A Framework for Visually guided Haptic Exploration with Five Finger Hands

A. Bierbaum, K. Welke, D. Burger, T. Asfour, and R. Dillmann
University of Karlsruhe
Institute for Computer Science and Engineering (CSE/IAIM)
Haid-und-Neu-Str. 7, 76131 Karlsruhe, Germany
Email: {bierbaum,welke,asfour,dillmann}@ira.uka.de

***Abstract*—In order for humanoid robots to enter human-centered environments, it is indispensable to equip them with the ability to recognize and classify objects in such an environment. A promising way to acquire the object models necessary for object manipulation appears in the supplement of the information gathered by computer vision techniques with data from haptic exploration. In this paper we present a framework for haptic exploration which can be used for both visually guided exploration with a five-fingered humanoid robot hand as well as with a human hand. We present experiments and results on 2D contour following and haptic exploration of 3D objects by the human hand. Volumetric shape data is acquired by a human operator hand using a data glove. The exploring human hand is located by a stereo camera system, whereas the finger configuration is calculated from the glove data.**

I. Introduction

In humans, different types of haptic exploratory procedures (EPs) for perceiving texture, weight, hardness, contact, size and the exact shape of a touched object are known [1]. These EPs require the exploring agent to initiate contact with the object and are therefore also referred to as active touch sensing.

In this paper, the contour following EP for shape recovery is subject of interest. A volumetric object model composed this way delivers a rather high amount of information for discriminating between objects. Also, volumetric object data is most suitable for supplementing and verifying geometric information in multimodal object representations.

Several approaches have been proposed for acquiring object shape information by robots through haptic exploration. An early, comprising experimental setup was presented in [2]: Here, the Utah/MIT dextrous robot hand, one of the first of its kind, was mounted to a manipulator arm and used for performing shape recovering, haptic exploratory procedures on unknown objects. The hand probed contact by closing around the object at predefined positions. The points of contact between fingers and object were calculated indirectly from proprioceptive information, i.e. joint angle position combined with crossing of a force limit from the tendon force readings. The resulting sparse point clouds were fitted to superquadric models defined by a set of five shape parameters. In addition, spatial rotation and translation of the superquadrics were estimated. The shape parameters were successfully used for recognizing several convex objects by comparison to a parameter database. The used superquadric model could not reflect non-convex bodies, therefore recognition and representation was limited to convex bodies in this approach.

In addition to the contact locations, the contact normal information gathered during haptic exploration was used in [3] to determine a set of intersecting planes which compose a polyhedral model as volumetric object representation. For object recognition the Euclidian distances of the polyhedral surface points to the borders of a surrounding cubic workspace box were measured at equidistant coordinates and matched to those of synthetic models. The approach was evaluated only in simulation. Object recognition was successful also for a limited translation of the object within the workspace box. The method appeared not appropriate for non-convex bodies. Beside contact probing, several procedures for contour following of object features have been investigated [4], [5]. Other approaches in active contact sensing utilizing contour following concentrate on the detection of local surface features [6].

In our approach, a framework has been developed that allows haptic exploration of objects for recovery of the exact global shape using different types of manipulators equipped with contact sensing devices. As we are interested to integrate the framework as basis for haptic exploration in our humanoid robot system [7] which is eqquiped with two five-fingered human-like and human-sized hands, we focus in this study on the application of exploring with five finger hands. In particular, the developed framework allows us to use the human hand of an operator as exploring manipulator by deploying a data glove with attached tactile sensors. This will also provide interesting possibilities for immediate comparison of haptic exploration results by a human versus a humanoid robot hand.

The exploratory process can be used in an unstructured environment, but it is currently limited by the constraint that the object being explored is in a fixed pose. Basically the approach for exploration is not limited to convex bodies though at this stage of our work results related to non-convex objects are not given, as we first want to concentrate on basic properties.

This paper is organized as follows. In the next section the relevant details and components of our system for haptic exploration focusing on object shape recovery are described.

This includes a description of the human hand model and the visual tracking of operator's hand. In section III we describe the experiments performed so far for evaluating the system and report on the results obtained. Finally we give a conclusion and an outlook on our future work is given in SectionIV.

