2010 IEEE International Conference on Robotics and Automation
Anchorage Convention District
May 3-8, 2010, Anchorage, Alaska, USA

Visual tracking of a jaw gripper based on articulated 3D models for grasping

José J. Sorribes, Mario Prats and Antonio Morales

Abstract— **Robust grasping of objects in uncertainty conditions can be achieved with the visual monitoring of the interaction between the robot hand and the object.**

In this paper we propose a new approach for the visual tracking of a robot hand suitable to observe the interaction between robot and object. It consists on the continuous vision-based recovery of the articular pose of the robot hand. It is based on the principles of virtual visual servoing, which allows to deal with articulated bodies and occlusions.

Its suitability is shown by tracking a parallel jaw gripper under different conditions such as self-occlusion, articulated motion and important changes in the point-of-view. The potential applications range from estimating the robot hand articulated pose under poor hand-eye calibration and joint feedback, until detecting deficient contact configurations, incipient slips, etc.

I. INTRODUCTION

Vision has a key role in robot grasping and manipulation tasks. The process of grasping an object in low-structured scenarios by a robot hand can be roughly divided in three main stages: planning, approaching and execution. In the planning stage, the robot decides how to grasp and approach an object. This approaching is done in the next phase while avoiding obstacles . Finally, in the execution stage the fingers are closed over the object, according to the planned grasp, and first contacts are made. Execution phase can also include corrections with more contacts and releases until a satisfactory grip on the object is reached.

Vision is used at different levels on this process. Firstly, it has been used to locate and identify target objects on the scene. The main approaches differ on whether there exists previous information about the shape or appearance of the object [1], [2], or not [3], [4].

Second, visual input from the objects has been used to plan potential grasp on the target objects and approaching paths to them. Some of these planners assume to have the pose and shape of the object [5], [6]. Other approaches do not rely on the whole shape of the object but use vision to identify and extract specific features that allow the planning of a grasp [7], [8].

And third, visual feedback is also used when the arm tries to reach the object. For an instance, Murphy et al. uses visual techniques to correct the orientation of four-finger hand while approaching an object to allow better contact locations [9], and Namiki et al. uses a fast control schema in combination with tactile feedback to cage an object [10].

Little attention has been put on the use of vision when the robot hand makes the first contacts on the objects. Most closed-loop hand controllers simply rely on contact-based sensors to obtain feedback at this stage [11], [12].

However, there are a number of reasons that motivate the use of vision in the control-loop of grasp execution. First, some arms and hands do not have suitable sensor feedback for providing accurate position information. In other cases, hand-eye calibration is poor and does not allow for open-loop accurate hand-object alignment. In these cases, vision could potentially provide the pose of the objects and the the pose and configuration of the hand. Finally, when contacts occur vision information could be combined with contact sensor modalities to provide richer information about the grasp, such as the contact configuration, object sliding, etc. Such information is of great interest for exploration and learning systems.

In this paper we propose a new approach for the visual tracking of a robot hand based on the continuous vision-based recovery of the articular pose of the hand by means of Virtual Visual Servoing [13]. The main applications of this method in the context of robotic grasping are (i) the capability of tracking the hand/object interaction without the need for special markers [14], (ii) the direct estimation of the hand pose in sensor-less hands, and (iii) the possibility to detect contact points from vision. The two main difficulties that such a tracking solution must overcome are two. In the first place, it must deal with occlusions, both self-occlusions and those produced by the object. In the second place, it must deal with articulated bodies.

Its suitability is shown by tracking a parallel jaw gripper under different conditions such as self-occlusion, camera motion, articulated motion and important changes in the point-of-view.

A. *Pose estimation work*

The proposed approach is based on the pose tracking of the robot hand. This technique has been mostly used in robot manipulation to track target objects [1], [15], [4], but not robot parts.

