

國立交通大學

電機學院 電機與控制學程

碩士論文

光線適應性背景模型於前景主體抽取

Lighting Adaptive Background Modeling
for Foreground Extraction



研究生：梁仲廷

指導教授：張志永

中華民國九十七年七月

光線適應性背景模型於前景主體抽取

Lighting Adaptive Background Modeling for Foreground Extraction

學 生：梁仲廷 Student: Chuang-Ting Liang
指導教授：張志永 Advisor: Jyh - Yeong Chang



Submitted to College of Electrical and Computer Engineering
National Chiao Tung University
in partial Fulfillment of the Requirements
for the Degree of
Master of Science
in
Electrical and Control Engineering
July 2008

Hsinchu, Taiwan, Republic of China

中華民國九十七年七月

光線適應性背景模型於前景主體抽取

學生：梁仲廷

指導教授：張志永博士

國立交通大學 電機學院 電機與控制學程碩士班

摘要

利用串流影像資訊於前景人物/物體的擷取能在許多地方應用，如：人機介面、安全監控、居家安全照護等系統。本論文中將提出一個可以針對光線改變的適應性背景模型於前景主體抽取的系統，在一般光線亮度改變差別不大時，可以簡單的使用 HSV 色彩模型或利用 GMM 於 RGB 色彩模型將前後景分離；但當背景光線突然有較大的改變時，我們無法使用上述二者系統將完整的前景資訊分離，因此我們使用適應性的背景模型，克服背景光線的突然改變，達到前、後景的分離。且當人物在原處停留一段時間時，我們亦使用適應性的歷史資料背景模型判斷人物是靜止或移動。

在本論文中，為了能快速適應亮度變化，我們必須先偵測光線的改變，在另一方面，有關人物行為的改變，如：移動人物的停止、靜止人物的開始移動，也都包含在本論文中。最後針對我們提出的適應性背景模型進行實驗驗證，當背景光線改變時，此模型皆能有效的將前景主體抽取出。

Lighting Adaptive Background Modeling for Foreground Extraction

STUDENT: Chung-Ting Liang

ADVISOR: Dr. Jyh-Yeong Chang

Degree Program of Electrical and Computer Engineering
National Chiao Tung University

ABSTRACT

Foreground subject/object extraction from video streams has a wide range of application such as human-machine interface, security surveillance, home care system, etc. The objective of this thesis is to provide an adaptive background modeling for foreground extraction to count for lighting and subject moving change. With a little lighting change, the foreground subject/object can be extracted easily in the HSV color space, on our proposed or in the RGB color space, by a Gaussian Mixture Model. With a large lighting change, we cannot extract the foreground image completely both systems above. To solve this, we propose an adaptive background modeling framework. When a moving subject stops for a while, we also use the History-Based Background Adaptation to classify the pixels whether it is to stop of moving subject or not.

In this framework, a lighting change detection scheme facilitates the quick adaptation to the intensity change. On the other hand, a subject action change, for example, from moving to a stop or from a stop to moving, is also detected and adapted to our framework. Our proposed adaptive background modeling demonstrates consistently excellent foreground extraction despite the various changes encountered.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my advisor, Dr. Jyh-Yeong Chang for his valuable suggestions, guidance, support, and inspiration. Without his advice, it is impossible to complete this research. Thanks are also given to all of my laboratory members for their suggestions and discussions. Finally, I would like to express my deepest gratitude to my family, specially my wife, Yi-Fang Lin for their concern, supports and encouragements.



Content

摘要	i
ABSTRACT	ii
ACKNOWLEDGEMENTS	iii
Content	iv
List of Figures	vii
List of Tables	x
Chapter 1 Introduction	1
1.1 Motivation	1
1.2 Background Modeling	3
1.3 The Background Model Modification	4
1.4 Foreground Subject Extraction	5
1.5 Thesis outlines	7
Chapter 2 Basic Concept	8
2.1 The HSV Color Space	8
2.2 The RGB Color Space	10
2.3 Converting Colors from RGB to HSV	11
2.4 Gaussian Mixture Models (GMMs)	12

Chapter 3 Background Subtraction System	15
3.1 Object Extraction Using HSV Model.....	15
3.1.1 The Intensity of The Image	15
3.1.2 Background Model in HSV Space	18
3.1.3 Adaptive Background Model with A Lighting Change Compensation.....	20
3.1.4 Foreground Subtraction Extraction and Shadow Suppression	22
A. Extract Foreground Subject	23
B. Shadow Suppression	24
3.2 Object Extraction Using RGB Model	25
3.2.1 GMM Background Model in RGB Space.....	25
3.2.2 Adaptive Background Model with A Lighting Change Compensation.....	26
3.2.3 Foreground Subtraction Extraction	27
A. Extract Foreground Subject by GMMs.....	28
B. Moving History-Based Background Adaptation to the Stop of Moving Subjects	30
Chapter 4 Experimental Result	31
4.1 Object Extraction Using HSV Model.....	33
4.1.1 Background Model Construction Using HSV Model.....	33
4.1.2 Foreground Subtraction Extraction Using HSV Model	35

4.1.3	Foreground Extraction with Lighting Change Compensation.....	38
4.2	Object Extraction Using RGB Model	40
4.2.1	Background Model Construction Using RGB Model	40
4.2.2	Foreground Subtraction Extraction Using RGB Model	42
4.2.3	Foreground Extraction with Lighting Change Compensation.....	44
4.2.4	Foreground Subject Extraction with Moving History-Based Background Adaptation	46
4.3	The Extraction Rate of Object.....	48
Chapter 5	Conclusion	50
References.....		51



List of Figures

Fig. 1.1	The flowchart of our subject extraction system.	2
Fig. 2.1	The HSV Cone.	9
Fig. 2.2	The RGB Color Cube.	10
Fig. 3.1	The comparison between frame ratio and frame difference. (a) Background image, (b) image frame with a human, (c) frame difference, (d) frame ratio, (e) histogram of frame difference, (f) histogram of frame ratio, (g) foreground pixels of frame difference after simply taking a threshold, and (h) foreground pixels of frame ratio after simply taking a threshold	17
Fig. 3.2	The framework for foreground subject/object extraction.	22
Fig. 3.3	The framework for foreground subject/object extraction.	27
Fig. 3.4	The proposed 3D ball model in the RGB color space.	28
Fig. 3.5	2D projection of the 3D ball model from RGB space onto the RG space.	29
Fig. 4.1	The experimental environment.	31
Fig. 4.2	Various image frames of three kind of lights. (a) “Normal” lighting, (b) “dark” lighting, and (c) “bright” lighting cases.	32
Fig. 4.3	Background images. (a) Background image in the H component, (b) Background image in the S component, and (c) Background image in the V component	33
Fig. 4.4	H, S, and V variations versus frame index of background video image frame 1 to frame 300. (a) H at (10, 10), (b) H at (120, 160), (c) S at (10, 10), (d) S at (120, 160), (e) V at (10, 10), and (f) V at (120, 160)	34
Fig. 4.5	An example of foreground extraction at different k_V thresholds. (a) An image frame with subject’s clothing color different from the background, (b) – (f) foreground detected images, (b) $k_V = 1.2$, (c) $k_V = 1.3$, (d) $k_V = 1.4$, (e)	

	$k_V = 1.5$, and (f) $k_V = 1.6$	36
Fig. 4.6	The example of the shadow suppression.....	37
Fig. 4.7	The example concerning lighting change. (a) An image frames in which the subject “walking” with “normal” lighting, (b) a foreground subject without lighting change compensation from (a), and (c) a foreground subject with lighting change compensation from (a). (d) An image frames in which the subject “walking” with “dark” lighting, (e) a foreground subject without lighting change compensation from (d), and (f) a foreground subject with lighting change compensation from (d). (g) An image frames in which the subject “walking” with “bright” lighting, (h) a foreground subject without lighting change compensation from (g), and (i) a foreground subject with lighting change compensation from (g).	39
Fig. 4.8	Background images. (a) Background image in the R component, (b) Background image in the G component, and (c) Background image in the B component.....	40
Fig. 4.9	R, G, and B variations versus frame index of background video image frame 1 to frame 300. (a) R at (10, 10), (b) R at (120, 160), (c) G at (10, 10), (d) G at (120, 160), (e) B at (10, 10), and (f) B at (120, 160)	41
Fig. 4.10	An example of foreground extraction at different k_G thresholds. (a) An image frame with subject’s clothing color different from the background, (b) – (f) foreground detected images, (b) $k_G = 1.2$, (c) $k_G = 1.3$, (d) $k_G = 1.4$, (e) $k_G = 1.5$, and (f) $k_G = 1.6$	43
Fig. 4.11	The example concerning lighting change. (a) An image frames in which the subject “walking” with “normal” lighting, (b) a foreground subject without lighting change compensation from (a), and (c) a foreground subject with lighting change compensation from (a). (d) An image frames in which the subject “walking” with “dark” lighting, (e) a foreground subject without lighting change compensation from (d), and (f) a foreground subject with lighting change compensation from (d). (g) An image frames in which the subject “walking” with “bright” lighting, (h) a foreground subject without lighting change compensation from (g), and (i) a foreground subject with lighting change compensation from (g).	45

Fig. 4.12 The example of moving History-Based Background Adaptation.47



List of Tables

TABLE I	COMPARISON RESULT OF THE PIXEL ACCURACY RATES OVER 30 IMAGES IN THE RGB COLOR SPACE	49
TABLE II	COMPARISON RESULT OF THE PIXEL ACCURACY RATES OVER 30 IMAGES IN THE HSV COLOR SPACE	49



Chapter 1 Introduction

1.1 Motivation

Object extraction from video streams has many applications such as home care system, human-machine interface, and automatic surveillance, etc. However, there is no rigid syntax and well-defined structure in object extraction system; therefore, it makes object extraction a very challenging task, especially in lighting change detection.

Several object extraction methods have been proposed in the past few years. There have been some significant projects on detecting, tracking people and recognizing their activities. W^4 [1] is one of them. W^4 can detect people (single person or people in group) by adopting an adaptive background model and identify the activities by finding the body parts on the silhouette boundary. Stauffer and Grimson [2] adaptive background mixture models for real-time tracking. Cucchiara, *et al.* [3] improving shadow suppression in moving object detection with HSV color information.

The objective of this thesis is to provide a background model to extract objects from video streams. This system can tell where the foreground subject is in an image frame. In this paper, we use two kinds of system to present. The first system, embedded in the HSV model is described in Sec. 3.1. The second system, Gaussian Mixture Model embedded in the RGB model is described in Sec. 3.2.

The flowchart of our proposed subject project is illustrated in Fig. 1.1, and the system can be separated into three components. The first component is background model training. The second component is the background model adaptation to a large lighting change. The third component is foreground subject extraction.

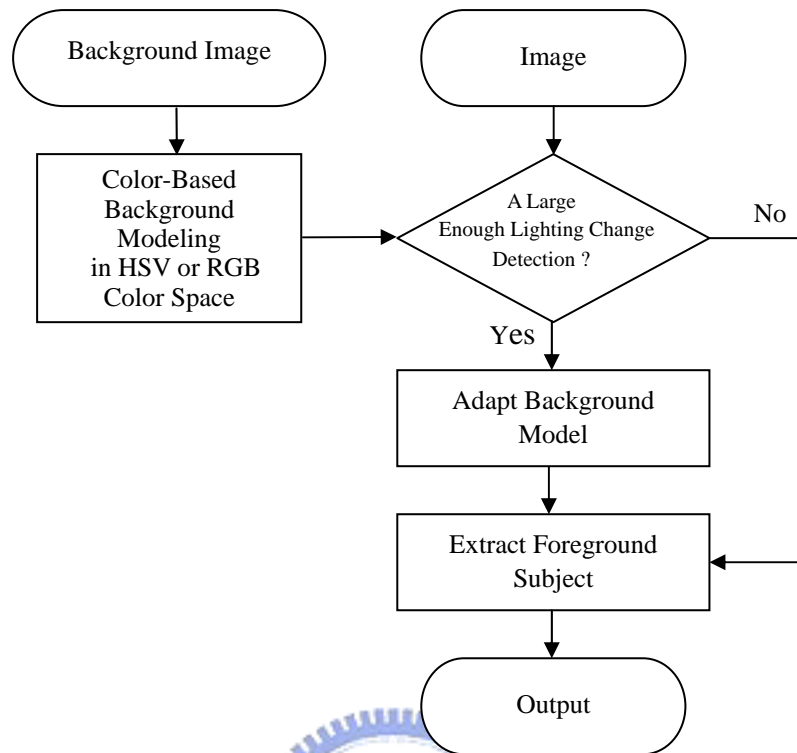


Fig. 1.1 The flowchart of our subject extraction system.



1.2 Background Modeling

In the HSV color space, we use the luminance information (Intensity V component) and the chromatic information (Saturation and Hue components) to describe the background model.

In the RGB color space, a reference image is generally used to perform background subtraction. The simplest approach to obtain a reference image is by averaging a period of frames [4]. However, it is not suitable to apply time averaging on the home-care applications because the foreground subjects (especially for the elderly people or children) usually move slowly and the household scene changes constantly due to light variations from day to night, switches of fluorescent lamps and furniture movement, etc. In short, deterministic methods such as the time averaging have been found to have limited success in practice. For indoor environments, a good background model must also handle the effects of illumination variation, and the interference from background and shadow. Furthermore, if the background model cannot respond the fast or slow variations from sunlight or fluorescent lamps, the entire image will be regarded as foreground. That is, a single model cannot represent the distribution of pixels with twinkling values. Therefore, it is necessary to describe a background pixel by a Gaussian Mixture Model, instead of a single model, in home-care applications of the real world.

