A Robust Real-Time 3D Tracking Approach for Assisted Object Grasping

Claudio Loconsole*, Fabio Stroppa*, Vitoantonio Bevilacqua§, and Antonio Frisoli*

*PERCRO, TeCIP Scuola Superiore Sant'Anna , Pisa
§Dipartimento di Ingegneria Elettrica e dell'Informazione (DEI), Politecnico di Bari, Bari

Abstract. Robotic exoskeletons are being increasingly and successfully used in neuro-rehabilitation therapy scenarios. Indeed, they allow patients to perform movements requiring more complex inter-joint coordination and gravity counterbalancing, including assisted object grasping. We propose a robust RGB-D camera-based approach for automated tracking of both still and moving objects that can be used for assisting the reaching/grasping tasks in the aforementioned scenarios. The proposed approach allows to work with non pre-processed objects, giving the possibility to propose a flexible therapy. Moreover, our system is specialized to estimate the pose of cylinder-like shaped objects to allow cylinder grasps with the help of a robotic hand orthosis.
To validate our method both in terms of tracking and of reaching/grasping performances, we present the results achieved conducting tests both on simulations and on real robotic-assisted tasks performed by a patient.

Keywords: Active exoskeletons, 3D Tracking, Reaching/Grasping task

1 Introduction

Hemiparesis of upper extremity represents a common impairment affecting patients after stroke [1]. Rehabilitation robots can be successfully employed since the earlier phases of recovery from stroke [2]. They are capable of performing (and assisting) movements in 3D real world and can overcome some of the major limitations of traditional assisted training [3].

Rehabilitation conducted in real embeddings (see Fig. 1) requires the robot assisting and guiding patient's reaching/grasping movements towards a real object with the correct object pose, in order to grasp it. To accomplish this objective, the therapy setting should be endowed with a robust, smart and fast system for automated tracking of moving objects. Moreover, the tracking system should: (a) be as much transparent as possible to the therapist, to let his/her hands manipulate the objects without affecting the system performance; (b) manage different kind of objects that can be even not pre-processed by the system for having a flexible therapy framework. As an example, in [4], we proposed a computer-vision based approach to let an upper limb exoskeleton reach and track objects in the robot workspace.

RGB-D cameras represent an appropriate hardware tool to implement such a system. In literature, several methods have been proposed to track real objects in images/video streams, most of which combine 2D tracking algorithms with 3D information and make use of model-based approach (known or pre-processed objects) [5–7].

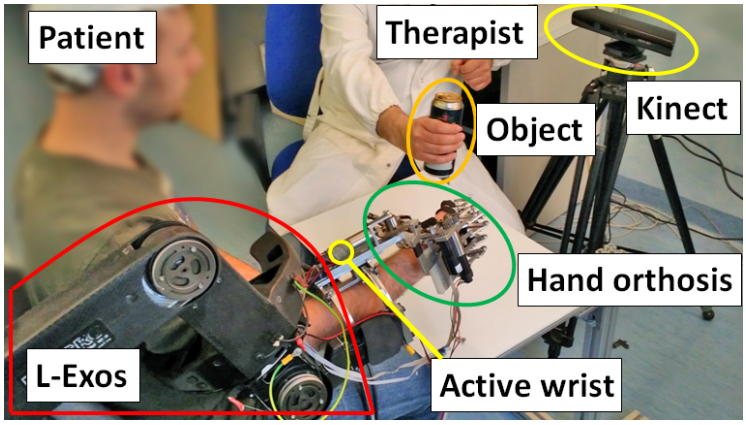


Fig. 1. The envisaged neuro-rehabilitation scenario

However, considering the state-of-the-art approaches [5, 8, 9], in order to achieve a robust tracking module which is able to locate and track real objects for automated robot-assisted reaching/grasping movements in stroke, there are several issues that need to be solved: (a) the required knowledge of the objects to be used; (b) the use of 2D features that worsen the performance of 3D feature-based methods; (c) the transparency of the tracking system; (d) the lack of object reconstruction techniques due to occlusions; (e) the use of techniques which do not merge different types of object feature to improve the robustness of the tracking. To overcome the above issues, in this paper, we propose a robust approach for robot-assisted object grasping which exploits a 3D tracking algorithm for general non pre-processed objects and can be deployed in neuro-rehabilitation scenarios.