Pose estimation techniques can be classified in appearance-based or model-based approaches [16]. Appearance-based methods work by comparing the 2D image of the object with those stored in a database containing previously acquired views from multiple angles. The main advantage of these methods is that they do not need a 3D object model, although a previous process must be performed in order to include a new object in

---

Authors are with the Computer Science and Engineering Department, Jaume-I University, 12071 Castellón, Spain. `{jsorribe,mprats,morales}@uji.es`

978-1-4244-5040-4/10/$26.00 ©2010 IEEE

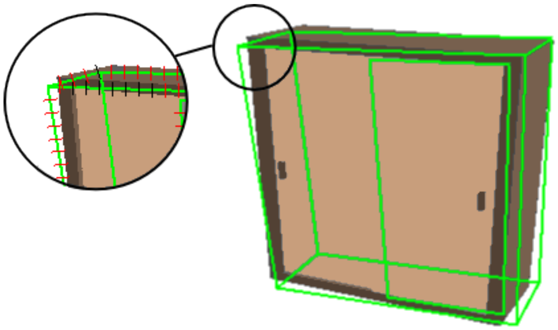


Fig. 1. A feature vector is built from the distances between the projected edges and high-gradient points searched along the edge normals, at the sampling interval. The goal of the non-linear minimization is to reduce all the distances to zero.

the database. Model-based methods obtain better accuracy and robustness, because of the use of model information for anticipating events like object self-occlusions. Some approaches consider a combination of both methods, like [1], where an appearance-based method is used first for getting an initial pose estimation, which is then used as initialization for a model-based algorithm.

Although vision has been widely adopted for detecting and tracking the objects to be manipulated, very few approaches have considered the use of vision for tracking the robot hand.

II. ARTICULATED VIRTUAL VISUAL SERVOING

There are two main methods in the literature for model-based pose estimation and tracking of articulated objects, both based on full-scale non-linear optimization. The first, developed by Drummond and Cipolla [17], is formulated from the Lie algebra point of view, whereas the second, proposed by Comport et. al. [18], [19], is based on the Virtual Visual Servoing (VVS) method [13]. Both methods implement robust estimation techniques and have shown to be very suitable for real-time tracking of common articulated objects in real environments. A comparison between both approaches is reported in [20], where it is shown that both formulations are equivalent, although some differences in performance can appear at run time. In our system, the VVS approach has been implemented [19], [13], mainly for its computational efficiency and because it is based on a solid background theory, i.e. 2D visual servoing, which convergence conditions, stability, robustness, etc. have been widely studied in the visual servoing community [21]. In addition, almost any kind of visual feature can be used and combined with this approach (points, lines, ellipses, etc.), as long as the corresponding interaction matrix can be computed. Different examples of the interaction matrix for the most common features are shown in [22].

A. The concept

The concept of the VVS approach, developed in [13], is to apply visual servoing techniques to a virtual camera, so that a set of object features projected in the virtual image from a model, match with those extracted from the real image. Under this approach, the pose estimation and tracking problem can be seen as equivalent to the problem of 2D visual servoing [18], which has been extensively studied in the visual servoing community [21]. Taking as input an object model, and an initial estimation of the camera pose in object coordinates, denoted as a pose vector, $\mathbf{r}$, the idea is to project a set of 3D features of the object model into a virtual image of the object, taken from the virtual camera position, $\mathbf{r}$. This virtual image is compared with the real one, and a vector of visual features is generated, denoted by $\mathbf{s}(\mathbf{r})$.

In our particular implementation, we make use of the point-to-line distance feature, as in [18], although any kind of geometric feature could be used as long as the interaction matrix can be computed. The edges of the object model, projected as lines in the virtual image, are sampled at regular intervals, and a search for a strong gradient is performed in the real image, in a direction perpendicular to the projected line, as shown in Figure 1. For each match, the point-to-line distance is computed and stored in the feature vector. The desired feature vector is given by $\mathbf{s}^* = 0$, which represents the case when all the edges of the object model are projected on strong gradients, and, ideally, the virtual camera position corresponds to the real one. The control law governing the virtual camera motion is given by:

$$\mathbf{v}_r = -\lambda \left(\widehat{\mathbf{D}}\widehat{\mathbf{L}_s}\right)^+ \widehat{\mathbf{D}}(\mathbf{s}(\mathbf{r}) - \mathbf{s}^*) \qquad (1)$$

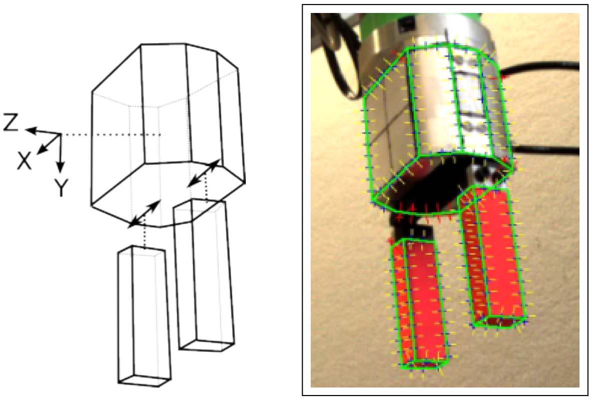
where $\mathbf{v}_r$ is the virtual camera velocity, λ is a control gain, $\widehat{\mathbf{L}_s}$ is the interaction matrix for the point-to-line distance feature, and $\widehat{\mathbf{D}}$ is a diagonal weighting matrix computed by iteratively re-weighted least squares, which is a robust estimator for dealing with outliers [18].