II. SYSTEM DESCRIPTION

Figure 1 gives a system overview with the components involved in acquisition of contact points during contour following EP.

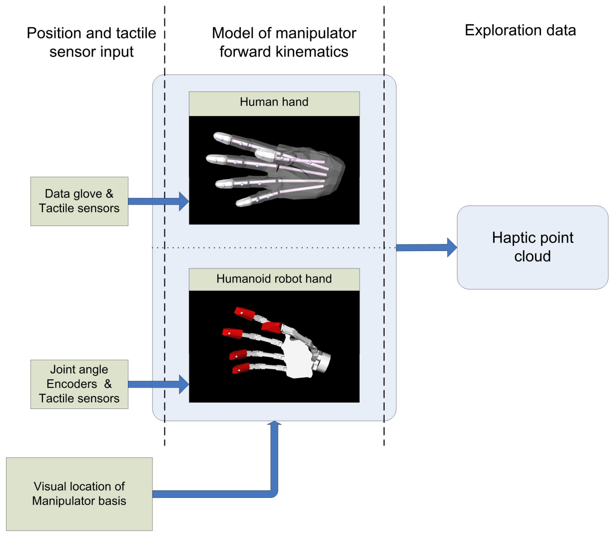


Fig. 1. System for acquisition of object shape data from haptic exploration using a human or a humanoid robot hand as exploring manipulator.

During the haptic exploration with the human hand, the subject wears a data glove that serves as an input device for calculating the joint angle configuration of the hand. The data glove we use is equipped with binary micro switches at the distal phalanges. When touching an object, the switches are actuated when local contact force exceeds a given threshold. The actuation also provides the operator with a mechanical sensation, the clicking of the switch, that helps to control the contact pressure during exploration. The data glove is made of stretch fabric and uses 23 resistive bend sensors that provide measurement data of all finger joint angle positions and the orientation of the palm.

Before starting exploration the operator needs to calibrate the data glove sensors with the forward kinematics of the underlying hand model. Currently we calibrate all fingers except for the thumb, but use only the tip of the index finger for exploration. We use a linear relation for the transformation from raw data glove sensor readings x to joint angles θ, following

$$\theta = C \cdot x + B \,.$$

For calibration the subject has to form a flat hand shape and a fist shape with the data glove respectively. From sensor readings and the corresponding model joint configurations the transformation matrices C and B can be determined.

Wrist position and orientation are determined in the reference frame of the camera coordinate system as described section II-B. During exploration the subject visually guides the finger tips along the desired contours. When the micro switch is actuated by touch, the current location of the sensor in the global reference frame is registered as a point in the 3D object point cloud. The sensor position is calculated using the forward kinematics of the hand model as described in the following section.

For exploration with our humanoid robot platform Armar-III we use a model for the forward kinematics of the robot hand presented in [8]. The robot is equipped with an advanced version of this hand with joint angle encoders attached to all controllable degrees of freedom (DoF).

A. Human hand model

The forward kinematics of the human hand must be modeled accurately to transform the coordinates of the tactile sensor locations gathered from the data glove sensor readings to a global reference frame in which the resulting 3D contact point cloud is accumulated. Furthermore, the model is required to cover the entire common configuration space of the human hand and the data glove, so that the human operator is preferably not restricted in the choice of hand movements that can be projected to the model.

A complex hand model deploying 27 DoFs was introduced in [9] and used in several studies requiring exact models ([10], [11]) for hand pose estimation from sensor input. The Carpometacarpals joints (CMC) were fixed, assuming the palm to be a rigid part of the hand. The CMCs are also often referred to as trapeziometacarpal joints (TM). The fingers were modeled as serial kinematic chains, attached to the palm at the metacarpophalangeal joints (MCPs). Interphalangeal joint(IP), distal interphalangeal joints (DIP) and proximal interphalangeal joints (PIP) have one DoF for flexion and extension. All MCPs joints have two DoFs, one for flexion and extension and one for abduction and adduction. The CMC joint of the thumb is modeled as a saddle joint. Several variants of this model exist in the literature (see [12], [13]).

