Details on the Multi-Modal Mean Fields algorithm for Probabilistic Occupancy Maps

Here, we illustrate the behaviour of MMMF in an ambiguous people detection case.

Input We use as input two background subtracted images from two cameras with overlapping fields of view, as depicted in Fig. 1. The cameras are calibrated and synchronized. Tracking is performed on a grid representing a discretized ground-plane. POM then estimates the probabilities of occupancy at every discrete location as the marginals of a product law minimizing the KL divergence from the "true" conditional posterior distribution, by defining an energy function. Its value is computed by using a generative model: It represents humans as simple cylinders projecting to rectangles in the various images. Given the probability of presence or absence of people at different locations and known camera models, this produces synthetic images whose proximity to the corresponding background subtraction images is measured and used to define the energy.

Multi-Modal Mean Fields for POM Using the background-substracted images, MMMF produces several modes, which correspond to several potential scene interpretations, with a different number of people in each. Fig. 2 represents 4 modes, where we concatenated vertically the images from the two cameras. On each image, the background subtraction appears in red and the projection of the MAP estimates in blue. We see that there is an ambiguity (circled in green), about whether the observed foreground corresponds to a single person or to two. In this specific case, the "right" mode is the second one, represented by the top-right image, where two persons are actually detected inside the green ellipsoid, instead of only one for the other modes.

Recovering the right solution by temporal consistency This "right" mode is actually recovered using temporal consistency, by our K-Shortest Path method for Multi-Modal POMs, the final output is presented in Fig. 3. Our new K-Shortest Path does not only choose the right mode, but it jointly optimize for the best tracks inside the chosen sequence of best modes.

