments, these matrices are based either on Euler angles or quaternions. He uses a non-linear optimization approach for the rotational part and a linear least squares approach for the translational part. The method by Neubert and Ferrier [12] uses Lie theory to model the problem and solves simultaneously for rotation and translation using a linear least squares approach. Horaud and Dornaika [13] present two methods, both of them based on quaternion representations. One is a closed-form approach using Lagrange multipliers. The other one is a non-linear least squares approach which solves for all unknowns at once. A completely different approach is presented by Young [14]. He does not refer to the $AX = XB$ and $AX = ZB$ formulations. Instead he uses a combination of a modified Denavit-Hartenberg (DH) convention and screw theory. No optimization is required. The method calibrates one joint at a time and can be used for any type of kinematic chain.

There are several papers which provide comparisons of the mentioned approaches to determine which method produces the most accurate results. Horaud and Dornaika [13] compared a linear and a non-linear least squares approach as well as an approach using Lagrange multipliers. Li [11] compared linear least squares approaches by Dornaika [15] and Tsai [8] with his own non-linear least squares approach. The results can be summarized as follows. The representation of the unknown rotation does not seem to be essential. In comparison the choice of the method to solve for the unknowns has a more significant impact on calibration precision. Non-linear least squares methods yield the most accurate results which is attributed to the degrading performance of linear least squares methods in the presence of noise. The downside of approaches that solve separately for rotation and translation is the fact that in these two-step methods the error propagates from the first part to the second part. Therefore it is reasonable to estimate all unknowns simultaneously.

Consequently, an advantage of the $AX = XB$ and $AX = ZB$ formulations is the fact that one does not have to resort to sophisticated representations of rotations. Simple representations like Euler angles provide very good results. However, there are some serious disadvantages. As proved several times ([8],[9],[16]), the equations $AX = XB$ and $AX = ZB$ have two degrees of freedom. A unique solution can only be found if two rotations around non-parallel axes of rotation are performed. This means that single joints with only one degree of freedom can not be calibrated. Each joint must have at least two degrees of freedom. An elegant solution to this issue is to combine two joints with one degree of freedom each and to treat them as one single joint with two degrees of freedom. Although most authors do not explicitly state it, this is only possible if the axes of the two respective joints intersect, because only this way a single common coordinate frame for both joints can be assigned to the point of intersection. Therefore the class of kinematic chains that can be calibrated using this approach is restricted. But even kinematic chains which according to the design schematics fulfill this condition may, due to production imprecisions, in practice not be accurate enough. In this case the simplifying assumption of intersecting joint axes is in fact a methodical error, resulting in inferior calibration accuracy.

The approach by Young [14] is more universal. Based on the Denavit-Hartenberg convention, it can be applied to any kind of kinematic chain. No simplifying assumptions are made. Therefore it should yield more accurate results than the approaches described above. Moreover, no optimization of any kind is used. However, the Denavit-Hartenberg convention always assigns the z axes to the axes of rotation or translation, which might not always be desirable.

In this paper an approach is suggested that combines the advantages of $AX = XB$ based methods and a DH-based approach. In contrast to the DH convention, the rotation axes are not necessarily assigned to the z axes which allows to choose arbitrary coordinate frames for each joint. No simplifying assumptions are made concerning the relationships between adjacent joints. The proposed method avoids to consider two or more distinct joints as one joint with multiple degrees of freedom. Instead, every single joint is calibrated separately. That way the proposed approach can be applied to a wider class of kinematic chains. An $AX = XB$ based formulation is used to derive a non-linear target function that is minimized using the method by Levenberg and Marquardt [17]. For each joint to be calibrated, all necessary unknowns are estimated simultaneously, avoiding error propagation between them.

III. System Configuration

The Karlsruhe Humanoid Head used for our work has seven rotational degrees of freedom (see Fig.1). Four of them are used to move the head: neck roll, neck yaw, neck pitch and head tilt. The eyes are actuated by a common tilt joint and two independent pan joints. The vision system consists of two stereo camera pairs and mimics the foveated structure of the human eye. Therefore each eye contains one perspective camera with a wide angle of view and a foveal camera with a small angle of view. In both cases Dragonfly cameras from PointGrey are mounted which are accessed via an IEEE1394 interface and provide a maximum frame rate of 30 fps at a resolution of 640 × 480 pixels [1]. For the experiments presented in this paper the perspective cameras were used, which were outfitted with lenses with a focal length of 4 mm.

