In human grid, we're the cogs

October 15 2007

Images from a research shopping trip with GroZi a Grocery Shopping Assistant for the Visually Impaired developed by UC San Diego computer science professor Serge Belongie. On October 15, 2007 Belongie presented a paper at an interactive computer vision conference and described how people posting comments on blogs could provide data critical for this project. Credit: Serge Belongie

Before you can post a comment to most blogs, you have to type in a series of distorted letters and numbers (a CAPTCHA) to prove that you are a person and not a computer attempting to add comment spam to the blog.

What if – instead of wasting your time and energy typing something meaningless like SGO9DXG – you could label an image or perform some other quick task that will help someone who is visually impaired do their grocery shopping?

In a position paper presented at Interactive Computer Vision (ICV) 2007 on October 15 in Rio de Janeiro, computer scientists from UC San

Diego led by professor Serge Belongie outline a grid system that would allow CAPTCHAs to be used for this purpose – and an endless number of other good causes.

"One of the application areas for my research is assistive technologyfor the blind. For example, there is an enormous amount of data that needs to be labeled for our grocery shopping aid to work. We are developing a wearable computer with a camera that can lead a visually impaired user to a desired product in a grocery store by analyzing the video stream. Our paper describes a way that people who are looking to prove that they are humans and not computers can help label still shots from video streams in real time," said Belongie.

The researchers call their system a "Soylent grid" which is a reference to the 1973 film Soylent Green (see more on this reference at the end of the article).

"The degree to which human beings could participate in the system (as remote sighted guides) ranges from none at all to virtually unlimited. If no human user is involved in the loop, only computer vision algorithms solve the identification problem. But in principle, if there were an unlimited number of humans in the loop, all the video frames could be submitted to a SOYLENT GRID, be solved immediately and sent back to the device to guide the user," the authors write in their paper.

From the front end, users who want to post a comment on a blog would be asked to perform a variety of tasks, instead of typing in a string of misshapen letters and numbers.

"You might be asked to click on the peanut butter jar or click the Cheetos bag in an image," said Belongie. "This would be one of the so called 'Where's Waldo' object detection tasks."

The task list also includes "Name that Thing" (object recognition), "Trace This" (image segmentation) and "Hot or Not" (choosing visually pleasing images).

"Our research on the personal shopper for the visually impaired – called Grozi – is a big motivation for this project. When we started the Grozi project, one of the students, Michele Merler – who is now working on a Ph.D. at Columbia University – captured 45 minutes of video footage from the campus grocery store and then endured weeks of manually intensive labor, drawing bounding boxes and identifying the 120 products we focused on. This is work the soylent grid could do," said Belongie.

From the back end, researchers and others who need images labeled would interact with clients (like a blog hosting company) that need to take advantage of the CAPTCHA and spam filtering capabilities of the grid.

"Getting this done is going to take an innovative collaboration between academia and industry. Calit2 could be uniquely instrumental in this project," said Belongie. "Right now we are working on a proposal that will outline exactly what we need – access to X number of CAPTCHA requests in one week, for example. With this, we'll do a case study and demonstrate just how much data can be labeled with 99 percent reliability through the soylent grid. I'm hoping for people to say, 'Wow, I didn't know that kind of computation was available.'"

Source: University of California - San Diego