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 a third person video, our knowledge on the action being performed, and the person performing it is mostly based on our understanding on 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 toward the action, or an 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. This thesis explores this relationship in several aspects. We explore supervised and unsupervised approaches for relating these two views seeking different objectives such as identification, temporal alignment, and action classification.