B. Virtual Visual Servoing on articulated objects

Comport et al. presented in [19] an approach for pose estimation and tracking of articulated objects based on the VVS method and the kinematic set concept. In their approach, the articulated pose is estimated directly from the visual observation of the object parts, leading to an efficient method that eliminates the propagation of errors through the kinematic chain. The only condition is that joint parameters must be decoupled in the minimization of the objective function. This can be accomplished by performing the minimization in object joint coordinates instead of in the camera space. Let $\mathbf{s}_1(\mathbf{r}_1)$ and $\mathbf{s}_2(\mathbf{r}_2)$ represent the perceived visual features on both parts of an articulated object composed of two links and one joint, and $\mathbf{s}_1^*$ and $\mathbf{s}_2^*$ be the desired values for those features, with $\widehat{\mathbf{L}_{s1}}$ and $\widehat{\mathbf{L}_{s2}}$ representing the corresponding interaction matrices. Then, the articular pose can be estimated by applying the following image-based control law:

$$\begin{pmatrix} \mathbf{v}_1 \\ \mathbf{v}_2 \end{pmatrix} = -\lambda \widehat{\mathbf{A}} \left(\widehat{\mathbf{D}}\widehat{\mathbf{H}}\right)^+ \widehat{\mathbf{D}} \begin{pmatrix} \mathbf{s}_1(\mathbf{r}_1) - \mathbf{s}_1^* \\ \mathbf{s}_2(\mathbf{r}_2) - \mathbf{s}_2^* \end{pmatrix} \qquad (2)$$

$$\widehat{\mathbf{H}} = \begin{pmatrix} \widehat{\mathbf{L}_{s1}} & \mathbf{0} \\ \mathbf{0} & \widehat{\mathbf{L}_{s2}} \end{pmatrix} \widehat{\mathbf{A}}$$



(a) (b)

Fig. 2. The 3D model of the hand and its registration into a real view.

$$\widehat{\mathbf{A}} = \begin{pmatrix} \widehat{{}^{C}\mathbf{W}_{O}}\mathbf{S} & \widehat{{}^{C}\mathbf{W}_{O}}\mathbf{S}^{\perp} & \mathbf{0} \\ \widehat{{}^{C}\mathbf{W}_{O}}\mathbf{S} & \mathbf{0} & \widehat{{}^{C}\mathbf{W}_{O}}\mathbf{S}^{\perp} \end{pmatrix}$$

where $\widehat{{}^{C}\mathbf{W}_{O}}$ represents the twist transformation matrix from the camera frame to the object joint frame, and $\mathbf{S}^{\perp}$ is a constraint matrix which depends on the type of joint [19]. Finally, the virtual camera velocities, one for each link, are given by $\mathbf{v}_1$ and $\mathbf{v}_2$.

III. MODEL-BASED TRACKING OF A PARALLEL JAW GRIPPER

A. *Jaw gripper model*

For this particular work, we consider a parallel jaw gripper as the one shown in Figure 2b. It consists of a box-shaped base containing the electronics, and two jaws actuated by a single motor. The hand can receive commands for opening, closing and stopping, and returns feedback only if the jaws are completely opened or closed. However, it does not contain sensors that provide the exact grip aperture, which is one of the reasons that motivate the use of visual information.

We define a 3D model of the gripper composed of the most distinguishable 3D edges, as shown in Figure 2a. The model is composed of three different parts: the base and the pair of jaws. The base and the pincers are kinematically linked through a prismatic joint along the base X axis. This joint is modeled with the holonomic constraint matrix $\mathbf{S}^{\perp} = (1, 0, 0, 0, 0, 0)^T$, and, as the motion of the pair of jaws is coupled and controlled by a single motor, the articulation matrix takes the form of:

