Markerless Self-Recognition and Segmentation of Robotic Manipulator in Still Images

ICRA 2013 Mobile Manipulation Workshop on Interactive Perception

Damien Teney
University of Liège, Belgium

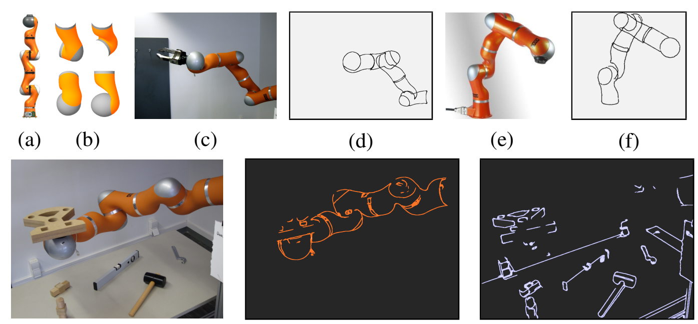
Dadhichi Shukla
University of Burgundy, France

Justus Piater
University of Innsbruck, Austria

Vision is a crucial capability for enabling robots to perceive and interact with their environment, e.g. manipulating or grasping objects. A current trend is bringing closer the aspects of interaction and perception, on the one hand by integrating visual information directly in the control process, and on the other hand, using interaction itself to help perception, allowing robots to explore their environment. In the context of manipulation, physical parts of the robot are then likely to appear in the observations, and an important capability emerges as the recognition of those parts, in order to separate the observations of the scene from those of the robot itself. Identifying the robot's own body parts in input images has been used before in different ways, helping obstacle avoidance or control directly (through visual servoing [1]). However, this is usually performed via indirect methods, tracking fiducial markers purposely attached to the robot [2], which imposes undesirable (e.g. visibility) constraints. Some recent work adresses the pose estimation of a robot manipulator *directly* [3], [4], but these methods focus on tracking the manipulator between consecutive frames, whereas the initial recognition is considered as the harder part. We propose a method for markerless, monocular recognition and pose estimation of an articulated robot arm, dealing with single images without initialization, allowing its use with unknown hand-eye calibration, imprecise kinematics or missing position feedback.

We use an existing appearance-based recognition method [5], that relies on object edges and contours, allowing the recognition of objects with few characteristic visual features (i.e. non-textured). The system is trained separately for each articulated link of the robot arm, using synthetic images of that link in known poses. The recognition of those elements proves extremely challenging, as their appearance may not offer very distinguishable visual clues, and because of the typical unstructured environment (background clutter) and possible (self-)occlusions. The initial recognition produces a set of candidate poses for each link, which are then combined, enforcing the known kinematic constraints between the links. These constraints are modeled as pairwise compatibility functions in a classical Markov random field, with a node for each link. Inference is carried out with an algorithm inspired by non-parametric belief propagation. We efficiently limit the evaluation of densities in the pose space to the discrete points proposed by the initial recognition step, thereby ensuring adequate efficiency. The algorithm ultimately recovers the pose of all the elements (links) of the arm. We can then classify the image features of the input scene as belonging to the scene or to the robot manipulator itself, simply by measuring their similarity with the training templates of the links in the identified poses. The poses of the links can also be used to recover the angles at each joint, together with the cartesian position of each element of the robot relative to the camera.

The system was implemented and tested with a Kuka Lightweight Robot arm. We considered the five internal links, of which four are completely identical in appearance, which constitutes an additional challenge. The four joints (revolute, both axial and hinge-like) are specified by the alignment of the joint axes of the adjacent links. Training images of the two types of links were generated with CAD software, at viewpoints about 30° apart. Although the recognition of individual links cannot be relied upon for any practical purposes, the pose ultimately recovered for the whole arm by the proposed algorithm was correct during most of our tests. The probabilistic inference can handle missing detections of links to some degree, as can happen with (self-)occlusions. The classification of the image features of the test image as robot/non-robot parts, as mentioned above, proved effective, and superior to using intermediate segmentation masks. Indeed, our procedure can handle occlusions, for example when the robot is manipulating an object. The capabilities offered by the whole system should help and make more robust the subsequent processing of visual data in the context of joint perception/manipulation scenarios. Future work will aim at relaxing the assumptions of known link and joint geometries, e.g. reusing existing work on autonomous learning of articulated models [6].



(a) The Kuka LWR used in our experiments (b) Synthetic training images of individual links (c, e) Test images and (d, f) rendering of the training templates as recognized for each link (g-i) After recognition of the robot arm, image features are classified as "robot" (orange) and "non-robot" (blue).

Research supported through European Commission's FP7 projects Xperience and IntellAct, and by the Belgian National Fund for Scientific Research.

[1] M. Prats, P. Martinet, A. del Pobil, and S. Lee, "Robotic execution of everyday tasks by means of external vision/force control," *Intelligent Service Robotics*, vol. 1, no. 3, pp. 253–266, 2008.

[2] K. Welke, J. Issac, D. Schiebener, T. Asfour, and R. Dillmann, "Autonomous acquisition of visual multi-view object representations for object recognition on a humanoid robot," in *ICRA*, 2010.

[3] J. J. Sorribes, M. Prats, and A. Morales, "Visual tracking of a jaw gripper based on articulated 3D models for grasping," in *ICRA*, 2010.

[4] X. Gratal, J. Romero, and D. Kragic, "Virtual visual servoing for real-time robot pose estimation," in *IFAC*, 2011.

[5] D. Teney and J. Piater, "Continuous pose estimation in 2D images at instance and category levels," in *Comp. and Rob. Vis.*, 2013.

[6] D. Katz and O. Brock, "Extracting planar kinematic models using interactive perception," in *Unifying Perspectives In Computational and Robot Vision*, no. 8. Springer Verlag, May 2008, pp. 11–23.