IV. Kinematic Calibration

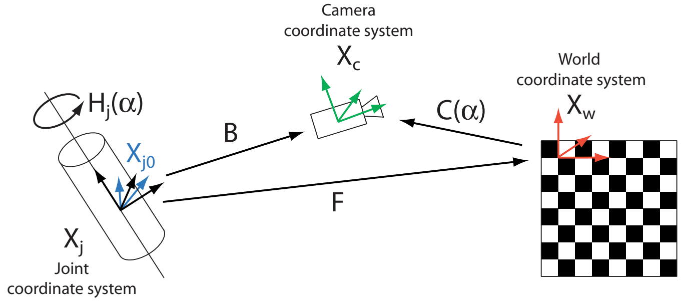


Fig. 2. Coordinate systems and transformations involved in the kinematic calibration procedure.

In this section, a detailed formal description of the proposed method is given. We proceed in the following way:

First, all involved coordinate systems and transformations are introduced. Then the acquisition of data for the necessary extrinsic camera calibration is explained. Based on this data the kinematic calibration problem is formulated as a non-linear least squares problem, which is solved using the method by Levenberg and Marquardt. Having determined the kinematic calibration of both the left and right eye pan joints, a stereo calibration for arbitrary angles of these joints can be calculated.

A. Coordinate Systems and Transformations

Throughout this paper, the following conventions for coordinate systems and transformations will be used (see figure 2):

- X_w denotes the world coordinate system. It is a fixed system which is used as a reference to determine the extrinsic camera calibrations. The world coordinate system is defined by the calibration pattern.
- X_{j_0} is the coordinate system in the joint j at zero position. It is fixed.
- X_j denotes the rotated coordinate system of the joint j after actuation.
- X_c is the camera coordinate frame. As the camera is rigidly attached to the pan joint, the camera system moves when the joint is actuated.
- $H_j(\alpha)$ describes the coordinate transformation from the fixed joint coordinate frame X_{j_0} to the rotated joint coordinate frame X_j. As the considered joints have one degree of freedom, $H_j(\alpha)$ describes a rotation around the one rotation axis of the respective joint by an angle α which is obtained from the encoder readings.
- B denotes the transformation from the rotated joint coordinate system X_j to the camera coordinate system X_c. As the joint movement is already described by the transformation H_j, B remains constant, independent of the actual joint position. The goal of the calibration process is to determine this transformation.
- F denotes the transformation from the fixed joint coordinate frame X_{j_0} to the world coordinate system X_w. As none of these systems moves, F is constant.
- $C(\alpha)$ is the transformation from the world coordinate frame X_w to the camera coordinate frame X_c. It depends on the camera position and therefore on the joint angle α. C is usually called the extrinsic camera calibration.

B. Extrinsic Camera Calibration Data Acquisition

The kinematic calibration process is performed on the basis of a set of extrinsic camera calibrations $C(\alpha_1)...C(\alpha_n)$ at different rotations of the joint to be calibrated. The matrices C are stored together with the corresponding joint angles α. As prerequisite for the calculation of extrinsic calibration data, the intrinsic camera parameters for each camera have to be determined. For this purpose the intrinsic camera calibration procedure proposed by Zhang [18] is used.

C. Kinematic Calibration

Once extrinsic camera calibration data has been aquired at different angle positions of the joint to be calibrated, the goal of the approach consists in determining the kinematic calibration matrix B. In our approach, the matrix B is calculated using a non-linear least squares minimization technique. More precisely the Levenberg-Marquardt algorithm [17] is deployed to determine B from the set of extrinsic calibrations C and the corresponding joint angles α. The Levenberg-Marquardt algorithm minimizes a target function which is derived in

[1] http://www.ptgrey.com

section IV-C.1. The representation and parameterization of the calibration matrix are discussed in section IV-C.2.

1) The Target Function: The target function which is minimized in order to determine the kinematic calibration matrix B is formulated using homogeneous matrices for all necessary coordinate transformations. A geometric interpretation of these transformations is given in figure 2.

