Snapshots of the future: Tool learns to predict user's gaze in headcam footage

September 12 2018

Credit: CC0 Public Domain

The miniaturization of video cameras has led to an explosion in their use, including their incorporation into a range of portable devices such as headcams, used in scenarios ranging from sporting events to armed

combat. To analyze tasks performed in view of such devices and provide real-time guidance to individuals using them, it would be helpful to characterize where the user is actually focusing within footage at each moment in time, but the tools available to predict this are still limited.

In a new study reported at the 15th European Conference on Computer Vision (ECCV 2018), researchers at The University of Tokyo have developed a computational tool that can learn from footage taken using a headcam, in this case of various tasks performed in the kitchen, and then accurately predict where the user's focus will next be targeted. This new tool could be useful to enable video-linked technologies to predict what actions the user is currently performing, and provide appropriate guidance regarding the next step.

Existing programs for predicting where the human gaze is likely to fall within a frame of video footage have generally been based on the concept of "visual saliency," which uses distinctions of features such as color, intensity, and contrast within the image to predict where a person is likely to be looking. However, in footage of human subjects performing complex tasks, this visual-saliency approach is inadequate, as the individual is likely to shift their attention from one object to another in a sequential, and often predictable, manner.

To take advantage of this predictability, in this study the team used a novel approach combining visual saliency with "gaze prediction," which involves an artificial intelligence learning such sequences of actions from existing footage and then applying the obtained knowledge to predict the direction of the user's gaze in new footage.

"Our new approach involves the construction of first a 'saliency map' for each frame of footage, then an 'attention map' based on where the user was previously looking and on motion of the user's head, and finally the combination of both of these into a 'gaze map,'" Yoichi Sato says. "Our

results showed that this new tool outperformed earlier alternatives in terms of predicting where the gaze of the headcam user was actually directed."

Although the team's results were obtained for footage of chores in a kitchen, such as boiling water on a stove, they could be extended to situations such as tasks performed in offices or factories. In fact, according to lead author Yifei Huang, "Tools for evaluating so-called egocentric videos of this kind could even be applied in a medical context, such as assessing where a surgeon is focusing and offering guidance on the most appropriate steps to be taken next in an operation."

The article "Predicting Gaze in Egocentric Video by Learning Task-dependent Attention Transition" is published in the proceedings of European Conference on Computer Vision (ECCV 2018) and as an arXiv paper at arxiv.org/abs/1803.09125 .

 More information: Predicting Gaze in Egocentric Video by Learning Task-dependent Attention Transition, arxiv.org/abs/1803.09125

Provided by University of Tokyo