The model is organized hierarchically and consists of fixed links and joints. The structure is tree-like with the wrist as root. The position of the wrist is aligned to the origin of the hand model reference frame. Every link connected to the wrist is described by a vector in a local reference frame. This scheme is used recursively to describe the links in lower levels of the hierarchical structure in their local coordinate system. The origin of the hand model is aligned in the global reference frame by the pose estimation of the wristband.

The hand model which we use in the presented framework is shown in Fig. 2. We have added two modifications to the basic model to improve the representation of real human hand kinematics. The first modification affects the modeling of the thumb's CMC joint. Following [13], the first metacarpal of the thumb performs a constrained rotation around a third orthogonal axis in the CMC joint, which contrasts the CMCs joint

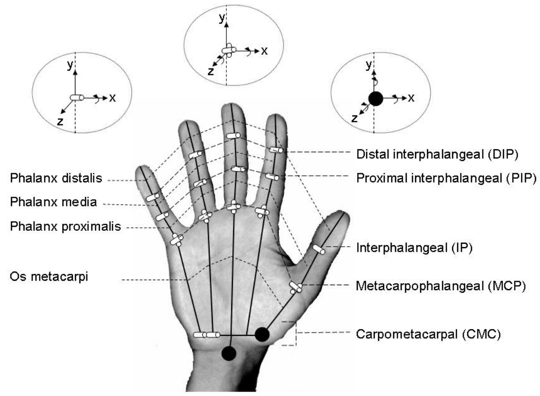


Fig. 2. The hand model for haptic exploration by a human operator with a data glove.

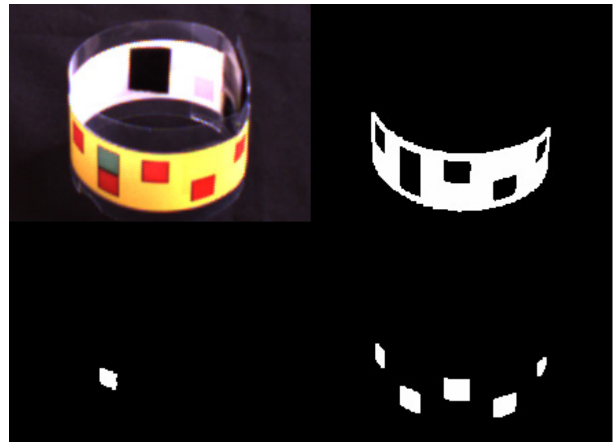


Fig. 3. The wrist band used for pose estimation of the data glove and segmentation results of the HSV segmentation for the colors yellow, green, and red.

model as a two DoF saddle joint. For reasons of simplicity we model this joint as a three DoF joint.

The second modification is to overcome the inadequate representation of the palm as a rigid body. As we want to incorporate the distal phalanges of the ring and little finger in the exploration process, we have extended the model by adding one DoF at the CMCs of these fingers respectively. By doing this the ability of the palm is reflected to fold and curve, when the little finger is moved towards the thumb across the palms inner side [14]. It is important to model this behavior as a human operator will occasionally utilize these types of movement when touching an object with the whole hand.

The resulting hand model consists of 26 DoFs. The four fingers have 4 DoFs each at the DIP, 4 DoFs at the PIP and 8 DoFs at the MCPs. The thumb is modeled with 1 DOF at its IP, its MCP is modeled with 2 DoFs and its CMC, as mentioned before, with a 3 DoF joint. Additionally we model the palm with 2 DoFs representing the CMCs of the little and ring fingers and add 2 DoFs for the wrist movement.

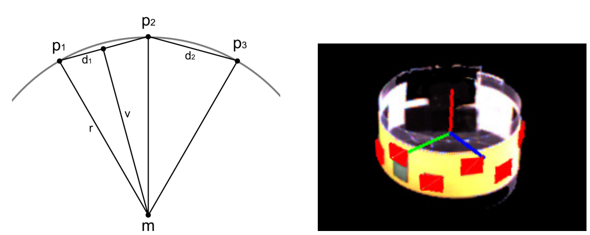
We have used the Open Inventor™[1] standard for constructing the hand model. This 3D modeling package allows the implementation of local reference frames as described before in a transparent way.