The coordinate transformation from the fixed joint frame to the rotated joint frame is

$$X_j = H_j X_{j_0}. \tag{1}$$

A transformation from the rotated joint frame to the camera coordinate frame can be written as

$$X_c = B X_j. \tag{2}$$

The transformation from the world coordinate system to the camera coordinate system depends on the position of the camera and therefore on the angle α of the joint the camera is attached to. This can be formulated as

$$X_c = C(\alpha) X_w. \tag{3}$$

The transformation from the fixed joint coordinate system to the world reference system is

$$X_w = F X_j. \tag{4}$$

By combining equations (1), (2), (3) and (4), F can also be expressed as

$$F = C(\alpha)^{-1} B H_j(\alpha). \tag{5}$$

As the world coordinate frame X_w and the joint coordinate frame at zero position X_{j_0} never change, different transformations F_i and F_k with

$$F_i = C(\alpha_i)^{-1} B H_j(\alpha_i) \tag{6}$$

and

$$F_k = C(\alpha_k)^{-1} B H_j(\alpha_k) \tag{7}$$

can be calculated for different joint angles α_i and α_k, but the condition

$$F_i = F_k \tag{8}$$

always holds.

In practice however, the extrinsic camera calibrations and the joint encoder readings are not entirely accurate. Due to these and other errors it is impossible to find a B which satisfies equation (8). Instead it is the goal to find a B which minimizes the error

$$||F_i - F_k||_f = ||C(\alpha_i)^{-1} B H_j(\alpha_i) - C(\alpha_k)^{-1} B H_j(\alpha_k)||_f. \tag{9}$$

In this context $||.||_f$ denotes the Frobenius norm [19]. Let N be the number of extrinsic camera calibrations determined using the process described in section IV-B. Furthermore, let $\vec{x}$ be a parameterization of B and $G_r(\vec{x})$ and $G_t(\vec{x})$ be the minimization functionals which express the rotational and translational difference of two transformations F_k and F_{k+1} belonging to two external camera calibrations at two adjacent joint positions α_k and α_{k+1} , i.e,

$$\begin{aligned} G_r(\vec{x})_k &= angle(F_{k+1}, F_k) \\ G_t(\vec{x})_k &= translation(F_{k+1}, F_k) \end{aligned}$$

Each pair $G_r(\vec{x})$,$G_t(\vec{x})$ describes the rotational and translational difference between the homogeneous matrix F_k and F_{k+1}. To find a solution for B means to solve the minimization problem

$$\min_{\vec{x}} \sum_{k=1}^{N-1} ||w_r G_r(\vec{x})_k + w_t G_t(\vec{x})_k||, \tag{10}$$

where w_r and w_t are weighting factors for the rotational and translational parts of the error.

2) Representation and Parameterization of the Calibration Matrix: The goal of the kinematic calibration procedure is to assign a coordinate system to the rotation axis of the actuated joint. In this context two decisions have to be made:

- The representation of the calibration matrix B
- The parameterization of the calibration matrix B

The representation of B describes the mathematical means used to model the rotational part of the coordinate transformation B whereas the parameterization of B deals with the question which components of B actually have to be estimated in order to find a meaningful solution to the kinematic calibration problem. Both the choice of a representation and a parameterization of B is necessary to compute the minimization functionals $G_r(\vec{x})$ and $G_t(\vec{x})$ introduced in section IV-C.1.

For this work a three-angle representation was used for the rotational part of B. The three elementary rotations were concatenated using the Roll-Pitch-Yaw convention.

According to [8], [9] and [16], two independent axes of rotation are necessary to determine all six parameters of B. When using only one rotation around a single rotation axis, the problem is under-determined. When doing rotations around one axis and calculating the extrinsic camera calibrations at different angle positions, these extrinsic calibrations contain sufficient information to identify the rotation axis. However, there is not enough information to determine all components of the position and orientation of the joint coordinate frame on this axis. The origin of the coordinate frame on the axis is not uniquely determined. Furthermore the orientation of the coordinate frame is only restricted in a way that one coordinate axis points in the direction of the joint axis, while the other two coordinate axes can be chosen in a way that the resulting coordinate frame is a right-handed system.

In order to determine the stereo calibration the exact positions and orientations of the coordinate frames on the joint axes are not necessary. For this purpose the partial solution explained above is sufficient. It is even possible to calibrate all joints of the head-eye system using this type of partial solution. If the complete head-eye system is to be calibrated it is necessary to registrate the last coordinate system in the head with the world coordinate system. However, the registration with the world coordinate system is always necessary, no matter if partial or complete solutions were determined for the individual joints' kinematic calibrations. Regarding the considerations above, arbitrary values can be used for the two undetermined components of the transformation, always resulting in a valid calibration matrix B.

