國立臺灣大學電機資訊學院電機工程學研究所

碩士論文

Department of Electrical Engineering

College of Electrical Engineering and Computer Science

National Taiwan University

Master Thesis

兩段式背景相減法偵測移動物體

Moving Object Detection Based on Two-Staged Background Subtraction Approach

蘇恕德

Su, Shu-Te

指導教授：陳永耀 博士

Advisor: Chen, Yung-Yaw Ph.D.

中華民國 98 年 6 月

June, 2009

誌謝

兩年的研究生涯轉眼而過，這段期間由衷感謝的是指導教授陳永耀博士給予指導及諄諄教誨，不僅提供良好及完善的研究環境和充分的研究資源讓研究工作能順利推行，並以培育學生獨立思考、主動積極的態度為目標，透過一次次會議討論及給予建議逐步實現研究的成果。感謝在於學識專業上的細心指導，使研究所兩年生活，扎實的學習如何解決問題及如何進行研究的態度及方法。除此之外，更感謝口試委員交大電控所的林進燈教授、台大電機所的傅立成教授與台大機械所的顏家鈺教授，於口試期間的悉心指導與寶貴建議，使學生的論文更佳的完善。

感謝智慧型精密運動控制實驗室的成員們。感謝學長姐，聲仰、凱翔、黃璿、小傑、易道、世康，在研究過程中提供寶貴的建議和研究生活中照顧；並且感謝兩年來一起為研究打拼的同學們，鈺堂、文謙、育民、鈞凱，有你們相伴，為研究生涯寫下許多難忘的回憶，珍惜與你們同在實驗室的這兩年時光；還有感謝學弟，少旋、群富、政賢、恩慶，在研究上的討論與協助，更要特別感謝助理，郁文，不遺餘力的協助我們，使得實驗室的運作更加完善。謝謝你們於實驗期間的幫助，以及口試時的協助，由衷感謝。

最後，要感謝提供我接受良好教育的家人，在我遭遇挫折及需要鼓勵的時刻，給於深深的關懷和支持，讓我能夠順利完成論文。研究所這兩年感謝的人實在太多，無法一一列出，但你們對我的幫助，我會永遠牢記於心。最後將這份榮譽，獻給諸位幫助我的人。

Abstract

Moving object detection based on two-staged background subtraction approach is proposed in this thesis. Background subtraction in image sequence is a popular approach for detecting moving objects, especially in a relatively static scene. However, there are some problems for background subtraction approach, such as the scene with varying illumination, the new static object in the background, and the captured frame with noise that caused by environment or camera. Moreover, sometimes the results in moving object detection have false detections due to near color in corresponding pixels of moving object and background. To solve above-mentioned problems, *Two-Staged Background Subtraction* (Two-StaBaS) approach is then proposed. The first stage is background modeling and background subtraction, the suggested background is *Improved Running Average Background* (IRAB). IRAB shows its improvements on the first three above-mentioned problems. The second stage is the modification of detection errors in the first stage, namely, it solves the near color problem. Background subtraction based on the weighting map in moving object candidate region, a number of false detections in background subtraction approach decreases. Finally, two-staged background subtraction approach has the advantages of fast computational speed, low memory requirement, and good accuracy for detecting moving objects.

Keywords: moving object detection, background subtraction, background update, auto-thresholding algorithm, foreground analysis, shadow detection.

摘要

本論文提出兩段式背景相減法偵測移動物體。在靜止的場景中使用背景相減法來偵測移動物體是常用的方法，然而此方法有許些問題需要解決，例如：場景中有光線的變化、新物體加入背景當中、場景或攝影機產生雜訊...等，甚至還有移動物體與背景有相近的顏色而導致偵測錯誤的情況。為了解決上述問題，本論文提出兩段式背景相減法偵測移動物體。第一階段是利用改善式移動平均背景建立背景與背景相減，改善式移動平均背景可以改善上述前三項問題。第二階段是修正第一階段的偵測錯誤，利用移動物體候選區的權重圖重新背景相減，即可改善上述的最後一項問題。最後，兩段式背景相減法有快速的計算效率、低記憶體需求、移動物體偵測的高準確率，同時亦能解決移動物體和背景顏色相近時的偵測錯誤。

關鍵字：移動物體偵測、背景相減法、背景更新、自動取閥值演算法、前景分析、陰影偵測。

Contents

List of Figures

List of Tables

Chapter 1 Introduction

1.1 Motivation

Moving object detection means a process of detecting moving objects in image sequences. That is, it is a markerless vision-based approach to detect moving objects. There are a lot of applications for moving object detection, such as home care system, video surveillance, human posture recognition and human behavior understanding, etc. It is more and more popular study area for researchers due to vision-based moving object detection is a non-invasive approach to detect moving objects. The objective in this thesis for detecting moving objects based on vision system is high accuracy, low memory requirement, fast computational speed, and solve the problem of near gray intensities in corresponding pixels of moving object and background.

1.2 Problem Definition

The research in this thesis focuses on background subtraction approach for moving object detection. The basic requirements for background subtraction approach are high accuracy, fast computational speed, low memory requirement, background updating, and shadow removal, etc. For the basic requirements, the achievement is based on solving the problems about how to reduce illumination variation and noise, how to update backgrounds if there are new static objects in the scene, and how to reduce the false detections of near color in corresponding pixels of moving object and background.

1.3 Previous Approach

The process for moving object detection usually first select object features, then design moving object detection approach based on selected object features. The general-utilized object features are shape (silhouette and contour), edges, reconstruction model, appearance (color and texture), and motion. The feature selection depends on the features of desired-detected moving objects.

The most popular approaches for detecting moving objects are categorized into four approaches. There are point feature approach, segmentation approach, pattern classification approach, and background subtraction approach. The first three approaches are used for detecting specific feature or rigid objects; the last one is for detecting general moving objects in captured frames.

Point feature approach focuses on the interest points which have expressive texture in moving object image. The most popular methods are Moravec's detector [10], Harris detector [11], KLT detector [12], *Scale Invariant Feature Transform* (SIFT) [13], and Affine Invariant Point Detector [14].

Segmentation approach partition the captured frame into perceptually similar regions. There are two problems for segmentation approaches [16]. The first one is the criteria for good segmentation due to there may be more than one correct answer. The second aspect is that the partitioning is inherently hierarchical. The most popular approaches are mean shift method [15][57], normalized cuts method [16], and Active contours [17].

Pattern classification approach is a supervised learning mechanism. By learning different object views from a set of examples, it can be used for recognizing specific-learned objects. The well-known methods are *Adaptive Boosting* (Adaboost) [20][42][43], *Support Vector Machines* (SVM) [18], neural network [19] and so on.

Background subtraction approach is modeling the scene as the background, and then detecting the difference of captured frame and background to find moving object regions. The most popular backgrounds are *Running Average Background* (RAB) [1][2][47], *Mixture of Gaussians* (MoG) [3][52], *Kernel Density Estimation* (KDE) [6], eigenbackground [4], dynamic texture background [7], and Wallflower [5], etc.

1.4 New Approach

The new approach suggested in this thesis, *Two-Staged Background Subtraction* (Two-StaBaS) approach, for detecting moving object is based on background subtraction approach. Two-StaBaS has two stages. The first stage is moving object detection using *Improved Running Average Background* (IRAB). IRAB is proposed for background modeling. Detecting the difference of captured frame and background, the foreground region is obtained. The *Detected Moving Object* (DMO) region is the remainder region of foreground region by deleting ghost region and shadow region.

In the second stage of Two-StaBaS, the works focus on the modification of false detections in the first stage, that is, reduce the false detections in the first stage result. First select a *Moving Object Candidate* (MOC) region via the morphological dilation of DMO region. The MOC region is the region that all pixels in the region are possible belong to *Moving Object* (MO) region, namely, MOC region includes MO region. A weighting map in MOC region is constructed for representing the possibility of MO region. Finally, background subtraction in MOC region based on the weighting map, its detection errors are decreased.

IRAB has the advantages of Good accuracy (reduce the variation of illumination and noise), fast computational speed, and low memory requirement. It also updating backgrounds and removing shadows. Two-StaBaS can reduce the false detections in the moving object pixel which corresponding pixel of moving object and background has near color.

1.5 Thesis Overview

The organization of this thesis is described in the following.

Chapter 2 shows an overview of the related researches on moving object detection. It briefly introduces point feature approach, segmentation approach, pattern classification approach, and background subtraction approach. Finally, the comparisons of four approaches and popular backgrounds are discussed.

Chapter 3 expresses the moving object detection procedure. The new method, *Improved Running Average Background* (IRAB), for detecting moving objects is discussed.

Chapter 4 describes the new approach, *Two-Staged Background Subtraction* (Two-StaBaS) approach, to solve a problem in background subtraction approach. The problem is the false detections due to the corresponding pixels of moving object and background has near color.

Chapter 5 shows the results in moving object detection using Two-StaBaS approach.

Chapter 6 gives the conclusions for Two-StaBaS approach, and discusses its advantages and drawbacks. The summary of this thesis contribution is also mentioned.

Chapter 2 State of the Art

2.1 Introduction

Moving object detection is a process of detecting moving objects in captured frame or frames. The moving object is defined as human beings or any object that moves in the scene. There are a lot of approaches for detecting moving objects. The approaches can be divided into marker-based approach and markerless-based approach. The marker-based approach utilizing markers to detect moving objects. Relative to marker-based approach, the markerless-based approach is cheap and non-obtrusive [8]. In this thesis, the discussion is focus on markerless vision-based moving object detection.

Vision-based moving object detection is a highly active research area due to its potential applications and inherent complexity. Its applications are roughly divided into three titles: surveillance, control, and analysis [9]. *Surveillance* covers the automatically monitoring and understanding of moving objects, such as people counting, crowd flux, car flux on highway, shopping behavior understanding, abnormal activities detection and person identification, etc. *Control* means the estimated motion or object shape parameter are used to control something, such as *Human-Computer Interface* (HCI). *Analysis* automatically diagnoses something, such as gait analysis of patient, the performance of athletes, car industry, sleeping detection, pedestrian detection and lane following.

The potential applications of moving object detection entice researchers study this area. The main approaches for detecting moving object can be categorized into four approaches; there are point feature approach, segmentation approach, pattern classification approach, and background subtraction approach. The suitable features for detecting moving object depend on the feature of moving object or selected detection approach. In this chapter, the first issue is feature selection, and then discusses detection approaches.

Table 1 Object detection categories and representative work [44]

Categories	Representative Work
Point Feature Approach	Moravec's detector [10] Harris detector [11] KLT detector [12] Scale Invariant Feature Transform [13] Affine Invariant Point Detector [14]
Segmentation Approach	Mean shift method [15][57] Normalized cuts method [16] Active contours [17]
Pattern Classification Approach	Support Vector Machines [18] Neural Networks [19] Adaptive Boosting [20]
Background Subtraction Approach	Running Average Background [1][2] [47] Mixture of Gaussians [3][52] Eigenbackground [4] Wall flower [5] Kernel Density Estimation [6] Dynamic texture background [7]

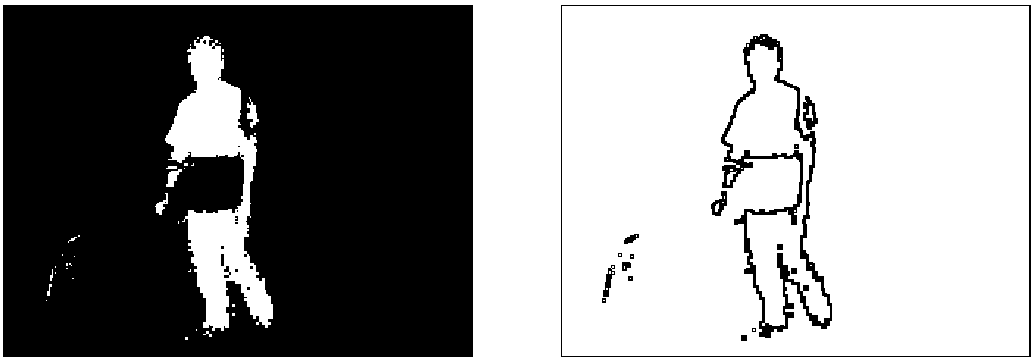
2.2 Feature Selection

Feature selection plays an important role in object detection. Every object has different feature, a correct feature selection helps accurate object detection. The general object features are categorized into shape (silhouette and contours), edges, reconstruction model, appearance (color and texture), and motion. The feature can be used for transforming image space to feature space to detect objects.

I. Shape: silhouette and contours

Silhouette is a drawing consisting of the outline of object with solid color, shown in Figure 1(a). Contour is the outline of silhouette, illustrated in Figure 1(b). The two features are relatively robust from images and insensitive to variation in surface such as color and texture. However, its performance is limited due to artifacts (such as object shadows and near color of moving object and background). It is also lack of depth information.

There are many methods to transform shape from image space to feature space, for example, it can be encoded using central moments [21] and Hu moments [22], or encoded using a combination of turning angle metric and Chamfer distance [23] or shape contexts [24]. It compared base on deformation cost [25].



(a) (b)

Figure 1 (a) The silhouette of a human subject (b) The contour of a human subject

II. Edge

Edge is a set of discontinuities in gray level of image, shown in Figure 2. It can be extracted robustly (invariant to lighting condition) and at low cost, but unsuitable when dealing with cluttered scenes or textured objects. It also lack of depth information. It can be utilized to extract silhouette [26][27][29] or object model [28][29].