2 The proposed approach for assisted object grasping

The proposed approach for assisted object grasping makes use of two main modules: the *active robot exoskeleton* for the upper limb, supporting the patient during reaching/grasping tasks and the *tracking system module*, modeling the environment around the exoskeleton and the patient, in order to let them interact with real objects.

2.1 The Active Robot Exoskeleton

The Active Robot Exoskeleton supports the patient in the reaching/grasping tasks conducted in neuro-rehabilitation therapies. It is composed by three main submodules (see Fig. 1): 1) the Light Exoskeleton featuring 4 Degrees of Freedom (DOFs) used to support the patient's impaired arm for the reaching part of the task; 2) the BRAVO hand orthosis featuring 2 DOFs and 3) the active wrist featuring 2 DOFs which support the patient's hand and wrist movements during the grasping part of the task. Exploiting the redundancy of the kinematic chain of the L-Exos for reaching tasks (that is considering only spatial position and not the orientation of the end-effector), as reported in a previous work [10], it is possible to mimic the human behavior. The two DOFs of the active wrist

allow the prono-supination and the flexion-extension of the wrist (so it does not support the adduction/abduction). It follows that, keeping decoupled the L-Exos (for redundancy exploitation) and the active wrist kinematics, it is not possible to fully align (that is having the same orientation) the hand and the cylinder axis to perform the grasping task. Moreover, for the selected type of graspable objects, the flexion-extension results to be useless. For this aim, we set up a specific procedure for the best-effort alignment of the hand with respect to the cylinder-shaped graspable object.

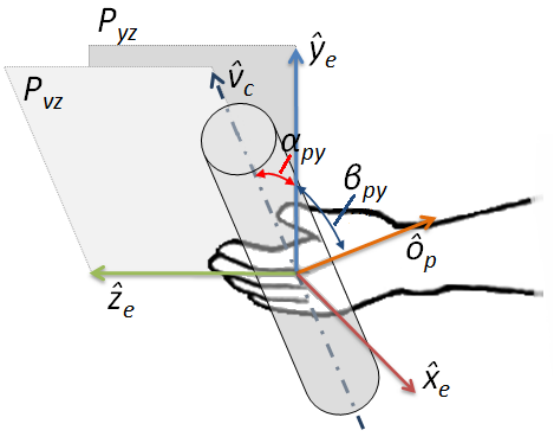


Fig. 2. The geometrical representation of the wrist alignment for grasping.

Wrist alignment for grasping With respect to the end-effector reference framework ($\hat{x}_e$, $\hat{y}_e$, $\hat{z}_e$) as shown in Fig. 2, the unit vector of the cylinder orientation axis is defined as:

$$\hat{v}_c = \langle v_{c_x}, v_{c_y}, v_{c_z} \rangle$$

In order to grasp the object with the hand orthosis, the condition of coincidence between $\hat{y}_e$ and the cylinder axis $\hat{v}_c$ would be required. However, as mentioned in the previous sections, to perform the grasp of the object, we will use only one DOF of the active wrist corresponding to its prono-supination (rotation around $\hat{z}_e$). Hence, taking into account this limitation, the best achievable alignment results to be the one obtained when $\hat{y}_e$ lies on the plane P_{vz} identified by $\hat{v}_c$ and $\hat{z}_e$.

