CHAPTER 2 BACKGROUND MODELING AND SINGLE PERSON DETECTION

This chapter focuses on the detection of single persons. For single person detection, a statistical representation of the background scene is built for supporting the detection of foreground regions. After foreground region detection, whether a foreground region contains a single person is determined. Section 2.1 expresses the color space used in our system. Section 2.2 describes the background modeling. In section 2.3, the detection of foreground regions is specified in detail. The method used to determine whether a foreground region contains a single person is explains in section 2.4.

2.1 Perceptually Uniform Color Space

In terms of color, each foreground region detected by our system should be able to be perceived by human. If a foreground region can not be perceived by human, it seems “false positive”. Because of this, we should make the ability of our system to detect foreground regions as well as humans. In addition, the term *skin color* is a subjective human concept. That is, humans perceive the skin color subjectively. Therefore, the color representation should be similar to the color sensitivity of humans to obtain a stable output similar to the one given by the human visual system. Such a color representation is called the perceptually uniform color system (UCS). Wu et al. [17] introduces a conversion method from the CIE color system to UCS that was proposed by the psychologist Farnsworth through psychophysical experiments in 1957 [18]. In this color system, the MacAdam ellipses that describe the noticeable chromatic difference become circles with approximately the same radius (see Fig. 2.1). This indicates that two colors with an equal distance as perceived by human viewers project with an equal distance in this color system.

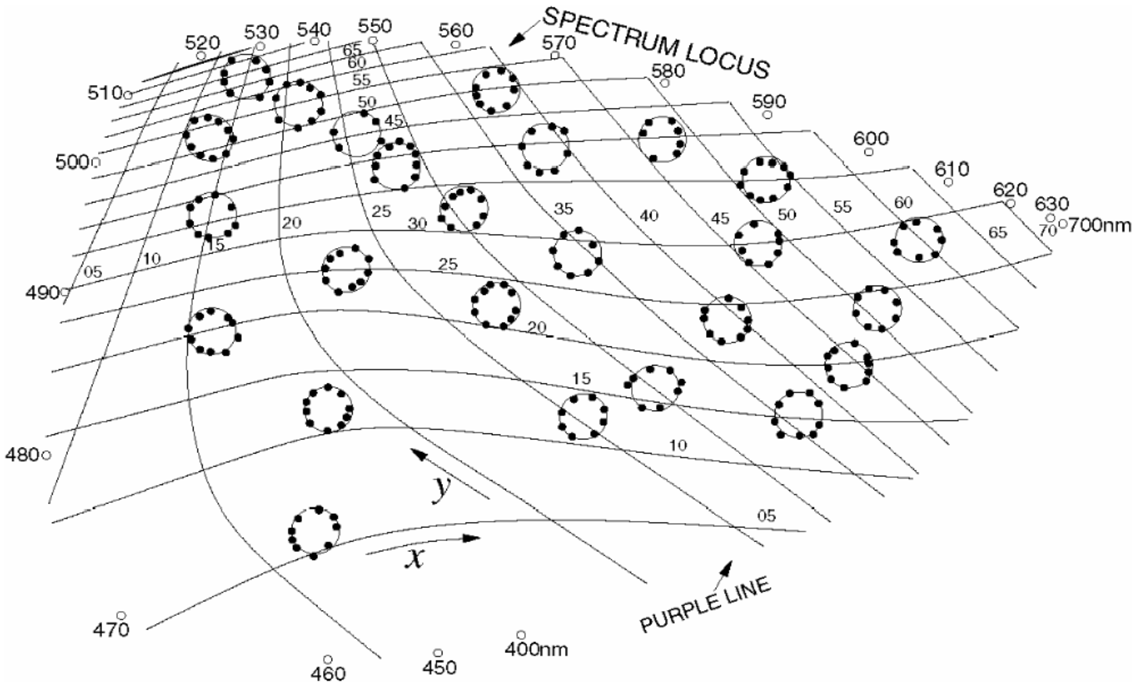


Fig. 2.1 Nonlinear transformation of the CIE chromaticity diagram where the transformed MacAdam's ellipses are circles with approximately the same radius (from Farnsworth, 1957).