1.3 The Background Model Modification

In the extraction system, a large lighting change is a very important phenomenon to consider in most cases. The lighting change result from changes in illumination, so lighting change has similar chromaticity but different brightness. To tackle this, we utilize the luminance component to detect this lighting change. In the HSV color space, we modify the intensity value, but in the RGB color space, we denote a brightness value I_{ij} of each pixel as given by

$$I_{ij} = \sqrt{R_{ij}^2 + G_{ij}^2 + B_{ij}^2}, \quad (1)$$

where I_{ij} is the i -th pixel's brightness of the j -th image frame. The brightness I_{ij} is employed the background model.



1.4 Foreground Subject Extraction

Foreground extraction is an important step of the subject/object extraction system. Many authors have developed methods of detecting people in images. Park and Aggarwal subtracted foreground pixels from background by computing Mahalanobis distance in each pixel in the HSV color model [5]. Leung and Yang built a human body outline labeling system [6]. Jabri and Duric [7] used color and edge information to improve the quality and reliability of the results. They all try to find out the real poses a human did by human body outline or by silhouettes.

Background subtraction is widely used for detecting moving objects from image frames of static cameras. Most of this work has been based on background subtraction using luminance or color information. In these approaches, difference between the system and the background image is performed to detect foreground subject/object.

If we only use the luminance information to do background subtraction, we cannot detect a foreground pixel correctly when the foreground color is similar to that of the background. To make fully use of the spectrum of a pixel, it is imperative to do the segmentation in the color domain. To the end, foreground subject/object extraction is done in the HSV or RGB color spaces. We can have both the luminance information and the chromatic information in the background subtraction task.

Background subtraction is extremely sensitive to dynamic scene differences due to a large illumination change. In order to solve the artifact of varying luminance conditions, we develop a method which is robust to the illumination changes. The method is changed each pixel's intensity matching to the closest lighting of the background model, which is selected according to the luminance mean of the image.

Furthermore, the moving cast shadows mostly exhibit a challenge for accurate

foreground subject detection. A lot of attempts have been developed to tackle the shadow suppression [8]–[13] encountered in background subtraction. Horprasert *et al.* [8] and Cucchiara *et al.* [9] utilized the rationale that shadows have similar chromaticity, but lower brightness than the background model. Under the proposed frame work in the HSV and RGB color spaces, we can effectively identify the shadow existent in our detected foreground subject.



1.5 Thesis Outlines

The thesis is organized as follows. Before introducing the technique of subject/object extraction system, the basic concepts concerning the HSV and RGB color spaces, including transformation between RGB and HSV spaces. Gaussian Mixture Models (GMMs) are also introduced in Chapter 2. Chapter 3 describes background subtraction system in detail. In Chapter 4, the experiment results of our extraction systems are shown. At last, we conclude this thesis with a discussion in Chapter 5.



Chapter 2 Basic Concept

In this chapter, we briefly explain the basic concepts of HSV and RGB color spaces and converting color from RGB to HSV. Then GMM concept is introduced.

2.1 The HSV Color Space

The HSV (hue, saturation and value) color space corresponds closely to the human perception of color. Conceptually, the HSV color space is a cone. Viewed from the circular side of the cone, the hues are represented by the angle of each color in the cone relative to the 0° line, which is traditionally assigned to be red. The saturation is represent as the distance from the center of the circle. Highly saturation color are on the outer edge of the cone, whereas gray tones (which have no saturation) are at the very center. The brightness is determined by the colors vertical position in the cone. At the point end of the cone, there is no brightness, so all colors are blacks. At the fat end of the cone are the brightness colors.

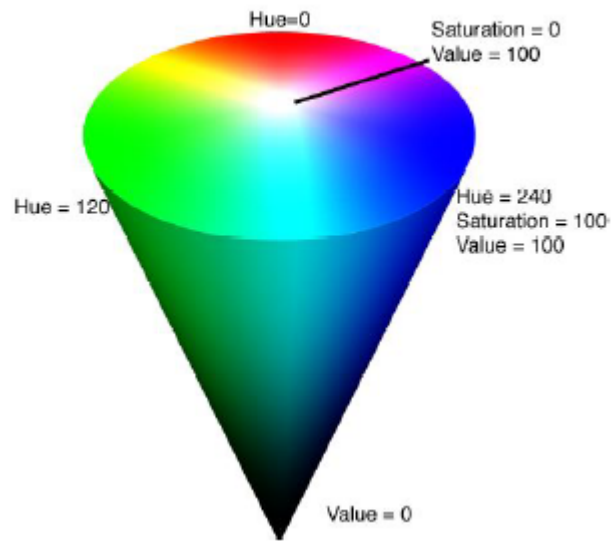


Fig. 2.1 The HSV Cone.

The hue parameter is the value which represents color information without brightness. Therefore, the hue is not affected by change of the illumination brightness and direction. Although hue is the most useful attribute, there are three observations in using hue attribute for color segmentation: (1) hue is meaningless when the intensity value is very low; (2) hue is unstable when the saturation is very low; and (3) saturation is meaningless when the intensity value is very low [11]. Accordingly, Ohba *et al.* [14] use these three criteria (*intensity, saturation, and hue values*) to obtain the more reliable hue value.

2.2 The RGB Color Space

In the RGB model, each color appears in its primary spectral components of red, green, and blue. This model is based on a Cartesian coordinate system. The color subspace of interest is the cube shown in Fig. 2.2, in which RGB values are at three corners; cyan, magenta, and yellow are at three other corners. Black is at the origin and white is at the corner farthest from the origin. In the model, the gray scale (points of equal RGB values) extends from black to white along the line joining these two points. The different colors in this model are points on or inside the cube, and are defined by vectors extending from the origin. For convenience, the assumption is that all color values have been normalized so that the cube shown in Fig. 2.2 is the unit cube. That is, all values of R , G , and B are assumed to be in the range $[0, 1]$.

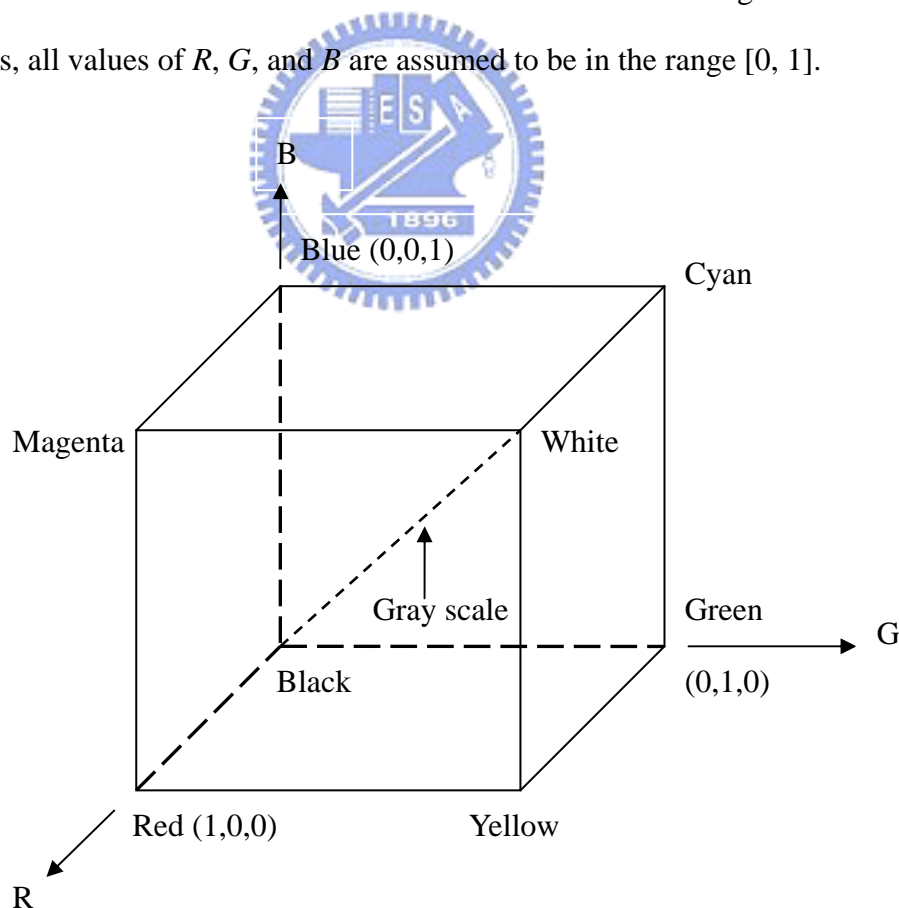


Fig. 2.2 The RGB Color Cube.

2.3 Converting Colors from RGB to HSV

Given an image in RGB color format, the V component of each pixel is obtained using the equation

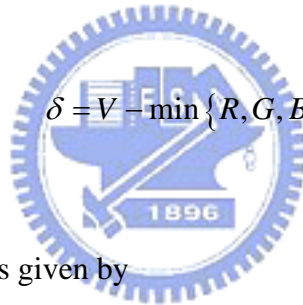
$$V = \max \{R, G, B\}. \quad (2)$$

The saturation component is given by

$$S = \frac{\delta}{V}, \quad (3)$$

with

$$\delta = V - \min \{R, G, B\}.$$



Finally, the value component is given by

$$\begin{aligned} H &= \frac{1}{6} \frac{G-B}{\delta}, & \text{if } R=V \\ H &= \frac{1}{6} \left(2 + \frac{B-R}{\delta} \right), & \text{if } G=V \\ H &= \frac{1}{6} \left(4 + \frac{R-G}{\delta} \right), & \text{if } B=V \end{aligned} \quad (4)$$

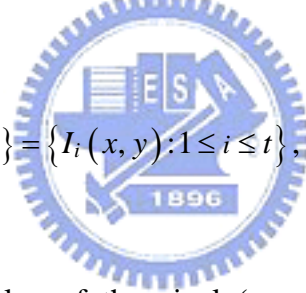
It is assumed that the RGB values have been normalized to the range $[0, 1]$ and that angle θ is measured with respect to the red axis of the HSV space. Hue, saturation and value are normalized to be in the range $[0, 1]$.

2.4 Gaussian Mixture Models (GMMs)

Although background subtraction approach is simple, it may be impractical in some real applications because backgrounds can change over time in some cases. Also, as mentioned earlier, it is difficult to have a clear and stationary background. Lighting can change the background subtly or the camera position may drift. An alternative approach is to find a way of adapting the background slowly such that changing background can be characterized in real-time. Such an approach is called adaptive background mixture models [2], [15]-[17].

Consider a history of pixels in a sequence of image frames. The history of a certain pixel (x, y) can be defined as a time series as

$$\{x_1, \dots, x_t\} = \{I_i(x, y) : 1 \leq i \leq t\}, \quad (5)$$



where x_i is the intensity value of the pixel (x, y) at time instant i . For color images x_i is a vector and for gray level image it is a scalar. The GMM is to use multiple Gaussian Models to model the probability of a pixel in a sequence of images. Assume k Gaussian probability density distributions are utilized for x_i . The probability to observe a certain pixel value at time t is determined as

$$P(x_t) = \sum_{i=1}^k w_{i,t} \times N(x_t, \mu_{i,t}, \Sigma_{i,t}), \quad (6)$$

with

$$N(x_t, \mu_{i,t}, \Sigma) = \frac{1}{(2\pi)^{\frac{d}{2}} |\Sigma|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(x_t - \mu)^T \Sigma^{-1} (x_t - \mu)\right\}, \quad (7)$$

where $w_{i,t}$ is the weight parameter that define the weights of the i -th Gaussian distribution used to describe the data. N is a Gaussian distribution that has two parameters: $\mu_{i,t}$ is the mean of the Gaussian distribution and $\Sigma_{i,t}$ is the covariance matrix at time instant t . Note that Stauffer and Grimson [2] use 3-5 distributions to describe the history of each pixel.

An on-line K-means approximation is then used to update the parameters of the distributions as new information is gained from new frames. For each frame every new pixel value is compared against the existing Gaussian distributions. A new pixel is said to match the current distribution if it is within 1.5 standard deviations from the mean of the distribution for normal light condition. If a pixel matched with one of the weighted Gaussian distributions the mean and the variance of this distribution are updated using the following equations.

$$\mu_t = (1 - \rho)\mu_{t-1} + \rho x_t, \quad (8)$$

$$\sigma_t^2 = (1 - \rho)\sigma_{t-1}^2 + \rho(x_t - \mu_t)^T(x_t - \mu_t), \quad (9)$$

with

$$\rho = \alpha N(X_t | \mu_k, \sigma_k), \quad (10)$$

and α is the learning rate that is defined by the user and is used to define the learning speed of the background updating. The mean and the variance of an unmatched distribution are not updated. Also the weight parameters of all distributions belonging to the certain pixel are updated as follows.

$$w_{k,t} = (1 - \alpha)w_{k,t-1} + \alpha M_{k,t}, \quad (11)$$

where $M_{k,t}$ is 1 for cases of matching the current distributions and 0 for cases of not matching the current distributions.

In order to define which of the k Gaussian distributions describing the history of a pixel result from background and which ones from the foreground, the distributions for each pixel are normalized by a factor w/σ . The distributions that describe the background are expected to have a large weight parameter and a small variance. Thus, B first distributions are marked as background distributions. B can be obtained with the help of the background threshold, which is defined by the user and simply indicates the minimum portion of the data that is considered to result from the background pixels. Thus,

$$B = \arg \min_b \left(\sum_{k=1}^b w_k > T \right). \quad (12)$$

If a pixel is matched with one of these B distributions, it is marked as a background pixel; otherwise it is a part of a moving object and thus it is marked as a foreground pixel.

