Studying Table-Top Manipulation Tasks: A Robust Framework for Object Tracking in Collaboration

Peter Lightbody
L-CAS, School of Computer Science
University of Lincoln, U.K.
plightbody@lincoln.ac.uk

Paul Baxter
L-CAS, School of Computer Science
University of Lincoln, U.K.
pbaxter@lincoln.ac.uk

Marc Hanheide
L-CAS, School of Computer Science
University of Lincoln, U.K.
mhanheide@lincoln.ac.uk

ABSTRACT

Table-top object manipulation is a well-established test bed on which to study both basic foundations of general human-robot interaction and more specific collaborative tasks. A prerequisite, both for studies and for actual collaborative or assistive tasks, is the robust perception of any objects involved. This paper presents a real-time capable and ROS-integrated approach, bringing together state-of-the-art detection and tracking algorithms, integrating perceptual cues from multiple cameras and solving detection, sensor fusion and tracking in one framework. The highly scalable framework was tested in a HRI use-case scenario with 25 objects being reliably tracked under significant temporary occlusions. The use-case demonstrates the suitability of the approach when working with multiple objects in small table-top environments and highlights the versatility and range of analysis available with this framework.

CCS CONCEPTS

• **Computing methodologies** → **Vision for robotics**; • **Human-centered computing** → *Collaborative interaction*; • **Computer systems organization** → *Robotics*;

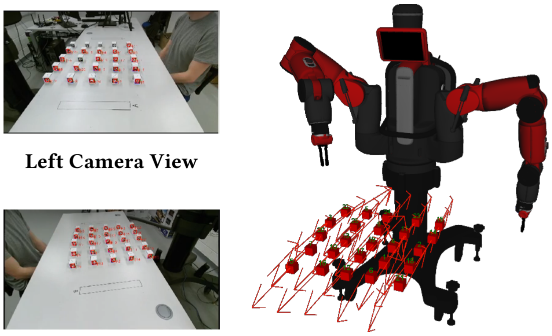
KEYWORDS

Fiducial Markers; Visual Tracking; Human Robot Collaboration

ACM Reference Format:
Peter Lightbody, Paul Baxter, and Marc Hanheide. 2018. Studying Table-Top Manipulation Tasks: A Robust Framework for Object Tracking in Collaboration. In *HRI '18 Companion: 2018 ACM/IEEE International Conference on Human-Robot Interaction Companion, March 5–8, 2018, Chicago, IL, USA.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3173386.3177045

1 INTRODUCTION

As the field of robotics progresses, the collaboration between humans and robots, rather than the replacement of humans by robots, will move to the forefront of the robotic world. The number of tasks being completed by a human-robot collaborative pair is increasing, but the frameworks to allow this to happen efficiently are not yet in place. This is in part due to the difficulty in accurately observing human-robot collaboration, which is ever more present when operating in smaller environments or in close proximity. Many of the attempts to mitigate these difficulties involve reducing the complexity of the task, the number of objects present or the dimensions in which the objects are monitored and tracked or by virtualising the interactions [2]. These methods can, however, often lead to systems which cannot scale to the complexity of real-world situations. Alternatively, some setups provide multi-camera systems accompanied by passive or active markers, which are able to track numerous items both in 3D and in real time. These commercial systems, such as the Vicon Motion Tracking System [10], are expensive, however, and suffer greatly from occlusion, which is particularly harmful in small tabletop environments.


Left Camera View

Right Camera View

Real-Time Visualisation

Figure 1: A snapshot visualisation of the setup

This paper presents a visual tracking system which can reliably identify and track a large set of objects. To evaluate the technical contribution of the real-time visual object identification and tracking, we present our framework in the context of a HRI study investigating the interpersonal differences between users performing a collaborative task with a robot, a hypothesis previously explored in [6]. The pre-study, conducted with 14 subjects in an object sorting task, demonstrates a framework which will significantly improve the efficiency of detection and tracking in tabletop human-robot collaboration.

2 SYSTEM COMPONENTS

The visual system is composed of two essential components: the Bayesian tracker, which merges individual detections into coherent, long-term estimates of the true pose of the objects, and the detection system, which in our case employs fiducial markers to allow the reliable discrimination of a fixed number of predefined markers.

Multi-Camera Bayesian Tracking: We have extended the Bayesian Tracking Framework described in [3] to allow a third dimension

HRI '18 Companion, March 5–8, 2018, Chicago, IL, USA

ACM ISBN 978-1-4503-5615-2/18/03.
https://doi.org/10.1145/3173386.3177045

when tracking, in addition to allowing the framework to distinguish between uniquely identified objects. This framework integrates the results from the object detection when multiple detectors are in use and is able to accurately estimate the 3D position of tracked items at a fixed frame rate, independently of the rate at which the sensors are operating.

An Unscented Kalman Filter, configured with a constant velocity motion model for prediction and fixed noise models for each sensor's observations, allows the framework to compensate for any temporary loss of detection, as described in [5], and limits the impact of occlusion. For each newly predicted observation, a gating procedure employs a validation region, which is relative to the target object [1] and which reduces the chance of assigning false positives. A Nearest Neighbour (NN) association algorithm then associates the detections with the correct target, resulting in only detections of the same object type being associated with the corresponding tracks. If no suitable target tracks are found, the detections are stored and eventually used to create a new track; providing their stability is shown over a predefined time frame. A further benefit is that this tracking-by-detection and recognition framework can be utilised alongside any object detection system.