In order to convert RGB color system to Farnsworth's UCS, the RGB color system is first converted to the CIE color system:

$$\begin{cases} X = 0.619R + 0.177G + 0.204B \\ Y = 0.299R + 0.586G + 0.115B \\ Z = 0.000R + 0.056G + 0.944B \end{cases}, \quad \begin{cases} x = \dfrac{X}{X+Y+Z} \\ y = \dfrac{Y}{X+Y+Z} \end{cases}$$

where Y carries the luminance information, and (x, y), which specifies the point on the CIE chromaticity diagram, describes the chromaticity. Then, the chromaticity (x, y) is converted to the Farnsworth's UCS with a nonlinear transformation.[1] The result of the conversion is represented by (u_f, v_f). The values of (u_f, v_f) of visible colors are in the range of:

$$\begin{cases} u_f \rightarrow [0,91] \\ v_f \rightarrow [0,139] \end{cases}.$$

1 A C program to perform the conversion is at: http://www.sys.wakayama-u.ac.jp/~chen/ucs.html.

So each visible color is represented by (Y, u_f, v_f) in the Farnsworth's UCS.

2.2 Background Modeling

A simple and common background modeling method involves subtracting each new image from a background model and thresholding the subtracting result to determine foreground pixels. One useful tool for building a background model is statistical modeling, where a process is modeled as a random vector in a feature space with an associated probability density function (pdf). The density function could be represented using parametric or nonparametric techniques. Parametric techniques represent the density function using a specified statistical distribution, which is assumed to approximate the actual distribution, with the associated parameters estimated from training data [1-5]. Nonparametric techniques estimate the density function directly from the data without any assumptions about the underlying distribution. This avoids having to choose a statistical distribution and estimating its parameters [6]. Nonparametric techniques are quite general and applicable to many vision problems where the underlying density is not known.

In our system, the *k* nearest neighbor (*kNN*) density estimation [19], which is a nonparametric density estimation technique, is utilized to build the background model. The background model uses color information as the feature for modeling the background. It keeps a sample of color values for each pixel and uses this sample to estimate the density function of the pixel color distribution. Therefore, the background model is able to estimate the probability of any newly observed color value. Note that the color space used is the Farnsworth's UCS, which is described in section 2.1.

2.2.1 Background Subtraction

Let $\{X_1, X_2, \cdots, X_N\}$ be a sample of color values representing the background for pixel

x, where X_i is $(Y_i(x), u_{fi}(x), v_{fi}(x))$ for $i = 1,2,\cdots,N$. Given this sample, the value of the pixel color probability density function at any color value can be estimated using the *k* nearest neighbor density estimation. Given the observed color value X_t, $(Y_t(x), u_{ft}(x), v_{ft}(x))$, at time *t*, the probability of this observation can be estimated as

$$\Pr(X_t) = \frac{k-1}{Nv(X_t)}$$

where $v(X_t)$ denotes the volume of $L(X_t)$, which is the local region extended around X_t until the kth nearest neighbor of X_t is found. The *k* nearest neighbors of X_t are selected from $\{X_1, X_2, \cdots, X_N\}$. The criterion that we use to select the *k* nearest neighbors of X_t is Euclidean distance. Notice that the information used to estimate the probability is (u_f, v_f). The reason why (u_f, v_f) is used instead of (Y, u_f, v_f) will be discussed later. Therefore, the Euclidean distance from X_t to the kth nearest neighbor of X_t, denoted by X_k, can be computed as

