RESEARCH2REALITY
Shining a light on research & innovation.

We're Not Seeing Eye to Eye with This New Normal

An Interview with **Niko Troje**, Vision Scientist, York University
Vision: Science to Applications (VISTA)

I'm a vision researcher, I've been working on many different things but what brings it together is what I call people perception.

So I've been working on face recognition many years ago, and then I think most of my colleagues know me for work that I did over the last 10 years on what we call biological motion. All that rich information that is contained in the way people move.

Over the last two, three years here at York, I became interested in telecommunication. There's many problems with systems such as Zoom or Skype that we know today. The main problem is that we are losing what I call directionality, so if someone happens to look into the camera, the person on the other end feels being looked at and there's no way to escape that gaze.

Or if someone is not looking into the camera for instance because they are looking at the screen where the other person is, we feel being looked at our chin or neck or something, but we can't catch the other person's gaze.

We can't just go down and catch the gaze because of course the camera that takes the footage of the other person is fixed, and what we are doing currently is to learn, how people look on a slightly different viewpoint angles and apply that in real time to a video screen.

We have a demo system, which functions beautifully but it's based on computer graphics. So the person I'm talking to, I see represented as an avatar and the other person sees me represented as an avatar.

Now our current work aims to replace computer graphics with photorealistic video again. Which means that we have to either really physically move

cameras on the other person's end, which of course would be very awkward and probably not feasible for most applications, or we have to simulate that

with much more photorealistic experiences, where we still get what I call mp-depth, so mp-depth from motion parallax, back into video conferencing.