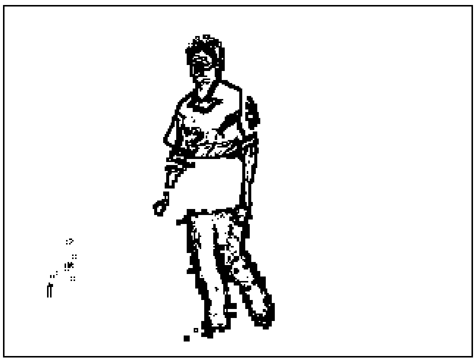


Figure 2 The edge of a human subject

III. Reconstructions Model

The depth information is known, the self-occlusions of objects can be detected. However, multiple cameras are required. The applications: 3D reconstruction can be create from silhouettes extracted in each view individually: volume intersection [30], 3D reconstruction can be create from silhouettes extracted in each view individually, vowel-based approach [31][32], stereometry triangulation [33][34], and stereometry optional aid of projected light patterns [35].

IV. Appearance (Color and Texture)

Appearance means the color or texture on the moving object, shown in Figure 3. If the appearance of moving object substantially unchanged, modeling the moving object based on color or texture is suitable. However, changing appearance due to clothing, illumination, and rotations makes bad features for extraction. Appearance is a popular utilized in skin color because it is a good cue for finding head and hands. The color can be modeled by Gaussian color distribution [2], color histogram [36], 3D appearance model [37], minimize the sum of pixel-wise intensity differences between image and synthesized model [38], and Sleeve, hem and sock lengths modeling by clothing parameters [39].

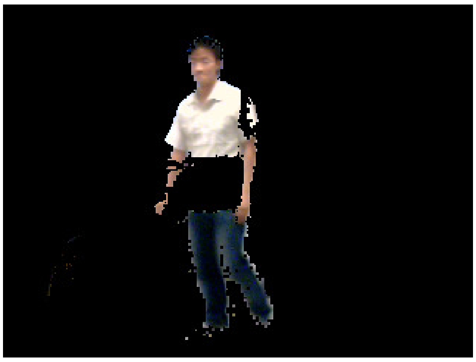


Figure 3 The appearance of a human subject

V. Motion

Motion can be measured by taking the difference between two consecutive frames. The brightness of the pixels that are part of the moving object in the image is assumed to be constant. The pixel displacement in the image is termed optical flow [40]. Namely, optical flow is the distribution of apparent velocities of movement of brightness patterns in an image [41], illustrated in Figure 4.

It is very common for assessing motion from a set of images. It is also an approximation of the two-dimensional flow field from image intensities.

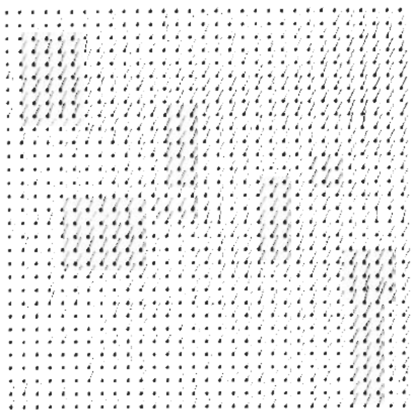


Figure 4 The optical flow is the distribution of apparent velocities of movement of brightness patterns in an image [41]

VI. Comparison of five features

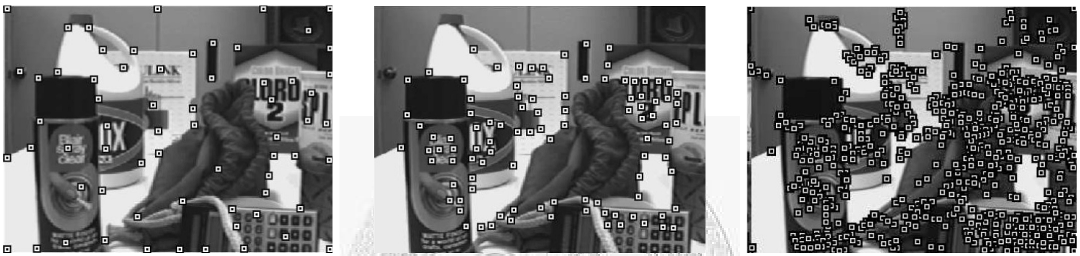
The five general-used image features are compared in the following table. Shape and edge are robust features for detecting moving objects. However, their performance is limited due to artifacts, cluttered backgrounds, textured clothing, and lake of 3D information. Shape, edge, appearance and motion are lack of depth information. Hence, the 3D information is applied in reconstruction model due to its requirement of multiple cameras. It is a trade of between 3D information and complexity of multiple cameras. It is helpful for good result in moving object detection if choosing appropriate image features.

Table 2 The comparison of five general-used image features

Feature	Pros	Cons
Shape	● robust feature	● performance is limited due to artifacts ● lack of depth information
Edge	● robust feature ● low cost	● deal with cluttered backgrounds ● deal with textured clothing ● lack of depth information
reconstruction model	● depth information is known ● solve self-occlusion problem	● multiple cameras
Appearance	● easy for extraction	● changing appearance
Motion	● moving vector is known	● long computational time

2.3 Point Feature Approach

The point feature approach is to find interest points in the moving object, and then detecting the moving objects by these interest points. The most popular methods are Harris detector [11], KLT detector [12], *Scale Invariant Feature Transform* (SIFT) [13], and Affine Invariant Point Detector [14]. The first three methods are illustrated in Figure 5.



(a) (b) (c)

Figure 5 Interest points which detected by applying (a) Harris detector, (b) KLT detector and (c) SIFT[44]

I. Harris Detector

Harris detector is a detector that detecting corners and edges. The corner detection method is based on the revision of the Moravec detector [10]. Computing the first order image derivatives in x and y directions to highlight the directional intensity variation. The matrix M is obtained:

$$M = \begin{bmatrix} A & C \\ C & B \end{bmatrix} = \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix} \qquad \text{Eq. (1)}$$

where $I_x \equiv \dfrac{\partial I}{\partial x} = I \otimes \begin{bmatrix} -1 & 0 & 1 \end{bmatrix}$, and $I_y \equiv \dfrac{\partial I}{\partial y} = I \otimes \begin{bmatrix} -1 & 0 & 1 \end{bmatrix}^T$. I is captured frame.

The trace of matrix M is $Tr(M) = \alpha + \beta = A + B$, the determinant of matrix M is $Det(M) = \alpha\beta = AB - C^2$, where α and β are the eigenvalues of M. Set the corner response R is equals $R = Det(M) - kTr(M)$, where k is a tunable value usually in the range from 0 to 0.25.

In the Figure 6 shows the auto-correlation principle curvature space. Base on the figure, thresholding the corner response R to get the corner and edge points. Namely, corner is the point which $R > R_{threshold}$, where $R_{threshold}$ is close to zero and fixed. The Harris detector method in [11], the capture frames shown in Figure 7, and the Harris detection results are illustrated in Figure 8.

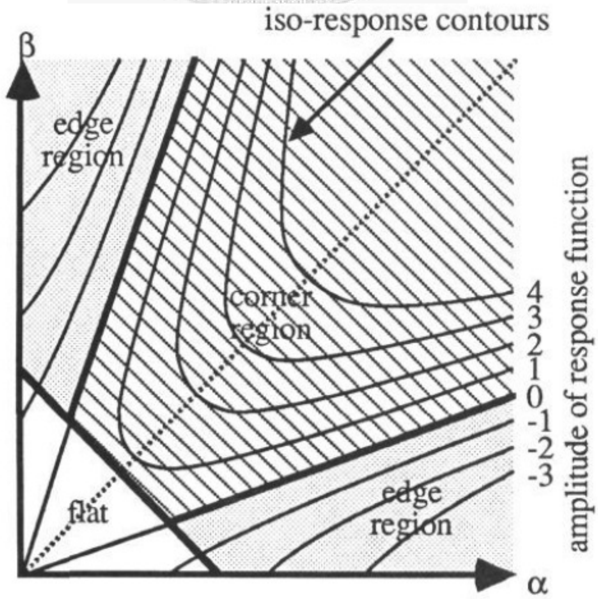


Figure 6 The auto-correlation principle curvature space - heavy lines give corner / edge / flat classification. Fine lines are equi-response contours [11]

Figure 7 Pair of images from an captured outdoor sequence [11]

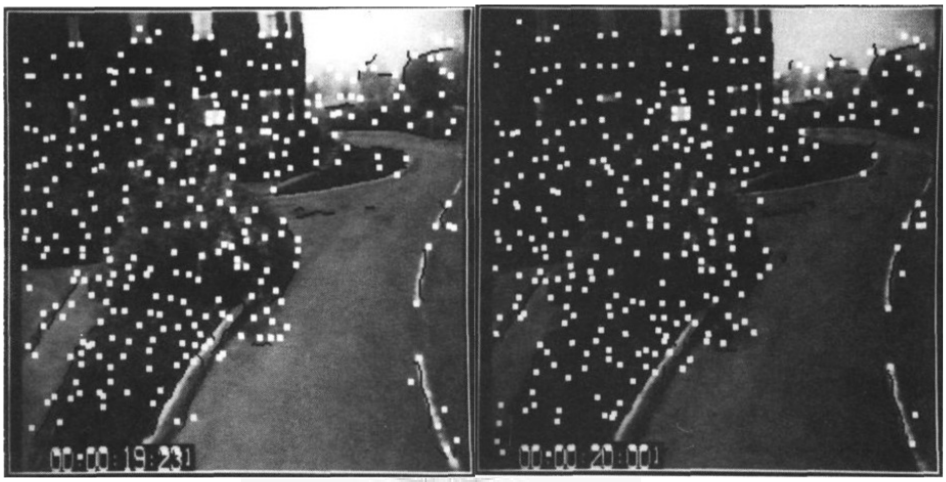


Figure 8 Completed Harris detector for the outdoor images [11]

II. KLT detector

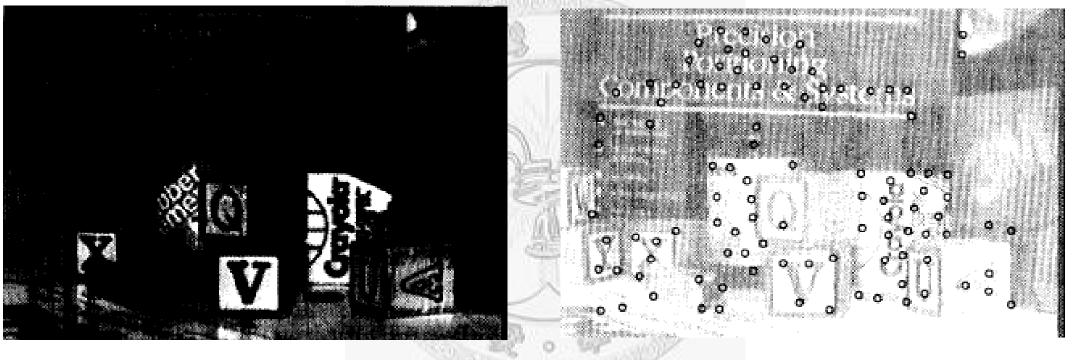
The KLT detector is similar to Harris detector. The same matrix M is given in Eq. (1). The eigenvalues of matrix M are λ_1 and λ_2, where

$$\lambda_1 = \frac{A+B}{2} + \frac{1}{2}\sqrt{(A-B)^2 + 4C^2} \qquad \text{Eq. (2)}$$

and

$$\lambda_2 = \frac{A+B}{2} - \frac{1}{2}\sqrt{(A-B)^2 + 4C^2} \qquad \text{Eq. (3)}$$

For each image pixel, if the eigenvalue $\lambda_2 = \min(\lambda_1, \lambda_2) \geq \lambda_{threshold}$, where $\lambda_{threshold}$ is constant, then save the pixel into a list, L. L is the corner candidate list. In order to find the interest point, corner point, first sorting list L in decreasing order. Then delete closing candidate in the local neighborhood. The neighborhood is tunable local window set by manual. The remainder is the interest points.



(a) (b)

Figure 9 (a) The captured frame (b) The feature selected according to KLT [12]

III. Scale Invariant Feature Transform

In theory, the matrix M (in Eq. (1)) is invariant to both rotation and translation. However, it is not invariant to affine transformations. To solve this problem, Lowe [13] introduced SIFT. This method includes four stages: detection of scale-space extreme, accurate key point localization, orientation assignment, and the local image descriptor.

1. Detection of scale-space extreme

The first stage of computation searches over all scales and image locations. It is implemented efficiently by using a difference-of-Gaussian function to identify potential interest points that are invariant to scale and orientation.

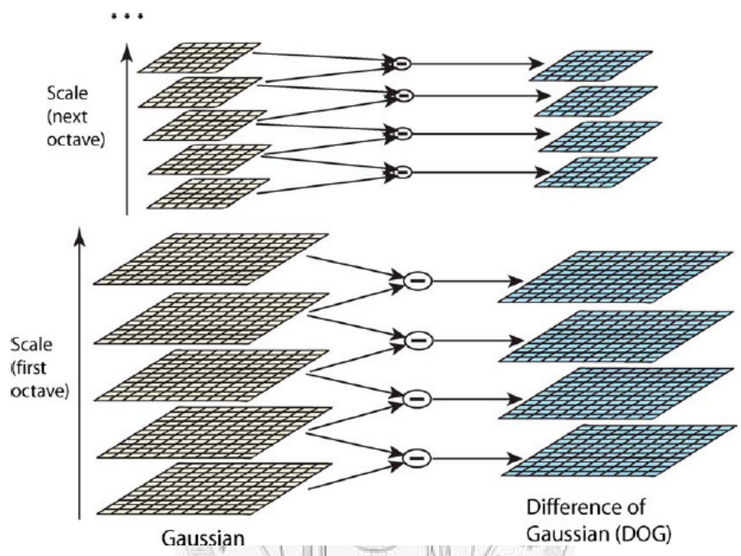


Figure 10 For each octave of scale space, the initial image repeatedly convolved with Gaussians to produce the set of scale space images shown on the left. Adjacent Gaussian images are subtracted to produce the difference-of-Gaussian images on the right. After each octave, the Gaussian image is down-sampled by a factor of 2, and the process repeated. [13]

2. Accurate key point localization

At each candidate location, a detailed model is fit to determine location and scale. Keypoints are selected based on measures of their stability.