$$d(X_t, X_k) = \sqrt{(u_{ft}(x) - u_{fk}(x))^2 + (v_{ft}(x) - v_{fk}(x))^2}$$

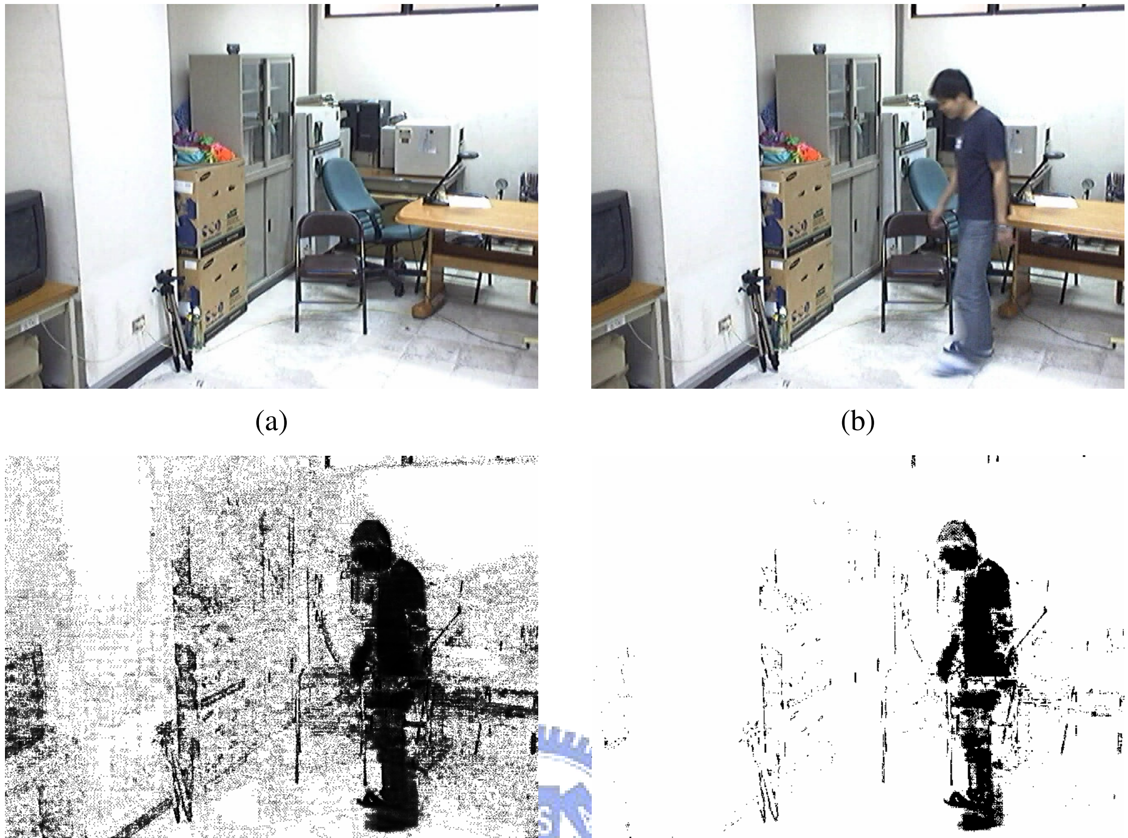
and $v(X_t)$ can be computed as

$$v(X_t) = \pi \cdot d^2(X_t, X_k).$$

Then the probability of X_t can be estimated as

$$\Pr(X_t) = \frac{k-1}{N\pi\left((u_{ft}(x) - u_{fk}(x))^2 + (v_{ft}(x) - v_{fk}(x))^2\right)}.$$

Using this probability estimation, pixel x is considered to be a foreground pixel at time *t* if $\Pr(X_t) < T$. The threshold T is a global threshold over all the images that can be adjusted to obtain a higher true positive rate with a lower false positives rate. Fig. 2.2(c) shows the estimated probability of each pixel of Fig. 2.2(b). Brighter pixels represent higher probability pixels. Fig. 2.2(d) shows a thresholding result of Fig. 2.2(c). Black pixels represent foreground pixels. Note that this background subtraction technique can be used with any color



(a) (b)

(c) (d)

Fig. 2.2 An example of background subtraction. (a) The background scene. (b) The observed image. (c) The estimated probability image of (b). (d) A thresholding result of (c).

space (e.g., HSI, YUV, etc.).

For the silhouette-based posture analysis, which will be described in chapter 3, the detection of shadows as part of the foreground regions could make the results of the shape-based posture analysis inaccurate. Then the inaccurate results could make human action recognition incorrect. Therefore, it is necessary to discriminate between objects and their shadows in our system. Chromaticity, which is extracted from color, is useful to suppress shadows from the detection [6]. Since chromaticity has no information about luminance, it is not sensitive to illumination changes that arise due to shadows. In the Farnsworth's UCS, Y carries the luminance information, and (u_f, v_f) describes the chromaticity information. To suppress shadows from the detection, (u_f, v_f) is used instead of (Y, u_f, v_f). Fig. 2.3 shows the results of detection using the color information (R, G, B) and the chromaticity information (u_f, v_f). Fig. 2.3(d) shows that using the chromaticity information (u_f, v_f) can detect the object without its shadow.