$$\widehat{\mathbf{A}} = \begin{pmatrix} \widehat{{}^{C}\mathbf{W}_{O}}\mathbf{S} & \widehat{{}^{C}\mathbf{W}_{O}}\mathbf{S}^{\perp} & \mathbf{0} \\ \widehat{{}^{C}\mathbf{W}_{O}}\mathbf{S} & \widehat{{}^{C}\mathbf{W}_{O}}\mathbf{S}^{\perp} & \widehat{{}^{C}\mathbf{W}_{O}}\mathbf{S}^{\perp} \\ \widehat{{}^{C}\mathbf{W}_{O}}\mathbf{S} & \widehat{{}^{C}\mathbf{W}_{O}}\mathbf{S}^{\perp} & -\widehat{{}^{C}\mathbf{W}_{O}}\mathbf{S}^{\perp} \end{pmatrix}$$

B. *Tracking*

The hand is tracked by iteratively applying equation 3 adapted to the case of three components and using the previous articulation matrix, i.e.:

$$\begin{pmatrix} \mathbf{v}_1 \\ \mathbf{v}_2 \\ \mathbf{v}_3 \end{pmatrix} = -\lambda \widehat{\mathbf{A}} \left(\widehat{\mathbf{D}} \widehat{\mathbf{H}} \right)^{+} \widehat{\mathbf{D}} \begin{pmatrix} \mathbf{s}_1(\mathbf{r}_1) - \mathbf{s}_1^* \\ \mathbf{s}_2(\mathbf{r}_2) - \mathbf{s}_2^* \\ \mathbf{s}_3(\mathbf{r}_3) - \mathbf{s}_3^* \end{pmatrix} \qquad (3)$$

$$\widehat{\mathbf{H}} = \begin{pmatrix} \widehat{\mathbf{L}_{s1}} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \widehat{\mathbf{L}_{s2}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \widehat{\mathbf{L}_{s3}} \end{pmatrix} \widehat{\mathbf{A}}$$

The distance feature vector is computed by sampling points in the projected edges at regular intervals and looking for strong gradients in the perpendicular direction. An oriented gradient search is thus performed, as in [18], by applying a gradient convolution mask along a linear path perpendicular to the specific edge on each sampled point. The length of the search line depends on the expected object variation between two consecutive frames. For slow object/camera motion and/or high frame rates, little image variation between two consecutive frames is expected. Therefore, the search line can be of a few pixels around the sampled point, thus increasing the tracker efficiency. On the contrary, for low frame rate and/or high camera/object velocity, the search space should be increased.

Points are sampled only on those edges that belong to visible faces. Face visibility is computed at each iteration by checking the position of the camera with respect to the planes defined by each face normal. Being A, B, C and D the parameters of the plane equation corresponding to a specific face, with the face normal pointing towards outside, the condition for face visibility can be computed as:

$$A \cdot r_x + B \cdot r_y + C \cdot r_z + D > 0$$

where r_x, r_y and r_z are the translational components of the pose vector $\mathbf{r}$ that contains the camera pose with respect to the object.

This approach allows to check face visibility in a very efficient way. However, object self-occlusions cannot be detected. This is not a major problem in our experiments, since outlier rejection via the weighting matrix $\mathbf{D}$ is able to deal with these situations, as long as the occluded edges are only a small part of the object. As future improvements, we would like to deal with self-occlusions via binary space partition trees, like in [18].

C. *Camera motion*

In grasping situations where the robot hand has to reach for an object, it is important to adopt an active vision approach in which the camera follows the hand motion. This allows to obtain a detailed view of the robot hand at the same time that it is always kept inside the image.

For this purpose, we attach the camera to a pan/tilt unit that allows to point the viewing direction towards any interesting point. The pan/tilt unit is a PTU-46-70 model from Directed Perception Inc., and communicates with the host computer via a RS-232 port. The camera is a standard Firewire camera providing 30 frames per second at the resolution of 640×480. The complete system can be seen in Figure 3.

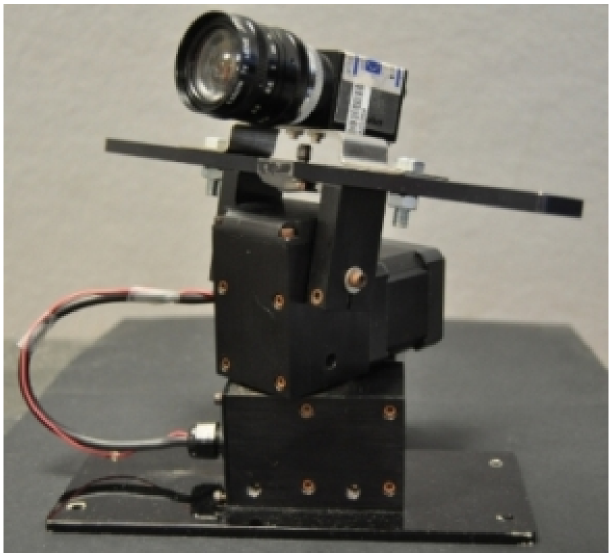


Fig. 3. A pan/tilt unit used for keeping the hand inside the camera field of view.