B. Wrist tracking

As mentioned earlier, the absolute position and the orientation of the data glove are determined using vision. In order to track the data glove in a robust manner, we use a marker wristband which is attached to the wrist of the subject wearing the data glove. The wristband color is yellow in the background, with twelve red squared markers and one green squared marker overlayed . The red markers are distributed along two circles in an alternating manner to improve the stability of the tracker. The green marker is required to register the wristbands coordinate system. During tactile exploration, the wristband is tracked with a calibrated stereo camera setup. The cameras generate color images with 640×480 pixels resolution.

1) Preprocessing: The relevant markers are identified by color segmentation performed in the current scene. We use segmentation based on the hue saturation value (HSV) color model. For each of the colors yellow, red, and green we specify the corresponding H value, a tolerance in the H channel, and valid ranges for the S and V channels. Points within the specified ranges are marked in the corresponding segmentation masks. Fig. 3 shows the segmentation results for all three colors.

2) Initialization: Initially, the position and rotation of the wrist band is unknown. In this case the system will locate the wristband in the current input images. Once the pose in the current frame has been calculated, it can be utilized as input for the tracking algorithm. In order to accomplish the initialization, an algorithm is required which determines the wristband pose without any prior knowledge.



(a) Geometry in the plane (b) Coordinate system

Fig. 4. Calculations in the plane of the three markers used for initialization, and the coordinate system after initialization.

In order to increase the robustness of the initialization process only red and green markers inside yellow areas are accepted. This step is performed for left and right image separately. For all accepted markers the 3D position is calculated by identifying corresponding markers in the left and right image and recovering the 3D information using the camera calibration. The epipolar line in the right image for

[1] http://oss.sgi.com/projects/inventor/

each marker in the left image is determined using the camera parameters from the offline calibration of the left and the right camera. For each marker from the left image the corresponding marker in the right image, which has the same color and minimal distance to the epipolar line, is determined. The 2D centroids of both markers are used to calculate the 3D position of the marker using the camera calibration.

During initialization, the 3D positions of the green marker $\mathbf{p_2}$ and the second and third closest red markers $\mathbf{p_1}, \mathbf{p_3}$ are calculated (see Fig. 4(a)). Since these markers lie in the same plane, the plane normal can be calculated:

$$\mathbf{d_1} = \mathbf{p_1} - \mathbf{p_2} \quad (1)$$

$$\mathbf{d_2} = \mathbf{p_3} - \mathbf{p_2} \quad (2)$$

$$\mathbf{n} = \mathbf{d_1} \times \mathbf{d_2} \quad (3)$$

Once the plane normal is determined, the center of the circle described by the three considered markers can be calculated for both pairs of markers $(\mathbf{p_1}, \mathbf{p_2})$ and $(\mathbf{p_2}, \mathbf{p_3})$. The radius r of the wristband is known, which allows to perform calculations for both pairs separately to retrieve two hypotheses for the center. For both distances $\mathbf{d_1}, \mathbf{d_2}$ the following calculations are performed:

$$l = \sqrt{|\mathbf{d_1}|^2 + r^2} \quad (4)$$

$$\mathbf{v} = \frac{\mathbf{d_1} \times \mathbf{n}}{|\mathbf{d_1} \times \mathbf{n}|} * l \quad (5)$$

$$\mathbf{m} = \frac{\mathbf{p_1} + \mathbf{p_2}}{2} + \mathbf{v} \quad (6)$$

where $\mathbf{m}$ denotes the center of the circle. Figure 4(a) shows the geometry in the plane spanned by the three considered markers. The above calculations are performed for both point pairs, which results in two centers $\mathbf{m_1}$ and $\mathbf{m_2}$. We use the mean value of both centers, if the difference is below a threshold, otherwise the initialization will return without success.

The coordinate system deployed in the initialization and tracking phase is shown in Fig. 4(b). The x-axis (denoted in red) can be derived from plane normal $\mathbf{n}$ and center $\mathbf{m}$. The y-axis (denoted in green) is calculated from the difference of the green marker $\mathbf{p_2}$ and the center $\mathbf{m}$. The z-axis is calculated with the cross product of x- and y-axis.

