We adopt the Soft Cascade detector architecture [8] which allows for efficient trade-off between accuracy and speed.

Figure 3 shows an infrared image frame with a number of pedestrian detections. The false alarm rate is very low, and persistent false alarms can easily be handled by the tracking filter, or ignored if the detection location is in unreasonable areas according to prior information of the buildings and environment. Non-persistent clutter is handled by a suitable initiator logic that prevents the false alarms to give rise to new tracks.

2.4 Related Research

Visual surveillance and crowd analysis in dynamic scenes with humans are very active research topics in computer vision [16, 50]. The possible applications are numerous, and so are the number of publications in the area.

This article’s focus is on the object tracker part of the surveillance system, see [49] and the references therein for an overview. The study [18] is an early publication where a particle filter is used for visual contour tracking. In [31] a mixture particle filter and an Ada-boost detector is used to track multiple objects (hockey players) in a video stream. Visual tracking is often performed in the image plane with the benefit of keeping the state dimension low and avoiding the calibration of extrinsic camera parameters, i.e., the location and orientation of the camera relative a world reference frame. In this work, tracking is performed in global coordinates which simplifies the motion model of the target and also makes it easier to combine with other tracking systems and contextual knowledge about the environment. Tracking in global coordinates with a vision sensor is essentially equivalent to tracking with a bearings-only sensor which has been traditionally treated in the target tracking community, see [35, Chapter 6] and the references therein.