As stated above, the parameterization of B also depends on which coordinate axis is assigned to the joint's axis of rotation. If its axis of rotation is the y axis, as shown in figure 3, the parameterization is $\vec{x} = (\alpha, \gamma, t_x, t_z)$, where α and γ denote elementary rotations around the x and z axes and t_x and t_z describe translations along the respective axes. The β and t_y components are not estimated and set to zero.

D. Stereo Calibration

The stereo calibration is required to enable methods of stereo vision on the Karlsruhe Humanoid Head. In order to use epipolar geometry to recover 3D positions of corresponding points from a stereo camera pair, the stereo calibration is required. For static cameras, the stereo calibration is usually calculated using the extrinsic camera calibrations of both cameras. The relative position of both camera coordinate systems H_{stereo} can be derived directly from the extrinsic calibrations. Having performed the kinematic calibration as described above H_{stereo} can be determined for arbitrary camera poses. This allows to perform stereo vision if the eye joints are actuated. First, in order to calculate H_{stereo} the transformation $H_{epl2epr}$ is calculated (see fig. 3) in the following way:

$$\begin{aligned} H_{epl2epr} \quad = \quad & H_{epr}^{-1}(\alpha_R) \cdot B_R^{-1} \cdot C_R(\alpha_R) \cdot \\ & C_L^{-1}(\alpha_L) \cdot B_L \cdot H_{epl}(\alpha_L). \end{aligned} \qquad (11)$$

The matrices B_R and B_L represent the kinematic calibrations of both eye pan joints. $H_{epr}(\alpha_R)$ and $H_{epl}(\alpha_L)$ model the rotations of the respective joints by certain angles α_R and α_L. The external camera calibrations depend on the same angles. In theory the transformation $H_{epl2epr}$ could be determined at any position of the two joints. In practice however, modeling the joints' movements using $H_{epr}(\alpha_R)$ and $H_{epl}(\alpha_L)$ introduces errors. Therefore, the most accurate result can be obtained with both joints at their home positions at $\alpha_L = \alpha_R = 0$ where $H_{epr}(\alpha_R)$ and $H_{epl}(\alpha_L)$ become identity.

The stereo calibration H_{stereo} for arbitrary joint angles can then be calculated using the following equation:

$$H_{stereo} = B_L \cdot H_{epl}(\alpha_L) \cdot H_{epl2epr}^{-1} \cdot H_{epr}(\alpha_R)^{-1} \cdot B_R^{-1} \qquad (12)$$

V. Experimental Results

Prior to the experiments, the optimal distance of the calibration pattern for the kinematic calibration was determined. Therefore different calibrations were performed with different positions of the calibration pattern. We used a pattern with 9×7 squares, side lenght 3.63 cm each, at distances of 0.50 m, 1.00 m, and 1.35 m from the eye system. For each calibration, the translational error Δt between the measured positions of the calibration pattern and the positions calculated using the calibrated kinematic model was measured. Table I shows the results of the experiments. As expected the accuracy of the calibration decreases with increasing distance of the calibration pattern. For the following experiments, we used a calibration distance of 0.80 m. With this distance the calibration pattern was still visible with eye pan actuations between -15 and 15 degrees and eye tilt actuations between -10 and 10 degrees. For the kinematic calibration we collected extrinsic data at steps of 1.5 degrees for the pan joints and 1.0 degrees for the tilt joint. Furthermore we evaluated the best weighting between rotational and translational error for the optimization. The best results could be achieved for the values $w_r = 0.5$ and $w_t = 1.0$. With these settings, a rotational error of 2 degrees corresponds to a translational error of 1 mm.

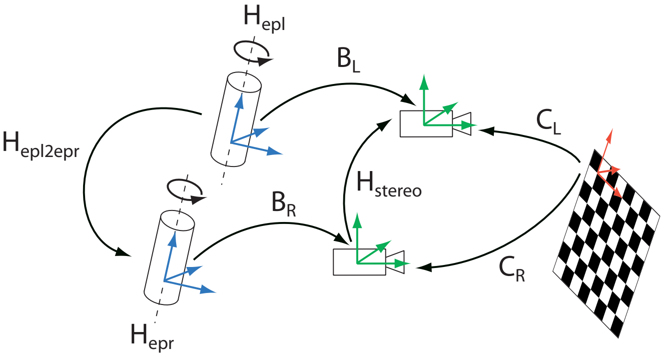


Fig. 3. The coordinate transformations necessary to determine the stereo calibration.