In order to keep the robotic hand inside the camera field of view, an image-based controller has been implemented. At each iteration the 3D center of the hand model is projected into the image according to the estimated hand pose. Its 2D distance to the center of the image is computed and fed back to a simple proportional controller that sends a pan/tilt velocity that moves the viewing direction towards the hand center.

IV. RESULTS

The hand tracker has been validated with three different experiments that reproduce real situations that will commonly occur during manipulation:

1) Hand rotation that involves self-occlusion and appearance/disappearance of faces.
2) Articulated motion of the pincers.
3) Simultaneous motion of the hand and the robot camera.

In the first of the experiments, the robot end-effector was placed at a fixed position, and the pan/tilt unit in a configuration in which the robot hand was centered in the image. A rotation velocity was set around the hand axis, as shown in Figure 4a. This motion made some faces of the hand appear and disappear, and also generated self-occlusions, specially on the pincers. The hand tracker was able to deal with these situations by dynamically selecting the visible faces. Self-occlusions, not detected by the visibility check described in the previous section, generate wrong matches in the image. However, as long as the number of wrong matches is a small amount compared with those features correctly matched, the robust estimator implemented with the tracker is able to classify them as outliers and reject them.

In the second experiment, both the robot end-effector and the pan/tilt unit were kept at a fixed position. The tracker was initialized and the hand pincers were commanded to open and close repeatedly. The tracker was able to follow this articulated motion even in the presence of self-occlusions, as shown in the top row of Figure 4b.

It is worth mentioning that this capability is specially interesting for this particular robotic hand that does not provide joint feedback. Therefore, vision can be used here in order to provide a direct estimation of the opening distance, and eventually control it to a desired configuration. Another experiment dealt with the case in which the hand joints were manually actuated by a human, as shown in the bottom row of Figure 4b.

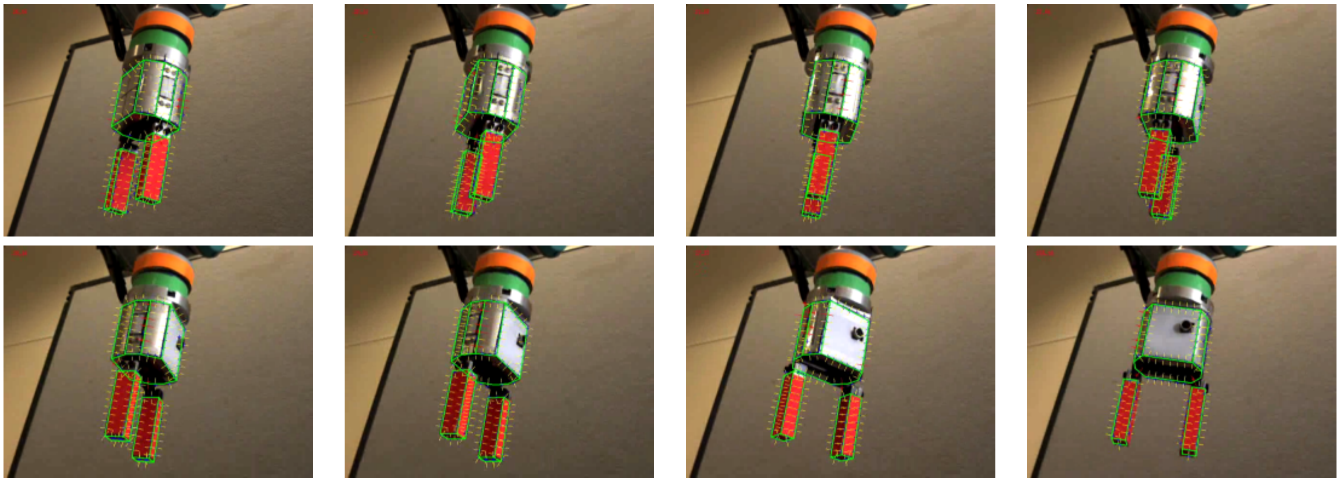
Finally, in the last experiment, a joint velocity was sent to the manipulator elbow, generating both translational and rotational motion of the robot hand. The pan/tilt controller was activated in order to keep the hand inside the camera view, even if part of the hand was outside the image limits, as shown in Figure 4c. The tracker also performed successfully in this situation where both robot and camera motion was performed simultaneously. Finally, it is worth mentioning that the tracker runs at video rate, as it can be observed in the video accompanying this paper.