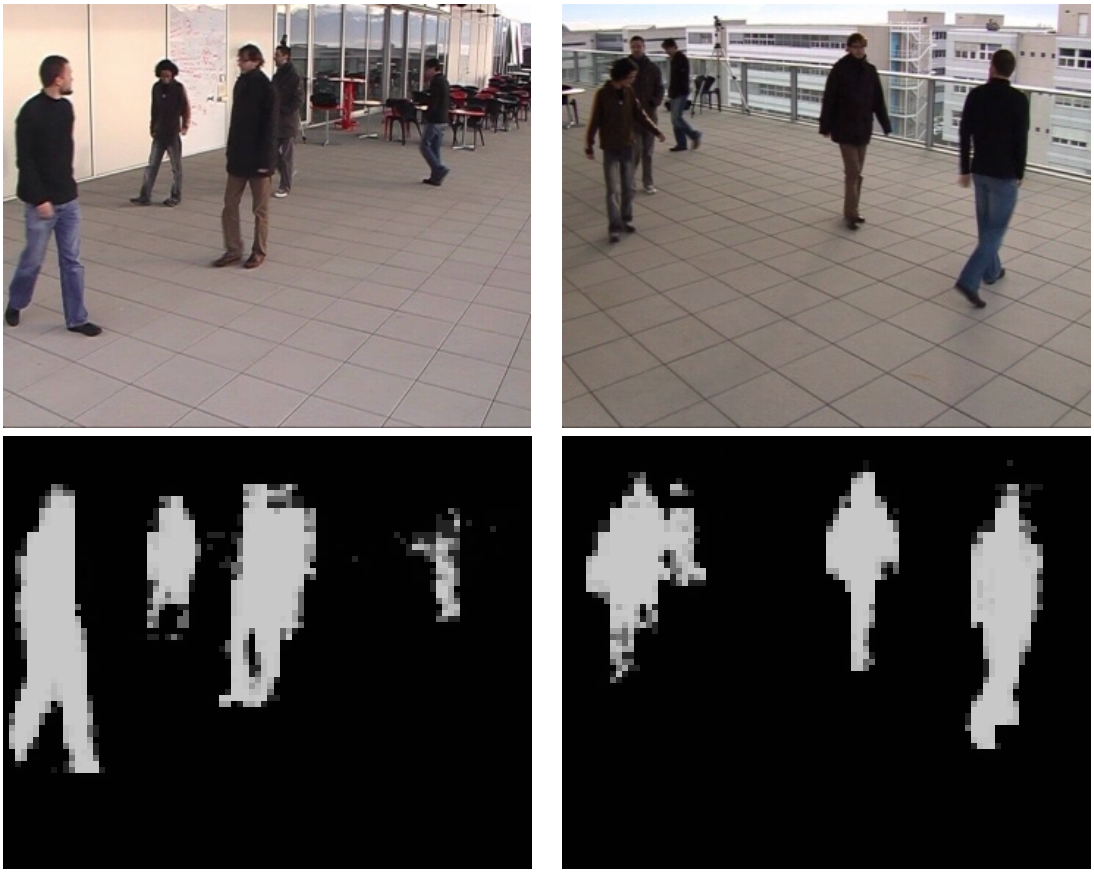

Figure 1: Input to MMMF. Top: Original Images. Bottom: Background subtraction images.