Chapter 3 Background Subtraction System

3.1 Object Extraction Using HSV Model

3.1.1 The Intensity of The Image

We assume the intensity of the image captured by a camera can be described as

$$I_i(x, y) = S_i(x, y)r_i(x, y), \quad (13)$$

where I_i is the intensity of the image, S_i is the spatial distribution of source illumination, r_i is the distribution of scene reflectance, (x, y) is the location of a pixel in the image, and i is the image sequence index. Now we can compare the difference caused by illumination change between frame difference and frame ratio. If we hold the camera still with no foreground subjects pass by, the reflectance of this background should be the same at any time. That is,

$$r_i(x, y) = r(x, y). \quad (14)$$

Although the reflectance is not changed, the effect of illumination is still going on. The frame difference and frame ratio between two consecutive frames can respectively be written as

$$\begin{aligned} I_i^d(x, y) - I_{i-1}^d(x, y) &= S_i^d(x, y)r(x, y) - S_{i-1}^d(x, y)r(x, y) \\ &= (S_i^d(x, y) - S_{i-1}^d(x, y))r(x, y), \end{aligned} \quad (15)$$

$$\begin{aligned}
\log\left(\frac{I_i^r(x, y)}{I_{i-1}^r(x, y)}\right) &= \log\left(\frac{S_i^r(x, y)r(x, y)}{S_{i-1}^r(x, y)r(x, y)}\right) \\
&= \log\left(\frac{S_i^r(x, y)}{S_{i-1}^r(x, y)}\right) \\
&= \log(S_i^r(x, y)) - \log(S_{i-1}^r(x, y)),
\end{aligned} \tag{16}$$

where I^d is the intensity of scene captured by camera of frame difference, S^d is the spatial distribution of source illumination of frame difference, and I^r and S^r is of frame ratio. Comparing Eqs. (15) and (16), we can find that the problems cause by reflectance still remains in the frame difference approach; nevertheless, the influence of reflectance is eliminated in the frame ratio approach.

Fig. 3.1 shows a comparison between frame ratio and frame difference. Fig. 3.1(a) is a background image and Fig. 3.1(b) is an image frame with a human. By using frame difference and frame ratio approach, we obtain Figs. 3.1(c) and 3.1(d), respectively. Gray level of the resulting images distributed from 0 to 255. Fig. 3.1(e) is the histogram of Fig. 3.1(c) and Fig. 3.1(f) is the histogram of Fig. 3.1(d). Comparing the histograms of Figs 3.1(d) and 3.1(e), we find out that there was less noise in the region of low gray level by using frame ratio method. Figs. 3.1(g) and 3.1(h) are respectively the binary image of extraction images which simply took a threshold value 15 at gray level of Figs. 3.1(c) and 3.1(d).

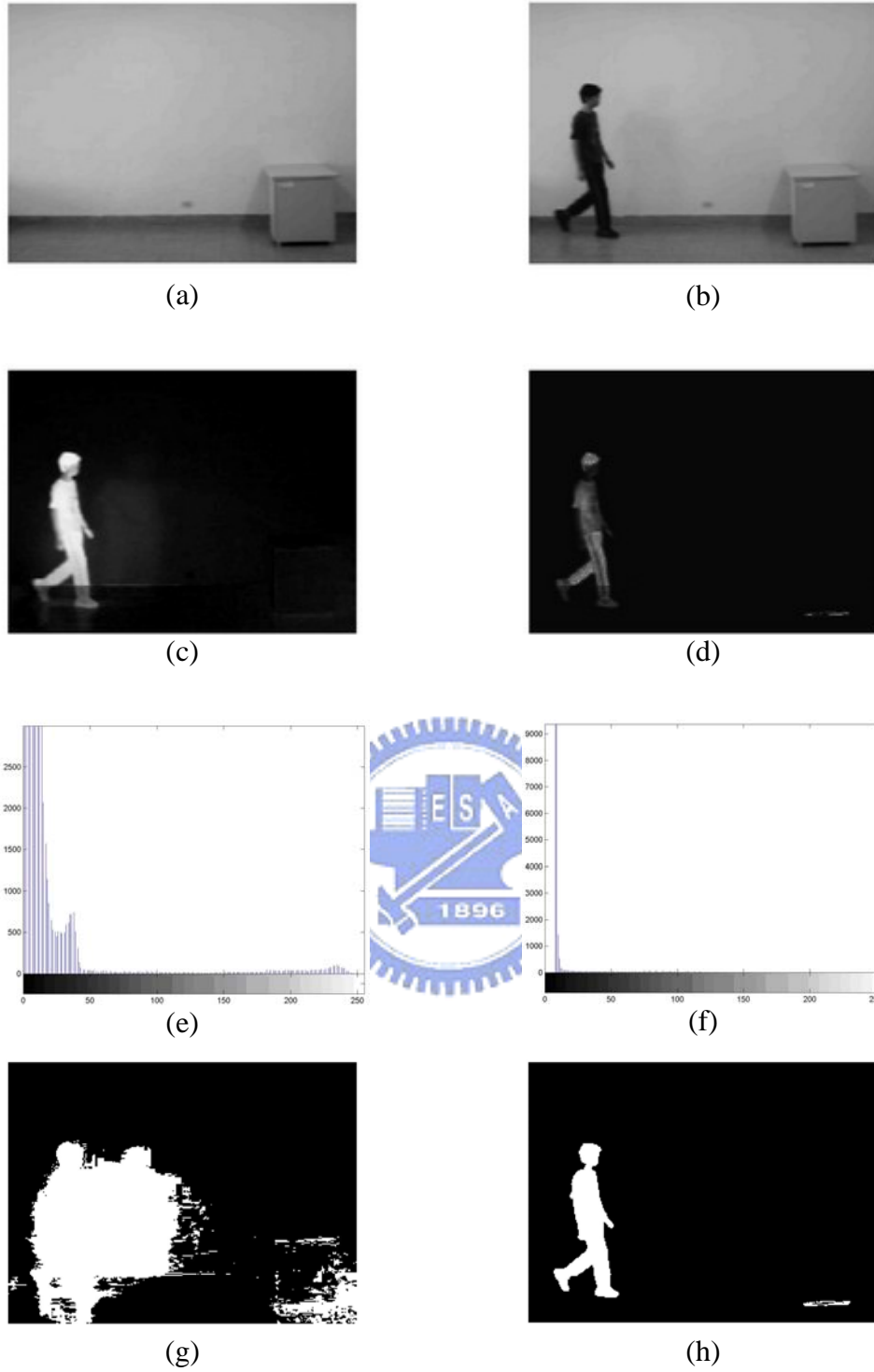


Fig. 3.1 The comparison between frame ratio and frame difference. (a) Background image, (b) image frame with a human, (c) frame difference, (d) frame ratio, (e) histogram of frame difference, (f) histogram of frame ratio, (g) foreground pixels of frame difference after simply taking a threshold, and (h) foreground pixels of frame ratio after simply taking a threshold.

3.1.2 Background Model in HSV Space

If we use only the luminance component to do background subtraction, we cannot detect reliably those foreground pixel whose luminance component close to background pixel. In order to solve this problem, we build our background model in the HSV color space. The HSV color space corresponds closely to the human perception of color. We can have the luminance information and the chromatic information simultaneously during background subtraction.

In the previous section, we have seen the advantage of using frame ratio approach to counter the luminance change. Hence, we propose to utilize the frame ratio to build the background model in the luminance component. We build our background model with the minimum value ($[n^H(x, y), n^S(x, y), n^V(x, y)]$) and maximum value ($[m^H(x, y), m^S(x, y), m^V(x, y)]$) in each HSV domain. Besides, we also record the inter-frame ratio in the brightness components and the inter-frame difference in the chromatic components.

We need a background video, without any moving objects, for background model training. Suppose that the observed image frame sequence contains N consecutive images. Let $I_i^H(x, y)$ be the pixel's hue value at (x, y) of the i -th image frame. Let $I_i^S(x, y)$ be the pixel's saturation value at (x, y) of the i -th image frame. Let $I_i^V(x, y)$ be the pixel's brightness value at (x, y) of the i -th image frame. The background model of a pixel is obtained by

$$\begin{bmatrix} m^H(x, y) \\ n^H(x, y) \\ d^H(x, y) \end{bmatrix} = \begin{bmatrix} \max_i \{I_i^H(x, y)\} \\ \min_i \{I_i^H(x, y)\} \\ \max_i \{|I_i^H(x, y) - I_{i-1}^H(x, y)|\} \end{bmatrix}, \quad (17)$$

$$\begin{bmatrix} m^S(x, y) \\ n^S(x, y) \\ d^S(x, y) \end{bmatrix} = \begin{bmatrix} \max_i \{I_i^S(x, y)\} \\ \min_i \{I_i^S(x, y)\} \\ \max_i \{|I_i^S(x, y) - I_{i-1}^S(x, y)|\} \end{bmatrix}, \quad (18)$$

$$\begin{bmatrix} m^V(x, y) \\ n^V(x, y) \\ d^V(x, y) \end{bmatrix} = \begin{cases} \begin{bmatrix} \max_i \{I_i^V(x, y)\} \\ \min_i \{I_i^V(x, y)\} \\ \max_i \{I_i^V(x, y)/I_{i-1}^V(x, y)\} \end{bmatrix} & \text{if } I_i^V(x, y)/I_{i-1}^V(x, y) \geq 1 \\ \begin{bmatrix} \max_i \{I_i^V(x, y)\} \\ \min_i \{I_i^V(x, y)\} \\ \max_i \{I_{i-1}^V(x, y)/I_i^V(x, y)\} \end{bmatrix} & \text{otherwise} \end{cases}, \quad (19)$$

where $i = 1, 2, \dots, N$.



3.1.3 Adaptive Background Model with A Lighting Change Compensation

Besides considering foreground and background regions, lighting change is an important factor that should be considered in image segmentation. Lighting change results in changes in illumination. In case of lighting change, the image pixel will have similar chromaticity but will increase its intensity component if brighter, and vice versa. With this concept in mind, we should utilize the luminance component instead of the chromatic components to reduce the lighting change artifact in foreground detection.

First, we compute the mean value of intensity by the use of statistics from N color background image frames, assuming without moving subject/object. Let M denote the number of pixels of each background image frame. I_{ij} be the i -th pixel's brightness value of the j -th image frame. The mean value μ_b of N background image frames is given by

$$\mu_b = \frac{1}{NM} \sum_{j=1}^N \sum_{i=1}^M I_{ij}. \quad (20)$$

Second, we classify the entire image frames with several average intensity levels in each frame, and get the mean value for each class. In this thesis, we assume the intensities of image frames contain normal, dark and bright lighting cases, denoted by $k=1, 2$ and 3 . Let the k -th case image contains N_k image frames. We can get the mean intensity μ_k of each class given by

$$\mu_k = \frac{1}{N_k M} \sum_{j=1}^{N_k} \sum_{i=1}^M I_{ij}, \quad k = 1, 2, \text{ and } 3. \quad (21)$$

Third, we make use of the luminance information by Eq. (22) to check whether the input t -th image frame existing a large enough lighting change with respect to the previous frame. If the entire frame detect the intensity changes given below, and the lighting change flag $LC(t)$ is set to 1 as given by

$$LC(t) = \begin{cases} 1, & \text{if } \frac{|I_t^V - I_{t-1}^V|}{I_t^V} \geq 5\% \quad \text{and} \quad \frac{|I_t^V - \mu_k|}{\mu_k} \geq 15\% \\ 0, & \text{otherwise} \end{cases} \quad (22)$$

Each time when $LC(t)$ is set to 1, we must modify the background model to move model's intensity component to the mean intensity of nearest lighting case detected, and the other two chromatic components are unchanged. The adaptive background model of a pixel is thus modified by

$$\begin{bmatrix} m^V(x, y) \\ n^V(x, y) \\ d^V(x, y) \end{bmatrix} = \begin{cases} \begin{bmatrix} m^V(x, y) \times \mu_k / \mu_b \\ n^V(x, y) \times \mu_k / \mu_b \\ d^V(x, y) \end{bmatrix} & \text{if } P(k) = 1 \\ \begin{bmatrix} m^V(x, y) \\ n^V(x, y) \\ d^V(x, y) \end{bmatrix} & \text{otherwise} \end{cases}, \quad (23)$$

where $i = 1, 2, \dots, N$, $j = 1, 2, 3$.

3.1.4 Foreground Subject Extraction and Shadow Suppression

Fig. 3.2 shows the framework we propose for foreground subject extraction. Our framework of foreground subject extraction is composed of three components. The first component is the background model adaptation to a large lighting change. The second component is foreground subject extraction. And the last component is shadow suppression.