V. DISCUSSION AND FUTURE LINES

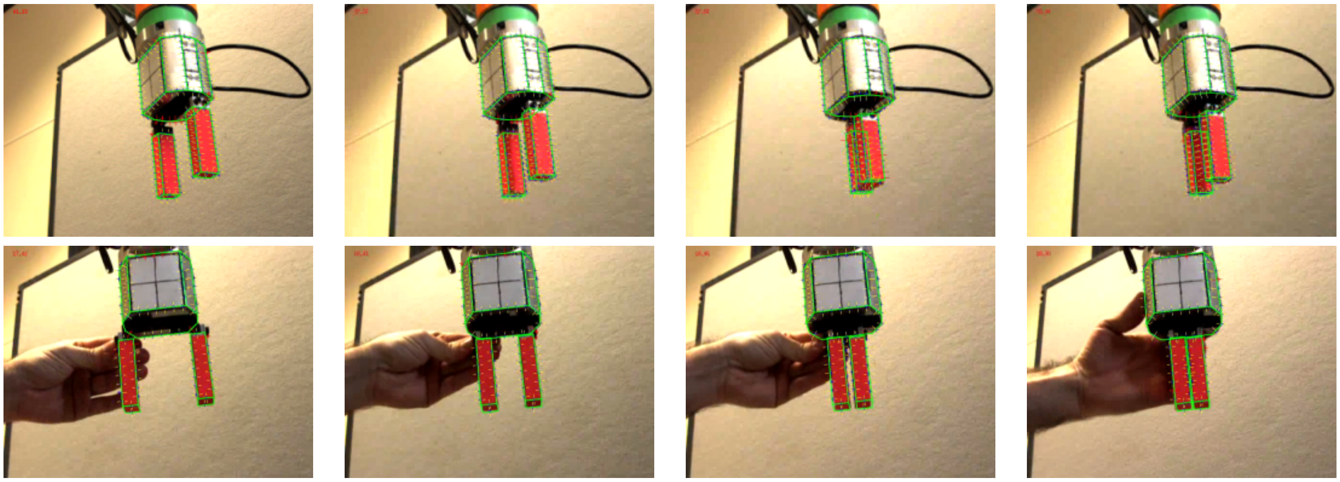
The approach proposed still presents several practical problems that need to be addressed before having a full-working version. The first one is the robustness of point matching. The current implementation requires small changes between two consecutive captured images. This can be accomplished either by a high capture rate or by limiting the hand/camera relative velocity. If the change of the object position between two sequential images is too large, the search distance has to be increased, and the tracker runs more slowly. In this case point tracking problems are frequently experienced. Also, if the movement of the camera or the robot hand is too fast, the tracker may lose the reference. There are several solutions to improve the tracking robustness. On one side is it possible to include forward prediction either using signals coming from arm and hand position controllers or simply using visual cues. A second option is to use a more robust and efficient point matching algorithm. In addition, the use of a kalman filter would also improve considerably this method.

The method has not been yet tested with complex hands. The parallel-jaw gripper that has been used has only one joint that actuates two different parts. Advanced robot hands usually have many more free joints. We have plans for applying this approach into a Barrett Hand, which also lacks position feedback on some joints. In general the method requires a set of strong visual features that can be efficiently tracked.

