

ABSTRACT

An object tracking apparatus based on meta learning receives inputs of a video frame and a target initialization sentence, specifies, when the input video frame is a first video frame of a video sequence, a target from the first video frame based on the target initialization sentence, detects, when the input video frame is not a first video frame, objects that may interact with the specified target from the input video frame, determines a target and a related object from the corresponding video frame by predicting an interaction between the specified target and the detected objects, determines the target and the related object until the last video frame of the video sequence, and predicts trajectories of the target and the related object.

FIG. 1