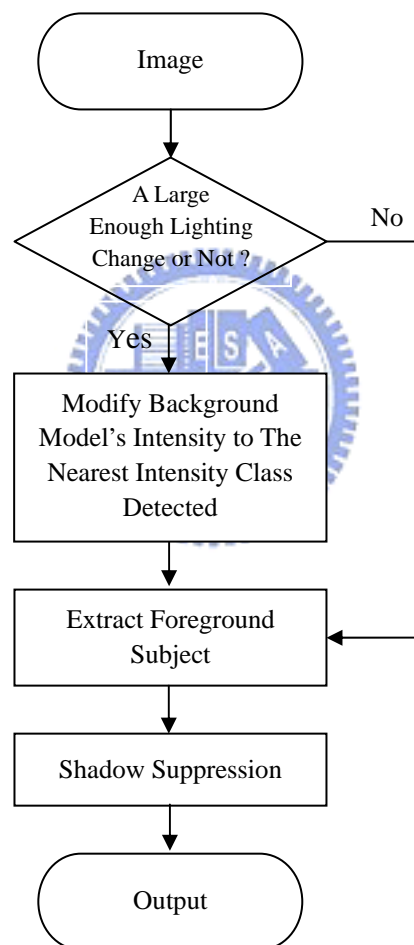


Fig. 3.2 The framework for foreground subject/object extraction.

A. Extract Foreground Subject

Foreground objects are required to be segmented from every frame of the video stream. Each pixel of the video frame is classified to either a background or a foreground pixel by the difference between the background model and a captured image frame. In line with W⁴ [1], we utilize the maximum luminance $m^V(x, y)$, minimum luminance $n^V(x, y)$ and maximum inter-frame luminance ratio $d^V(x, y)$ of the training background model to segment the foreground pixel by

$$B(x, y) = \begin{cases} 0, & \text{if } I_i^V(x, y)/m^V(x, y) < k_v d^V(x, y) \\ & \text{or } I_i^V(x, y)/n^V(x, y) < k_v d^V(x, y) \\ 255, & \text{otherwise} \end{cases}, \quad (24)$$

where $I_i^V(x, y)$ is the intensity of a pixel of the i -th image which is located at (x, y) , $B(x, y)$ is the gray level of a pixel in a binary image, and k_v is a threshold, determined by light sufficiency of the scene. The value of k_v is normally set to 1.3 for normal light condition, and k_v will be reduced for in-sufficient light condition and increased otherwise.

B. Shadow Suppression

The pixels of the moving cast shadows are easily detected as the foreground pixel in normal condition. Because the shadow pixels and the object pixels share two important visual features: motion model and detectability. For this reason, the moving shadows cause object merging and object shape distortion. Horprasert *et al.* [8] and Cucchiara *et al.* [9] utilize the rationale that shadows have similar chromaticity, but lower brightness than the background model. Adapting this concept, we can detect the shadow from foreground subject in the HSV color space. We analyze only points belonging to possible moving object that are detected in the above Step A. We define a shadow mask S for each (x, y) point as follows:

$$S(x, y) = \begin{cases} \text{shadow,} & \text{if } I_i^V(x, y) - n^V(x, y) < 0 \\ & \text{and } |I_i^H(x, y) - m^H(x, y)| < k_H d^H(x, y) \\ & \text{and } |I_i^S(x, y) - m^S(x, y)| < k_S d^S(x, y) \quad , \\ \text{object,} & \text{otherwise} \end{cases} \quad (25)$$

where $I_i^H(x, y)$, $I_i^S(x, y)$, and $I_i^V(x, y)$ are respectively the HSV channel of a pixel located at (x, y) , and $S(x, y)$ is the shadow mask to class the pixel in the moving cast shadow. Values k_S and k_H are selected threshold values used to measure the similarities of the hue and saturation between the background image and the current observed image. We can utilize the shadow mask $S(x, y)$ to change the shadow pixels into background in $B(x, y)$ of Eq. (24).

3.2 Object Extraction Using RGB Model

In this section, we propose to use RGB model for moving object segmentation. The background subtraction and the lighting change detection are common approaches for object extraction in the literature.

3.2.1 GMM Background Model in RGB Space

In this section, we will exploit GMM model, which is presented in detail in Sec. 2.4 by Eqs. (5)-(7), to build our background model in the RGB color space. Moreover, Eqs. (8)-(11) are utilized to adapt the background model slowly to the environment.



3.2.2 Adaptive Background Models with A Lighting Change Compensation

In the HSV color space, we utilize the illumination component to reduce the lighting change artifact in foreground detection. In the RGB color space, it is ready to compute the intensity value reliably.

We will check the current image's luminance average by Eq. (22) and if $LC(t)$ is set to 1, we will change each pixel's intensity matching to the luminance mean of the closest lighting cases, and the other two chromatic components are kept unchanged. In this case, each pixels adaptive background models of pixel is modified by

$$\begin{bmatrix} R(x, y) \\ G(x, y) \\ B(x, y) \end{bmatrix} = \begin{cases} \begin{bmatrix} R(x, y) \cdot \mu_k / \mu_b \\ G(x, y) \cdot \mu_k / \mu_b \\ B(x, y) \cdot \mu_k / \mu_b \end{bmatrix}, & \text{if } LC(t) = 1 \\ \text{unchanged} & \text{otherwise} \end{cases} \quad (26)$$

where $i = 1, 2, \dots, N$, $j = 1, 2, 3$.

3.2.3 Foreground Subject Extraction

Fig. 3.3 shows the framework we apply to foreground subject extraction. Our framework of foreground subject extraction is composed of four components. The first component is lighting change detection and modify background model. The second component is the foreground subject detection by luminance. And the last component is the moving history-based background adaptation.

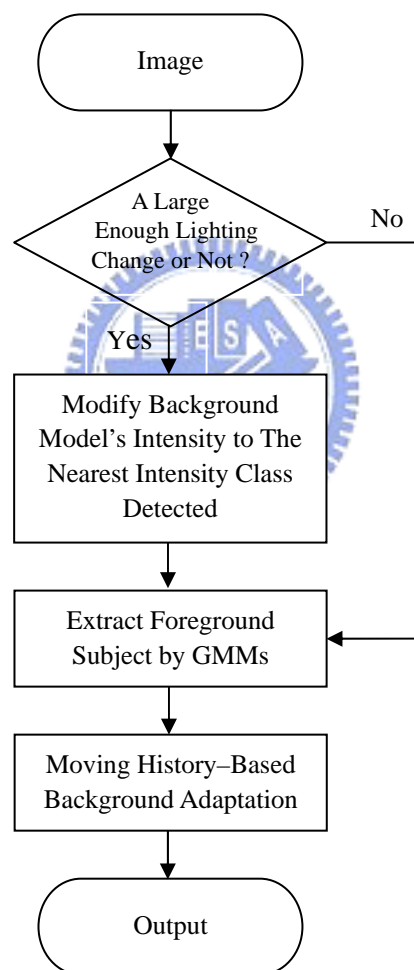


Fig. 3.3 The framework for foreground subject/object extraction.

A. Extract Foreground Subject by GMMs

In this section, we utilize a color-based background subtraction. This work proposes a 3D ball model that is similar to the pillar model proposed by Hoprasert [62]. In the RGB space, a Gaussian distribution of the background model becomes a ball whose center is the mean of the Gaussian component, and the length of each principle axis equals k_G standard deviations of the Gaussian component, where k_G is a threshold. A new pixel $I(R,G,B)$ is considered to belong to background if it is located inside the ball. Fig. 3.4 illustrates the 3D ball model in the RGB color space.

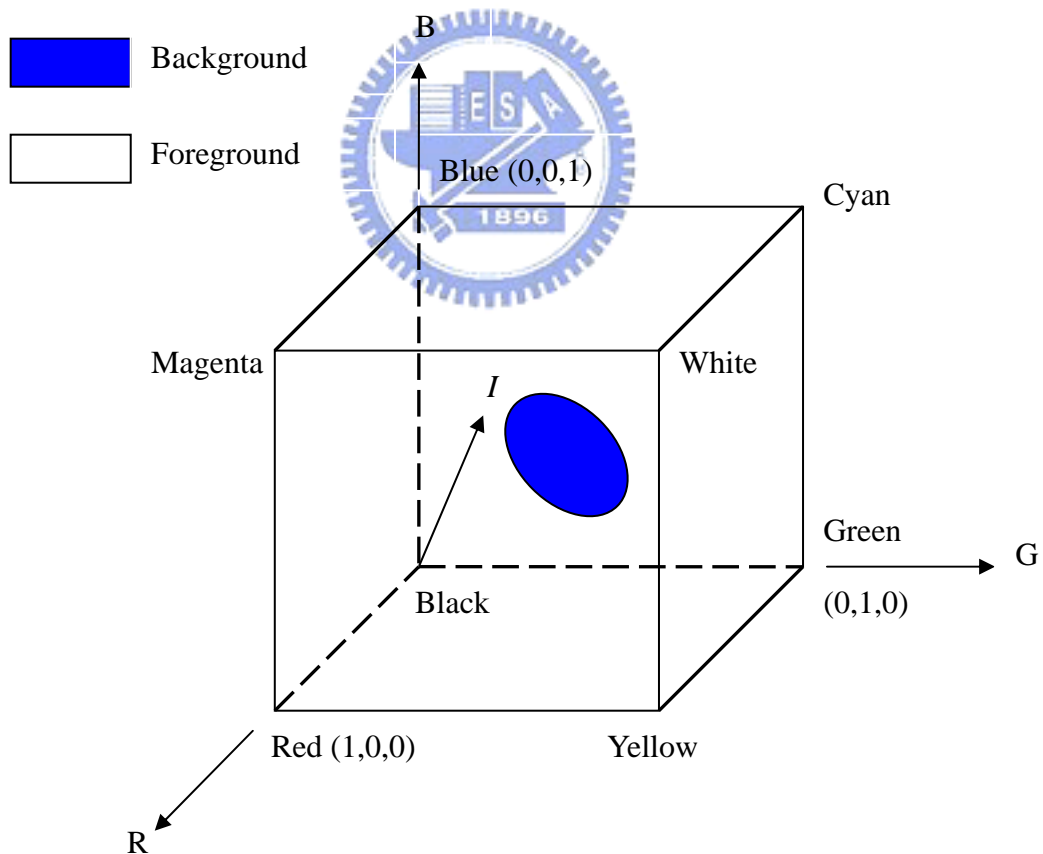


Fig. 3.4 The proposed 3D ball model in the RGB color space.

It is difficult to classify the pixel using the standard deviations in the 3D space. The 3D ball is projected onto the 2D space to classify a pixel using the standard deviations. Fig. 3.5 illustrates the projection of the 3D ball model onto the RG 2D space.

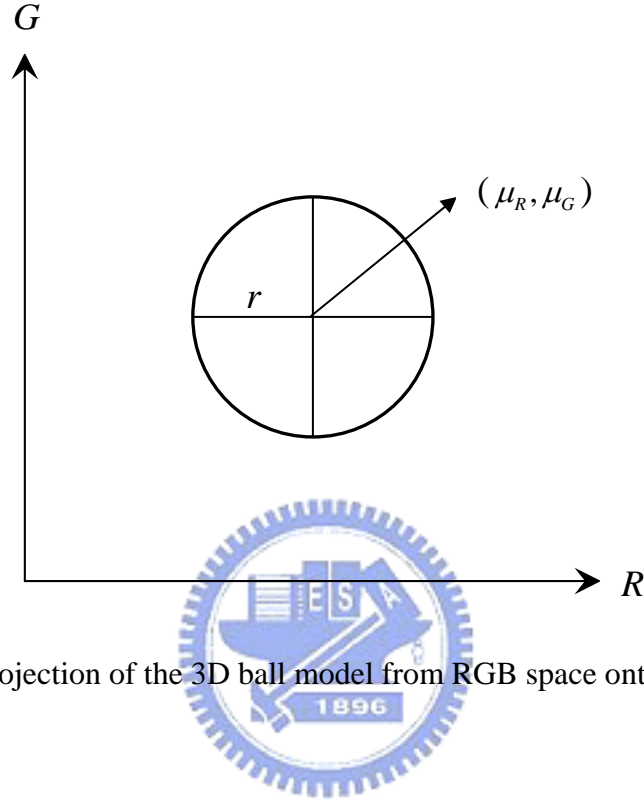


Fig. 3.5 2D projection of the 3D ball model from RGB space onto the RG space.

Let r denote the length of radius axis of the circle, where $r = k_G \cdot \sigma$. The center of the circle is (μ_R, μ_G) . A matching result set is given by $F_b = \{f_{bi}, i=1,2,3\}$, where f_{bi} is the matching result of a specific 2D space. A pixel vector $I = [I_R, I_G, I_B]$ is then projected onto the 2D spaces of R-G, G-B, and B-R. The pixel matching result is set to 1 when the pixel located inside the circle and the image pixel is classified as background using the matching result set F_b by Eq. (27)

$$B(x, y) = \begin{cases} 0, & \sum F_b = 3 \\ 255, & \text{otherwise.} \end{cases} \quad (27)$$

B. Moving History-Based Background Adaptation to the Stop of Moving Subjects

Sometimes, a moving subject may stop for a while, for example chatting with friends during his moving of walking, so we must detect this situation whether the moving subject is stop. Thus, we will utilize the Moving History-Based Background Adaptation to solve this problem. In this system, we will operate on the segmented images.

The ideal behind this method is based on a *motion support map* (mS) which represents the number of times a pixel location is classified as a moving pixel by subtracting in the last N frames by Eq. (28).

$$mS(x, y, t) = \begin{cases} mS(x, y, t-1) + 1, & \text{if } (x, y) \text{ is foreground pixel,} \\ mS(x, y, t-1), & \text{if } (x, y) \text{ is background pixel,} \end{cases} \quad (28)$$

where $t = 1, 2, \dots, N$. During tracking, if pixel that cumulates a large enough image frame as foreground, then it could be becoming stationary, and should be included into the background model. The new extraction image frame is determined by Eq. (29)

$$B(x, y) = \begin{cases} 128 & \text{A moving subject may stop for a while,} & \text{if } (mS(x, y) > k \cdot N) \\ 0 & \text{A moving subject,} & \text{if } (mS(x, y) < k \cdot N), \end{cases} \quad (29)$$

where k is typically 0.8.