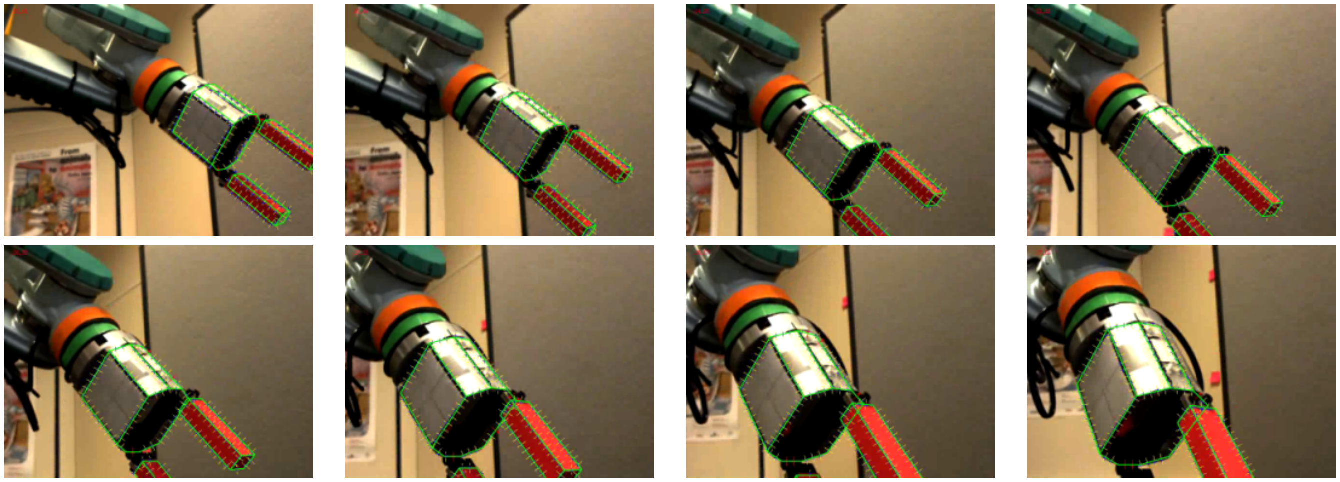
Finally, a crucial aspect of the tracking method is initialization. Currently, a human operator clicks on a camera image to project the initial model of the gripper. However an automatic method needs to be developed and implemented to solve this issue. One possibility is to make use of an initial estimation of the hand position, computed from a coarse hand-eye calibration, in case it is available. If not, an



(a) Occlusions and new faces appearing



(b) Opening and closing the actuator



(c) Hand tracking via head motion

Fig. 4. Results of the model-based hand tracker.

appearance-based pose estimation method can be adopted in order to provide a coarse initialization.

The approach proposed in this paper deals with a problem that has not been successfully addressed yet in the literature. The results presented in this paper are promising but still some improvements are needed before a robust robot hand tracker is obtained. Such a solution would allow a robust control of the hand while contacting an object, and in the last term allow robust grasping of objects under uncertainty.

VI. CONCLUSIONS

Visual tracking of a robot hand offers important potential applications. First is that it allows to estimate the hand configuration when the hand joints do not provide any feedback, or it is inacurate. In addition, it potentially avoids the use of external hand-eye calibration, thus being very suitable in situations where the robot camera position is not accurate, or the kinematics relating the camera and the hand systems is poor. It also allows to detect and correct any hand-object misalignment during and after the grasp execution. This is particularly interesting for the detection of deficient grasps or incipient slip. Finally, the tracking of the robot hand while contacting and object allows the fusion with contact sensor data (force/torque, tactile, pressure), and opens new possibilities for robot control, object exploration and robot learning.

This paper has described an approach for estimating and tracking the articular pose of a robotic hand. This approach is based on the method of virtual visual servoing, and allows to estimate the articular 3D pose of an object from a model and natural object features. It has been tested with a parallel jaw gripper, and under several conditions that are normally present in any manipulation task. Although the method can be improved in many different ways, it already represents a valid solution for simple hands like a parallel jaw gripper. In the future we would like to validate this approach with a more complex hand and during real grasping actions.

VII. ACKNOWLEDGEMENTS

This research was partly supported by the European Commission's Seventh Framework Programme FP7/2007-2013 under grant agreements 215821 (GRASP project), 217077 (EYESHOTS project), and 248497(TRIDENT Project), by Ministerio de Ciencia e Innovación (DPI-2008-06636; and DPI2008-06548-C03-01), and by Fundació Caixa Castelló-Bancaixa (P1-1B2008-51; P1-1A2006-11; and P1-1B2009-50).

REFERENCES

[1] D. Kragic and H.I. Christensen. Model based techniques for robotic servoing and grasping. *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 1:299–304, October 2002.

[2] P. Azad, T. Asfour, and R. Dillmann. Combining appearance-based and model-based methods for real-time object recognition and 6D-localization. In *International Conference on Intelligent Robots and Systems (IROS)*, Beijing, China, 2006.