TABLE I

IMPACT OF THE DISTANCE TO THE CALIBRATION PATTERN ON THE MEAN MAXIMUM TRANSLATION ERROR Δt FOR DIFFERENT DISTANCES.

Calibration distance (m)	**Error** Δt (mm)
0.50	1.49
1.00	3.41
1.35	5.10

In the following we present experiments on the kinematic accuracy, the stereo accuracy and the accuracy of open-loop control.

A. Kinematic Calibration Accuracy

In a first series of experiments we investigated how accurate the kinematic model of the two pan joints is determined with the proposed method. Therefore we used a smaller calibration pattern with 5×4 squares, side length 4.5 cm. In the experiments, we performed arbitrary eye pan movments in the calibrated range of the eyes. The test pattern was positioned at distances ranging from 60 cm up to 140 cm from the eye system. For each distance 50 random test eye poses were recorded.

In order to determine the accuracy of the kinematic model, we located the 3D pose of the test pattern in the left and in the right camera using a model-based approach. Based on the calibrated kinematic model, both poses were transformed

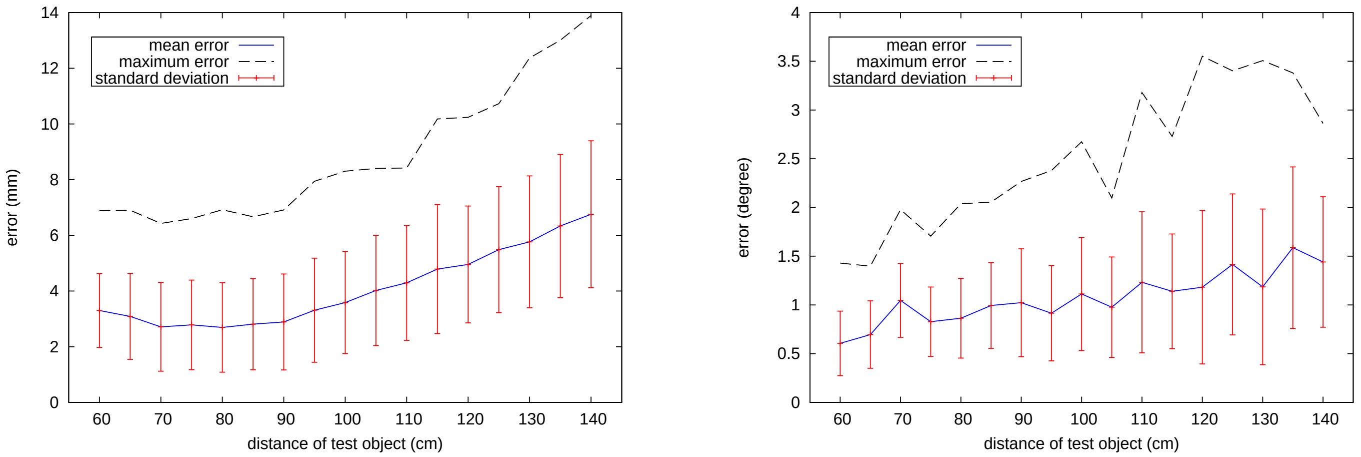


Fig. 4. Accuracy of the proposed kinematic calibration. The 3D pose of a test pattern was determined in the left and right perspective camera using a model-based approach. Both poses were transformed to a common coordinate frame using the calibrated model. The plots show the translational and rotational error for different distances of the test pattern.

into a common coordinate system and the translational and rotational errors were measured. Fig. 4 shows the results of this experiment. The plots illustrate the mean error, the standard deviation of the error and the maximum error for each distance and for the rotational and translational parts of the error. As can be seen, the major trend is a decreasing accuracy of the kinematic model with increasing distance of the test pattern. The plot for the mean of the translational part has its minimum of 2.34 mm at 80 cm - the distance where the kinematic calibration was calculated. The mean error reaches its maximum in the distance of 140 cm (6.85 mm mean error). The rotational error ranges from 0.58 up to 1.62 degrees.

