Chapter 2
Background/Foreground Detection

2.1 Introduction

With the acquisition of an image, the first step is to distinguish objects of interest from the background. In surveillance applications, those objects of interest are usually humans. Their various shapes and different motions, including walking, jumping, bending down, and so forth, represent significant challenges in the extraction of foreground pixels from the image.

To tackle that problem, we consider two approaches: one is based on the fact that the background is stationary in the image captured by a fixed monocular camera, while the other is based on the assumption that the foreground contains objects of interest that are moving. The next sections elaborate on the two approaches in detail.

2.2 Pattern Classification Method

A popular detection methodology is based on the concept that once the background is extracted, the foreground can conveniently be obtained by subtracting the viewed image from the background model [162, 164].

2.2.1 Overview of Background Update Methods

Current background extraction methods include the multi-frame average, selection, selection-average, and random update methods [1–7] and Kalman filter-based method [8, 9]. These are briefly elaborated and compared below.

---

1 Portions reprinted, with permission from Xinyu Wu, Yongsheng Ou, Huihuan Qian, and Yangsheng Xu, A Detection System for Human Abnormal Behavior, IEEE International Conference on Intelligent Robot Systems. ©[2005] IEEE.
2.2.1.1 Multi-Frame Average Method

The multi-frame average method can be described as follows. For an image sequence $B_i \ i = 1, \ldots, n$,

$$B_n = \frac{\sum_{i=1}^{n} B_i}{n} \quad (2.1)$$

Here, if the number of frames, $n$, is as large as needed, then $B_n$ should be the background. However, in this method, too much memory is required to store the image sequence $B_i$. Hence, an approximate method is developed, which can be presented as

$$B_{pt} = kB_{pt-1} + (1-k)C_{pt-1} \quad 0 < k < 1 \quad (2.2)$$

where

- $B_{pt}$ — a pixel in the current background;
- $B_{pt-1}$ — a pixel in the last background;
- $C_{pt-1}$ — a pixel in the current image; and
- $k$ — a threshold value.

The disadvantages of this method are: 1) the threshold value $k$ is difficult to determine, and 2) $k$ needs to be adjusted, depending on the degree of environmental change. When there are many objects in an active area, the noise in that area can be great, because the objects can also be deemed as background.

2.2.1.2 Selection Method

In this method, a sudden change in pixels is deemed to be the foreground, which will not be selected as the background.

$$\text{If } |C_{p1} - C_{pt-1}| > T \quad \text{Then } B_{pt} = B_{pt-1} \quad (\text{don't update}) \quad (2.3)$$

$$\text{Else } B_{pt} = C_{pt-1} \quad (\text{update})$$

where

- $B_{pt}$ — a pixel in the current background;
- $B_{pt-1}$ — a pixel in the last background;
- $C_{pt}$ — a pixel in the current image of the image sequence;
- $C_{pt-1}$ — a pixel in the last image of the image sequence; and
- $T$ — a threshold value.

The disadvantage of this method is that determining the threshold value of $T$ is difficult. If $T$ is too small, then the background will be insensitive to change. If $T$ is too large, then the background will be too sensitive to change.