POPULAR SCIENCE SUMMARY

How Robots Learn to Pour by Seeing and Feeling

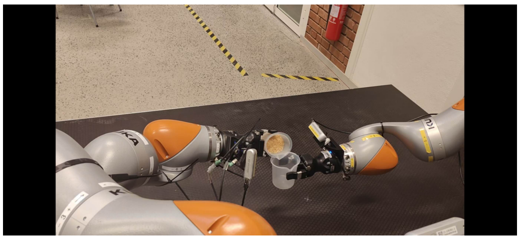
Authors: Wenrui Xie and Jiacheng Zhang

Supervisor: Volker Krueger

Examiner: Maj Stenmark

Pouring is easy for people but difficult for robots. Combining camera views with joint signals helped a two-armed robot reach the pouring stage in 95% of tests, up from 55% with vision alone.

Pouring liquids is a common laboratory task, but even a small mistake can waste samples, disrupt experiments or expose staff to hazardous substances. A useful laboratory robot must therefore do more than move one container above another. It must grasp both containers, coordinate two arms, follow the progress of the pour and stop at the right moment.



The dual-arm robot pours rice using camera views and joint-torque feedback.

Cameras provide important information about position and alignment, but they do not reveal everything. A gripper can block the view, and a full container may look almost identical to an empty one. This project investigated whether a robot could complement vision with signals already available inside its joints. These joint-torque signals change when the load or physical interaction changes, giving the robot a simple form of physical awareness without requiring extra force sensors.

The robot learned the task from 150 demonstrations performed by a human operator using virtual-reality controls. The learning method, called a diffusion policy, does not predict only the next isolated movement. Instead, it generates short sequences of coordinated actions, which is useful for a smooth task such as pouring. Rice was used instead of liquid during the experiments. It is safer and still requires controlled flow, accurate alignment and careful timing, although it does not reproduce liquid effects such as splashing, viscosity or surface tension.

Three versions of the system were tested 20 times each. A robot using only one overhead camera completed the full task once. Adding two wrist-mounted cameras improved the result to 7 successful trials. When joint-torque signals were also included, the robot reached the pouring stage in 19 of 20 trials and completed the whole task in 9.

One unexpected result was that estimating the exact changing weight of the container from processed force data was too unreliable. The robot's raw joint signals were more useful when supplied directly to the learning system. The approach is not yet fully robust: when the containers collided, unusual torque signals could confuse the policy and cause unexpected motions.

The results show that seeing the task from several angles is important, but vision alone does not always reveal whether a grasp has succeeded or how far pouring has progressed. Combining vision with physical signals could help future robots perform delicate laboratory and manufacturing tasks more safely and consistently. Further work must test larger datasets, different containers and real liquids before the method can be used in practical laboratory automation.