The Eyes Have It... But What Is It?

Richard Wildes, Computational Vision Scientist, York University, Associate Director, Vision: Science to Applications (VISTA)

We're trying to understand, from a fundamental principles point of view, what it means to see, be you a computer system or be you a biological system. With an artificial system, say you start with a camera as the input, and use computers to do the processing. With a biological system, you start with eyes as the input, and use a brain to do the processing.

But in either case, the artificial system and the natural system, the biological system, are living in the same physical world, and in both cases are trying to get the same information out. You're trying to recognize faces, you're trying to recognize actions, you're trying to recognize the scene in which the actions are occurring. It makes sense that the processing in between is the same at a certain level of abstraction.

So for example, if we're taking a video, say, of somebody being interviewed, how can we actually analyze that video to be able to say, here's a man, he's answering questions, he's telling a story about his research.

What are the applications of your research?

The single thing that I'm most excited about in my research I would put into the category of video understanding, or recognition from video. So this will allow us to be able to take computers and technology to be able to mine and categorize the vast online video repositories, to be able to automatically recognize what actions are occurring in the videos. What are the scenes in which they're occurring? And these are places where we are very close to

being able to have real world deployable technologies and really understand the basic scientific principles that underlie them. A recent statistic says that there's over 300 hours of video uploaded to YouTube every minute. And that's just one online video repository. So it seems unlikely that one person, or even one group of people is going to watch all these videos, analyze them and categorize them for us. So wouldn't it be great if we could have an artificial vision system, say a system that was capable of true video understanding, working in our service?