3. Orientation assignment

One or more orientations are assigned to each keypoint location based on local image gradient directions. All future operations are performed on image data that has been transformed relative to the assigned orientation, scale, and location for each feature, thereby providing invariance to these transformations.

4. The local image descriptor

The local image gradients are measured at the selected scale in the region around each keypoint. These are transformed into a representation that allows for significant levels of local shape distortion and change in illumination.

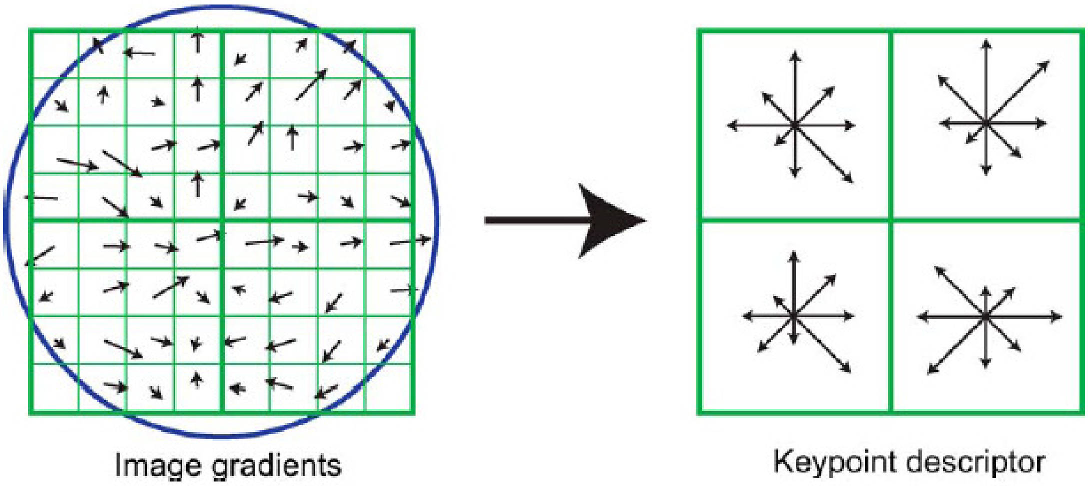


Figure 11 A keypoint descriptor is created by first computing the gradient magnitude and orientation at each image sample point in a region around the keypoint location, as shown on the left. These are weighted by a Gaussian window, indicated by the overlaid circle. These samples are then accumulated into orientation histograms summarizing the contents over 4x4 subregions, as shown on the right, with the length of each arrow corresponding to the sum of the gradient magnitudes near that direction within the region. This figure shows a 2 × 2 descriptor array computed from an 8 × 8 set of samples, whereas the experiments in this paper use 4 × 4 descriptors computed from a 16 × 16 sample array.

IV. Comparison of three methods

Harris detector and KLT detector are closely related and based on the local structure matrix M. Harris detector is frequently used by Europe, and KLT is by U.S. They are invariant with rotational and translation. SIFT is not only invariant with rotation, scaling and translation, but also invariant to illumination variation and occlusions. Many features can be generated for even small objects. However, the SIFT method needs long time for computation. The comparison of three methods is shown in Table 3.

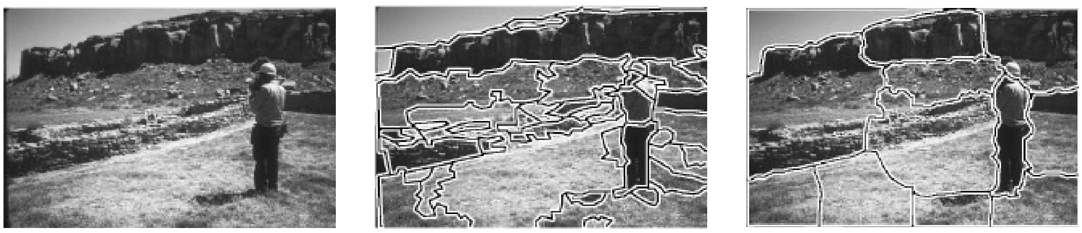
Table 3 The comparison of three most popular feature point detection methods

categories	pros	cons
Harris detector KLT detector	High computational speed Invariant to rotation Invariant to translation	Small object may no feature
SIFT	Many feature is extracted	Low computational speed

2.4 Segmentation Approach

The objective of segmentation approach is to partition the image frame into perceptually similar regions. Every segmentation algorithm has to solve two main problems [16], described in the following. In this section, two relevant methods, the mean shift method [15][57] and the normalized cut method [16], are introduced.

1. What is the criterion that one wants to optimize?
2. Is there an efficient algorithm for carrying out the optimization?



(a) (b) (c)

Figure 12 (a) Original captured image frame for segmentation (b) Segmentation using mean shift method (c) Segmentation based on the normalized cut method [44]

I. The mean shift method [15][57]

The mean shift method has two parameters controlling the resolution in the spatial and range domains. Its aim is to cluster the same property regions in the frame by tuning the two parameters. The steps of algorithm are shown as follows:

Step 1. Initialization: give a large number of hypothesized cluster centers that randomly chosen from the data in the image frame.

Step 2. Find mean shift vector: the mean shift vector is the vector from cluster center to the mean of data in its cluster.

Step 3. Iteration: the mean shift vector is computed iteratively until the cluster centers are converged.

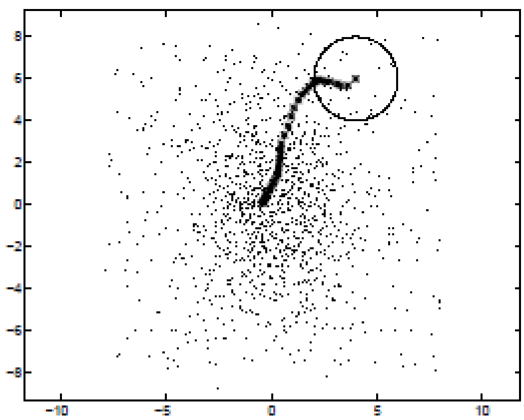


Figure 13 Successive computations of the mean shift define a path leading to a local density maximum [15]

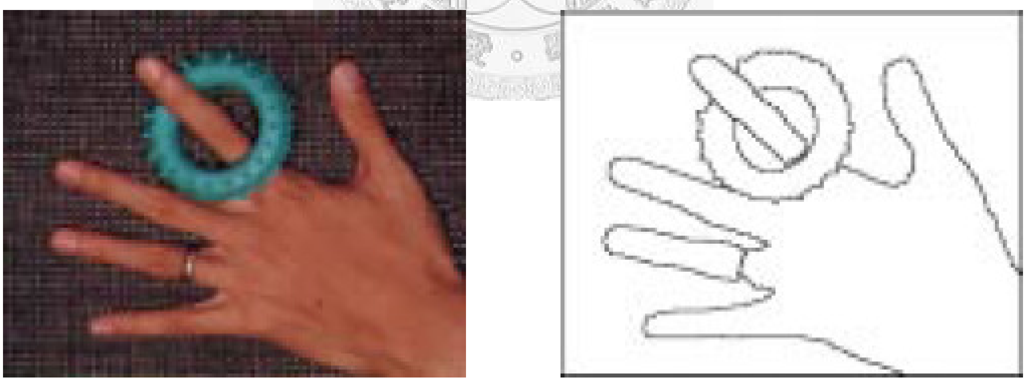
Step 4. Merge: if the clusters of convergence points are closer than a predefined parameter, then merge the clusters.

Step 5. Optimal: eliminate spatial regions smaller than M regions, where M is predefined parameters.

The mean shift method requires fine tuning of several parameters to obtain better segmentation.

(a) (b)

Figure 14 (a) Original captured frame (b) The mean shift segmented method with variable $(\sigma_s, \sigma_r, M) = (16, 7, 40)$ [15]



(a) (b)

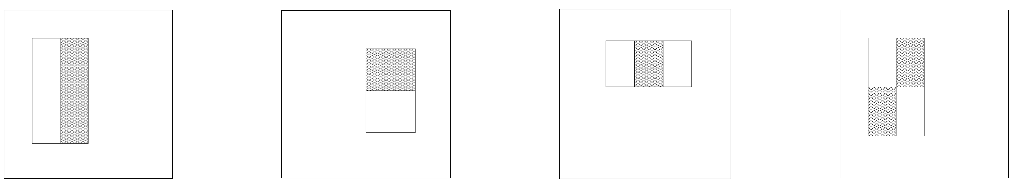
Figure 15 (a) Original captured frame (b) The mean shift segmented method with variable $(\sigma_s, \sigma_r, M) = (16, 19, 40)$ [15]

II. The normalized cut method [16]

For every image frame, G, can be represented as a set of pixels, V, and divided into N disjoint regions, A_i, where $\bigcup_{i=1}^{N} A_i = V$ and $A_i \cap A_j = \phi$. The image frame is divided into N disjoint regions by pruning the weighted edge of the image. The total weight of the pruned between two regions is called a cut. The weight of normalized cut method is computed by the product of the color similarity and the special proximity for segmentation. The each new segment, the process is recursively performed until a threshold is reached.

2.5 Pattern Classification Approach

Pattern classification approach classifies objects by means of learning the features of moving object. Give a large number of examples for the learning mechanism, the supervised learning classifier maps the input to the desired output. The most popular methods are *Adaptive Boosting* (Adaboost) [20][42][43], *Support Vector Machines* (SVM) [18], and neural network [19]. We only introduce Adaboost method here.



(a) (b) (c) (d)

Figure 16 Example of base classifier [42]

Figure 17 The selected base classifier for human eye recognition [42]

Adaboost method is an iterative method of finding a very accurate classifier by combining many base classifiers, shown in Figure 16. For example, its application can be used in face detection, and its base classifiers can be which illustrated in Figure 17. By collecting a huge number of examples being training examples, such as many human face, shown in Figure 18, and non-human face examples. In the training phase of Adaboost, the learning mechanism constructs the distribution of weight for base classifier according to the proportional of accuracy, illustrated in Figure 19. Then the classifier can be utilized for mapping new input image into accurate output, human and non-human face, illustrated in Figure 20. Its detailed introduction and applications can be found in [60][61][62].

Figure 18 Collecting a huge number of examples being training examples for Adaboost [42]

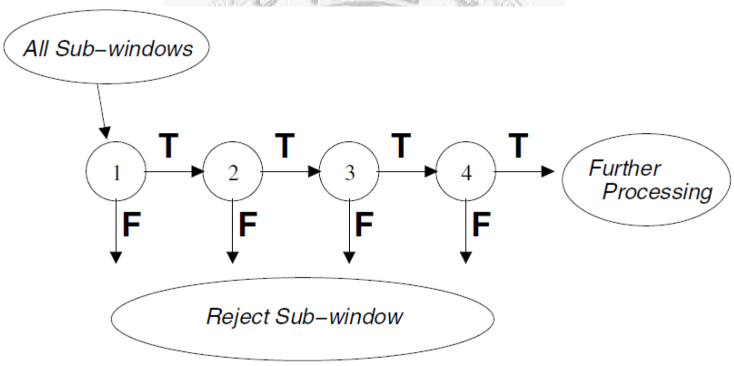
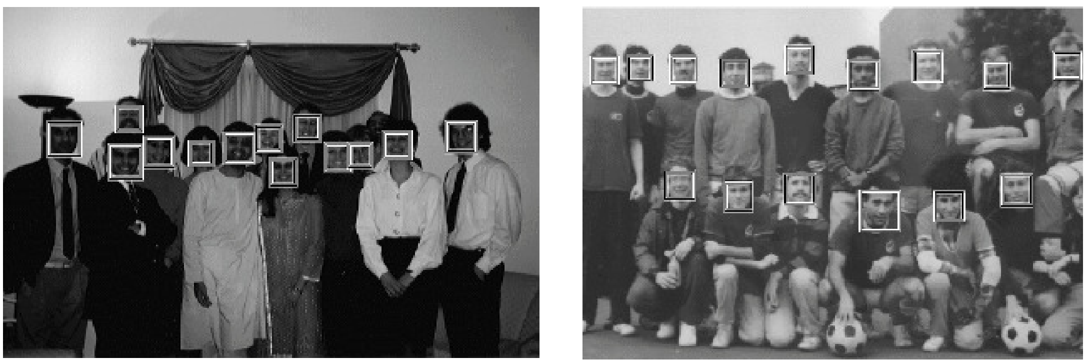


Figure 19 In the training phase of Adaboost, the learning mechanism constructs the distribution of weight for base classifier according to the proportional of accuracy [42]



(a) (b)

Figure 20 Adaboost classifier can be utilized for mapping new input image into accurate output [42]

2.6 Background Subtraction Approach

The general process of background subtraction approach is first constructing background by captured frame or frames, then detecting the difference of captured frame and background. The different region is defined as foreground region. The moving object is assumed in the foreground region.

2.6.1 Popular Backgrounds

The most popular backgrounds are collected in a recent review paper [46]. They are *Running Average Background* (RAB) [1][2][47], *Mixture of Gaussians* (MoG) [3][52], Eigenbackground [4], Wall flower [5], Kernel Density Estimation [6], and Dynamic texture background [7].