Chapter 4 Experimental Result

In our experiment, we tested our system on videos taken by digital camera. We took the video in our laboratory at the 5th Engineering Building in NCTU campus. The camera has a frame rate of thirty frames per second and image resolution is 320×240 pixels. The experimental environment is shown in Fig. 4.1.



Fig. 4.1 The experimental environment.

The background is not complex. The light source is fluorescent lamps and is stable. We assume the intensities of image frames contain three situations: “normal,” “dark” and “bright” lighting cases. The person performed walking “from left to right” and “from right to left.”

We test the foreground detection capability in two cases depending on each background model. That subject extraction using HSV model is the first case. And the second case concerns with subject extraction using GMMs implemented in RGB domain. When the environment with a large enough lighting change in the both cases, a moving subject, such as human body, may not be segmented correctly from image frames. We compare the result in these two cases including with or without

lighting change compensation. Our foreground subtract extraction system demonstrates eminent improvement in the segmentation quality. We classify our entries image frames into three cases and Fig. 4.2 shows sampled image frames in the experiment. Fig. 4.2(a) contains ten image frames in which the subject “walking from left to right” with “normal” lighting; Fig. 4.2(b) includes ten image frames in which the subject “walking from left to right” with “dark” lighting; and Fig. 4.2(c) has ten image frames in which the subject “walking from right to left” with “bright” lighting cases.

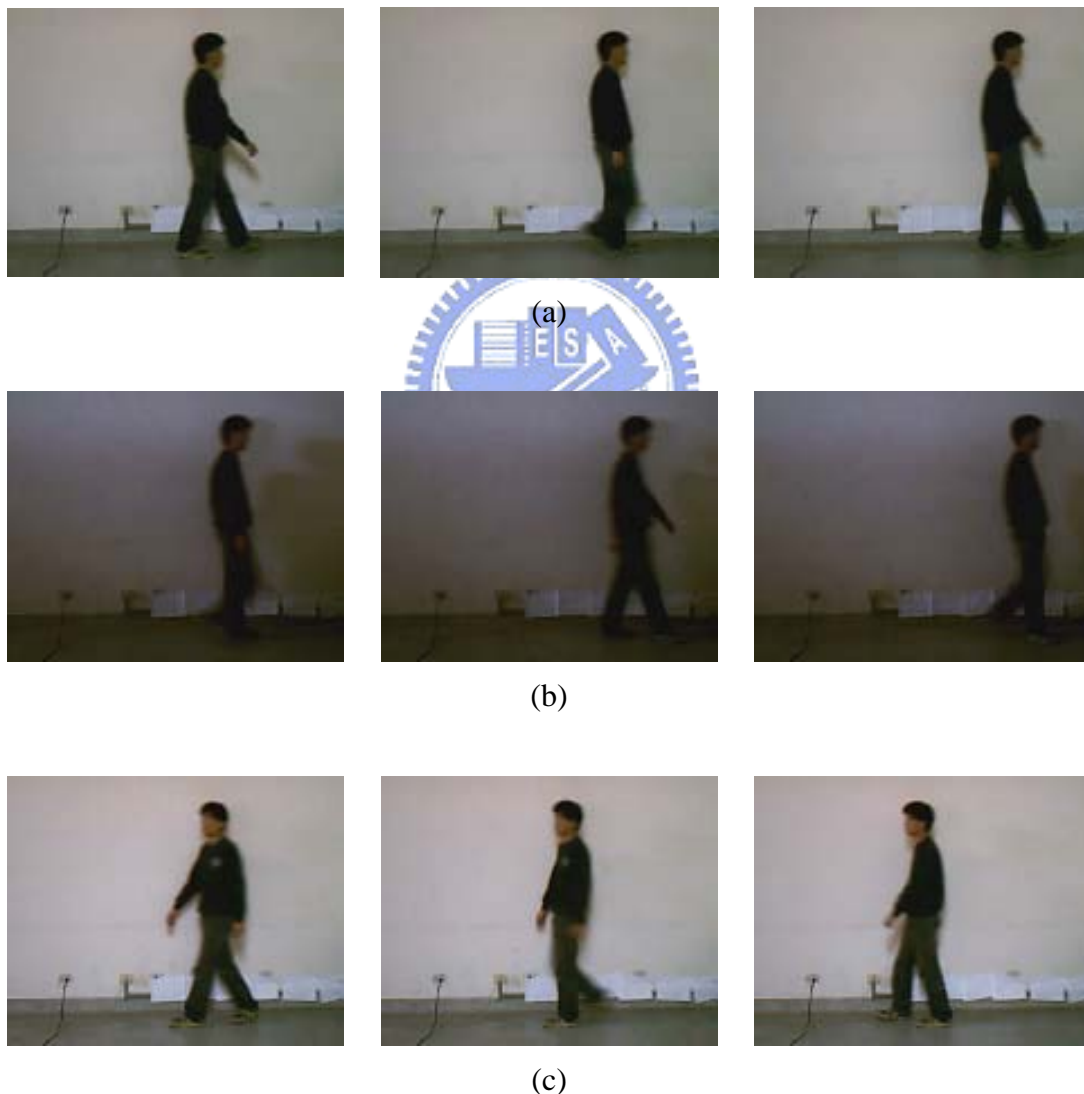


Fig. 4.2 Various image frames of three kind of lights. (a) “Normal” lighting, (b) “dark” lighting, and (c) “bright” lighting cases.

4.1 Object Extraction Using HSV Model

4.1.1 Background Model Construction Using HSV Model

We built the background model in the HSV color space. The value of H or S or V is between 0 and 255. Figs. 4.3(a), 4.3(b), and 4.3(c) show the background image in the H, S, and V component, respectively. We can find from these three figures that the hue value is relatively unstable when the saturation is close to zero. We make an experiment to test the changes in the HSV components in constructing the background model. Fig. 4.4 represents the H, S, and V variations of two pixels at coordinates $(x, y) = (10, 10)$ and $(x, y) = (120, 160)$ during the first 300 frames in the background video. From Fig. 4.4, we can see that V component is most stable of the background model. H and S components are less stable than V.

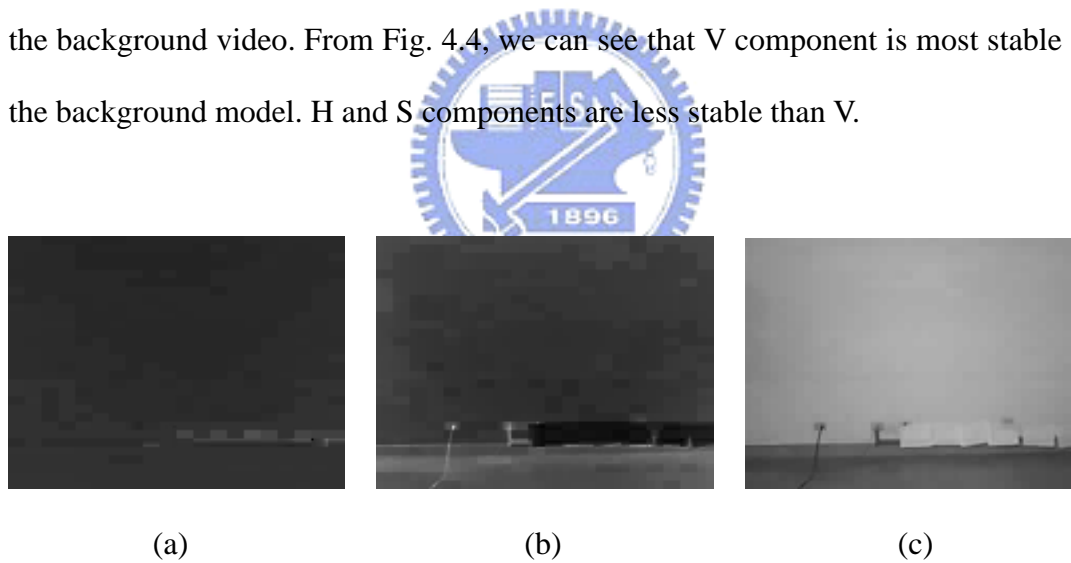


Fig. 4.3 Background images. (a) Background image in the H component, (b) Background image in the S component, and (c) Background image in the V component.

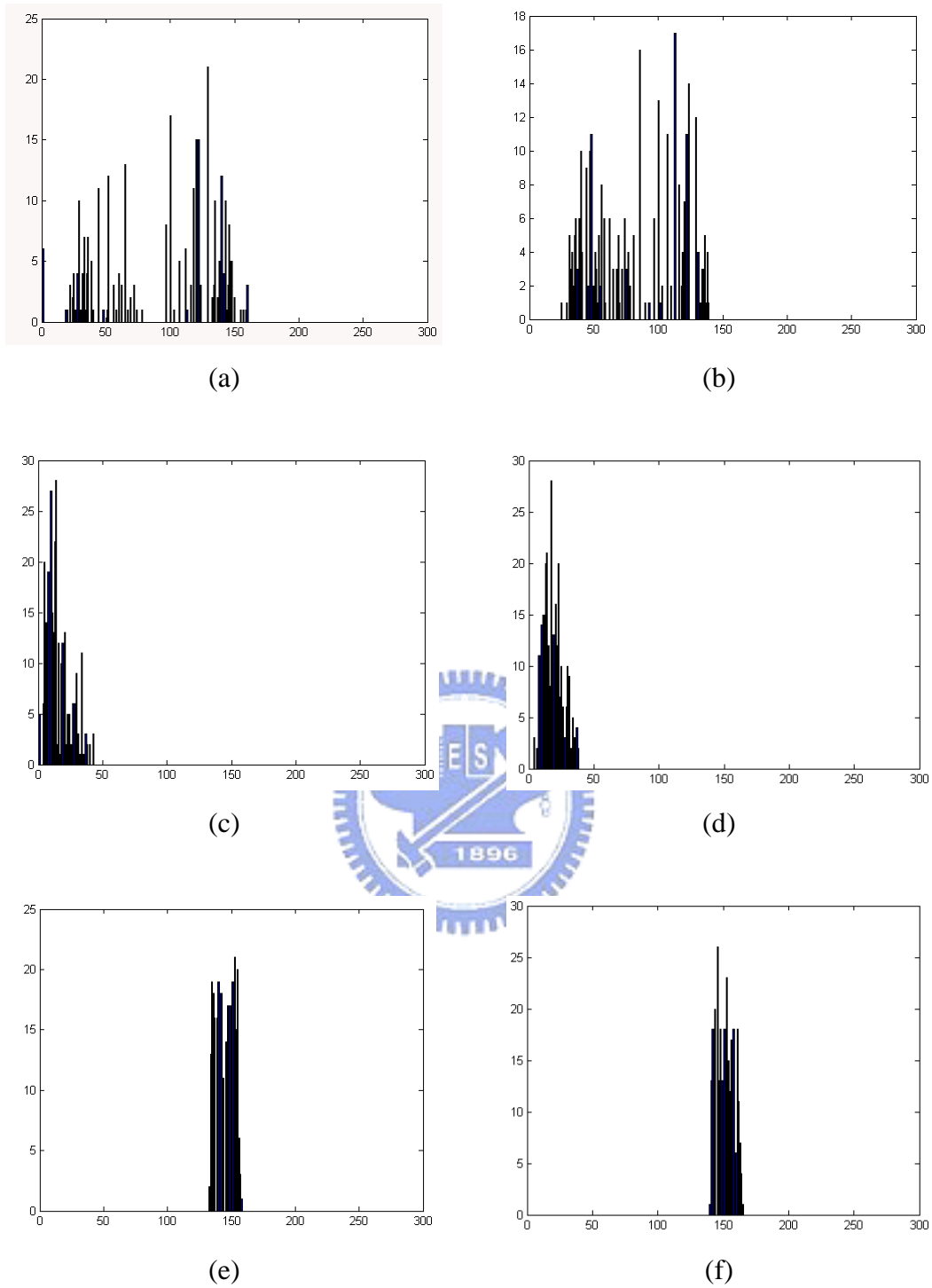


Fig. 4.4 H, S, and V variations versus frame index of background video image frame 1 to frame 300. (a) H at (10, 10), (b) H at (120, 160), (c) S at (10, 10), (d) S at (120, 160), (e) V at (10, 10), and (f) V at (120, 160).

4.1.2 Foreground Subjects Extraction Using HSV Model

In segmenting the images, the V component is usually stable and reliable, but it has two drawbacks: the V component is insensitive to the similar, especially light, color such as yellow, pink, and light blue. When the subjects wear the clothing with the color different from the background, we can do background subtraction well in the V color component.

In the first step, we use the frame ration in the V color component to get the binary image $B(x, y)$ in Eq. (24) described in Sec. 3.1.4. The value k_V is chosen by experiments and varies with different trials. Hence, we ran a series of experiments to determine the optimal threshold k_V . When the subject's clothing color is different from the background, Fig. 4.5 shows the binary image $B(x, y)$ obtained by different k_V 's. After the experiment, we set $k_V = 1.5$ in our system.



