Sometimes, the ability to analyze something also gives us the ability to synthesize it, and vice versa. Baiget et al. work under the assumption that this is true for modeling three-dimensional animated virtual agents (often referred to as avatars).

The paper describes an approach to generating augmented video sequences for the interaction between real agents (that is, human beings) and virtual ones. On one side of the communication chain, human motion is estimated in each time step, by means of multi-object tracking, to determine the positions of the extremities: hands, elbow, neck, head, hip, knees, and so on. On the other side, the system automatically generates and controls avatars based on behavior models that consider their interaction with the environment and with other agents, based on the tracked extremities. The paper describes the underlying algorithms and shows the performance of the proposed approach in an indoor and outdoor scenario that simulates human and vehicle agents.

Even though it seems that many of the introduced techniques have already been presented elsewhere, and despite some minor English grammar flaws, this paper is fun to read. I recommend it to people who want to get an introduction to the topic.