I. Running Average Background

Running Average Background (RAB), also called single Gaussian background, models the captured scene by a single Gaussian distribution in every pixel. The whole frame is represented by the mean and variance, (μ_k, σ_k). For the purpose of background updating, the adaptive filter is then given.

$$\mu_{k+1} = \alpha F_k + (1-\alpha)\mu_k \qquad \text{Eq. (4)}$$

$$\sigma_{k+1}^2 = \alpha(F_k - \mu_k)^2 + (1-\alpha)\sigma_k^2 \qquad \text{Eq. (5)}$$

where F_k is captured frame in k-th frame.

and its threshold method depends on Eq. (6).

$$|F_k - \mu_k| > threshold \qquad \text{Eq. (6)}$$

The threshold can be chosen as $n\sigma_k$, where n is a constant.

Figure 21 Detecting single human subject using IRAB [2]

II. Mixture of Gaussians

Due to the captured frame includes many components, such as background, moving objects, moving object shadow. In this method, every pixel in the frame is modeled with about from 3 to 5 mixture Gaussians distributions $(\mu_k, \sigma_k, \omega_k)$, where $k = 1 \sim 5$ is the Gaussian distribution number, μ_k, σ_k and ω_k are the mean, variance and weighting, respectively. The *Mixture of Gaussians* (MoG) exactly models both the foreground and the background. Pick up the background by ranking according to their $\frac{\omega_k}{\sigma_k}$ value. The first one is chosen as the background distribution. The others are the foreground, includes moving objects and moving object shadows, etc.

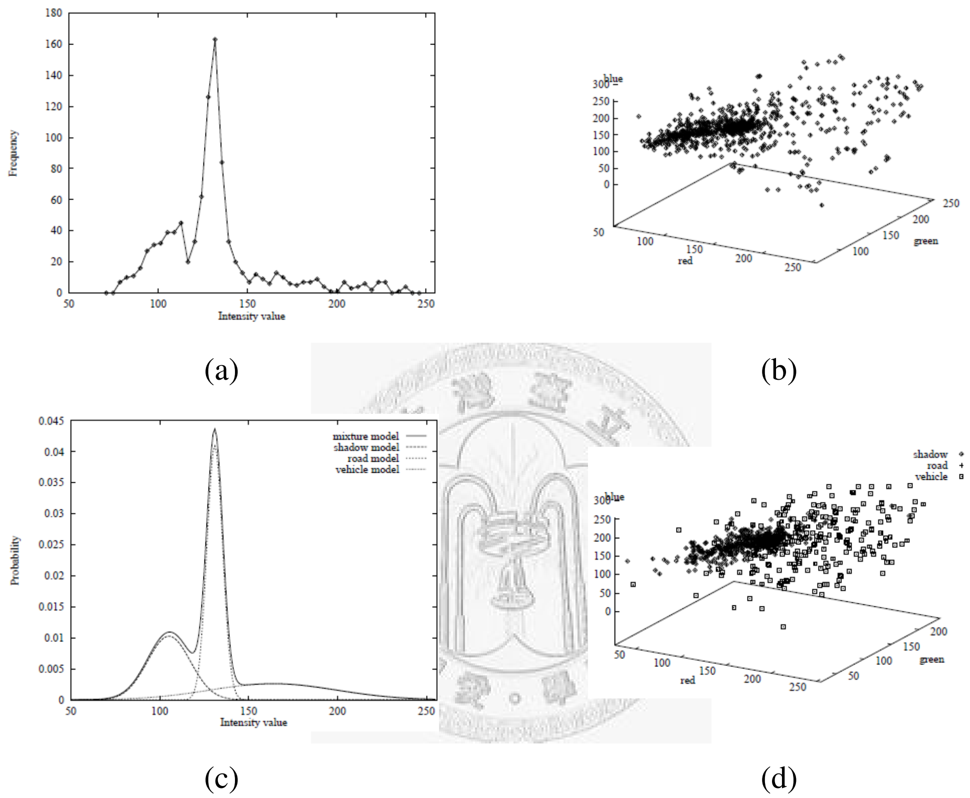
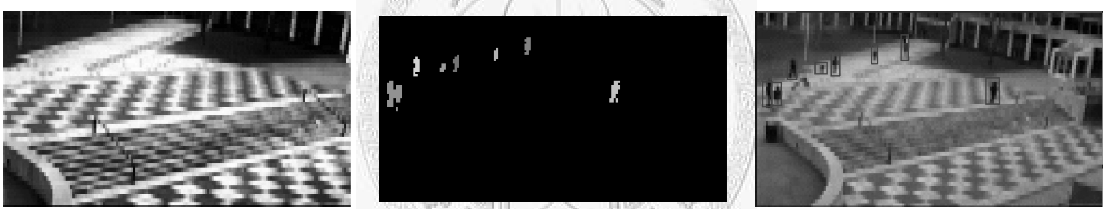


Figure 22 (a) Empirical distribution of intensity values for pixel point (160, 170) over 1000 frames. (b) Scatter plot of RGB values for the same pixel. (c) Fitted three component mixture of Gaussians for the data in (a). (d) Scatter plot of 1000 randomly generated data points from a little three component mixture of Gaussians for the data in (b). [48]

III. Eigenbackgrounds

Eigenbackgrounds maps the image space to the eigen space using *Principal Component Analysis* (PCA)[49]. That is, PCA by way of eigenvector decomposition is a method to reduce the dimensionality of image space.

For the frame resolution is $m \times n$. Collect n frames for the sample of background, and rearrange the n frames as the columns of a matrix, S, where $S \in R^{mn \times N}$. The mean matrix and covariance matrix are μ_b and $C_b = SS^T$. From the covariance matrix, C_b, the diagonal matrix of its eigenvalues, L_b, and the eigenvector matrix, Φ_b, are obtained. On the other hand, $L_b = \Phi_b C_b \Phi_b^T$. Select the m eigenvectors are kept, corresponding to the m largest eigenvalues to give a Φ_{M_b} matrix. Then the eigenbackground is $B_k = \Phi_{M_b}^T \left(I_k - \mu_b \right)$, where I_k is current captured frame. By thresholding $D_k = \left| I_k - B_k \right|$, the foreground region is then determined.



(a) (b) (c)

Figure 23 (a) The mean matrix for the sample images (background mean image) (b) Moving object detection using eigenbackgrounds (c) Captured frame with blob bounding boxes [4]

IV. Comparison of backgrounds

The objective for background modeling is to real-time detecting foreground (in the condition of fixed camera). However, the background is not fixed. It requires adapting to illumination variation, noise reduction and new fixed object in the scene. It also needs fast computational speed, low memory requirement and high accuracy. The table shown in the Table 4 describes the pros and cons of most popular three backgrounds.

Table 4 The comparison of three most popular backgrounds

background	RAB	MoG	eigenbackgrounds
illumination variation reduction	yes	yes	no
online background update	yes	yes	no
memory requirement	low	high	high
computational speed	high	middle	middle
accuracy	acceptable	good	good

2.6.2 Subtraction Operator

The well-known subtraction operator [4], $D_k = |I_k - B_k|$, is the absolute value of the pixel value in the frame I_k subtracts the corresponding pixel of B_k. This subtraction operator is linear.

For some applications, the suitable subtraction operator is nonlinear form. For example, if the background is almost dark or white, the background subtraction operator is then suggested as Eq. (7) [50][51].

$$d = 1 - \frac{\sqrt{(i+1)(b+1)}}{\frac{(i+1)+(b+1)}{2}} \frac{\sqrt{(256-i)(256-b)}}{\frac{(256-i)+(256-b)}{2}} \qquad \text{Eq. (7)}$$

where i is the pixel value in the captured frame, and b is the corresponding pixel in the background. The background subtraction result is d. For the small or large b, the subtraction operator in Eq. (7) is sensitive to i.

2.6.3 Thresholding Algorithm

For the sake of detecting the foreground region in the difference frame, thresholding the difference frame is then introduced. If the difference of captured frame and background is "large", the region is assumed to be foreground region. On the contrary, the region with "small" difference is assumed to be background region. The process of selecting the threshold to define "large" and "small" is thresholding.

The well-known methods for threshold selection can be categorized into deterministic methods and probabilistic methods. Deterministic methods select threshold by algorithm without non-parametric assumption, usually choosing the appropriate algorithm according to the histogram of the difference frame. Probabilistic methods make parametric assumptions about foreground region and background region distributions and then derive optimal thresholds.

Deterministic methods include P-tile method, peak and valley method, and triangle algorithm, etc. Probabilistic methods include Kullback information distance and other methods. The introduction for these methods [45] is in the following.

I. P-tile method

First sketch the histogram (probability density function, pdf) of the difference frame, number of pixels versus difference of intensity, shown in Figure 24. And its cumulative density function (cdf) is illustrated in Figure 25.

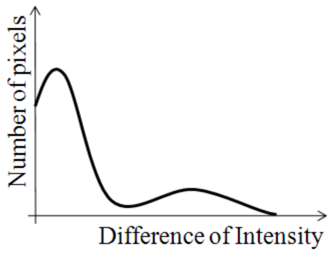


Figure 24 The histogram (pdf) of the difference frame

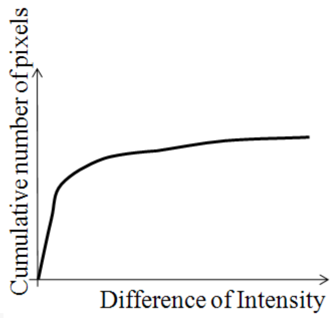


Figure 25 The cdf of the difference frame

In some situation, the approximately percentage, p, of foreground region in the captured frame is known. That is, the "large" difference part is foreground region; the "small" difference part is background region. By setting the threshold using cdf, the threshold divides 1-p% for background region, p% is foreground region.

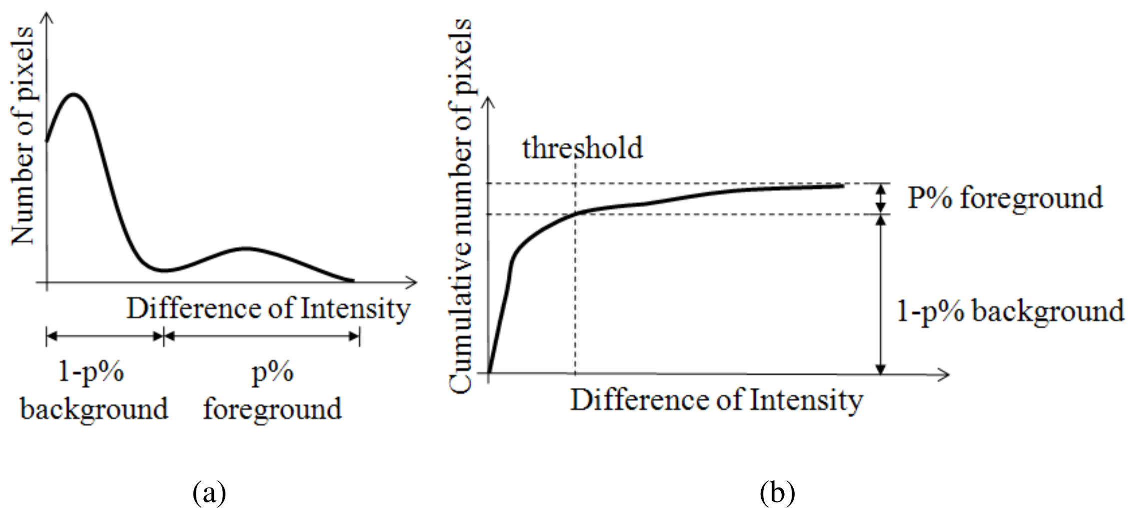


Figure 26 Determine the threshold using p-tile method (a) the area of foreground part occupies p% in the whole frame (b) determine threshold in 1- p% of cdf

II. Peak and valley method

Based on the histogram shown in Figure 27, first find the two prominent peaks in the figure. The two highest peaks are g_1 and g_2, respectively, where $g_1 < g_2$. Namely, $pdf(g_1)$ and $pdf(g_2)$ are local maxima.

$$pdf(g_1) > pdf(g_1 \pm \Delta g) \quad \text{Eq. (8)}$$

$$pdf(g_2) > pdf(g_2 \pm \Delta g) \quad \text{Eq. (9)}$$

Find the deepest valley, g, between g_1 and g_2.

$$pdf(g) \le pdf(g') \quad \text{Eq. (10)}$$

, where $g, g' \in [g_1, g_2]$

Then assign

$$threshold = g \quad \text{Eq. (11)}$$

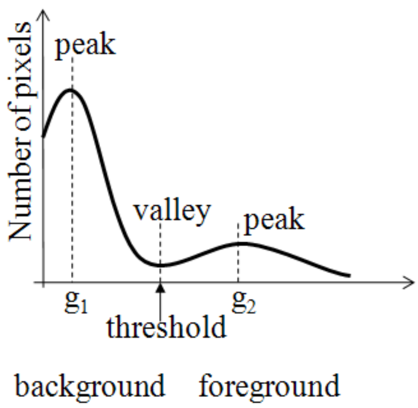


Figure 27 Determine the threshold using peak and valley method, the deepest valley between two highest peaks is assigned to be the threshold

III. Triangle algorithm

Based on the histogram shown in Figure 28, first connect a line between global maxima, g_{max}, and global minima, g_{min}. Then find the maximal distance between the line and the histogram, D_{max}. Then the threshold is determined as its pdf value, $pdf\left(threshold\right)$, to the line is D_{max}.

