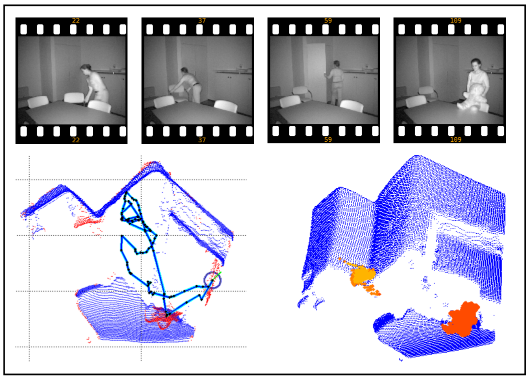
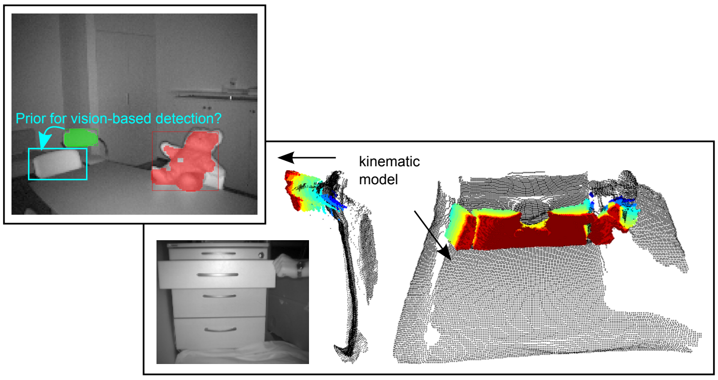
The Articulated Scene Model: Model-less Priors for Robot Object Learning?

Agnes Swadzba, Niklas Beuter, Sven Wachsmuth, and Franz Kummert
Applied Informatics, Bielefeld University
aswadzba@techfak.uni-bielefeld.de

Abstract

Human analysis of dynamic scenes consists of two parallel processing chains [2]. The first one concentrates on *motion* which is defined as variation of location while the second one processes *change* which is the variation of structure. The detection of a scene change is realized phenomenologically by comparing currently visible structures with a representation in memory. These psychological findings have motivated us to design an *articulated scene modeling* approach [1] which enables a robot to extract articulated scene parts through observing the spatial changes caused by their manipulation. This approach processes a sequence of 3D scans taken from a fixed view point which captures a dynamic scene where a human moves around and manipulates the environment by, e.g., replacing chairs or opening doors. It estimates per frame $\mathcal{F}_t$ the moving entities $\mathcal{E}_t$, the so far static scene background $\mathcal{S}_t$, and movable objects $\mathcal{O}_t$. The moving entities are tracked using a particle filter with a weak cylinder model. Static and movable scene parts are computed by a comparison of the current frame with the background model $\mathcal{S}_{t-t}$ estimated from the previous frames. For dense depth sensors, like the SwissRanger camera or the Kinect camera, such a comparison can be implemented as pixel-wise subtraction of $\mathcal{S}_{t-t}$ from $\mathcal{F}_t$. Using the fact that per pixel the farthest static depth measurements define the static background, arbitrary movable objects (like a replaced chair or an opened cupboard door) can be extracted model-less from depth measurements as they emerge in front of a known static background. The video [1] [2] shows for an Swissranger sequence the emerging of the static background (in blue), the movable objects (in orange), and the trajectories of an entity (in cyan and green) for two view points. The scene modeling part of our approach can also be presented on site in real-time on Kinect data. The development of cameras like the Kinect camera which combine dense depth measurement with a normal color camera in an elegant way, opens up new possibilities for interactive object learning. Future work could concentrate on the question whether the extracted movable objects (like chair) can be used to compute suitable features that can be used to detect, for example, other chairs in the scene which have not been moved, so far. Further, a history of positions of an articulated object like a drawer can be used to learn its kinematic model [3].





References

[1] N. Beuter, A. Swadzba, F. Kummert, and S. Wachsmuth. Using Articulated Scene Models for Dynamic 3D Scene Analysis in Vista Spaces. *3D Research*, 3(4), 2010. http://dx.doi.org/10.1007/3DRes.03(2010)04.

[2] R. A. Rensink. Change Detection. *Annual Review of Psychology*, 53:245–277, 2002.

[3] J. Sturm, V. Predeep, C. Stachniss, C. Plagemann, K. Konolige, and W. Burgard. Learning Kinematic Models for Articulated Objects. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pages 1851–1856, Pasadena, CA, USA, 2009. AAAI Press.

[1] http://www.youtube.com/watch?v=VVc9bngjB2s
[2] http://aiweb.techfak.uni-bielefeld.de/content/6d-scene-analysis