Therefore, this can be pursued if the wrist rotates around $\hat{z}_e$ of the angle α_{py} between P_{vz} and $\hat{y}_e$. The latter angle is calculated as the complementary angle β_{py} between $\hat{y}_e$ and the vector $\boldsymbol{o}_p$ which, in turn, is perpendicular to P_{vz}. The unit vector $\hat{o}_p$ of $\boldsymbol{o}_p$ required for the angle calculation can be obtained through the cross product:

$$\hat{o}_p = \hat{z}_e \times \hat{v}_c = \langle 0, 0, 1 \rangle \times \langle v_{c_x}, v_{c_y}, v_{c_z} \rangle =$$

$$\langle (0 * v_{c_z}) - (1 * v_{c_y}), (1 * v_{c_x}) - (0 * v_{c_z}), (0 * v_{c_y}) - (0 * v_{c_x}) \rangle =$$

$$\langle -v_{c_y}, v_{c_x}, 0 \rangle$$

Once $\hat{o}_p$ has been calculated, the rotation angle α_{py} is equal to:

$$\alpha_{py} = \arcsin\left(\frac{\hat{o}_p \cdot \hat{y}_e}{\|\hat{o}_p\|\|\hat{y}_e\|}\right) = \arcsin\left(\frac{\langle -v_{c_y}, v_{c_x}, 0\rangle \cdot \langle 0,1,0\rangle'}{\|\langle -v_{c_y}, v_{c_x}, 0\rangle\|\|\langle 0,1,0\rangle\|}\right) = \arcsin(v_{c_x})$$

It is worth to notice that in order to obtain the correct sign for the angle α_{py}, the unit vector $\hat{v}_c$ is always considered having the v_{c_y} coordinate positive ($v_{c_y} \geq 0$), that is considering the orientation of the cylinder axis toward the upper hemisphere of the 3D space.

2.2 The Tracking System Module

The Tracking System Module, based on computer vision, uses a RGB-D camera (Microsoft Kinect) and aims to track general non pre-processed objects. Moreover, for cylinder-like shaped objects, it allows to perform special routines to manage occlusions and to estimate their orientation axis for grasping tasks. The module is a substantial improvement of the original computer vision-based system proposed in [4], both in terms of applied techniques and potential applications. The proposed fast and robust tracking algorithm is performed in 3D and is able to detect any generic shaped object. Moreover, due to the fact that the hand orthosis (see Fig. 1) can carry out only cylindrical grasps (constraint imposed by its mechanical structure), the tracking algorithm provides specific features for cylindrical-like shaped objects. The entire procedure is based on an iterative algorithm that can be logically sub-divided into two main parts:-*2D Pre-Processing phase* and the *3D Processing phase*. As mentioned before, for graspable (cylindrical) objects, the algorithm provides also their pose (position and orientation) and an "Occlusion Management procedure". Figure 3 shows the flow diagram of the entire Tracking System module.

2D Pre-Processing phase This phase allows to speed up the entire workflow of the tracking module removing useless data from the 2D image, since the 3D filtering requires high computational costs. It is composed by two functions: the *Coarse Depth Filtering* and the *Skin Removal*. The latter is necessary because we plan to remove the skin pixels such that the therapist's hands do not interfere with the object tracking, since our interface must be as transparent as possible to the therapist during the object handling phase. However, by removing the skin pixels from the image, the object can result "spatially divided" by the therapist's hand and it can be considered as two or more separate different objects. Of course, this requires a proper occlusion management procedure that has been realized in the 3D processing phase section.
3D Processing phase At the end of the 2D Pre-Processing phase, a 3D Point Cloud (PC) is obtained by the filtered 2D image and the depth map, through the use of the intrinsic parameters of the camera model. Then, the 3D Processing phase is performed, consisting of three sub-phases: *1) 3D Filtering and Clustering*, *2) Cylinder Recognition* and *3) 3D Tracking*. More in detail, the Cylinder Recognition step can be logically subdivided into *Cylinder Segmentation* and *Occlusion Management Procedure for Cylinders*. The latter sub-function allows to opportunely manage the two main occlusion cases respectively caused by