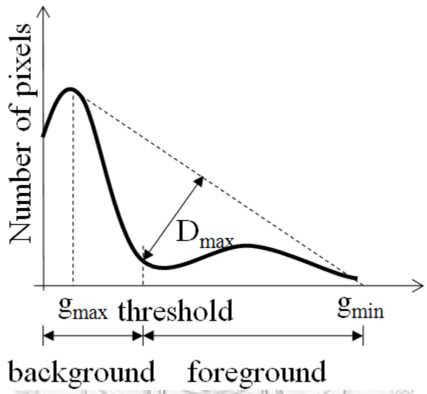


Figure 28 Determine the threshold using triangle algorithm, the threshold makes the maximal distance between the histogram to the line which connects g_{max} and g_{min}

IV. Kullback information distance

The observed histogram, Figure 29, is mixture of foreground part and background part. In other words, the components of histogram, background and foreground, can be assumed to be mixture of Gaussians, $pdf_b\left(i\right)$ and $pdf_f\left(i\right)$, to form parametric model.

$$pdf_b(i) = \frac{pdf(\mu_b)}{\sqrt{2\pi}\sigma_b} e^{-\frac{1}{2}\left(\frac{i-\mu_b}{\sigma_b}\right)^2} \qquad \text{Eq. (12)}$$

$$pdf_f(i) = \frac{pdf(\mu_f)}{\sqrt{2\pi}\sigma_f} e^{-\frac{1}{2}\left(\frac{i-\mu_f}{\sigma_f}\right)^2} \qquad \text{Eq. (13)}$$

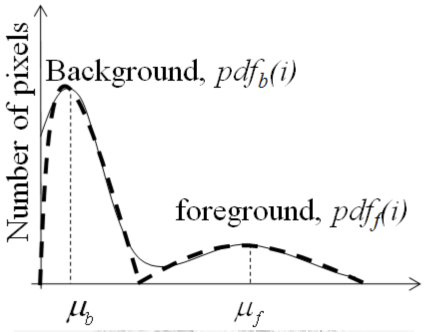


Figure 29 Determine the threshold using Kullback information distance with constructing mixture of Gaussians

2.6.4 Foreground Analysis and Shadow Removal

I. Foreground analysis

In the literature [53], the captured frame can be categorized into background region and foreground region. Foreground region can be divided into object part and shadow part. Object includes moving object region and ghost region. Similarly, shadow involves moving object shadow and ghost shadow. The taxonomy is illustrated in Figure 30.

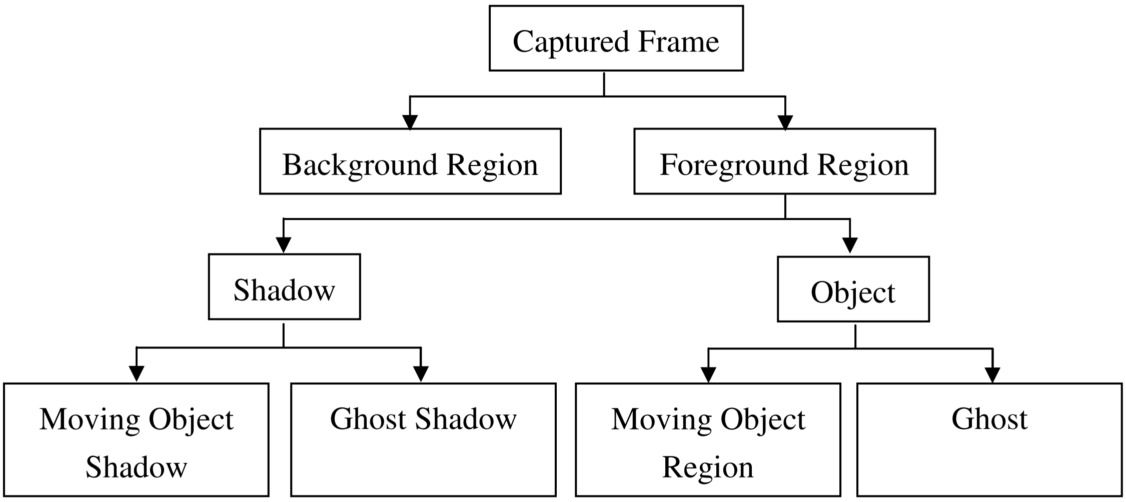


Figure 30 The taxonomy in the captured frames

II. Shadow removal in color space [53][58][59]

In order to detect the object region, shadow removal is required. In color space, for example, HSV (Hue, Saturation, and Value) space, the shadow is darker than the background in intensity. However, the chromatic information is basically invariant. That is, the shadow is invariant in chromatic information but lower in illumination information. Based on above-mentioned specification, the shadow detection in color space is presented as following three conditions.

(1) $D_k^H \le \tau_H$: Hue is one of chromatic information which is stationary, the hue difference of captured frame and background is $D_k^H = \min\left[\left|C_k^H \Theta B_k^H\right|, 360 - \left|C_k^H \Theta B_k^H\right|\right]$ should smaller than a constant τ_H, where Θ means subtraction operator.

(2) $D_k^S < \tau_S$: Saturation is also one of chromatic information which is stationary. The saturation difference of captured frame and background is $D_k^S = \left|C_k^S \Theta B_k^S\right|$ should smaller than a constant τ_S.

(3) $\tau_\alpha \le R_k^V = \dfrac{I_k^V}{B_k^V} \le \tau_\beta$: Value, also called intensity, which is illumination information.

Due to the intensity of shadow is darker than the background, the ratio of R_k^V is in the interval of $\left[\tau_\alpha, \tau_\beta\right]$, where τ_α and τ_β are in the interval of $[0,1]$.

III. Shadow removal in gray image

Here introduces the shadow cancellation operator to get the accurate silhouette. The morphological operator is used for removing the cast shadow of moving object. The operator for shadow removal in gray image and the characteristic are discussed in the following paragraph.

The morphological operator is used for shadow removal. The operator is simple and effective. The gradient operator is given [54] as following function.

$$GRA = (I \oplus B) - (I \Theta B) \qquad \text{Eq. (14)}$$

where I is captured frame and B is structuring element of morphological operation. Θ is morphological erosion operation which is given in Eq. (14), $\oplus$ is morphological dilation operation which given in Eq. (14). And GRA is the gradient image as Figure 31. The shadow removal algorithm, the cast shadow of moving object will be reduced, and the motion information is still kept.

Erosion:

$$I \Theta B = \{x \mid (B)_x \subseteq I\} \qquad \text{Eq. (15)}$$

where I is captured frame, B is structuring element of morphological operation, Θ is morphological erosion operation. The value of the output pixel is the minimum value of all the pixels in the input pixel's neighborhood. In a binary image, if any of the pixels is set to 0, the output pixel is set to 0.

Dilation:

$$I \oplus B = \left\{ x \mid \left[(\hat{B})_x \cap I \right] \subseteq I \right\} \quad \text{Eq. (16)}$$

where I is captured frame, B is structuring element of morphological operation, $\oplus$ is morphological dilation operation. The value of the output pixel is the maximum value of all the pixels in the input pixel's neighborhood. In a binary image, if any of the pixels is set to the value 1, the output pixel is set to 1.

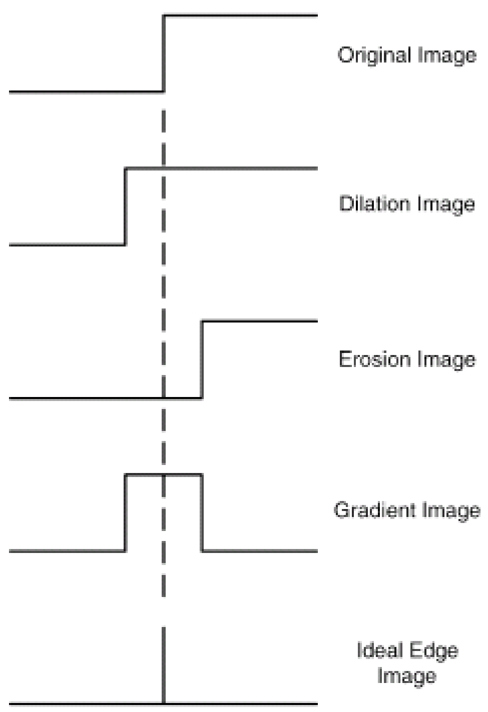


Figure 31 Edge thickening effect of morphological gradient filter [55]

The shadow cancellation effect is shown in Figure 31. Compared with ideal edge image, the edge of gradient image is thickened, which maybe make the segmentation result of boundary not accurate. The GRA operation still has two disadvantages:

1. Cannot deal with parallel and strong light source
2. may cause errors when the texture of background is significant

2.7 Comparisons and Summary

In the four categories of moving object detection methods, there are point feature approach, segmentation approach, pattern classification approach, and background subtraction approach. The comparisons of four approaches are shown in the Table 5.

Table 5 The comparison of four moving object segmentation approaches

approach	pros	cons
Point feature	invariant with orientation, scale, and location, even occlusion	specific object only rigid body only
segmentation	high computational speed	simple environment and object only
pattern classification	camera and scene are no need stationary high computational speed	specific learned feature object only need a mount of examples waste lots of time in learning
Background subtraction	object contour general moving objects high computational speed	camera and scene are stationary

For the aim of detecting general moving object, background subtraction method is chosen. In the Table 4, it shows three most popular backgrounds. The backgrounds, RAB, has can reduce illumination variation, online background update, low memory requirement, and high computational speed. However, its accuracy is only acceptable. The aim in this thesis is to construct a background that is high accuracy and has the advantages of RAB.

Chapter 3 Moving Object Detection Using Improved Running Average Background

Moving object detection based on *Two-Staged Background Subtraction* (Two-StaBaS) approach is proposed in this thesis. There are two stages in Two-StaBaS. The first stage is moving object detection using *Improved Running Average Background* (IRAB), which is proposed and discussed in this chapter. The second stage is background subtraction in *Moving Object Candidate* (MOC) region with weighting map, which is suggested and described in the next chapter.

The moving object is defined as the object which moves in every frame but temporarily static in few frames is also allowed. For example, two people walk in the scene, if they meet each other, they talk and stand without moving in few frames. In Two-StaBaS approach, the two standing people are still regarded as moving objects. If they remain too long, then they are regarded as background because they have been updated by Two-StaBaS. In the following sections, the first stage of Two-StaBaS is discussed.

3.1 Moving Object Detection Process

The process for detecting moving objects in image sequences is discussed in this section. The first step in the process of moving object detection is background modeling and background subtraction, and IRAB is proposed to be the background model. IRAB has two parameters for background updating, one is for background updating, and the other one is for reducing the effect of noise and illumination variation. In order to detect the moving object region, assume that most of the corresponding pixels of foreground and background have large difference in gray level intensities or color. The difference of captured frame and background is constructed to be difference frame, where the background subtraction operator is utilized for detecting the difference of captured frame and background, and triangular auto-thresholding algorithm is used for deciding whether the pixels in captured frame belong to the foreground region or not. There are moving object region, moving object shadow region, ghost region, and ghost shadow region in the foreground region. Shadow removal is introduced for the removing moving object shadow region and ghost shadow region. Then, the moving object region is the remainder in the foreground region. The flow chart of moving object detection using IRAB is illustrated in Figure 32.

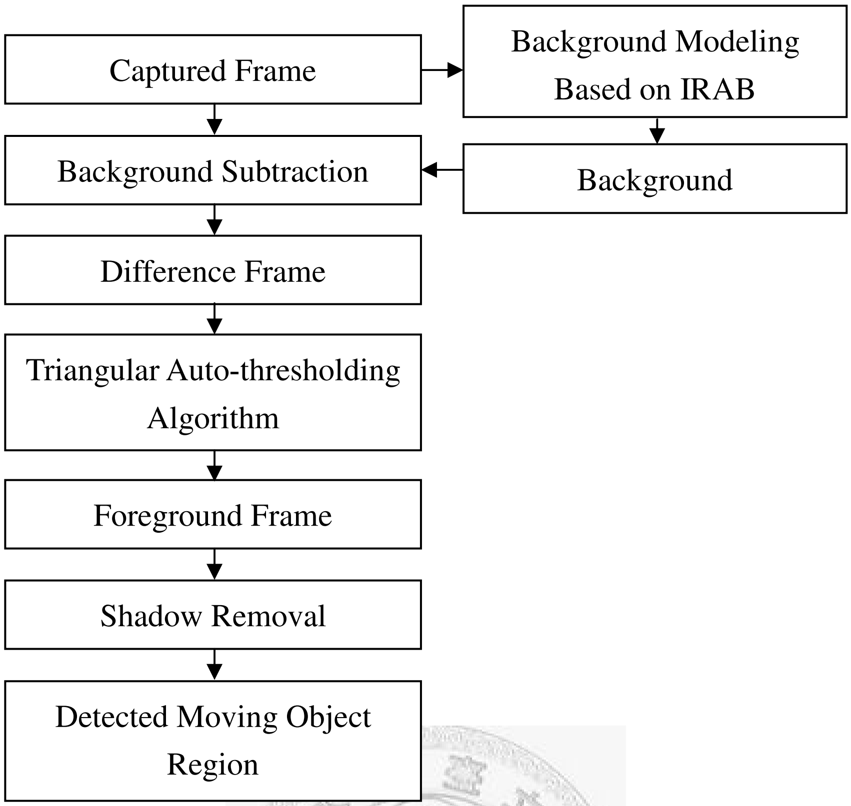


Figure 32 The flow chart of moving object detection using IRAB

3.2 Foreground Detection

Background is utilized for detecting foreground region which is obtained by thresholding the difference of captured frame and background. *Improved Running Average Background* (IRAB) is proposed for constructing the background, that is, IRAB is utilized for background modeling. IRAB reduces the effect of environment noise, camera noise, and the illumination variation, it also updates the background as there are new static objects in the scene. IRAB is not only a fast algorithm for foreground region detection, but also low memory requirement and good accuracy.

