Many computer systems people interact with on a daily basis require knowledge about certain aspects of the world, or models, to work. These systems have to be trained, often needing to learn to recognize objects from video or image data. This data often contains superfluous content that reduces the accuracy of models. So researchers found a way to incorporate natural hand gestures into the teaching process. This way, users can more easily teach machines about objects, and the machines can also learn more effectively.

You've probably heard the term machine learning before, but are you familiar with machine teaching? Machine learning is what happens behind the scenes when a computer uses input data to form models that can later be used to perform useful functions. But machine teaching is the somewhat less explored part of the process, of how the computer gets its input data to begin with. In the case of visual systems, for example ones that can recognize objects, people need to show objects to a computer so it can learn about them. But there are drawbacks to the ways this is typically done that researchers from the University of Tokyo's Interactive Intelligent Systems Laboratory sought to improve.

"In a typical object training scenario, people can hold an object up to a camera and move it around so a computer can analyze it from all angles to build up a model," said graduate student Zhongyi Zhou. "However, machines lack our evolved ability to isolate objects from their environments, so the models they make can inadvertently include unnecessary information from the backgrounds of the training images. This often means users must spend time refining the generated models, which can be a rather technical and time-consuming task. We thought there must be a better way of doing this that's better for both users and computers, and with our new system, LookHere, I believe we have found it."

Zhou, working with Associate Professor Koji Yatani, created LookHere to address two fundamental problems in machine teaching: firstly, the problem of teaching efficiency, aiming to minimize the users' time, and required technical knowledge. And secondly, of learning efficiency -- how to ensure better learning data for machines to create models from. LookHere achieves these by doing something novel and surprisingly intuitive. It incorporates the hand gestures of users into the way an image is processed before the machine incorporates it into its model, known as HuTics. For example, a user can point to or present an object to the camera in a way that emphasizes its significance compared to the other elements in the scene.