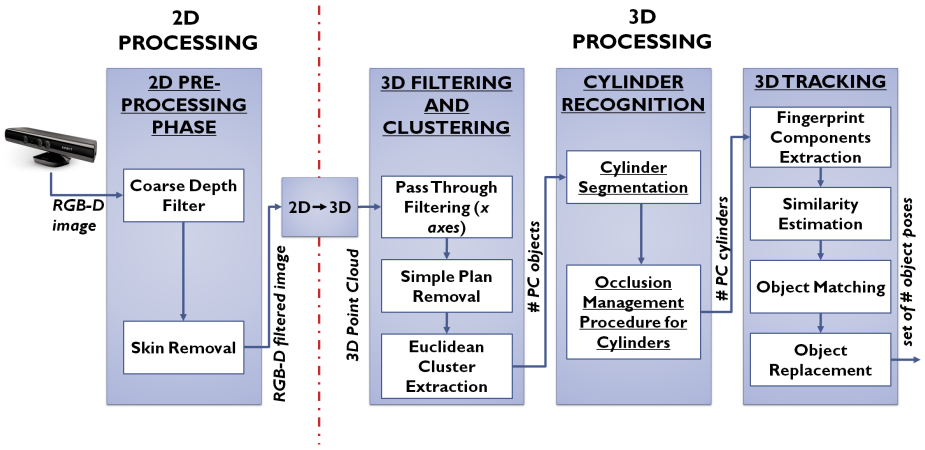


Fig. 3. The entire flow diagram of the Tracking System module

occlusions due to different objects and occlusions due to the object itself for perspective reasons.

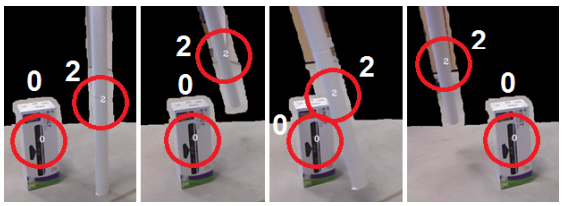
The last function of the 3D Processing phase uses an innovative and robust tracking algorithm based on Viewpoint Feature Histogram (VFH) 3D features. Recognition is performed by processing two consecutive frames of the video, on-line performing object matching and exploiting object fingerprint matching techniques based on four classes of features: 1) the VFH features, 2) the 3D position, 3) the average color and 4) the number of points (size) of the object PC. These four classes of features are considered as the *fingerprint components* of a particular object and are stored in memory to keep track of each object appeared during the video stream. Hence, for each couple of objects x,y (current object fingerprint, stored object fingerprint) we define and calculate the *Similarity Distance* based on the euclidean distances of the four aforementioned classes of features.

3 Experimental description and results

In order to validate the proposed approach, we conducted two series of tests: the first aims to evaluate the robustness of the tracking algorithm, whereas the second aims to demonstrate the overall system performances in reaching/grasping tasks. Next sections focus respectively on the first and the second series of tests.

3.1 Experimental validation of the tracking system module

In order to validate the tracking algorithm, several tests have been successfully conducted to evaluate: 1) the average of the computational time, 2) the maximum object speed to assure object tracking, 3) the robustness to: -changes of light conditions; -object roto-translations; -object occlusions (the latters due both to other objects 4(a) and to human occlusions 4(b)). Regarding the computational time estimation, the average of the time registered over 100 frames



(a) Occlusion among objects

(b) Occlusion due to a person

Fig. 4. Different examples of a complete occlusion management

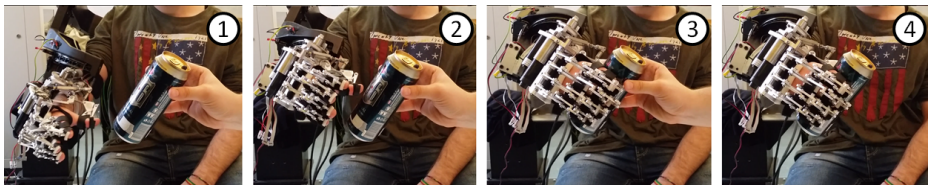


Fig. 5. The sequence of a reaching/grasping task example grabbed from the Kinect point of view.