3) Tracking: After initialization, the wristband is tracked using a particle filter approach [15]. For the particle filter we use a model of the wristband which comprises all 12 red markers. The configuration of the model is defined by the 6D pose of the band. In each iteration of the particle filter algorithm 100 new configurations are generated using a gaussian distribution with variance $\sigma^2 = 0.45$. Furthermore the movement of the model is estimated by taking into account the movement in the previous frames. In order to retrieve the estimated 6D pose of the wrist band, the model is projected into both camera images using the camera calibration. Only markers are projected which are visible from the specific camera. The visibility of a marker is calculated by the angle between principal axis of the camera and normal on the marker plane.

To validate each particle, a weighting function is used which compares the segmentation mask for the red colour with the model. Ideally, the model markers would cover all red pixels situated inside yellow regions. In order to derive a weight for each configuration, we count all red pixels inside yellow regions f and all red pixels, which overlap with the projected model m. The probability for each particle z and the current images i can then be formulated in the following way:

$$p(z|i) \propto exp\left(\lambda * \frac{m}{f}\right) \quad (7)$$

where λ defines the sector of the exponential function which is used. After all particles have been weighted according to equation 7, the 6D pose is estimated by the weighted sum of all configurations.

III. Experiments for object shape recovery

For evaluation of the system described above we have performed related to the exploration by a human subject. The experimental setup for both experiments is shown in Fig. 5.

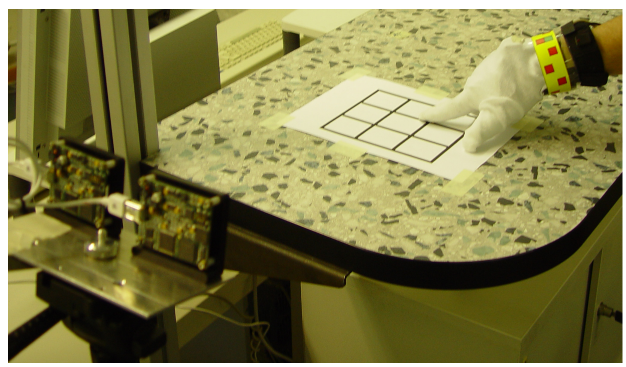


Fig. 5. Experimental setup. Here a planar grid structure is subject to contour following.

The region viewable by the stereo camera system defines the workspace in which the objects to be explored have to reside. The human operator wears a data glove with the wrist marker attached and performs a contour following EP with the tip of the index finger upon the object, i.e. the human operator visually guides the contact sensor along the contours. As mentioned earlier, the micro switch for contact sensing is also located at the distal phalanx. During the EP the object is fixed within the workspace. The human operator is allowed to move hand and fingers arbitrarily within the workspace as long as the marker bracelet are visually detected and localized. In case the markers localization fails, the subject needs to move the hand until it can be detected again.

A. 2D contour following in 3D space

As an initial experiment we chose to follow the visible contours of a planar structure to verify whether the exploration system delivers length and angle preserving point clouds.

These properties were inspected visually from the resulting point cloud data. We calculated the PCA for all points in the point cloud for quantitative determination of the plane the planar structures are located in. Further, we determined the standard deviation of the point locations in respect to this plane.

As planar shapes we chose a circle, an isoceles triangle and a 3×3 grid. The edge length of the bounding box for each of these shapes was set to $16cm$. The subject followed the contours of a printout of the respective shape with the index finger. The resulting point clouds for the triangle and the grid shape are shown in Fig. 6. The figures show that the contours of the test shapes are situated in different planes, which originates from a change in the position of the camera system between the two explorations. The duration of the exploration process was 40 seconds for the circle and triangle shapes and 2 minutes for the grid structure. Exploration speed was mainly limited by the performance of the wrist tracking algorithm.