B. *Stereo Calibration Accuracy*

The accuracy of the stereo calibration was tested in a similar way. Again the position of the test pattern was determined in the left camera using a model-based approach. Additionally, three corresponding corner points of the calibration pattern in the left and right image were determined and the 3D position of these points was calculated utilizing the epipolar geometry based on the calibrated transformation matrix H_{stereo}. From these three points, the pose of the test rig in the left camera was estimated. In that way the accuracy of the stereo triangulation could be evaluated with respect to the model based approach.

The experiments were performed for the test pattern located at distances from 60 cm up to 140 cm from the eye system, in each step 50 random test poses were recorded and evaluated. Fig. 5 shows the resulting translational and rotational error between the pose using the model-based approach and the pose calculated based on stereo triangulation. As can be seen the errors in the stereo triangulation accuracy show the same trend as the kinematic error, but are much larger. The increase of the error results from the fact, that small position errors (within the kinematic calibration) result in larger errors when performing stereo triangulation. The best results could again be achieved for small distances of the test pattern. Within a test pattern distance of 70 cm the mean translational error was measured with 8.7 mm and the minimum mean rotational error was measured with 1.72 degrees for the same distance.

C. *Inverse Kinematics Accuracy*

In the third series of experiments we tested the kinematic calibration in saccadic eye movement tasks. Therefore we calibrated the first three DoF of the head-eye system (namely eye pan left, eye pan right and eye tilt). For these joints, conventional differential inverse kinematics based on the inverse Jacobian could be deployed, since the kinematic system is not redundant (see also [4]). In the experiments in this section we again evaluated the accuracy at different distances of the test rig. For each distance, arbitrary camera poses were generated by moving three head joints (neck pitch, neck roll and neck yaw) to random positions in the interval of -10 to 10 degrees for each joint. The pose of the test pattern in the left camera was again determined using a model-based approach. Using the calibrated kinematic model, the inverse kinematic problem was solved and a saccade was performed in order to point the optical axes of the cameras towards the origin of the test pattern. In order to evaluate the accuracy, the distance between the corresponding corner of the test rig and the principal point of the camera in the image plane were measured, after performing the movement. Since left and right camera share a common tilt joint, the error in y direction will never be zero. The inverse kinematics module outputs the mean eye tilt actuation for left and right eye to minimize the overall error. In order to compensate for this effect, we used a modified error in y direction y_m to derive a more realistic result:

$$y_m = \frac{y_l + y_r}{2} \qquad (13)$$

Using the modified y_m, the position error for left and right camera was calculated.

Fig. 6 shows the results of these experiments. The error in the left camera decreases with increasing distance of the test pattern. This effect is caused by the fixed range for eye pan