has been calculated considering 3 different cases and achieving the following results: tracking of 1) single object (66.4 ms); 2) two objects (90.6 ms); 3) three objects (161.8 ms). The tracking module runs on an Intel Core2 Quad Q9550 (12M Cache, 2.83 GHz, 1333 MHz FSB) PC with 3 GB RAM and Windows 7 operating system. The maximum object speed, instead, has been calculated using an iterative method. The experiment consisted in evaluating the correct tracking between two adjacent camera frames of a moving object to be tracked. Object to be tracked has been put at a starting distance from the Kinect of 1500 mm. The object was, then, translated of several predefined offsets (re-starting always from the same point) along the Kinect x axis (i.e. the horizontal direction on the camera image plane). The offset has been iteratively incremented of 100 mm for each test. It has been observed that the maximum distance that an object can get without been lost by the algorithm is 600 mm. With reference to the computation time of a single object in the scene (66.4 ms), the maximum speed that an object may feature in its motion to be correctly tracked is $9,03$ m/s. Finally, we successfully test the robustness of the tracking algorithm to different light conditions, to object roto-translation and to occlusions.

3.2 Reaching/grasping task experiments

A healthy subject (male, 25yrs) was involved in this experiment. He was asked to remain completely passive during the experiment, in order to simulate the

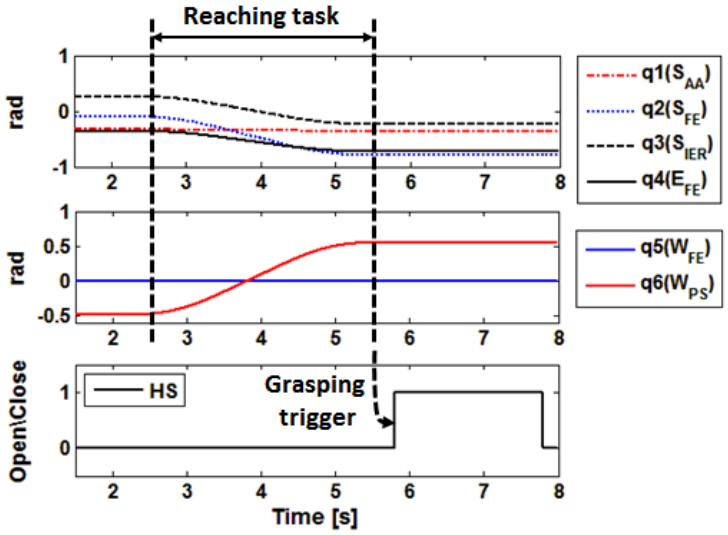


Fig. 6. The three plots respectively report the trends of the L-Exos and the wrist joints (1 and 2) and the Hand Signal that drives the opening/closing of the hand. With S_{AA} we refer to "*Shoulder adduction/abduction*", whereas with S_{FE} to "*Shoulder flexion/extension*", S_{IER} to "*Shoulder internal/external rotation*", S_{FE} to "*Elbow flexion/extension*", W_{FE} to "*Wrist flexion/extension*", W_{PS} to "*Wrist pronation/supination*", and HS to "*Hand Signal*".

impaired upper limb of a post-stroke patient. In Fig. 5, we show the sequence of a reaching/grasping task example grabbed from the Kinect point of view. The therapist shows a cylinder-shaped object (partially occluded from his hand) to the tracking system. It provides to the path planner of the active exoskeleton the object 3D pose (position and orientation). Then, the L-Exos correctly supports and drives the patient's arm toward the object 3D position. In the meanwhile, the active wrist supports the movements of the patient's wrist (in our case only its pronation/supination) to align with the cylinder axis orientation. Finally, once the object has been reached, the hand orthosis is triggered to help the patient to grasp the object closing his hand. It is worth to notice that the performed task allows the patient to hold the object without the help of the therapist (see Fig. 5.4).