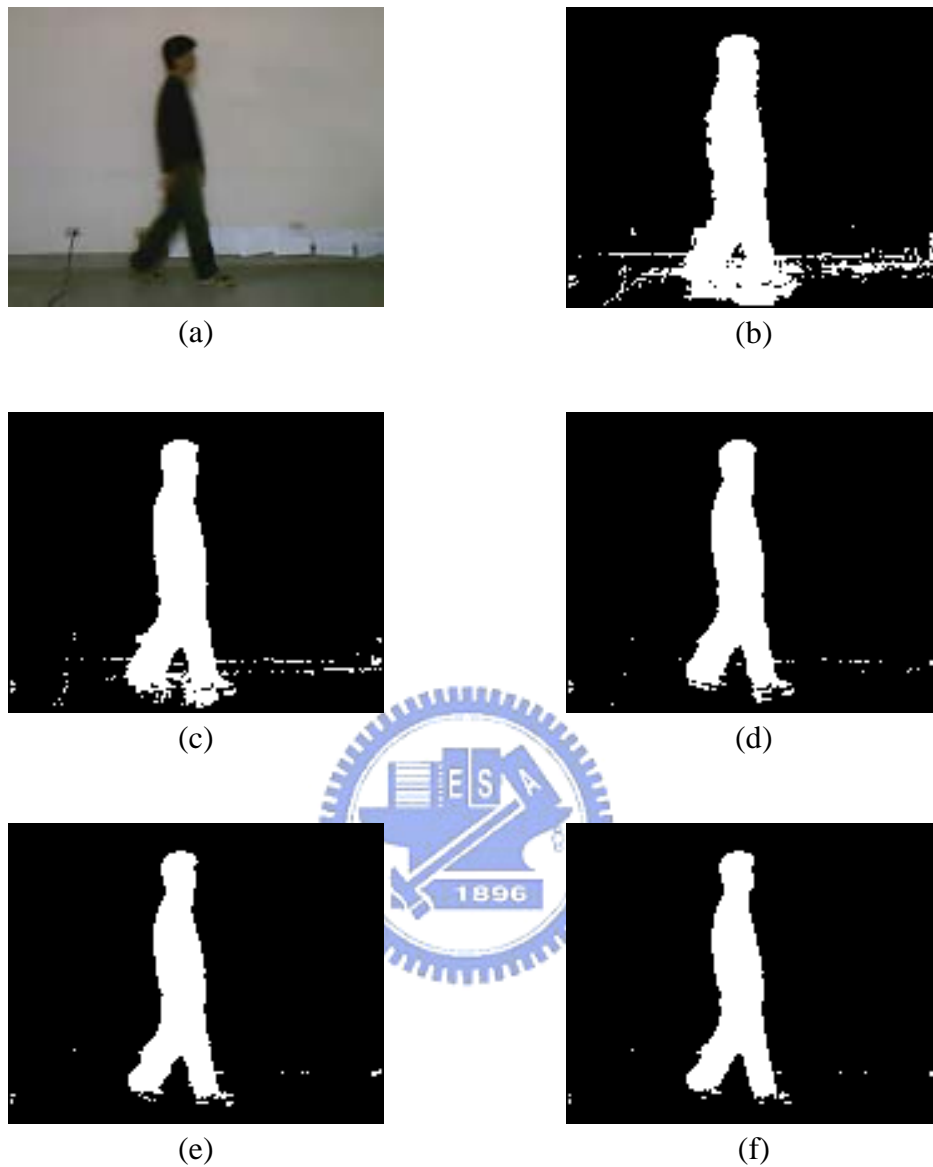


Fig. 4.5 An example of foreground extraction at different k_V thresholds. (a) An image frame with subject's clothing color different from the background, (b)–(f) foreground detected images, (b) $k_V = 1.2$, (c) $k_V = 1.3$, (d) $k_V = 1.4$, (e) $k_V = 1.5$, and (f) $k_V = 1.6$.

During the foreground extraction, the shadowing effect introduces artifact foreground subjects and deteriorates the recognition result. We use the shadow mask, which including the shadow's characteristic existing in HSV domains of Eq. (25) described in Sec. 3.1.4 to classify the pixels whether it is a shadow point or not. Fig. 4.6 shows the process result regarding shadow suppression. Figs. 4.6(a) and 4.6(b) are two input images. Figs. 4.6(c) and 4.6(d) are the foreground subject without shadow suppression. The foreground subject with shadow suppression is shown in Figs. 4.6(e) and 4.6(f), which improves greatly comparing with Figs. 4.6(c) and 4.6(d).

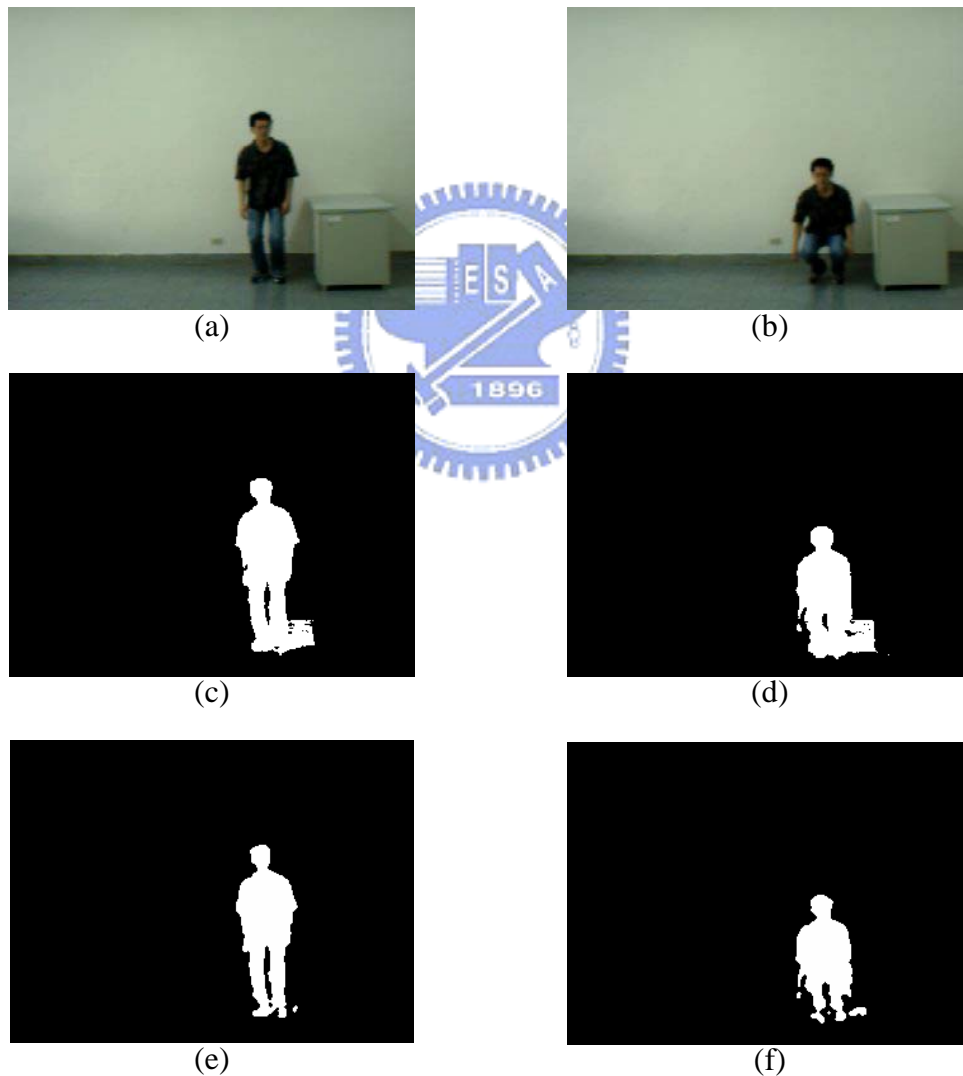


Fig. 4.6 The example of the shadow suppression.

4.1.3 Foreground Extraction with Lighting Change Compensation

During the foreground extraction, it is imperative that the background modeling should adapt to the lighting of the scene. Under the framework of lighting change detection scheme of Eqs. (20), (21), (22) and (23) described in Sec. 3.1.3, we detect the current images whether existing a considerate lighting change or not. Fig. 4.7 shows the process result concerning lighting change. Figs. 4.7(a), 4.7(d) and 4.7(g) are three input images. Figs. 4.7(b), 4.7(e) and 4.7(h) are the foreground subject without lighting change compensation. The foreground subject with lighting change compensation is shown in Figs. 4.7(c), 4.7(f) and 4.7(i), which improves greatly comparing with Figs. 4.7(b), 4.7(e) and 4.7(h).



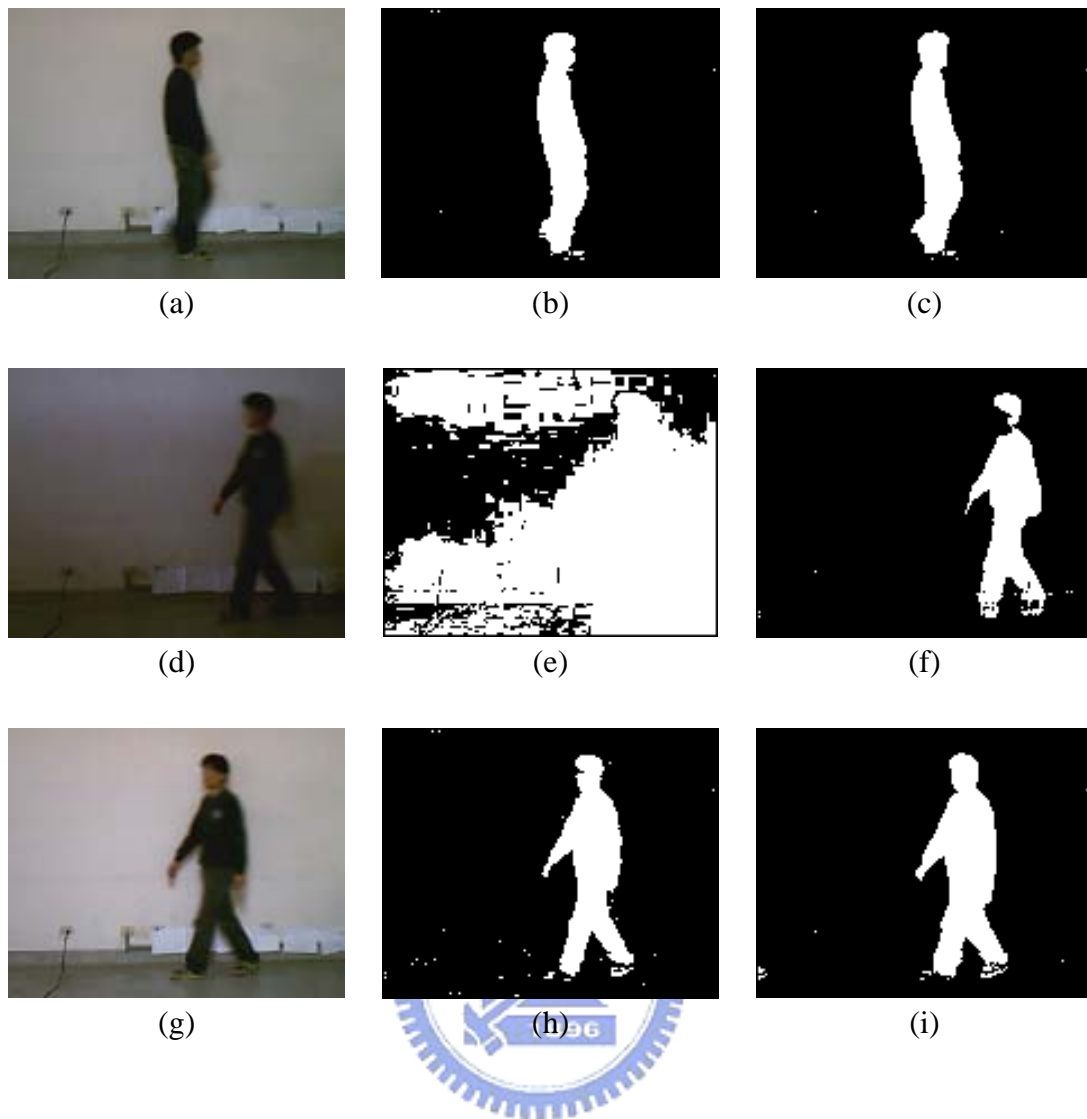


Fig. 4.7 The example concerning lighting change. (a) An image frames in which the subject “walking” with “normal” lighting, (b) a foreground subject without lighting change compensation from (a), and (c) a foreground subject with lighting change compensation from (a). (d) An image frames in which the subject “walking” with “dark” lighting, (e) a foreground subject without lighting change compensation from (d), and (f) a foreground subject with lighting change compensation from (d). (g) An image frames in which the subject “walking” with “bright” lighting, (h) a foreground subject without lighting change compensation from (g), and (i) a foreground subject with lighting change compensation from (g).

4.2 Object Extraction Using RGB Model

4.2.1 Background Model Construction Using RGB Model

We built the background model in the RGB color space. The value of R or G or B is between 0 and 255. Figs. 4.8(a), 4.8(b), and 4.8(c) show the background image in the R, G, and B component, respectively. We can find from these three figures that the R, G, and B are relatively stable, and Fig. 4.9 represents the R, G, and B variations of two pixels at coordinates $(x, y) = (10, 10)$ and $(x, y) = (120, 160)$ during the first 300 frames in the background video. From Fig. 4.9, we can see that R, G, and B components are stable of the background model. Hence, we utilize R, G, and B components to build background model using GMMs.