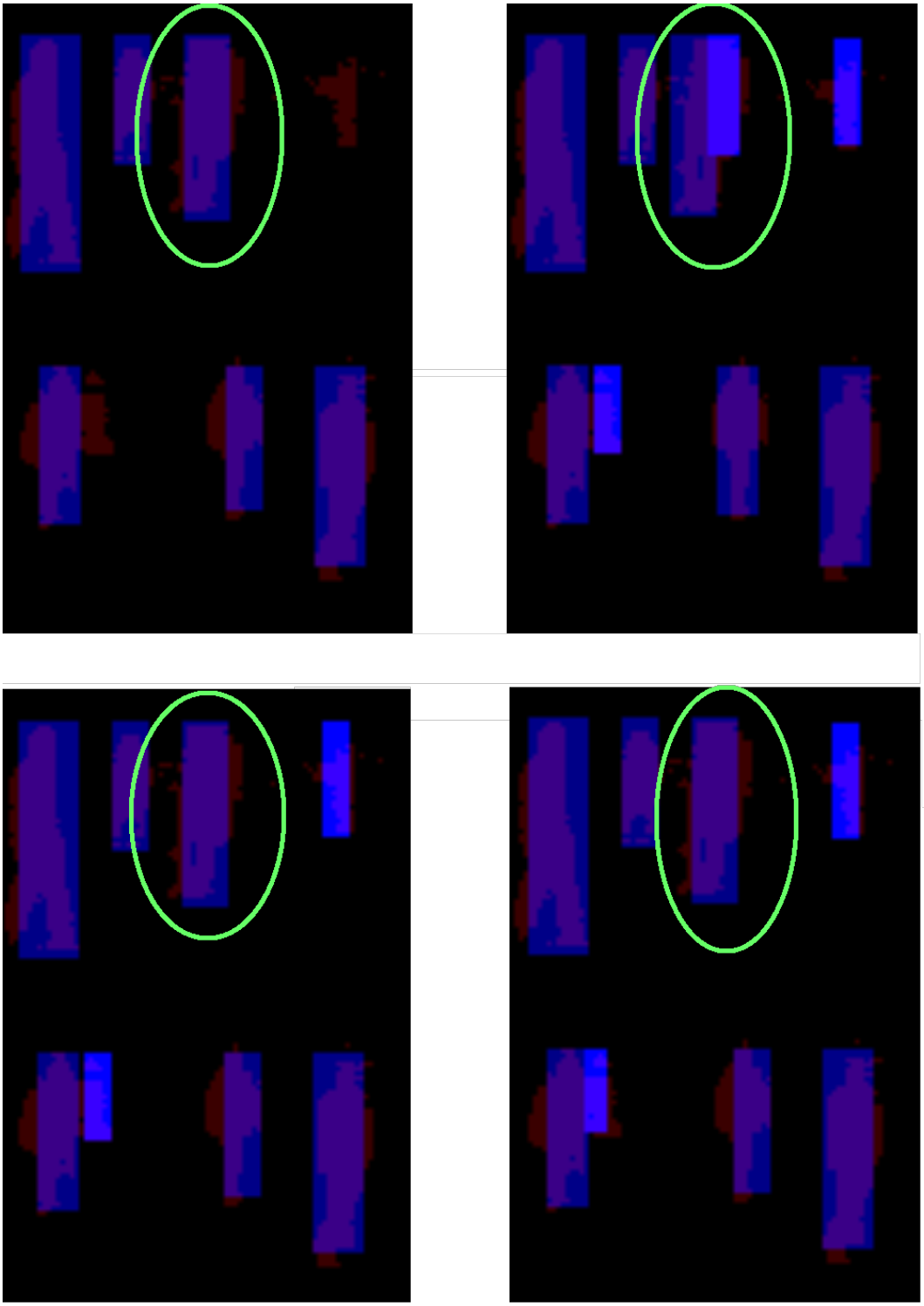

Figure 2: Four different MMMF modes, were detections corresponding to MAP estimates are projected in the camera planes. For each one, we superpose the images from the two cameras, The ellipse denotes an ambiguous part of the image that can be interpreted as one or two people. The light-blue corresponds to variables which have been clamped.

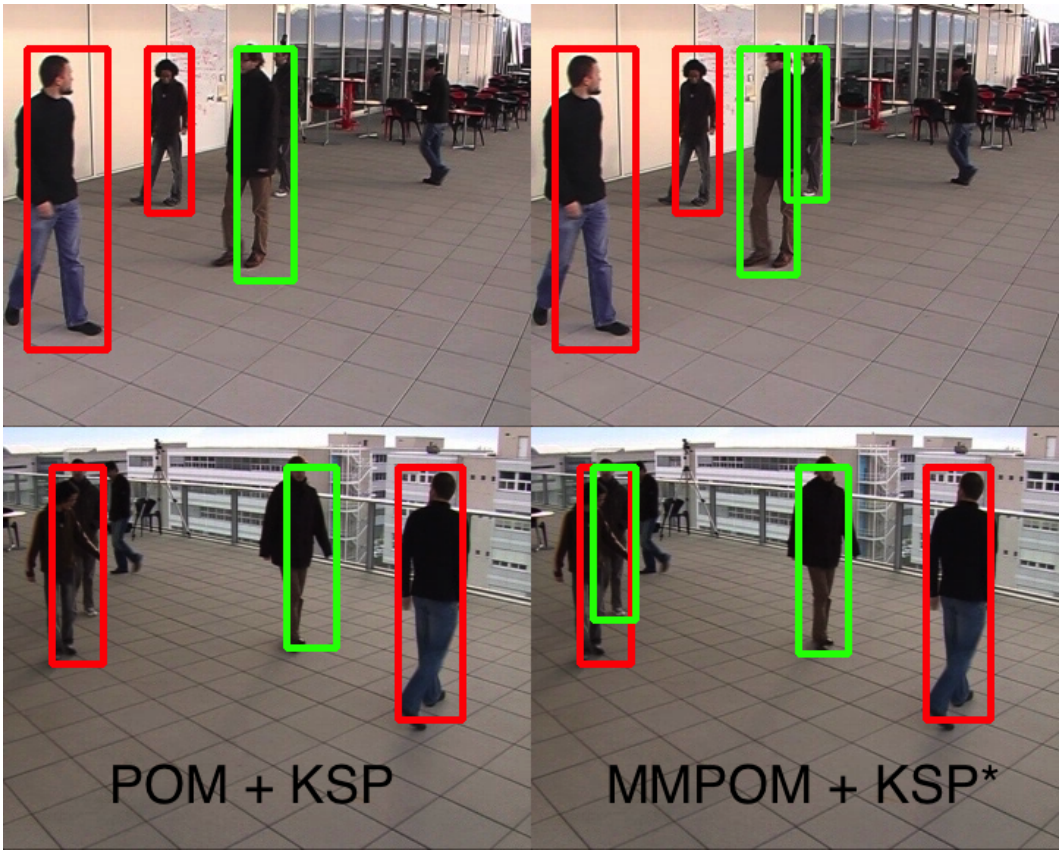

Figure 3: Left: POM + KSP misses two persons. Right: MMMF + KSP* recovers one more correct detection.