The disadvantage of using chromaticity is also that chromaticity has no information about luminance. Luminance is related to the difference between different colors that have the same chromaticity. For example, consider the case where the object is white and moves against a gray background. In this case, the object can not be detected only using chromaticity

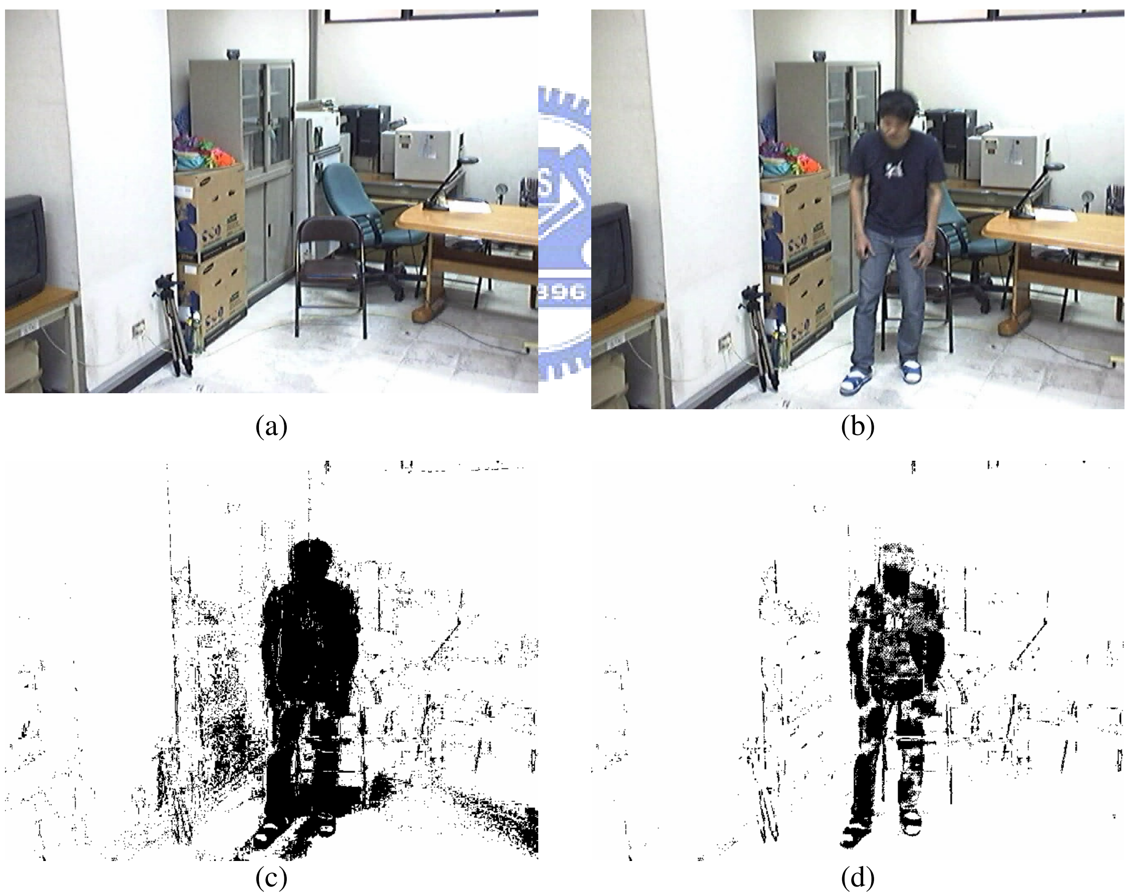


Fig. 2.3 An example of the detection results using color information and chromaticity information. (a) The background scene. (b) The observed image. (c) The result using the color information (R, G, B). (d) The result using the chromaticity information (u_f, v_f).

since both white and gray have the same chromaticity. To improve the disadvantage of using chromaticity, luminance still needs to be used in the background subtraction technique. Let $\{X_1, X_2, \cdots, X_N\}$ be a sample of color values representing the background for pixel x, where X_i is $(Y_i(x), u_{fi}(x), v_{fi}(x))$ for $i = 1, 2, \cdots, N$, and let X_t, $(Y_t(x), u_{ft}(x), v_{ft}(x))$, be the observed color value at time *t*. While selecting the *k* nearest neighbors of X_t to estimate the probability of X_t, each of the *k* nearest neighbors, denoted by X_j, have to satisfy an extra restriction,