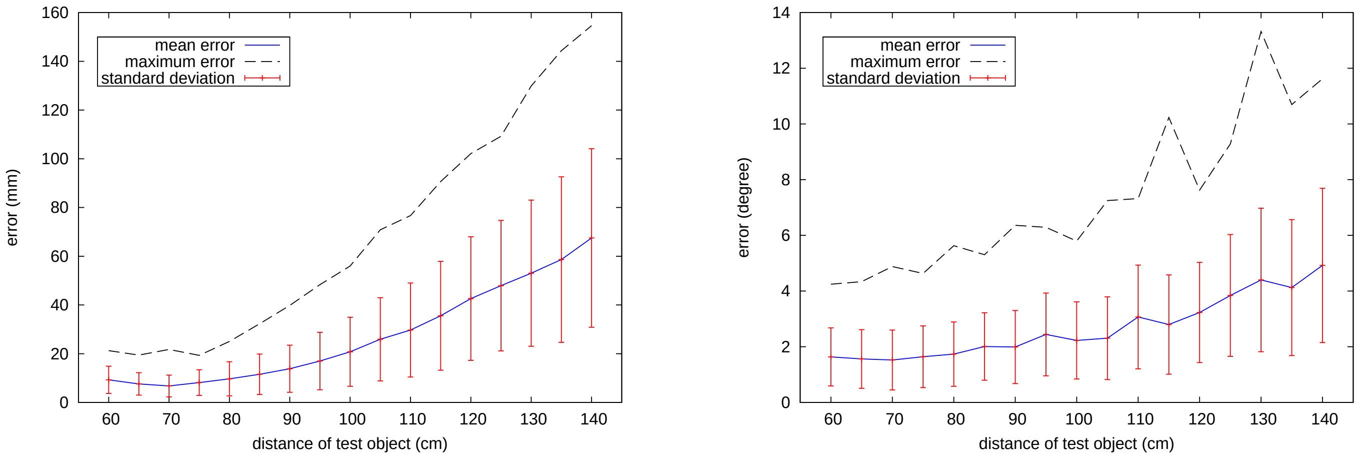


Fig. 5. Accuracy of the proposed stereo calibration. The 3D pose of a test pattern was determined in the left perspective camera using a model-based approach. Furthermore, the pose of the pattern was determined using stereo triangulation. The plots show the translational and rotational error between model based and triangulation based pose estimation for different distances of the test pattern.

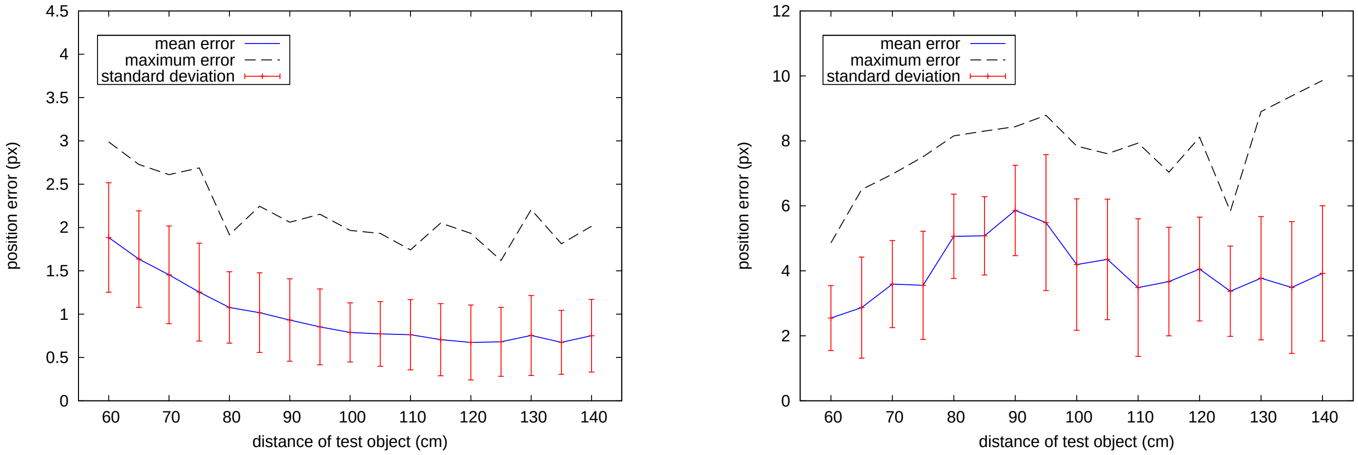


Fig. 6. Accuracy of saccadic eye movements. The 3D pose of a test pattern was determined in the left perspective camera using a model-based approach. The calibrated model was used for differential inverse kinematics in order to point the optical axes of the cameras towards the origin of the test pattern. The plots show the distance in pixels from principal point to the target corner of the test pattern in the image planes. Left: left camera. Right: right camera.

(-20 to 20 degrees) and tilt (-15 to 15 degrees) actuations. The maximum of the mean error for the left camera amounts to about 2 pixel for a distance of 60 cm. The plot for the right camera differs slightly from the results of the left camera. The increased error in the right camera is caused by the additional matrix H_{stereo} which has to be considered in the differential kinematics. In subsequent experiments, where the test pattern was located in the right camera, similar plots could be produced with more accurate results for the right camera and less accurate results for the left camera.

VI. Conclusions

In this paper we presented a new approach to solve the kinematic calibration problem for the Karlsruhe Humanoid Head's active camera system. The classical $AX = XB$ formulation of the head-eye calibration problem was combined with the benefits of a DH-based approach. The suggested method offers several advantages over common methods:

1) **Generality and accuracy:** As our method does not assume joint axes to intersect, it avoids a methodical error and allows for an improvement in calibration accuracy. Above that, it can be applied to a wider class of kinematic chains than most common methods.
2) **Robustness and accuracy:** In order to solve for the coordinate transformation from the joint to be calibrated to the camera coordinate frame, the desired calibration matrix is expressed as a non-linear least squares target function. Compared to solutions based on linear least squares, the non-linear approach presented here is less sensitive to noisy input data.
3) **Error propagation:** In contrast to many two-step approaches in the literature, the suggested method estimates all unknowns for one joint simultaneously.
4) **Verifiability:** Only one joint is calibrated at a time. This way the accuracy of an individual joint's kinematic calibration can be easily examined.