The plots reported in Fig. 6 show the trends of all the robot joints involved in the reaching/grasping task execution. The L-Exos joints (q_1, q_2, q_3 and q_4), as well as the wrist joint used for prono-supination (q_6, q_5 corresponding to wrist flexion/extension is not used for path planning and is kept fixed), follows the fully synchronized bounded jerk trajectory planning algorithm reported in [10] which allows to mimic human behavior in reaching/grasping tasks. Furthermore, in the same figure, it is possible to observe that the end of the reaching task triggers the closure of the hand orthosis (after a constant time delay)allowing the grasp and the hold of the object.

4 Conclusion and future works

In this paper, we presented a robust approach for assisted object grasping which exploits a RGB-D camera-based algorithm for real-time tracking of non pre-processed real objects. The proposed tracking module is also specialized to discriminate and provide specific geometrical features, as well as proper occlusion management procedures for cylinder-like shaped objects.

The conducted tests have demonstrated the robustness of the proposed approach, as well as its performance in a neuro-rehabilitation scenario through reaching/grasping task experiments. Regarding future works, the planned actions deal with the validation of the system by post-stroke patients.

Acknowledgments

This work has been partially funded from the EU FP7 project n. 601165 WEARHAP.

References

1. S. Beckelhimer, A. Dalton, C. Richter, V. Hermann, and S. Page, "Computer-based rhythm and timing training in severe, stroke-induced arm hemiparesis," *The American Journal of Occupational Therapy*, vol. 65, no. 1, p. 96, 2011.
2. H. Krebs, B. Volpe, D. Williams, J. Celestino, S. Charles, D. Lynch, and N. Hogan, "Robot-aided neurorehabilitation: a robot for wrist rehabilitation," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 15, no. 3, pp. 327–335, 2007.
3. G. Prange, M. Jannink, C. Groothuis-Oudshoorn, H. Hermens, and M. IJzerman, "Systematic review of the effect of robot-aided therapy on recovery of the hemiparetic arm after stroke," *Journal of Rehabilitation Research and Development*, vol. 43, no. 2, p. 171, 2006.
4. C. Loconsole, F. Banno, A. Frisoli, and M. Bergamasco, "A new kinect-based guidance mode for upper limb robot-aided neurorehabilitation," in *Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ International Conference on.* IEEE, 2012, pp. 1037–1042.
5. C. Choi, S.-M. Baek, and S. Lee, "Real-time 3d object pose estimation and tracking for natural landmark based visual servo," in *Intelligent Robots and Systems, 2008. IROS 2008. IEEE/RSJ International Conference on.* IEEE, 2008, pp. 3983–3989.
6. V. Bevilacqua, P. Casorio, and G. Mastronardi, "Extending hough transform to a points cloud for 3d-face nose-tip detection," in *Advanced Intelligent Computing Theories and Applications. With Aspects of Artificial Intelligence.* Springer, 2008, pp. 1200–1209.
7. A. D. Sappa, V. Bevilacqua, and M. Devy, "Improving a genetic algorithm segmentation by means of a fast edge detection technique," in *Image Processing, 2001. Proceedings. 2001 International Conference on*, vol. 1. IEEE, 2001, pp. 754–757.
8. P. Alimi, "Object persistence in 3d for home robotics," 2012.
9. L. Masson, M. Dhome, and F. Jurie, "Robust real time tracking of 3d objects," in *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, vol. 4. IEEE, 2004, pp. 252–255.
10. A. Frisoli, C. Loconsole, R. Bartalucci, and M. Bergamasco, "A new bounded jerk on-line trajectory planning for mimicking human movements in robot-aided neurorehabilitation," *Robotics and Autonomous Systems*, vol. 61, no. 4, pp. 404–415, 2013.