3.2.1 Improved Running Average Background (IRAB)

Background modeling is the first step in the process of background subtraction approach. The IRAB is suggested for background modeling. The color space is based on RGB (Red, Green, and Blue) space. In Chapter 2, *running average background* (RAB) is mentioned. RAB has high computational speed and low memory requirement, however, only acceptable accuracy for detecting moving objects. The IRAB is proposed for an improvement of the accuracy of RAB, and its detail is discussed in the following paragraphs.

The main cause in detection errors are noise, varying illumination, and new static object which is in the scene. The noise includes environment noise and camera noise. The varying illumination means that the illumination in the scene may be various if the light source is sun shine outdoors or fluorescent lamp indoors. The feature of noise is it changes rapidly in time with small various gray level intensities. The feature of varying illumination is it changes slowly in time with large various gray level intensities. However, in only few frames, the feature of noise and illumination variation are the same, they changes only in a little gray level intensities. In the above-mentioned conditions, a low pass filter is proposed for reducing the problems of noise and illumination variation. By weighted integrating the variation of noise and illumination, the captured frame and background has near properties in noise and illumination variation. The component of noise and illumination variation in the difference of captured frame and background is then reduced. It is illustrated as Figure 33. The main idea for weighted integration is that the past information which is closer to the current time is more important than the old. Hence, the past information closer to current time has larger weighting.

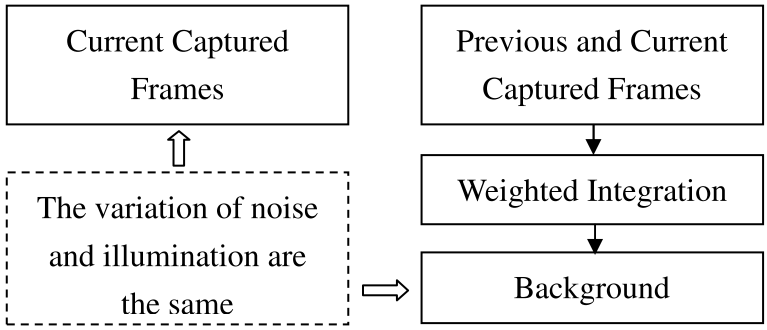


Figure 33 By means of the difference between the weighted integration of captured frame and background, the effect of noise and illumination variation is reduced

By reducing the effect of noise and illumination variation, the detection accuracy is improved. The suggested low pass filter is exponentially decay function (in time domain). It can be written as Eq. (17).

$$B_k = B_{k-1} + \alpha_2 \left(C_k \Theta B_{k-1} \right) \qquad \text{Eq. (17)}$$

where C_k is captured frame in k-th frame, B_k is background in k-th frame, Θ is background subtraction operator which is discussed in Subsection 3.2.3, and α_2 is background updating parameter which is discussed in Subsection 3.2.2.

For the case of new static objects in the scene, the static objects are required being updated in the background. RAB is the one of good solutions for background updating. In the condition that the captured frame and background in the corresponding pixels of new static object region has large difference, RAB is utilized for background updating. RAB is formulated as Eq. (18).

$$B_k = B_{k-1} + \alpha_1 \left(C_k \Theta B_{k-1} \right) \qquad \text{Eq. (18)}$$

where C_k is captured frame in k-th frame, B_k is background in k-th frame, Θ is background subtraction operator which is discussed in Subsection 3.2.3, and α_1 is background updating parameter which is discussed in Subsection 3.2.2.

To solve the problem of noise, illumination variation, and new static objects, IRAB is then proposed. It is the combination of Eq. (17) and Eq. (18).

$$B_k = \begin{cases} B_{k-1} + \alpha_1 \left(C_{k-1} \Theta B_{k-1} \right) & \text{,if } k > 0, \left| C_{k-1} \Theta B_{k-1} \right| > th \\ B_{k-1} + \alpha_2 \left(C_{k-1} \Theta B_{k-1} \right) & \text{,if } k > 0, \left| C_{k-1} \Theta B_{k-1} \right| \le th \\ C_0 & \text{,if } k = 0 \end{cases} \qquad \text{Eq. (19)}$$

where C_k is captured frame in k-th frame, B_k is background in k-th frame, Θ is background subtraction operator which is discussed in Subsection 3.2.3, α_1 and α_2 are background updating parameters which is discussed in Subsection 3.2.2. α_1 is smaller than α_2, and $0 < \alpha_1 < \alpha_2 < 1$. α_1 is the parameter utilized for background updating, and α_2 is the parameter used for reducing noise and illumination variation. The typical values are $\alpha_1 = 0.01$, $\alpha_2 = 0.05$. The suggested values for the two parameters, α_1 and α_2, are discussed in the next subsection. The sign "Θ" is background subtraction operator, it can be subtraction "-", then the IRAB can be written as Eq. (20) or Eq. (21).

$$B_k = \begin{cases} B_{k-1} + \alpha_1 \left(C_{k-1} - B_{k-1} \right) & ,if\ k > 0, \left| C_{k-1} - B_{k-1} \right| > th \\ B_{k-1} + \alpha_2 \left(C_{k-1} - B_{k-1} \right) & \text{,if } k > 0, \left| C_{k-1} - B_{k-1} \right| \le th \\ C_0 & \text{,if } k = 0 \end{cases} \qquad \text{Eq. (20)}$$

$$B_k = \left(1 - \alpha_{1,2}\right)^{k-1} C_0 + \sum_{r=1}^{k-1} \alpha_{1,2} \left(1 - \alpha_{1,2}\right)^{k-1-r} C_r \qquad \text{Eq. (21)}$$

It is an exponentially decay with frames, that is, the weighting of past captured frames are smaller if the frames are old. The advantages of IRAB are low memory requirement, fast computational speed, and good accuracy.

3.2.2 Background Updating Parameters

IRAB is discussed in the previous subsection. In this subsection, the discussion is focused on two parameters, α_1 and α_2, in Eq. (19). The first one, α_1, is for background updating, that is, updating new static objects in the background. The second one, α_2, is for reducing the variation of noise and illumination.

In the condition of there are new static objects in the scene, background updating is required. The faster background updating for new static objects is better. However, moving objects ought not to be updated. Hence, it is a tradeoff for choosing background updating parameters. Based on some constraints for background updating parameter α_1, IRAB updates new static object but not moving object.

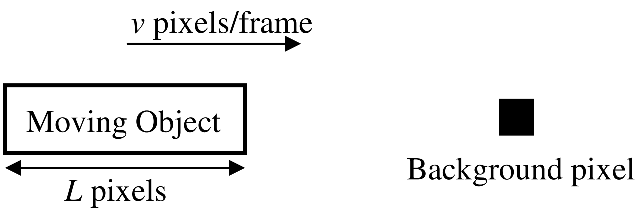


Figure 34 The length of the moving object is L pixels, and it moves straight towards the background pixel with constant velocity v pixels per frame, the moving object requires a good background updating parameter to prevent from being updated into the background

The discussion on how to prevent moving objects from being updated is in this paragraph. See the illustration in Figure 34, there is one moving object and one background pixel. Assume that the length of the moving object is L pixels, and it moves straight towards the background pixel with constant velocity v pixels per frame. Then the moving object occupies the background pixel L/v frames. In these occupied frames, the constraint is the moving objects should not be updated by IRAB. Namely, in the foreground detection algorithm results, moving objects should not belong to background region. Hence, the difference of gray level intensity of moving object and background should larger than threshold value.

Assume the set of the gray level intensities in moving object region and the gray level intensity sequence of background pixel are Eq. (22) and Eq. (23), respectively.

$$GLV_{MO} = \{g_1, g_2, ..., g_n\} \qquad \text{Eq. (22)}$$

$$\{g_{background,0}, g_{background,1}, g_{background,2}, ..., g_{background,k}, ...\} \qquad \text{Eq. (23)}$$

where the elements in the set GLV_{MO} are gray level intensities in moving object region. The minimal value in the set is $g_{\min}$, the maximal is $g_{\max}$.

$$g_{\min} = \min\left(GLV_{MO}\right) \qquad \text{Eq. (24)}$$

$$g_{\max} = \max\left(GLV_{MO}\right) \qquad \text{Eq. (25)}$$

The gray level intensity in background pixel is updated by IRAB, and its equation can be represented by Eq. (26).

$$g_{background,k} = g_{\max} + \left(g_{background,0} - g_{\max}\right)\left(1-\alpha_1\right)^k \qquad \text{Eq. (26)}$$

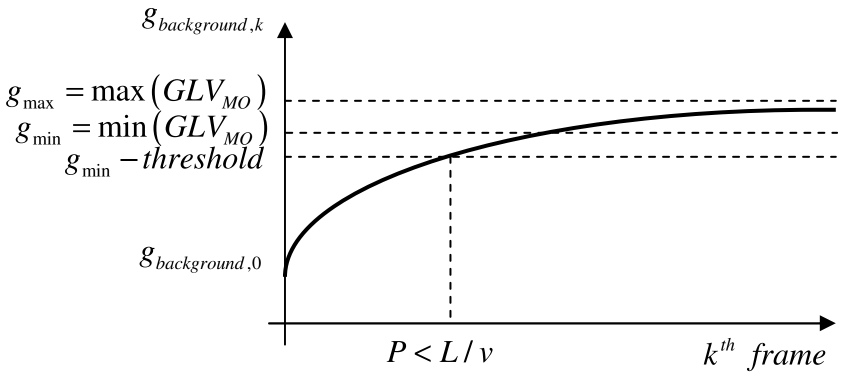


Figure 35 Choosing background updating parameters with physical constraints

For the reason that moving object region and background region are independent, then the gray level intensity in background pixel has the constraint in Eq. (27). That is, the moving object region is not belong to background region, the difference of moving object and background should larger than the threshold.

$$g_{background,\frac{L}{v}} < g_{\min} - threshold \qquad \text{Eq. (27)}$$

Substitute Eq. (26) into Eq. (27), obtain Eq. (28) or Eq. (29).

$$g_{\max} + \left(g_{background,0} - g_{\max}\right)\left(1-\alpha_1\right)^{\frac{L}{v}} < g_{\min} - threshold \qquad \text{Eq. (28)}$$

$$1 - \left(\frac{g_{\min} - g_{\max} - threshold}{g_{background,0} - g_{\max}}\right)^{\frac{v}{L}} > \alpha_1 \qquad \text{Eq. (29)}$$

In this result shown in Eq. (29), it ensures that α_1 should smaller than a value and the updating speed is limited so that the moving object will not being updated by IRAB. Hence, the suggested value for α_1 is a value smaller than the value in Eq. (29).

In the next, the discussion is focused on background updating parameter α_2. It is hard to model noise and illumination variation due to they occur randomly with unknown value that caused by environment, CCD camera, or artificiality. However, by observations of them, the sooner for background updating is better. Hence, the selection of background updating parameter α_2 is chosen by $\alpha_2 = 5\alpha_1$.

For the typical values, assume that the moving object length is 80 pixels, and its velocity is 16 pixels per frame. The minimal and maximal gray level intensities in moving object region are 145 and 240, respectively. The general threshold value is set to be 40. Then the background updating parameter α_1 is bout 0.05. And the second background updating parameter α_2 is 0.25.

3.2.3 Background Subtraction Operator

In the process of detecting foreground region, background subtraction operator plays an important role. Background subtraction operator is utilized for finding the difference of captured frame and background. For the different backgrounds, such as simple black or white background, complex color background, and crowded background, the suitable background subtraction operator depends on different backgrounds. In the following, the background subtraction operators and its suitable situation are discussed.

I. Subtraction

The basic background subtraction operator is subtraction, it means "the difference of captured frame and background" is equal to "captured frame subtracts background". It can be represented by Eq. (30).

$$C_k \Theta B_k = C_k - B_k \qquad \text{Eq. (30)}$$

It is a linear operator, hence, this operator is suitable for the background which is uniform distributed in gray level intensities or the intensity distributions of background and moving object is unknown. The background subtraction result is shown in Figure 36. The intensities of $C_k(p)$ and $B_k(p)$ are from intensity 0 to 255. And the result of background subtraction using "subtraction" is also from intensity 0 to 255. The intensity 0 is represented by color complete dark, and the intensity 255 is described by color complete white. The intensity is darker means the background subtraction result has small difference between $C_k(p)$ and $B_k(p)$, on the contrary, the intensity is lighter means the background subtraction result has large difference between $C_k(p)$ and $B_k(p)$.

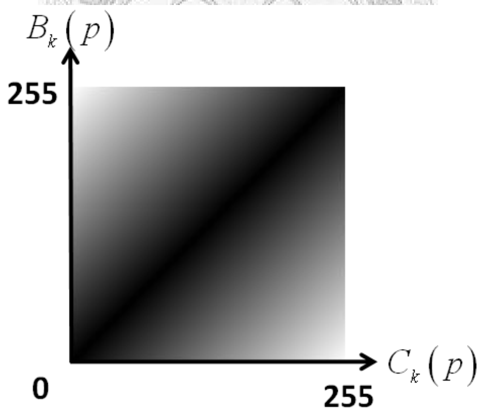


Figure 36 The difference curve of C_k and B_k using background subtraction operator "subtraction", the dark intensity means the difference is small, and the light intensity means the difference is large.

II. Geometric

The background subtraction operator is "geometric". It causes the background subtraction curve is non-linear. Its equation is described in Eq. (31).

$$C_k \Theta B_k = sign(C_k - B_k)255\left[1-\frac{2\sqrt{(C_k+1)(B_k+1)}}{(C_k+1)+(B_k+1)}\frac{2\sqrt{(256-C_k)(256-B_k)}}{(256-C_k)+(256-B_k)}\right] \quad \text{Eq. (31)}$$

It is a non-linear operator. By observation of the background subtraction result which is shown in Figure 37, this operator is suitable for the background which is dark or white due to it is sensitive to the background is bark or white. However, the sensitivity is bad if the intensity distribution of background is focused on the middle of gray level intensity.