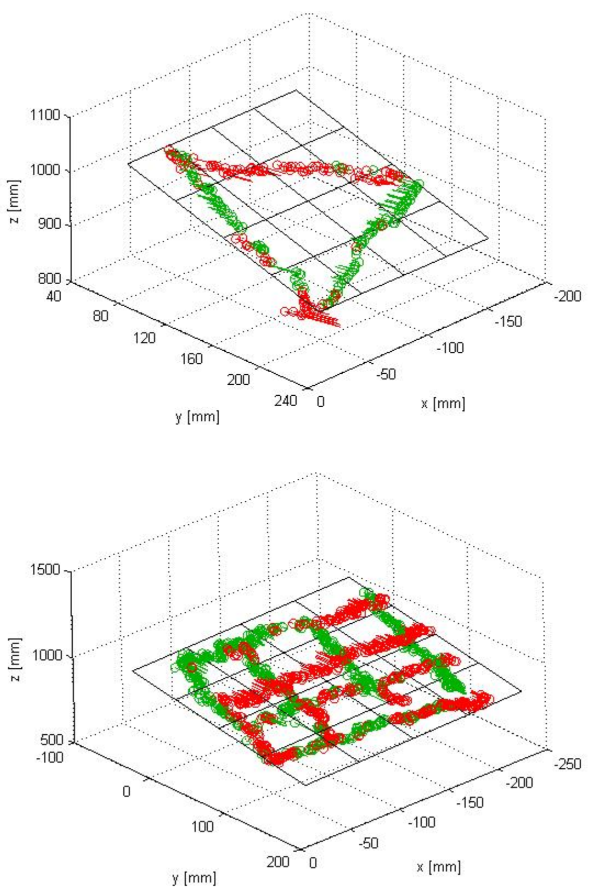


Fig. 6. Resulting point clouds for tactile contour following of planar shapes and corresponding fitted planes. Red points are situated below the plane, green points above.

During circle contour following 245 contact points were acquired, the standard deviation to the fitted plane was calculated to $\sigma = 5.37mm$. For the triangle contour following exploration 259 data points were acquired with $\sigma = 6.02mm$. For the grid exploration finally, 1387 data points were acquired with $\sigma = 5.86mm$.

B. *Contour following EP of a 3D object*

We further investigated the capability of the system to explore 3D objects in the workspace. In this experiment the subject had to track the edges of a rectangular box situated on a table in the workspace. The EP delivered 1334 data points, the resulting point cloud is depicted in Fig. 7. The box dimensions were $150 \times 50 \times 120mm$.

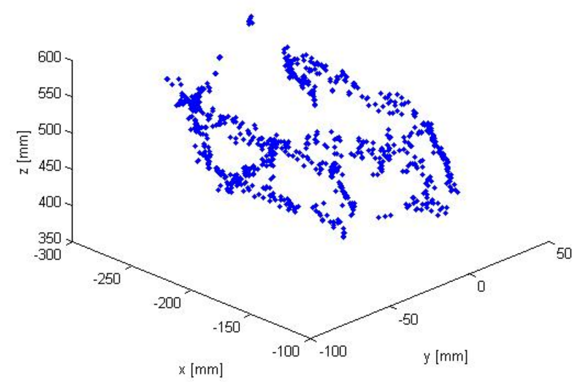


Fig. 7. Resulting point cloud for haptic contour following of a rectangular box.

IV. Conclusion and Future Work

In this paper we presented a framework for acquiring volumetric object models via haptic exploration by contour following. In a first evaluation results for haptic exploration of 2D and 3D contours by a human subject wearing a data glove in an unstructured environment are described.

It could be shown that the standard deviation of the acquired point clouds for 2D contours towards the estimated plane is within a constant and reasonable range. As next step we will extend shape data acquisition to all fingers of the hand by equipping all data love finger tips with a contact sensor, which will accelerate the overall exploration process. We will also address the evaluation of data acquired during contour following of 3D objects by fitting superquadric functions to the acquired point clouds [2], [3]. For complex objects this will also require decomposition of complex 3D structures to superquadric primitives.

Finally we will address the transfer of the haptic exploration framework to our humanoid robot platform. The platform already incorporates the same stereo vision system as used for the experiments described in this paper. A tactile sensor system for the robot hand has been developed that provides more information over the binary switches deployed with exploration by a human.

The focus of our work will further move to the development of autonomous and robust visually guided haptic exploration strategies for shape recovery by a humanoid robot with five-finged hands.

ACKNOWLEDGEMENT

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 and the German Humanoid Research project (SFB 588) funded by the German Research Foundation (DFG: Deutsche Forschungsgemeinschaft).