We presented experiments on the kinematic calibration accuracy, the stereo triangulation accuracy and the accuracy of inverse kinematics for saccadic eye movements. The experiments on stereo triangulation accuracy showed that at manipulation distance the pose of the test pattern could be determined with an error less then 1.5 cm. The experiments on the accuracy of the inverse kinematics showed that even at larger distances, the saccadic eye movement could be performed with a position error of less then 6 pixels in the image plane.

VII. Acknowledgments

The work described in this paper was conducted within the EU Cognitive Systems project PACO-PLUS (FP6-2004-IST-4-027657) funded by the European Commission.

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, p. 24782483.

[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,” 2003. [Online]. Available: http://www.automation.fujitsu.com

[4] T. Asfour, K. Welke, P. Azad, A. Ude, and R. Dillmann, “The Karlsruhe Humanoid Head,” in *IEEE-RAS International Conference on Humanoid Robots (Humanoids2008)*, 2008.

[5] A. Ude, C. Gaskett, and G. Cheng, “Foveated vision systems with two cameras per eye,” *Robotics and Automation, 2006. ICRA 2006. Proceedings 2006 IEEE International Conference on*, pp. 3457–3462, 15-19 2006.

[6] L. Itti and C. Koch, “Computational modelling of visual attention.” *Nature Reviews Neuroscience*, vol. 2, no. 3, pp. 194–203, March 2001.

[7] J. Tsotsos, “Visual attention: from covert search to foveating saccades,” in *Proc. Stockholm Workshop on Computational Vision*, 1993.

[8] R. Tsai and R. Lenz, “A new technique for fully autonomous and efficient 3D robotics hand/eye calibration,” *Robotics and Automation, IEEE Transactions on*, vol. 5, no. 3, pp. 345–358, June 1989.

[9] Y. Shiu and S. Ahmad, “Calibration of wrist-mounted robotic sensors by solving homogeneous transform equations of the form AX=XB,” *Robotics and Automation, IEEE Transactions on*, vol. 5, no. 1, pp. 16–29, February 1989.

[10] M. Li, D. Betsis, and J.-M. Lavest, “Camera calibration of the KTH head-eye system,” Computational Vision and Active Perception Lab., Dept. of Numerical Anaysis and Computing Science, Royal Institute of Technology (KTH), Tech. Rep., 1994.

[11] M. Li, “Kinematic calibration of an active head-eye system,” *Robotics and Automation, IEEE Transactions on*, vol. 14, no. 1, pp. 153–158, Feb. 1998.

[12] J. Neubert and N. Ferrier, “Robust active stereo calibration,” in *Robotics and Automation, 2002. Proceedings. ICRA ’02. IEEE International Conference on*, vol. 3, 11-15 May 2002, pp. 2525–2531.

[13] F. Dornaika and R. Horaud, “Simultaneous robot-world and hand-eye calibration,” *Robotics and Automation, IEEE Transactions on*, vol. 14, no. 4, pp. 617–622, Aug. 1998.

[14] G.-S. Young, T.-H. Hong, M. Herman, and J. Yang, “Kinematic calibration of an active camera system,” in *Computer Vision and Pattern Recognition, 1992. Proceedings CVPR ’92., 1992 IEEE Computer Society Conference on*, 15-18 June 1992, pp. 748–751.

[15] R. Horaud and F. Dornaika, “Hand-eye calibration,” *Int. Journal of Robotics Research*, vol. 14, No. 3, pp. 195–210, June 1995.

[16] F. Park and B. Martin, “Robot sensor calibration solving AX=XB on the euclidean group,” *Robotics and Automation, IEEE Transactions on*, vol. 10, no. 5, pp. 717–721, October 1994.

[17] D. Marquardt, “Algorithm for least-square estimation of non-linear parameters,” *Journal of the SIAM*, vol. 11, pp. 431–441, 1963.

[18] Z. Zhang, “A flexible new technique for camera calibration,” Microsoft Research, Tech. Rep., 1998.

[19] G. H. Golub and C. F. van Loan, *Matrix Computations*. North Oxford Academic, 1983.