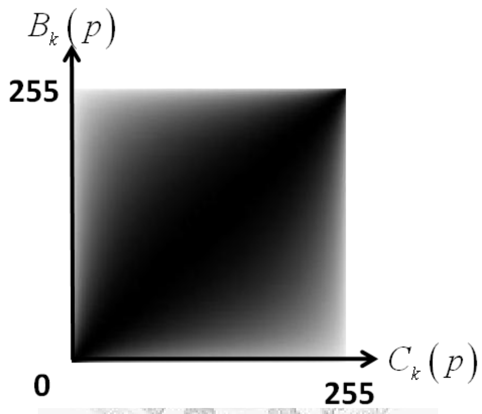


Figure 37 The difference curve of C_k and B_k using background subtraction operator "geometric", the dark intensity means the difference is small, and the light intensity means the difference is large

III. Exponential

The background subtraction operator is "exponential". It causes the background subtraction curve is also non-linear. The equation of "exponential" is represented by Eq. (32).

$$C_k \Theta B_k = sign(C_k - B_k)\frac{255}{e-1}\left(e^{\frac{|C_k - B_k|}{255}} - 1\right) \quad \text{Eq. (32)}$$

It is a non-linear operator. By observation of the background subtraction result which is shown in Figure 38, this operator is suitable for the background which has too much noise or illumination variation due to it is insensitive to the difference of $C_k(p)$ and $B_k(p)$. And its suitable background is which the distribution of gray level intensity is uniform or unknown.

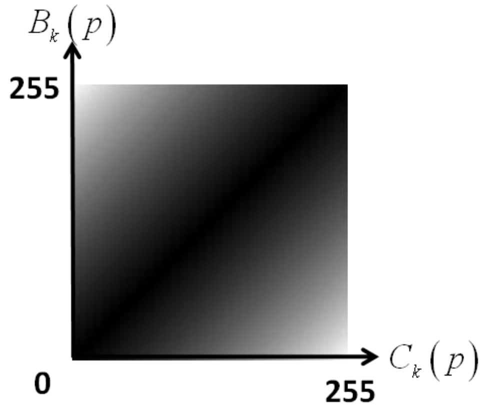


Figure 38 The difference curve of C_k and B_k using background subtraction operator "exponential", the dark intensity means the difference is small, and the light intensity means the difference is large

IV. Logarithmic

The background subtraction operator is "logarithmic". It causes the background subtraction curve is non-linear. The equation is shown in Eq. (33).

$$C_k \Theta B_k = sign(C_k - B_k)\frac{255}{\ln(256)}\ln\left[|C_k - B_k| + 1\right] \qquad \text{Eq. (33)}$$

It is a non-linear operator. By observation of the background subtraction result which is shown in Figure 39, this operator is suitable for the background which $C_k(p)$ and $B_k(p)$ has near color or gray level intensity due to the small difference of $C_k(p)$ and $B_k(p)$ causes larger difference result. And its suitable background is which the distribution of gray level intensity is uniform or unknown.

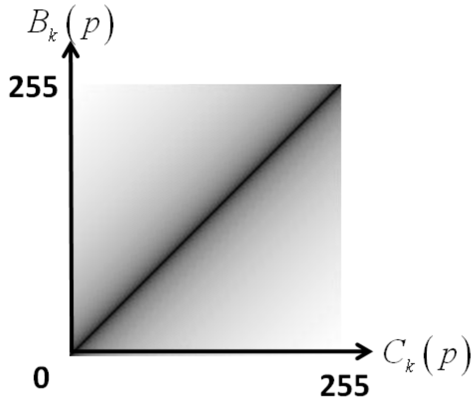


Figure 39 The difference curve of C_k and B_k using background subtraction operator “logarithmic”, the dark intensity means the difference is small, and the light intensity means the difference is large

Table 6 The comparison of four background subtraction operators

Operator	Linear / nonlinear	Suitable situations
Subtraction	Linear	background is uniform or unknown in intensity distribution
Geometric	Nonlinear	background is dark or white
Exponential	Nonlinear	background is uniform or unknown in intensity distribution and background has too much noise or illumination variation
Logarithmic	Nonlinear	background is uniform or unknown in intensity distribution and moving object and background have near color or gray level intensity

V. Comparison of four background subtraction operators

The four different background subtraction operators are suitable for different scenarios. In the following table, it points out the suitable situations for four background subtraction operators.

3.2.4 Triangular Auto-thresholding Algorithm

Background subtraction approach is a process of thresholding the difference of captured frame and background, that is, it is the thresholding of the difference frame. The auto-thresholding algorithm is discussed in this subsection.

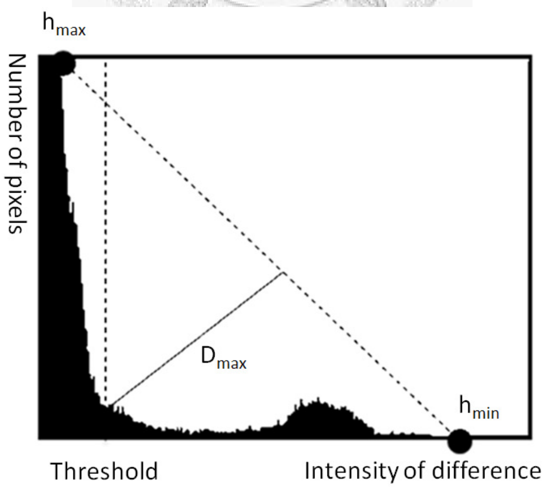


Figure 40 The histogram of difference frame

In general, the foreground region occupies only small region in the captured frame. The histogram of the difference frame is almost similar to Figure 40. Then, the auto-thresholding algorithm is suggested as follows.

(1) Find the absolute maximum point and absolute minimum point in the histogram. If the maximum point or minimum point is not only one, then choose the one that the intensity of difference is smaller.
(2) Connect the maximum point and minimum point by a straight line.
(3) Find the maximal distance between the straight line and the histogram. The intensity of difference of the point is assigned to be the threshold.

Assume that the intensity of difference in the difference frame is larger than threshold value, than the pixel belongs to foreground region; the intensity of difference is smaller than threshold value, then it belongs to background region.

$$\text{Background Region} = \{ p \mid value(p) \le thrshold \} \quad \text{Eq. (34)}$$

$$\text{Foreground Region} = \{ p \mid value(p) > thrshold \} \quad \text{Eq. (35)}$$

3.3 Foreground Analysis and Shadow Removal

Foreground is obtained by the thresholding of the difference frame. Foreground includes moving object, moving object shadow, ghost, and ghost shadow, which are discussed in Subsection 2.6.4. Ghost is a set of connected points detected as in motion by means of background subtraction, but not corresponding to any real moving object. For example, new static object in the scene, reflection, and something that detected with unknown reason. Ghost is updated by IRAB such that the effect of ghost is reduced. However, moving object shadow and ghost shadow should not be updated, and they also the main cause in moving object detection error. Hence, the removal for moving object shadow and ghost shadow is one of method for the improvement of moving object detection accuracy.

The feature of shadow (moving object shadow and ghost shadow) is discussed in the following. The chromatic information between shadow in captured frame and background are near, namely, the hue and saturation information have only a little difference. But its gray level intensity may have large difference because the shadow in the captured frame is darker than background. As the reason discussed above, transform the color space from the RGB space to HSV (Hue, Saturation, and Value) space, the shadow detection method is then represented by Eq. (36).

$$\begin{cases} AD_H = \min\left(\left|C_H \Theta B_H\right|, 360 - \left|C_H \Theta B_H\right|\right) < TH_H \\ AD_S = \left|C_S \Theta B_S\right| < TH_S \\ TH_{V,L} < \dfrac{C_V}{B_V} < TH_{V,H} \end{cases} \qquad \text{Eq. (36)}$$

where $\left(C_H, C_S, C_V\right)$ means the hue, saturation, and value of captured frame, $\left(B_H, B_S, B_V\right)$ is the hue, saturation, and value of background, AD_H and AD_S are the absolutely difference of captured frame and background in hue space and saturation space, respectively, TH_H, TH_S, $TH_{V,L}$, and $TH_{V,H}$ are the threshold manual assigned, and $0 < TH_{V,L} < TH_{V,H} < 1$.

If the pixel satisfies above equations, then the pixel belongs to shadow region. By the shadow detection, the almost shadow would be reduced, and the moving object can be found.

$$\text{Shadow Region} = \left\{ p \middle| p \in Shadow \right\} \qquad \text{Eq. (37)}$$

$$\text{DMO Region} = \left\{ p \middle| p \in \text{Foreground Region}, p \notin \text{Shadow Region} \right\} \qquad \text{Eq. (38)}$$

3.4 Discussions and New Idea

Detecting moving object is discussed in this chapter. First construct background using IRAB, which reduces noise and illumination variation, and updates backgrounds. It achieves high computational speed, low memory requirement and good accuracy. By using background subtraction method to detect foreground region, moving object is then obtained by removing the shadow part of foreground region, and the result is called DMO region.

However, the DMO region is not perfect. If near gray intensities in corresponding pixels of moving object region and background, then the detection errors occur. In order to solving this problem, second stage background subtraction is proposed. If background subtraction again in a moving object candidate region with some weighting, then the results may be better. The modification method for this detection error is suggested in the next chapter.

Chapter 4 Two-Staged Background Subtraction Approach

In Chapter 3, a proposed moving object detection approach is discussed. The background modeling method is based on *Improved Running Average Background* (IRAB). It reduces the effect of noise (comes from environment and camera) and illumination variation (caused by fluorescent lamp indoor and sunlight outdoor). It also updates new static object in the scene and removes shadows. The detecting result is called as *Detected Moving Object* (DMO) region. The advantages for this approach are good accuracy, high computational speed, and low memory requirement. However, there are detection errors in DMO region. As near color in the corresponding pixel of moving object and background, the result in background subtraction approach usually is the pixels not belong to *moving object* (MO) region. The result is false negative, namely, the pixel actually belong to MO region but opposite in detection result. In order to solve the above-mentioned problem, *TWO-STAged BAckground Subtraction approach* (Two-StaBaS) is then proposed. In Section 4.1 introduces the process of Two-StaBaS. In Section 4.2 and 4.3 describes the detail for Two-StaBaS.

4.1 The Process of Two-Staged Background Subtraction Approach

The aim for Two-StaBaS is the improvement of detection error of near color in the corresponding pixel of moving object region and background. Two-StaBaS has two stages for detecting MO region. The first stage is moving object detection using

IRAB that is discussed in Chapter 3, and its detection result is DMO region. The second stage in Two-StaBaS is weighted background subtraction in *Moving Object Candidate* (MOC) region. And its result is 2nd-staged DMO region. The process is illustrated in the following figure.

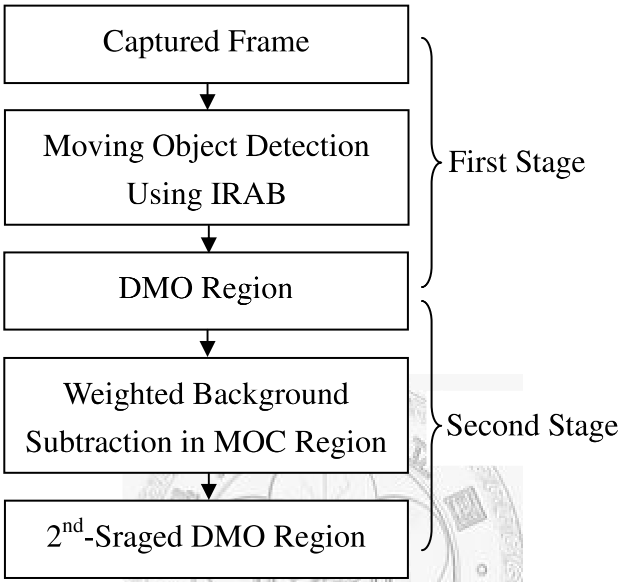


Figure 41 The process of Two-StaBaS has two stages

The DMO region sometimes has the false negative if near gray level intensity or color in corresponding pixels of moving object region and background. The idea for reducing this false negative is first constructing possible MO region, called *Moving Object Candidate* (MOC) region, and the MOC region includes MO region. It means that all pixels in MOC region are possible pixels in MO region. However, the pixel in MOC region but not belongs to MO region requires to be removed. Assume that the pixels close to the center of mass in MOC region are more probable belong to MO region than the boundary pixels, then the pixels closer to center of mass in MOC region, the weighting is larger. The weighted background subtraction is suggested. And its MO region detection result is called as 2nd-staged DMO region. It solves the problem of near gray level intensity or color in corresponding pixels of moving object region in the captured frame and background.

The method for constructing possible MO region and the weighting map are discussed in Section 4.2. The method for detecting 2^{nd}-staged DMO region using weighting map is discussed in Section 4.3.