[3] Beata J. Grzyb, Eris Chinellato, Antonio Morales, and Angel P. del Pobil. Robust grasping of 3D objects with stereo vision and tactile feedback. In *International Conference on Climbing and Walking Robots and the Support Technologies for Mobile Machines (CLAWAR)*, pages 851 – 858, Coimbra, Portugal, 2008.

[4] C. Dune, E. Marchand, C. Collewet, and C. Leroux. Active rough shape estimation of unknown objects. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 3622–3627, Nice, France, September 2008.

[5] B. Wang, L. Jiang, J.W. Li, H.G. Cai, and H. Liu. Grasping unknown objects based on 3D model reconstruction. In *IEEE/ASME International Conference on Advanced Intelligent Mechatronics*, Monterrey, California, July 2005.

[6] A. Morales, P.J. Sanz, A.P. del Pobil, and A.H. Fagg. Vision-based three-finger grasp synthesis constrained by hand geometry. *Robotics and Autonomous Systems*, 54(6):496–512, June 2006.

[7] Daniel Aarno, Johan Sommerfeld, Danica Kragic, Nicolas Pugeault, Sinan Kalkan, Florentin Wörgötter, Dirk Kraft, and Norbert Krüger. Early reactive grasping with second order 3D feature relations. In *IEEE Conference on Robotics and Automation*, Jeju Island, Korea, 2007.

[8] Ashutosh Saxena, Justin Driemeyer, and Andrew Y. Ng. Robotic grasping of novel objects using vision. *International Journal of Robotics Research*, 27(2):157–173, Feb 2008.

[9] T.G Murphy, D.M. Lyons, and A.J. Hendriks. Stable grasping with a multi-fingered robot hand: A behavior-based approach. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, volume 2, pages 867–874, Yokohama, Japan, July 1993.

[10] A. Namiki, Y. Nakabo, I. Ishii, and M. Ishikawa. High speed grasping using visual and force feedback. In *IEEE International Conference on Robotics and Automation*, Detroit, MI, USA, May 1999.

[11] R. Platt Jr., A. H. Fagg, and R. Gruppen. Nullspace composition of control laws for grasping. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 1717–1723, Lausanne, Switzerland, 2002.

[12] H. Kawasaki, T. Mouri, J. Takai, and S Ito. Grasping of unknown object imitating human grasping reflex. In *IFAC world congress*, Barcelona, July 2002.

[13] E. Marchand and F. Chaumette. Virtual visual servoing: a framework for real-time augmented reality. In *EUROGRAPHICS 2002*, volume 21(3), pages 289–298, Saarebrücken, Germany, 2002.

[14] M. Prats, P. Martinet, Angel P. del Pobil, and S. Lee. Robotic execution of everyday tasks by means of external vision/force control. *Intelligent Service Robotics*, 1(3):253–266, 2008.

[15] A. Stemmer, G. Schreiber, K. Arbter, and A. Albu-Schäffer. Robust assembly of complex shaped planar parts using vision and force. In *IEEE International Conference on Multisensor Fusion and Integration*, pages 493–500, Heidelberg, Germany, September 2006.

[16] V. Lepetit and P. Fua. Monocular model-based 3d tracking of rigid objects. *Foundations and Trends in Computer Graphics and Vision*, 1(1):1–89, 2005.

[17] T. Drummond and R. Cipolla. Real-time visual tracking of complex structures. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7):932–946, 2002.

[18] Andrew I. Comport, Éric Marchand, and François Chaumette. Robust model-based tracking for robot vision. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 692–697, 2004.

[19] Andrew I. Comport, Éric Marchand, and François Chaumette. Object-based visual 3d tracking of articulated objects via kinematic sets. In *Computer Vision and Pattern Recognition Workshop*, volume 1, page 2, Washington, DC, USA, 2004. IEEE Computer Society.

[20] Andrew I. Comport, Danica Kragic, Éric Marchand, and François Chaumette. Robust real-time visual tracking: Comparison, theoretical analysis and performance evaluation. In *IEEE International Conference on Robotics and Automation*, pages 2852–2857, Barcelona, Spain, 2005.

[21] S. Hutchinson, G. Hager, and P. Corke. A tutorial on visual servo control. *IEEE Transactions on Robotics and Automation*, 12(5):651–670, Oct 1996.

[22] B. Espiau, F. Chaumette, and P. Rives. A new approach to visual servoing in robotics. *IEEE Transactions on Robotics and Automation*, 8(3):313–326, Jun 1992.