Fiducial Marker Detection: To allow the identification of a large number of objects in real-time, and to provide the six degrees of freedom for each detected object, we incorporated fiducial markers in place of more common computer vision techniques. The marker detection algorithm utilised within this system was the AprilTags [9] detection system.

One problem, however, was that the implementation in a table-top environment meant that the markers had to be downsized to fit on the small objects used. Operating at this scale meant that in order to detect the AprilTag marker, the resolution of the cameras had to be increased from 640x480 to 960x540, which introduced a significant time delay in the real-time tracking, varying from two to sixteen seconds. To combat this, and as will be discussed below, a follow-up study is proposed which will utilise the round WhyCode Marker [7], an extension to the WhyCon Marker described in [4].

3 USE CASE: TABLE-TOP OBJECT SORTING

The experiment setup, shown in Fig 1, is comprised of a grid of 25 small square blocks, each depicting a single line which had both an orientation and a length. The human partner and a Baxter robot were positioned on opposite sides of the grid, monitored by two RGB cameras placed on either side of the collaborative pair. After being shown six examples that were not present in the dataset, the user was asked to categorise the blocks into two sets. Unbeknownst to the user, the categories were determined by the length of the line alone, with the orientation acting as a distractor. The robot, acting as our expert, periodically suggested blocks that belonged to a particular category. For each user, the robot operated on a certain policy; either suggesting blocks that fell close to the category boundary or alternatively suggesting blocks which sat far from the category boundary. Each policy was used on 50% of the participants and the task itself was repeated four times by each participant.

As mentioned previously, our study was undertaken in the context of identifying whether individuals complete similar tasks differently, but the versatility of the framework enables utilisation in a wide variety of setups. Even though the task itself involved many instances in which occlusion occurred, the implemented framework was able to accurately track and identify each of the 25 items within the scene and returned clear differences in each individual's completion of the task and collaboration with the robot.

4 RESULTS

The results of the pre-study, which for brevity can be accessed through https://goo.gl/JLFP2a, clearly demonstrate a tracking framework which alleviates many of the real-world tracking issues faced by HRI researchers and provides a system which is able to accurately and efficiently track multiple objects in small tabletop environments.

The performance statistics for each of the systems components, as well as a comparison to other systems in a more general context, can be found in [8].

5 FUTURE WORK

The current tracking delay within the system jeopardises the beneficial impact that this system could have on the field of HRI. We propose to repeat the previous study with the AprilTag system replaced with the WhyCode tracking system, which is significantly faster [7] and operates independently of the image size. This will allow the processing of higher resolution images without the delay, thus allowing us to provide further validation for our hypothesis.

ACKNOWLEDGEMENTS

The work is partially supported by the Research Investment Fund (RIF) of the University of Lincoln, and the European Community's funded projects 732773 'ILIAD'.

REFERENCES

[1] Yaakov Bar-Shalom and Xiao-Rong Li. 1995. *Multitarget-multisensor tracking: principles and techniques*. Vol. 19. YBs London, UK:.

[2] Paul Baxter, Rachel Wood, and Tony Belpaeme. 2012. A Touchscreen-Based 'Sandtray' to Facilitate, Mediate and Contextualise Human-Robot Social Interaction. In *7th ACM/IEEE International Conference on Human-Robot Interaction*. IEEE Press, Boston, MA, U.S.A., 105–106.

[3] Nicola Bellotto and Huosheng Hu. 2010. Computationally efficient solutions for tracking people with a mobile robot: an experimental evaluation of Bayesian filters. *Autonomous Robots* 28, 4 (2010), 425–438.

[4] Tomáš Krajník, Matías Nitsche, Jan Faigl, Petr Vaněk, Martin Saska, Libor Přeučil, Tom Duckett, and Marta Mejail. 2014. A practical multirobot localization system. *Journal of Intelligent & Robotic Systems* 76, 3-4 (2014), 539–562.

[5] X Rong Li and Vesselin P Jilkov. 2000. A Survey of Maneuvering Target Tracking: Dynamic Models. In *Proceedings of SPIE Conference on signal and data processing of small targets*, Vol. 6. FL, USA, 212–235.

[6] Peter Lightbody, Christian Dondrup, and Marc Hanheide. 2015. Make me a sandwich! Intrinsic human identification from their course of action. (2015).

[7] Peter Lightbody, Tomáš Krajník, and Marc Hanheide. 2017. An Efficient Visual Fiducial Localisation System. *SIGAPP Appl. Comput. Rev.* 17, 3 (Nov. 2017), 28–37. https://doi.org/10.1145/3161534.3161537

[8] Peter Lightbody, Tomáš Krajník, and Marc Hanheide. 2017. A Versatile High-performance Visual Fiducial Marker Detection System with Scalable Identity Encoding. In *Proceedings of the Symposium on Applied Computing (SAC '17)*. ACM, ACM, New York, NY, USA, 276–282. https://doi.org/10.1145/3019612.3019709

[9] Edwin Olson. 2011. AprilTag: A robust and flexible visual fiducial system. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 3400–3407.

[10] Vicon. [n. d.]. Vicon MX Systems. ([n. d.]). http://www.vicon.com