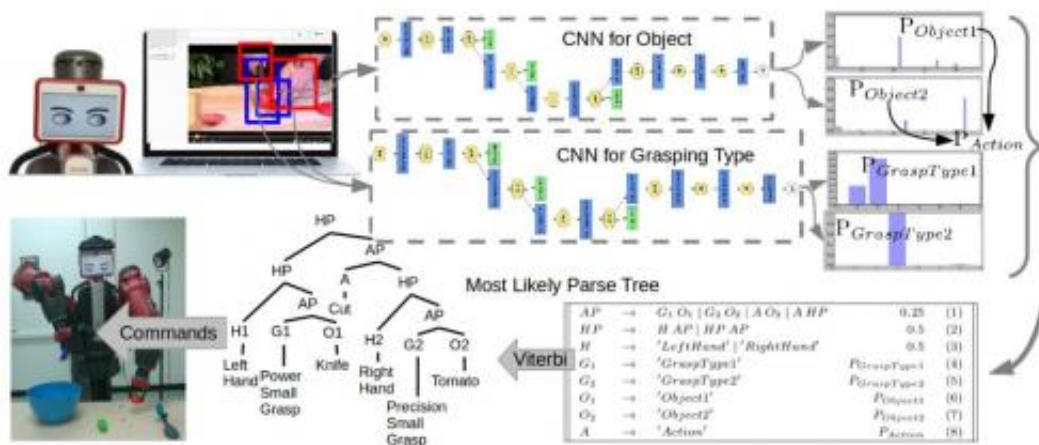# Robots do kitchen duty with cooking video dataset

January 5 2015, by Nancy Owano

The integrated system reported in this work. Credit: study paper

Now that we have robots that walk, gesture and talk, roboticists are interested in a next level: How can they learn more than they already know? The ability of these machines to learn actions from human demonstrations is a challenge for those working on intelligent systems or, in Eric Hopton's words, in writing for *redOrbit*, for instances where "you need it to do a new task that's not part of its database." Now researchers from the University of Maryland and the Australian NICTA (an information communications technology research center) have written a paper reporting they have succeeded in this area. They are to present their findings at the 29th annual conference of the Association for the Advancement of Artificial Intelligence later this month, from January 25

to 30, in Austin, Texas. They have explored what it takes for a self-learning robot to improve its knowledge about fine-grained manipulation actions –namely, cooking skills through its "watching" demonstration videos.

Their paper is titled "Robot Learning Manipulation Action Plans by 'Watching' Unconstrained Videos from the World Wide Web." In simple terms, they set a goal to see if they could build a robot that is self-learning and can improve its knowledge about fine-grained manipulation actions via demo videos.

Jordan Novet in *VentureBeat* said these researchers utilized convolutional neural networks, to identify the way a hand is grasping an item and to recognize specific objects. The system also predicts the action involving the object and the hand. The new robot-training system is based on recent advances in our understanding of "deep neural networks," said Hopton.

The authors wrote, "The lower level of the system consists of two convolutional neural network (CNN) based recognition modules, one for classifying the hand grasp type and the other for object recognition. The higher level is a probabilistic manipulation action grammar based parsing module that aims at generating visual sentences for robot manipulation."

They said their experiments showed the system was able to learn manipulation actions by 'watching' the videos with high accuracy.

To train their model, researchers selected data from 88 YouTube videos of people cooking. From there, the researchers generated commands that a robot could then execute. They said, "Cooking is an activity, requiring a variety of manipulation actions, that future service robots most likely need to learn." They conducted experiments on a cooking video dataset, YouCook. They said that data was prepared from 88 open-source

YouTube cooking videos with unconstrained third-person view. "Frame-by-frame object annotations are provided for 49 out of the 88 videos. These features make it a good empirical testing bed for our hypotheses."

The [YouCook](#) dataset, from researchers at the Department of Computer Science and Engineering, SUNY at Buffalo, explains what these videos are all about: They are downloaded from YouTube and are in the third-person viewpoint. They represent a more challenging visual problem than existing cooking and kitchen datasets.

 More information: — Robot Learning Manipulation Action Plans by "Watching" Unconstrained Videos from the World Wide Web (PDF) [www.umiacs.umd.edu/~yzyang/pap … Mani_CameraReady.pdf](#)

— YouCook: An Annotated Data Set of Unconstrained Third-Person Cooking Videos [www.cse.buffalo.edu/~jcorso/r/youcook/](#)