$$\alpha \le \frac{Y_t(x)}{Y_j(x)} \le \beta$$

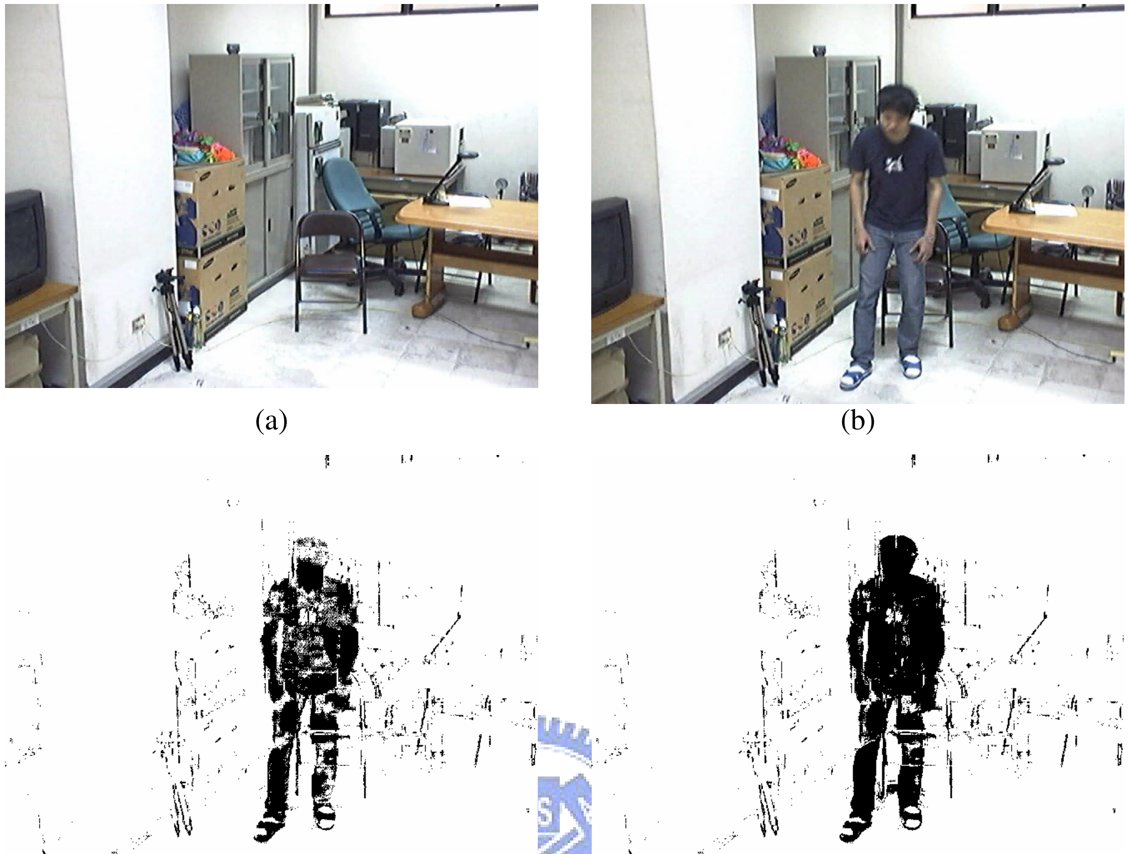
where $0 \le \alpha \le 1$ and $1 \le \beta$. The parameters α and β are constant over all the image. Fig. 2.4(c) shows the detection result without using the extra restriction, and Fig. 2.4(d) shows the result using the extra restriction.

2.2.2 Background Model Update

The background model cannot be expected to be used for a long period of time without update. There could be illumination changes, such as the change of sunlight, or physical changes, such as a deposited object. Illumination changes can cause false positives, and physical changes can cause false negatives.

Illumination change adaption: Add the observed color value to the background model and discard the oldest color value in the background model, if the observed color value is considered as the background.

Physical change adaption: The observed color value of pixel x starts to be added to the background model, if this pixel has been considered as the foreground successively for a long period of time. For each pixel, a foreground counter is used.



(a) (b)

(c) (d)

Fig. 2.4 An example of the influence of using the extra restriction. (a) The background scene. (b) The observed image. (c) The result without using the extra restriction. (d) The result using the extra restriction.

2.3 Foreground Region Detection

Foreground regions are segmented from each image of the image sequence by a five-stage process:

- thresholding,
- noise eliminating,
- morphological filtering,
- connected-component extracting, and
- hole filling.

