Abstract: Thanks to the availability and increasing popularity of wearable devices such as GoPro cameras, smart phones and glasses, we have access to a plethora of videos captured from the first person perspective. Capturing the world from the perspective of one’s self, egocentric videos bear characteristics distinct from the more traditional third-person (exocentric) videos.

In many computer vision tasks (e.g. identification, action recognition, face recognition, pose estimation, etc.), the human actor is the main focus. Thus, detecting, localizing, and recognizing the human actor is often incorporated as a vital component. In an egocentric video, however, the person behind the camera is often the person of interest. This would change the nature of the task at hand, given that the camera holder is usually not visible in the content of his/her egocentric video. In other words, our knowledge about the visual appearance, pose, etc on the egocentric camera holder is very limited. This suggests relying on other cues in first person videos.

In a third person video, our knowledge of the action being performed, and the person performing it is mostly based on our understanding of the foreground of the video. The pose and motion of the actor lead us to reason about the action being performed. In an egocentric video, however, we see the world from the actor’s perspective. Therefore, the foreground of the video cannot lead to the same information. In fact, the main cue is the change of (global) background motion patterns, hinting towards the action, or the identity. On another note, comparing the contents of an egocentric and an exocentric video is non-trivial. We know that a third-person video can contain the egocentric actor, but not everything visible in the exocentric video will be visible in the egocentric video. What would be visible depends on what is contained in the field of view of the egocentric camera holder. In other words, the content of the egocentric video should be compared to the content of the exocentric video which is in the field of view of the egocentric viewer.

First and third person videos have been studied in the past in the computer vision community. However, the relationship between first and third person vision has yet to be fully explored. During my Ph.D., I participated in projects exploring this relationship in several aspects such as identification, temporal alignment, and action classification.

Bio: Shervin Ardeshir is a Ph.D. candidate at Center for Research in Computer Vision (CRCV), at University of Central Florida (UCF). He obtained his B.Sc. degree in Electrical Engineering from Sharif University of Technology (Tehran/Iran), M.Sc. in Computer Science from University of Central Florida, and has defended his Ph.D. at CRCV (UCF). Working in the area of computer vision, his research interest includes location-aware image understanding, and exploring relationship between first-person and third-person vision. His work has been published in computer vision conferences CVPR’14, CVPR’15, ECCV’14, ECCV’16, ECCV’18, and computer vision journals PAMI and CVIU.

Homepage: www.shervin-ardeshir.com