Chapter 3
Segmentation and Tracking

3.1 Introduction

After foreground extraction, if nonsignificant merging occurs among people, then segmentation is necessary, and the merged blobs can be split into separate ones equal in number to the number of people in the merged blob. Thereafter, all of the blobs can be indexed and tracked for motion and behavior analysis.

3.2 Segmentation

The position and orientation of a camera influence the degree of occlusion. If the camera’s optical axis is horizontal, then occlusion will occur in the segment parallel to the horizontal axis of the image plane. Occluded blobs will be indistinguishable due to overlapping in depth.

To increase the segmentation capability from being one dimensional (1D) (horizontal axis of the image) to two dimensional (2D) (the horizontal and vertical axes of the image plane), the camera needs to be installed relatively high and tilted down for inspection.

We develop a hybrid segmentation methodology to split occluded blobs, using a histogram-based approach for horizontal occlusion and an ellipse-based approach for vertical occlusion.

---


To measure successful segmentation, evaluation rules must first be established. We introduce the following three rules:

- **R1)** The number of blobs in the segmentation results should be the same as the number of persons in that blob;
- **R2)** The areas of the segmented blobs in the histogram-based approach should be reasonably large; and
- **R3)** The shape of each blob after segmentation should be the filling rate of each ellipse in the ellipse-based approach, which should be larger than a certain threshold.

Before segmentation, it is necessary to filter out blobs of individuals, which do not need to be segmented, to reduce the computation cost and enhance real-time performance. The area threshold of a blob is not efficient because the area varies significantly, even though it contains only a single person, if the person makes great changes in posture or moves into different areas (depths) of the image.

We assume that in a less crowded environment, when entering the camera’s field of view, a person first appears as an individual blob. When blobs merge, we can compute the number of persons in the occluded blob according to the take-up rate of its circum rectangle by the number of individual blobs in the previous frame, as illustrated in Figure 3.1.

![Figure 3.1](image)

**Fig. 3.1** Human counting

The blue curves are the contours of two separate persons in frame number $k$, and the green rectangles are their corresponding circum rectangles. In frame number $k + 1$, they are occluded, and the black rectangle is the circum rectangle of that occluded blob. The occluded blob’s area is overlapped by the two separate rectangles