

Guiding robot planes with hand gestures

15 March 2012, by Larry Hardesty

Aircraft-carrier crew use a set of standard hand gestures to guide planes on the carrier deck. But as robot planes are increasingly used for routine air missions, researchers at MIT are working on a system that would enable them to follow the same types of gestures.

The problem of interpreting hand signals has two distinct parts. The first is simply inferring the body pose of the signaler from a [digital image](#): Are the hands up or down, the elbows in or out? The second is determining which specific [gesture](#) is depicted in a series of [images](#). The MIT researchers are chiefly concerned with the second problem; they present their solution in the March issue of the journal *ACM Transactions on Interactive [Intelligent Systems](#)*. But to test their approach, they also had to address the first problem, which they did in work presented at last year's IEEE International Conference on Automatic Face and Gesture Recognition.

Yale Song, a PhD student in MIT's Department of Electrical Engineering and Computer Science, his advisor, computer science professor Randall Davis, and David Demirdjian, a research scientist at MIT's Computer Science and Artificial Intelligence Laboratory (CSAIL), recorded a series of videos in which several different people performed a set of 24 gestures commonly used by [aircraft-carrier](#) deck personnel. In order to test their gesture-identification system, they first had to determine the body pose of each subject in each frame of video. "These days you can just easily use off-the-shelf Kinect or many other drivers," Song says, referring to the popular Microsoft Xbox device that allows players to control video games using gestures. But that wasn't true when the MIT researchers began their project; to make things even more complicated, their algorithms had to infer not only body position but also the shapes of the subjects' hands.

The MIT researchers' software represented the contents of each frame of video using only a few variables: three-dimensional data about the

positions of the elbows and wrists, and whether the hands were open or closed, the thumbs up or down. The database in which the researchers stored sequences of such abstract representations was the subject of last year's paper. For the new paper, they used that database to train their gesture-classification algorithm.

The main challenge in classifying the signals, Song explains, is that the input - the sequence of body positions - is continuous: Crewmembers on the aircraft carrier's deck are in constant motion. The algorithm that classifies their gestures, however, can't wait until they stop moving to begin its analysis. "We cannot just give it thousands of [video] frames, because it will take forever," Song says.

The researchers' algorithm thus works on a series of short body-pose sequences; each is about 60 frames long, or the equivalent of roughly three seconds of video. The sequences overlap: The second sequence might start at, say, frame 10 of the first sequence, the third sequence at frame 10 of the second, and so on. The problem is that no one sequence may contain enough information to conclusively identify a gesture, and a new gesture could begin halfway through a frame.

For each frame in a sequence, the algorithm calculates the probability that it belongs to each of the 24 gestures. Then it calculates a weighted average of the probabilities for the whole sequence. Gesture identification is based on the weighted averages of several successive sequences, which improves accuracy, since the averages preserve information about how each frame relates to those before and after it. In evaluating the collective probabilities of successive sequences, the algorithm also assumes that gestures don't change too rapidly or too erratically.

In tests, the researchers' algorithm correctly identified the gestures collected in the training database with 76 percent accuracy. Obviously, that's not a high enough percentage for an

application that deck crews - and multimillion-dollar pieces of equipment - rely on for their safety. But Song believes he knows how to increase the system's accuracy. Part of the difficulty in training the classification algorithm is that it has to consider so many possibilities for every pose it's presented with: For every arm position there are four possible hand positions, and for every hand position there are six possible arm positions. In ongoing work, the researchers are retooling the algorithm so that it considers arm position and hand position separately, which drastically cuts down on the computational complexity of its task. As a consequence, it should learn to identify gestures from the training data much more efficiently.

Philip Cohen, co-founder and executive vice president of research at Adapx, a company that builds computer interfaces that rely on natural means of expression, such as handwriting and speech, says that the MIT researchers' new paper offers "a novel extension and combination of model-based and appearance-based gesture-recognition techniques for body and hand tracking using computer vision and machine learning."

"These results are important and presage a next stage of research that integrates vision-based gesture recognition into multimodal human-computer and human-robot interaction technologies," Cohen says.

This story is republished courtesy of MIT News (web.mit.edu/newsoffice/), a popular site that covers news about MIT research, innovation and teaching.

Provided by Massachusetts Institute of Technology

APA citation: Guiding robot planes with hand gestures (2012, March 15) retrieved 10 April 2021 from <https://phys.org/news/2012-03-robot-planes-gestures.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.