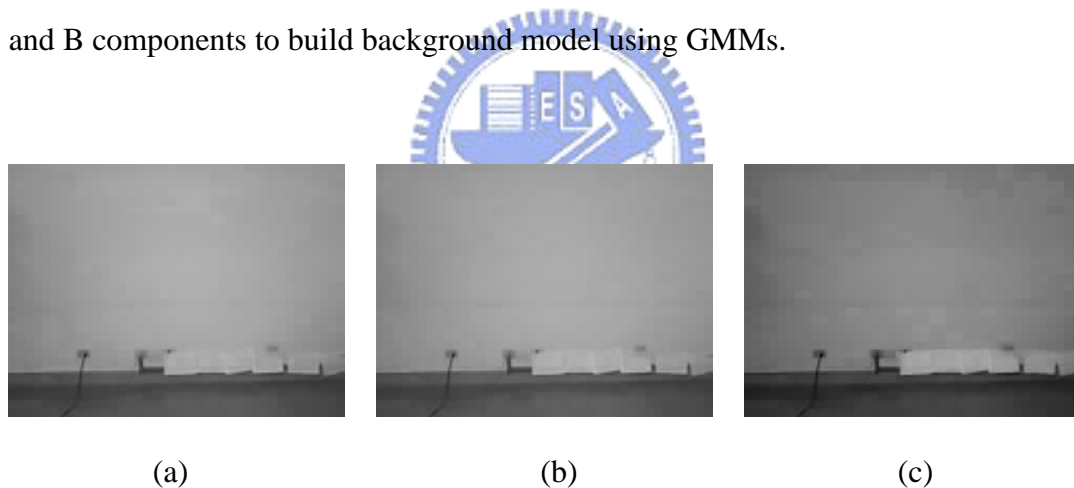


Fig. 4.8 Background images. (a) Background image in the R component, (b) Background image in the G component, and (c) Background image in the B component.

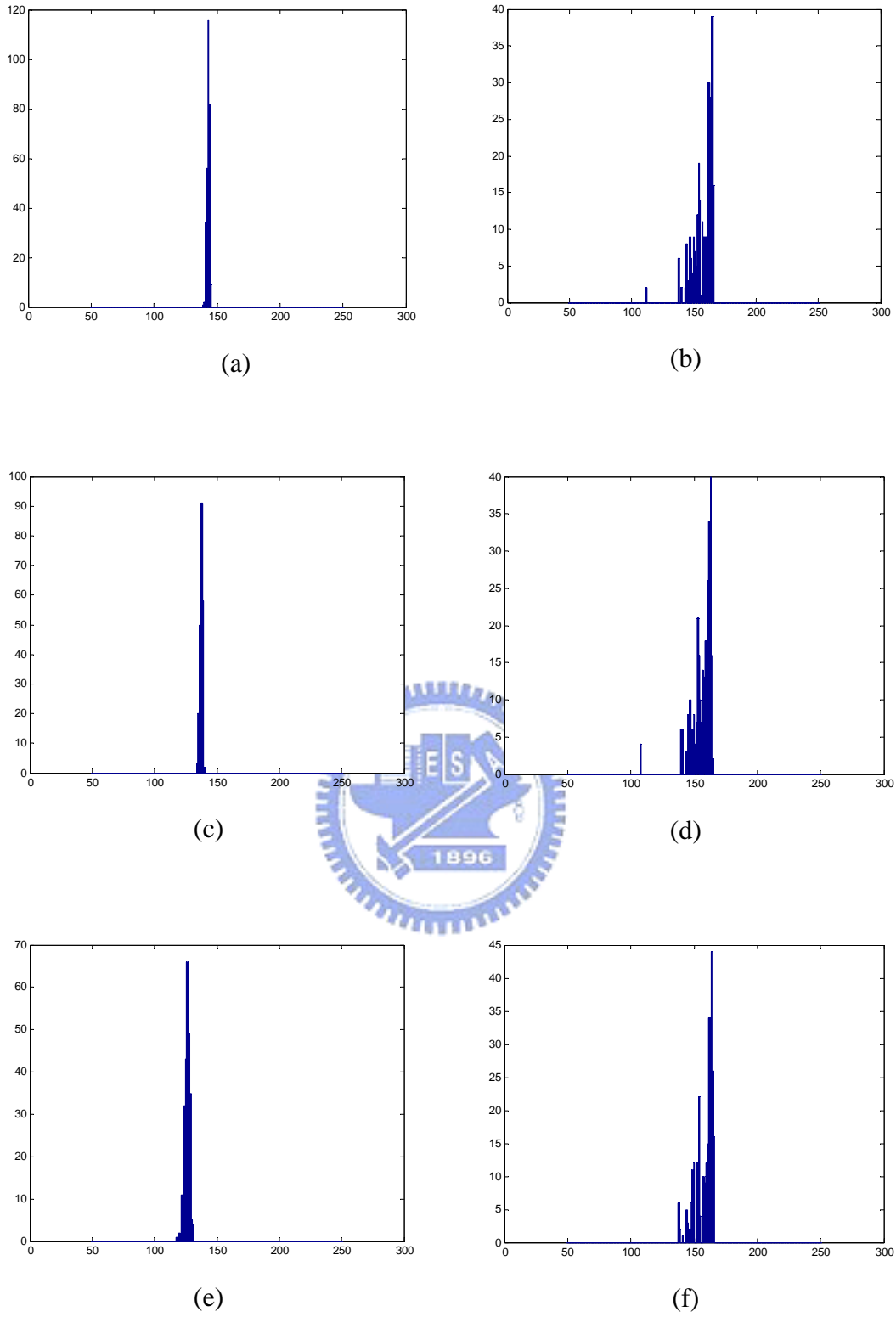


Fig. 4.9 R, G, and B variations versus frame index of background video image frame 1 to frame 300. (a) R at (10, 10), (b) R at (120, 160), (c) G at (10, 10), (d) G at (120, 160), (e) B at (10, 10), and (f) B at (120, 160).

4.2.2 Foreground Subjects Extraction Using RGB Model

The R, G and B color components are stable and reliable. So we utilized R, G, and B component to describe a background pixel by a GMMs instead of a signal model.

In the first step, the 3D space is projected onto the 2D space of R-G, G-B, and B-R to classify a pixel using the standard deviations to get the binary image $B(x, y)$ in Eq. (27) described in Sec. 3.2.3. The value k_G is chosen by experiments and varies with different trials. Hence, we ran a series of experiments to determine the optimal threshold k_V . Fig. 4.10 shows the binary image $B(x, y)$ got by different k_G with subject's clothing color different from the background. After the experiment, we set $k_G = 1.5$ in our system.



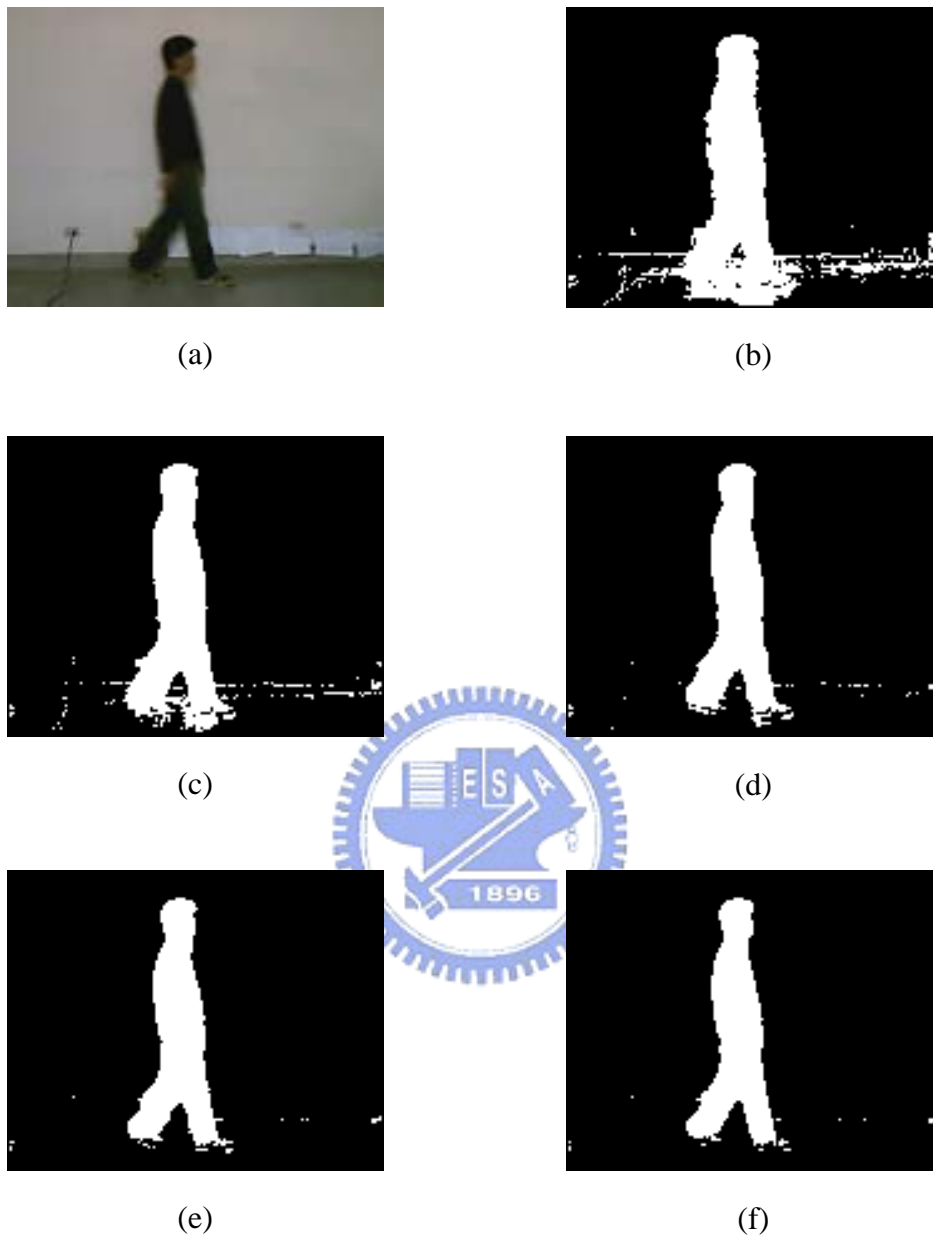
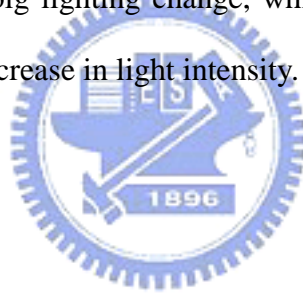


Fig. 4.10 An example of foreground extraction at different k_G thresholds. (a) An image frame with subject's clothing color different from the background, (b)–(f) foreground detected images, (b) $k_G = 1.2$, (c) $k_G = 1.3$, (d) $k_G = 1.4$, (e) $k_G = 1.5$, and (f) $k_G = 1.6$.

4.2.3 Foreground Extraction with Lighting Change Compensation

The environment with a large enough lighting change makes the foreground subjects distort and deteriorates the recognition result. We use the lighting change detection, compensation algorithm Eq. (22) described in Sec. 3.1.3, to determine the image frame pixel should be changed if these exists a large lighting change. Fig. 4.11 compares the results with/without lighting change compensation. Figs. 4.11(a), 4.11(d) and 4.11(g) are three input images. Figs. 4.11(b), 4.11(e) and 4.11(h) are the foreground subjects without lighting change compensation. Figs. 4.11(c), 4.11(f) and 4.11(i) are the foreground subjects with lighting change compensation. It is evident that foreground extraction with lighting change compensation can segment the subject correctly despite the big lighting change, while the subject is not detected, Fig. 4.11(e), for the sudden decrease in light intensity.