REFERENCES

[1] Susan J. Lederman and Roberta L. Klatzky, "Hand movements: A window into haptic object recognition," *Cognitive Psychology*, vol. 19, no. 3, pp. 342–368, 1987.

[2] P.K. Allen and K.S. Roberts, "Haptic object recognition using a multi-fingered dextrous hand," in *Robotics and Automation, 1989. Proceedings., 1989 IEEE International Conference on*, 14-19 May 1989, pp. 342–347 vol.1.

[3] S. Caselli, C. Magnanini, and F. Zanichelli, "Haptic object recognition with a dextrous hand based on volumetric shape representations," in *Multisensor Fusion and Integration for Intelligent Systems, 1994. IEEE International Conference on MFI '94.*, 2-5 Oct. 1994, pp. 280–287.

[4] P.K. Allen, "Mapping haptic exploratory procedures to multiple shape representations," in *Robotics and Automation, 1990. Proceedings., 1990 IEEE International Conference on*, 13-18 May 1990, vol. 3, pp. 1679–1684.

[5] N. Chen, H. Zhang, and R. Rink, "Edge tracking using tactile servo," in *Intelligent Robots and Systems 95. 'Human Robot Interaction and Cooperative Robots', Proceedings. 1995 IEEE/RSJ International Conference on*, 5-9 Aug. 1995, vol. 2, pp. 84–89.

[6] A.M. Okamura and M.R. Cutkosky, "Haptic exploration of fine surface features," in *Robotics and Automation, 1999. Proceedings. 1999 IEEE International Conference on*, 10-15 May 1999, vol. 4, pp. 2930–2936.

[7] T. Asfour, K. Regenstein, P. Azad, J. Schröder, A. Bierbaum, N. Vahrenkamp, and R. Dillmann, "ARMAR-III: An integrated humanoid platform for sensory-motor control," in *IEEE-RAS International Conference on Humanoid Robots (Humanoids 2006)*, Dec. 2006, pp. 169–175.

[8] A. Kargov, T. Asfour, C. Pylatiuk, R. Oberle, H. Klosek, S. Schulz, K. Regenstein, G. Bretthauer, and R. Dillmann, "Development of an anthropomorphic hand for a mobile assistive robot," in *Rehabilitation Robotics, 2005. ICORR 2005. 9th International Conference on*, 28 June-1 July 2005, pp. 182–186.

[9]] J. Lee and T. Kunii, "Constraint-based hand animation," in *Models and Techniques in Computer Animation*, Tokyo, 1993, pp. 110–127, Springer, Tokyo.

[10] Y. Yasumuro, Qian Chen, and K. Chihara, "3d modeling of human hand with motion constraints," in *3-D Digital Imaging and Modeling, 1997. Proceedings., International Conference on Recent Advances in*, 12-15 May 1997, pp. 275–282.

[11] A. Erol, G. Bebis, M. Nicolescu, R.D. Boyle, and X. Twombly, "A review on vision-based full dof hand motion estimation," in *Computer Vision and Pattern Recognition, 2005 IEEE Computer Society Conference on*, 20-26 June 2005, vol. 3, pp. 75–75.

[12] M. Bray, E. Koller-Meier, and L. Van Gool, "Smart particle filtering for 3d hand tracking," in *Automatic Face and Gesture Recognition, 2004. Proceedings. Sixth IEEE International Conference on*, 17-19 May 2004, pp. 675–680.

[13] B. Buchholz and T.J. Armstrong, "A kinematic model of the human hand to evaluate its prehensile capabilities," *Journal of Biomechanics*, vol. 25, no. 2, pp. 149–162, 1992.

[14] J.J. Kuch and T.S. Huang, "Vision based hand modeling and tracking for virtual teleconferencing and telecollaboration," in *Computer Vision, 1995. Proceedings., Fifth International Conference on*, 20-23 June 1995, pp. 666–671.

[15] Michael Isard and Andrew Blake, "ICONDENSATION: Unifying low-level and high-level tracking in a stochastic framework," *Lecture Notes in Computer Science*, vol. 1406, pp. 893–908, 1998.