Fig 2.5 shows an example of foreground region detection using the five-stage process. Note that Fig. 5.4(f) is called the foreground region silhouette.

2.4 Human Extraction

We determine whether a foreground region contains a single person according to the number of its pixels and the proportion of its skin color pixels to its pixels.

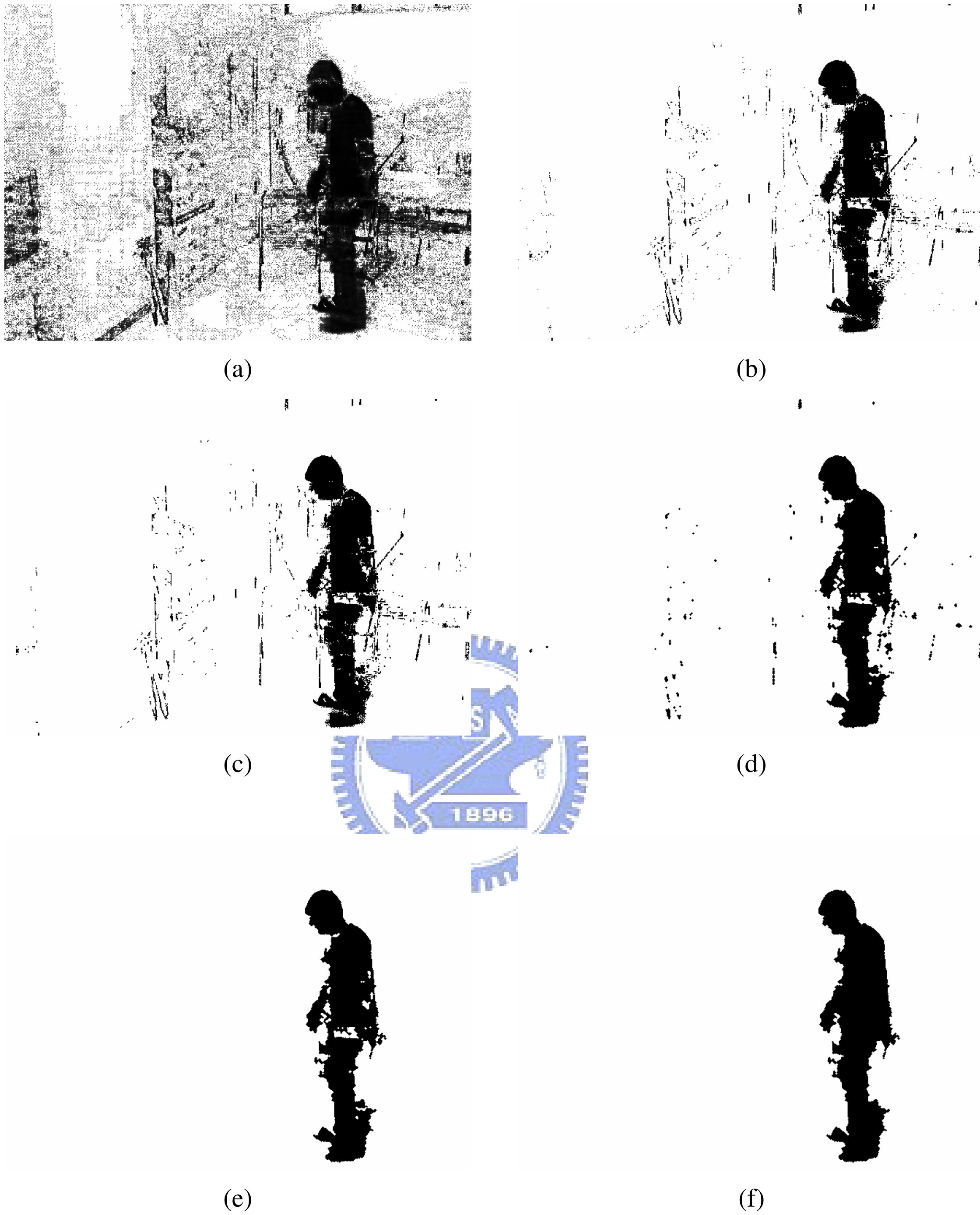


Fig. 2.5 An example of foreground region detection. (a) An estimated probability image. (b) Thresholding. (c) Noise Eliminating. (d) Morphological Filtering. (e) Connected-component Extracting. (f) Hole Filling.