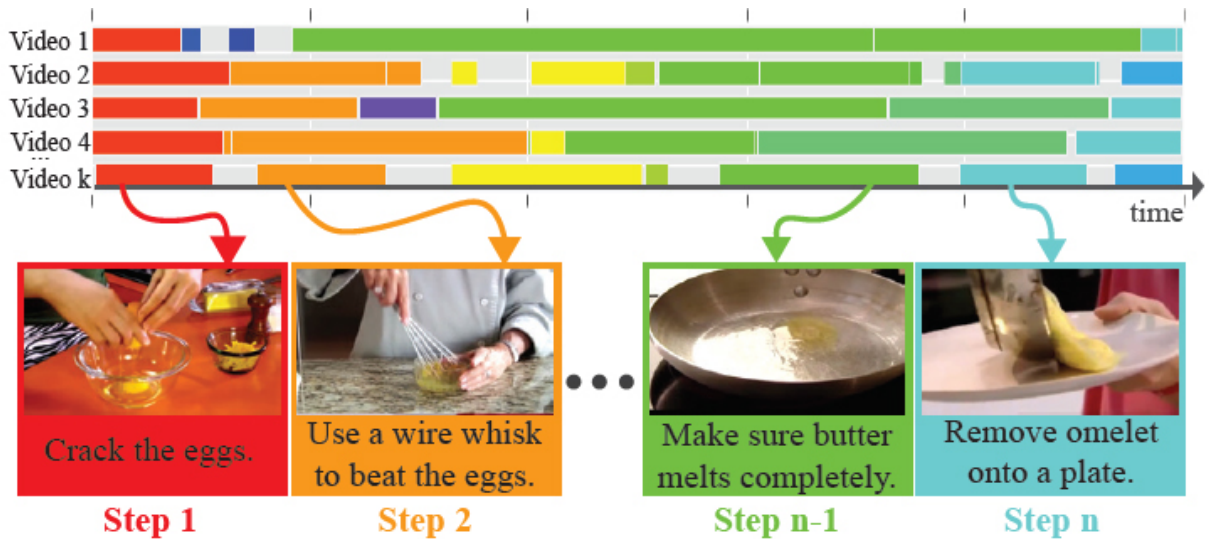


# Robots learn by watching how-to videos

December 21 2015, by Bill Steele



Scanning several videos on the same how-to topic, a computer finds instructions they have in common and combines them into one step-by-step series.

When you hire new workers you might sit them down to watch an instructional video on how to do the job. What happens when you buy a new robot?

Cornell researchers are teaching robots to watch instructional videos and derive a series of step-by-step instructions to perform a task. You won't even have to turn on the DVD player; the robot can look up what it needs on YouTube. The work is aimed at a future when we may have "personal robots" to perform everyday housework – cooking, washing dishes,

doing the laundry, feeding the cat – as well as to assist the elderly and people with disabilities.

The researchers call their project "[RoboWatch](#)." Part of what makes it possible is that there is a common underlying structure to most how-to videos. And, there's plenty of source material available. YouTube offers 180,000 videos on "How to make an omelet" and 281,000 on "How to tie a bowtie." By scanning multiple videos on the same task, a computer can find what they all have in common and reduce that to simple step-by-step instructions in natural language.

Why do people post all these videos? "Maybe to help people or maybe just to show off," said graduate student Ozan Sener, lead author of a paper on the [video](#) parsing method presented Dec. 16 at the International Conference on Computer Vision in Santiago, Chile. Sener collaborated with colleagues at Stanford University, where he is currently a visiting researcher.

A key feature of their system, Sener pointed out, is that it is "unsupervised." In most previous work, robot learning is accomplished by having a human explain what the robot is observing – for example, teaching a robot to recognize objects by showing it pictures of the objects while a human labels them by name. Here, a robot with a job to do can look up the instructions and figure them out for itself.

Faced with an unfamiliar task, the [robot](#)'s computer brain begins by sending a query to YouTube to find a collection of how-to videos on the topic. The algorithm includes routines to omit "outliers" – videos that fit the keywords but are not instructional; a query about cooking, for example, might bring up clips from the animated feature *Ratatouille*, ads for kitchen appliances or some old Three Stooges routines.

The computer scans the videos frame by frame, looking for objects that

appear often, and reads the accompanying narration - using subtitles – looking for frequently repeated words. Using these markers it matches similar segments in the various videos and orders them into a single sequence. From the subtitles of that sequence it can produce written instructions. In other research, robots have learned to perform tasks by listening to verbal instructions from a human. In the future, information from other sources such as Wikipedia might be added.

The learned knowledge from the YouTube videos is made available via [RoboBrain](#), an online knowledge base robots anywhere can consult to help them do their jobs.

Provided by Cornell University

Citation: Robots learn by watching how-to videos (2015, December 21) retrieved 23 April 2024 from <https://phys.org/news/2015-12-robots-how-to-videos.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.