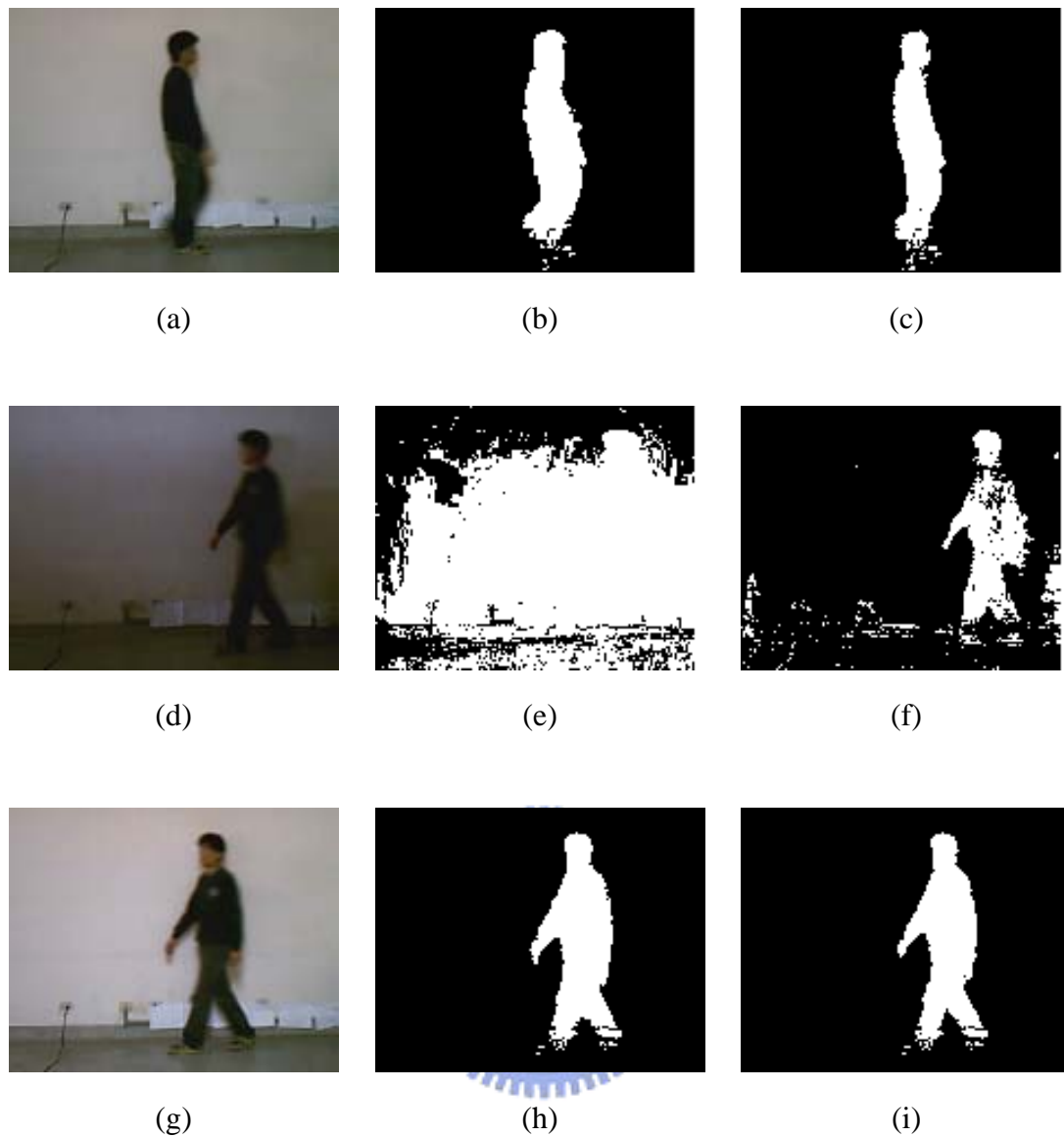


Fig. 4.11 The example concerning lighting change. (a) An image frames in which the subject “walking” with “normal” lighting, (b) a foreground subject without lighting change compensation from (a), and (c) a foreground subject with lighting change compensation from (a). (d) An image frames in which the subject “walking” with “dark” lighting, (e) a foreground subject without lighting change compensation from (d), and (f) a foreground subject with lighting change compensation from (d). (g) An image frames in which the subject “walking” with “bright” lighting, (h) a foreground subject without lighting change compensation from (g), and (i) a foreground subject with lighting change compensation from (g).

4.2.4 Foreground Subjects Extraction with Moving History-Based Background Adaptation

When a moving subject stops for a while, it should be included into the background. We use the History-Based Background Adaptation in Eqs. (28) and (29) described in Sec. 3.2.3 to classify the pixels whether they become to stop of moving subjects. Fig. 4.12 shows the process result.

Fig. 4.12 contains twelve image frames in which the subject walks “from the left to the right,” and the other one subject walks “from the left to the center,” stops “for a while” and then walks “from the center to the right.” Figs. 4.12(a), 4.12(b) and 4.12(c) are two subjects walking. Figs. 4.12(d), 4.12(e), 4.12(f), 4.12(g) and 4.12(h) are one of the subject stops for a while, and it becomes gray. Figs. 4.12(i), 4.12(j), 4.12(k) and 4.12(l) are the subject begins walking.



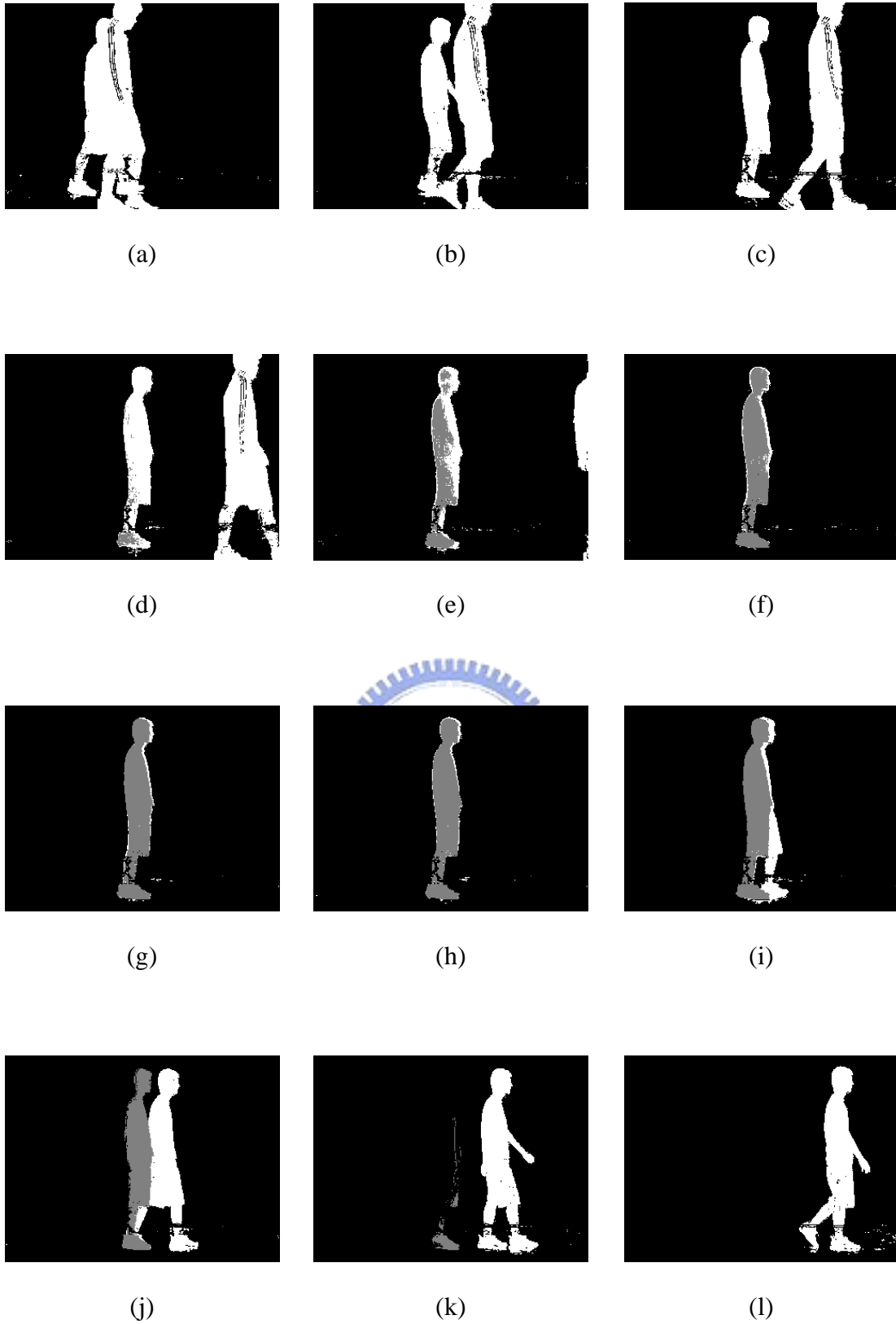


Fig. 4.12 The example of moving History-Based Background Adaptation.

4.3 The Extraction Rate of Object

We randomly selected 30 frames from the video sequence of the model with three kinds of lighting cases, and each frame with a subject wearing the clothing with the color different from the background. The “foreground subject ground truths” of these 30 frames were generated manually. Let A be a detected foreground subject region and B be the corresponding “ground truth.” Then we test the pixel accuracy rate [18] by

$$\text{Accuracy rate} = \frac{A \cap B}{A \cup B} \times 100\%, \quad (30)$$

this measure counts the percentage of the mutual positive pixels to expanded positive pixels. Table I shows the accuracy rates in HSV color space of 30 frames, and demonstrates the improvement of lighting change compensation over that without lighting change compensation. Similar to the result shown in Fig. 4.7, the segmentation accuracy is greatly improved from the “dark” case. Table II shows the accuracy rates in RGB color space of 30 frames, and demonstrates the improvement of lighting change compensation over that without lighting change compensation. Similar to the result shown in Fig. 4.11, the segmentation accuracy is greatly improved from the “dark” case.

TABLE I

COMPARISON RESULT OF THE PIXEL ACCURACY RATES OVER 30 IMAGES
IN THE HSV COLOR SPACE

Luminance information	Pixel Accuracy Rates (HSV color space)	
	Without lighting change compensation	With lighting change compensation
“Normal” lighting case	87.67%	87.86%
“Dark” lighting case	28.95%	81.09%
“Bright” lighting case	87.99%	89.28%
Average	68.20%	86.08%



TABLE II

COMPARISON RESULT OF THE PIXEL ACCURACY RATES OVER 30 IMAGES
IN THE RGB COLOR SPACE

Luminance information	Pixel Accuracy Rates (RGB color space)	
	Without lighting change compensation	With lighting change compensation
“Normal” lighting case	76.53%	77.43%
“Dark” lighting case	24.87%	70.63%
“Bright” lighting case	84.48%	84.99%
Average	61.96%	77.68%

Chapter 5 Conclusion

In this thesis, we have proposed an adaptive background modeling, either in the HSV or RGB color spaces, to improve the subject extraction. In these two color spaces, we can utilize not only the luminance component but also the chromatic component existent in the background image. The luminance change is detected to trigger the adaptation of our model to this change, and we can reliably extract the foreground subject. On the other hand, the statistics on subject moving/stopping status triggers our model's adaptation to the moving changes. Experimental results have shown that we have obtained consistently excellent results in the foreground subject extraction despite the changes encountered.

Some subjects wearing light color clothing, e.g., pink, still cannot be extracted well, which deserves to be investigated further. In addition, recognition from a different viewing direction, extensions of various test environments, more complicated surrounding, and more complicated activity are our future work.

References

- [1] I. Haritaoglu, D. Harwood, and L. S. Davis, “W⁴: real-time surveillance of people and their activities,” *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 809–830, 2000.
- [2] C. Stauffer and W. E. L. Grimson, “Adaptive background mixture models for real-time tracking,” in *IEEE Conference on Computer Vision and Pattern Recognition.*, vol. 2, June 1999.
- [3] R. Cucchiara, C. Grana, M. Piccard, and A. Prati, “Improving shadow suppression in moving object detection with HSV color information,” in *Proc. IEEE Intelligent transportation System Conference*, pp. 334–339, 2001.
- [4] N. Friedman and S. Russell, “Image segmentation in video sequences: a probabilistic approach,” in *Proc. Thirteenth Conf. Uncertainty in Artificial Intelligence*, pp.175–181, Aug. 1997,
- [5] S. Park and J. K. Aggarwal, “Segmentation and tracking of interacting human body parts under occlusion and shadowing,” in *Proc. of the Workshop on Motion and Video Computing*, pp.105–111, 2002.
- [6] M. K. Leung and Y. H. Yang, “First sight: a human-body outline labeling system,” *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 17, no. 4, pp. 359–377,1995.
- [7] S. Jabri, Z. Duric, H. Wechsler, and A. Rosenfeld, “Detection and location of people in video images using adaptive fusion of color and edge information,” in *Proc. Int. Conf. Pattern Recognition*, pp. 627–630, 2000.
- [8] T. Horprasert, D. Harwood, and L.S. Davis, “A statistical approach for real-time robust background subtraction and shadow detection,” in *Proc. IEEE*

ICCV'99, 1999.

- [9] R. Cucchiara, C. Grana, M. Piccardi and A. Prati, "Improving shadow suppression in moving object detection with HSV color information," in *Proc. IEEE Intelligent transportation System Conference*, pp. 334–339, 2001.
- [10] A. Prati, I. Mikic, M. Trivedi and R. Cucchiara, "Detecting moving shadows: algorithms and evaluation," in *Proc. IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 7, pp. 918–923, 2003.
- [11] B. Chen and Y. Lei, "Indoor and outdoor people detection and shadow suppression by exploiting HSV color information," *Fourth International Conference on Computer and Information Technology*, pp. 137–142, 2004.
- [12] S. Vitabile, G. Pilato, G. Pollaccia, and F. Sorbello, "Road signs recognition using a dynamic pixel aggregation technique in the HSV color space," in *Proc. 11th International Conference on Image Analysis and Processing*, pp. 572–577, 2002.
- [13] R. Cucchiara, M. Piccardi and A. Prati, "Detecting moving objects, ghosts, and shadows in video streams," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 10, pp. 1337–1342, 2003.
- [14] K. Ohba, Y. Sato, and K. Ikeuchi, "Appearance-based visual learning and object recognition with illumination invariance," *Machine Vision and Applications*, Vol. 12, No. 4, pp. 189–196, 2000.
- [15] S. J. Mckenna, Y. Raja, and S. Gong, "Tracking color objects using adaptive mixture models," *Image and Vision Computing* 17, pp. 225–231, 1999.
- [16] P. KaewTraKulPong, and R. Bowden, "An improved adaptive background mixture model for real-time tracking with shadow detection," in *Proc. 2nd European Workshop on Advanced Video Based Surveillance System, AVBS01. Sept 2001.*

- [17] D. S. Lee, “Effective gaussian mixture learning for video background subtraction,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.27, no.5, pp.827–832, 2005.
- [18] L. Li, W. Huang, I. Y. H. Gu, and Q. Tian, “Statistical modeling of complex backgrounds for foreground object detection,” *IEEE Transactions of Image Process*, vol.13, no.11, pp.1459–1472, 2004.