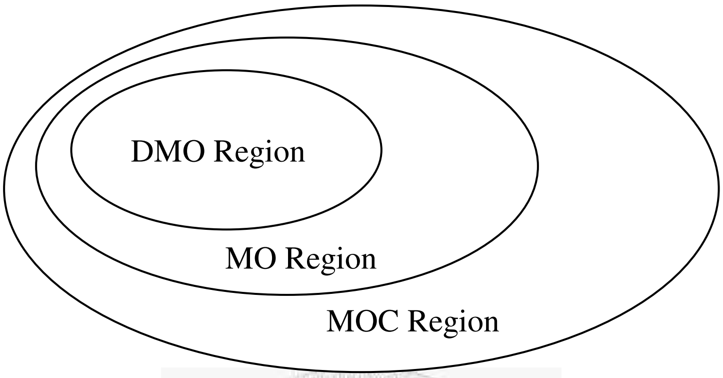


Figure 42 The relationship of DMO region, MO region and MOC region

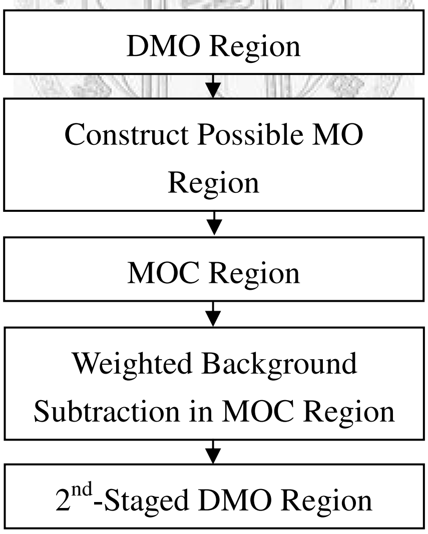


Figure 43 The process of second stage in Two-StaBaS

4.2 Weighting Map

4.2.1 Moving Object Candidate Region

The DMO region is obtained in the first stage background subtraction. Because DMO region occurs false negative due to near gray level intensity or color in corresponding pixels of moving object region and background. In order to solve the problem, a method is proposed. In the first of the method, the MOC region is constructed. It is a region that all pixels possible belong to MO region. Namely, assume DMO region belongs to *Moving object* (MO) region, shown as Eq. (39) or Eq. (40).

$$\forall p \in \text{MO region}, p \in \text{MOC region} \qquad \text{Eq. (39)}$$

$$\text{MO region} \subset \text{MOC region} \qquad \text{Eq. (40)}$$

In the reality, the results in the first stage background subtraction are not ideal. The detection error obviously occurs as the color in the pixels of moving object and background are near. The neighborhood of DMO region also possible belongs to MO region.

$$\forall p \in \text{DMO region, neighborhood}(p) \text{ is possible belongs to MO region}$$

Assume that the morphological dilation of DMO region is MOC region.

$$\text{MOC region} = Dilation(\text{MO region}) \quad \text{Eq. (41)}$$

However, the MOC region is not equal to MO region. Because MOC region may include the pixels not belong to MO region. The check of the pixel belongs to MOC region that also belongs to MO region is required. The checking method is discussed in the next subsection.

4.2.2 Weighting Map and Distance Transform

The comparison of the point close to the center and boundary of MOC region, the former is more probable belong to MO region. Therefore, the pixels closer to the center of mass in MOC region, the weighting is larger. In this subsection, weighting map is discussed.

Assume that there is no hole encompassed by the MOC region, that is, there are no background region pixel enclosed by MOC region. The weighting map is assigned by distance transform. The distance transform is an operator applied to binary images. The result of the transform is a gray level image that the MOC region becomes the distance to the closest boundary. The detail for distance transform is introduced in the next paragraph.

A. Distance Transform

The distance transform is an operator applied to binary images. That is, the distance transform of input image is a binary image. The result of distance transform is a gray level image that is similar to the input image, except that the gray level intensities of pixels inside foreground regions are changed to show the distance to the closest boundary from each pixel.

The calculation of distance between two pixels in an image usually has three methods. There are Euclidean distance, city block distance, and chessboard distance. The distance between two pixels (x_1, y_1) and (x_2, y_2) based on above-mentioned three methods are given in the following.

1. Euclidean distance

$$Dis_{Euclidean} = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \quad \text{Eq. (42)}$$

2. City block distance

$$Dis_{City\ Block} = |x_2 - x_1| + |y_2 - y_1| \quad \text{Eq. (43)}$$

3. Chessboard distance

$$Dis_{Chessboard} = \min(|x_2 - x_1|, |y_2 - y_1|) \quad \text{Eq. (44)}$$

For the computational speed, the last two methods are faster than the first one. In this thesis, chessboard distance is utilized.

An example for distance transform is illustrated in Figure 44. The left one in the figure is input image, it is a binary image. The right one is the distance transform of input image, it is a gray level image. The intensity depends on the distance to the closest boundary. The distance calculation is based on chessboard distance. Another example is shown in Figure 45.

0	0	0	0	0	0	0	0
0	1	1	0	1	1	1	0
0	1	1	1	1	1	1	0
0	1	1	1	1	1	1	0
0	0	1	1	1	1	1	0
0	0	0	1	1	1	0	0
0	0	0	0	1	0	0	0
0	0	0	0	0	0	0	0

(a)

0	0	0	0	0	0	0	0
0	1	1	0	1	1	1	0
0	1	2	1	2	2	1	0
0	1	1	2	3	2	1	0
0	0	1	2	3	2	1	0
0	0	0	1	2	1	0	0
0	0	0	0	1	0	0	0
0	0	0	0	0	0	0	0

(b)

Figure 44 An example for distance transform (a) is input binary image with object region represented as number 1 (b) the output image of distance transform which is gray image similar to input image

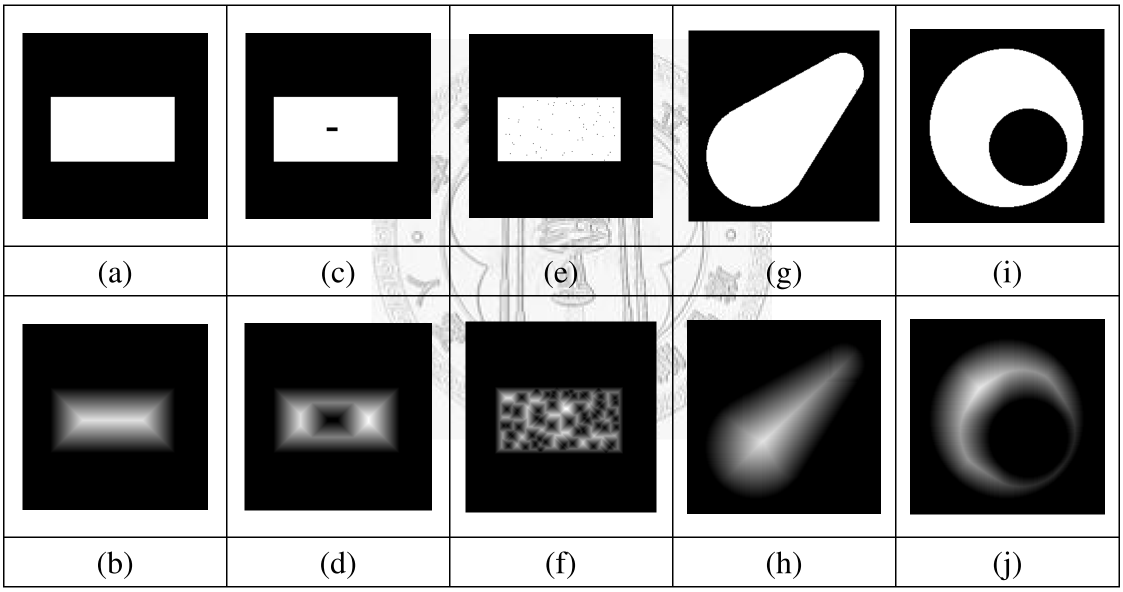


Figure 45 Five set of input image and its output image of distance transform [63]

B. Weighting Map

The distance transform of MOC region is a map with distances, called distance map. The distances in MOC region are decided by the closest chessboard distance to the boundary. The normalization of distance in distance map is then assigned to be the weighting; it is relative to the possibility of MO region. The weighting of the center in

MOC region is larger than other pixels, and the weighting of boundary pixel is smaller because it is relatively not important. The chessboard distances of all pixels in the MOC region form a set. The maximal distance and minimal distance are $d_{\max}$ and $d_{\min}$, respectively.

$$D_{MOC\ region} = \{d_1, d_2, d_3, \ldots, d_n\} \qquad \text{Eq. (45)}$$

$$d_{\max} = \max\{d_1, d_2, d_3, \ldots, d_n\} \qquad \text{Eq. (46)}$$

$$d_{\min} = \min\{d_1, d_2, d_3, \ldots, d_n\} \qquad \text{Eq. (47)}$$

The normalization of distance follows Eq. (48).

$$\overline{d}_i = \frac{d_i - d_{\min}}{d_{\max} - d_{\min}} \qquad \text{Eq. (48)}$$

The weighting map is then the normalization of distance map.

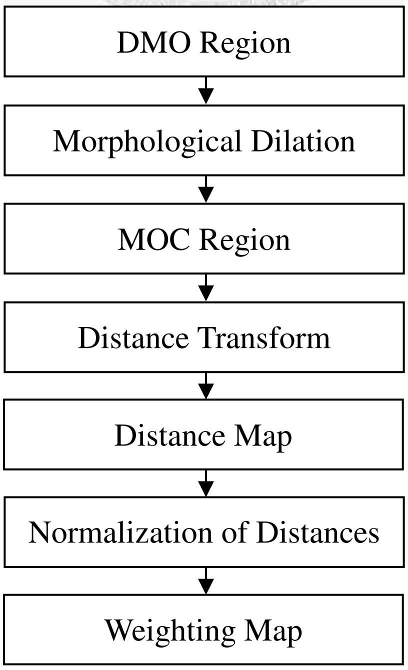


Figure 46 The process for computing weighting map

4.3 Second Stage Background Subtraction

The aim in second stage of Two-StaBaS approach is to find MO region without near color’s false detections. The suggested method is to detect MO region in MOC region with weighting map. The MOC region and weighting map have been discussed in Subsection 4.2.1 and Subsection 4.2.2.

The dilation of DMO region is MOC region, which is probable MO region and assumed that all pixels in MO region belong to MOC region. The weighting map is the normalization of distance transform of MOC region which weighting is corresponding to the importance of the pixel. The pixel close to the boundary has lower weighting. On the contrary, the pixel close to the center of mass of MOC region has higher weighting. Namely, the center of MOC region is emphasized by the weighting map, and the boundary is ignored.

In the second stage of Two-StaBaS, first find the difference map by detecting the difference of captured frame and background in MOC region, where the detection method is based on background subtraction operator. The difference of intensity in every pixel of difference map is multiplied by the corresponding weighting in the weighting map to form weighted difference map. Its equation can be written as Eq. (49).

$$D_{with\ difference} = \left(1+\bar{d}_i\right)\left(C_k \Theta B_k\right)\left(p\right) \qquad \text{Eq. (49)}$$

By smoothing the weighted difference map using morphological gray level dilation, and thresholding, then the 2nd-staged DMO region is finally detected, which threshold is the same as Eq. (34) and Eq. (35). In the detection result, the false detections in the corresponding pixel of moving object and background have near color or gray level value is then reduced. The process of second stage of Two-StaBaS is illustrated in Figure 47.

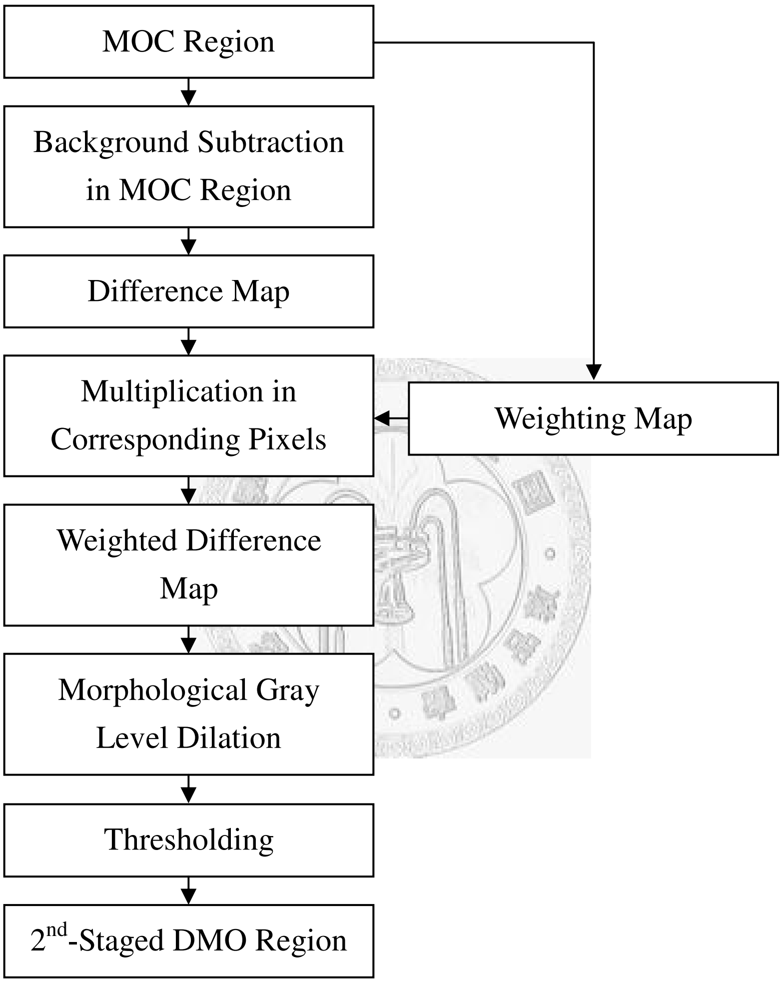


Figure 47 The detailed process in the second stage of Two-StaBaS

Chapter 5 Results

5.1 System Architecture

This subsection describes the setup of whole system. The image acquisition device which utilized in the experiments is a CCD camera. The image processing system is based on PC platform, and the simulation software is BCB 6.0. The experiment framework and device specifications show in Figure 48 and Table 7, respectively.

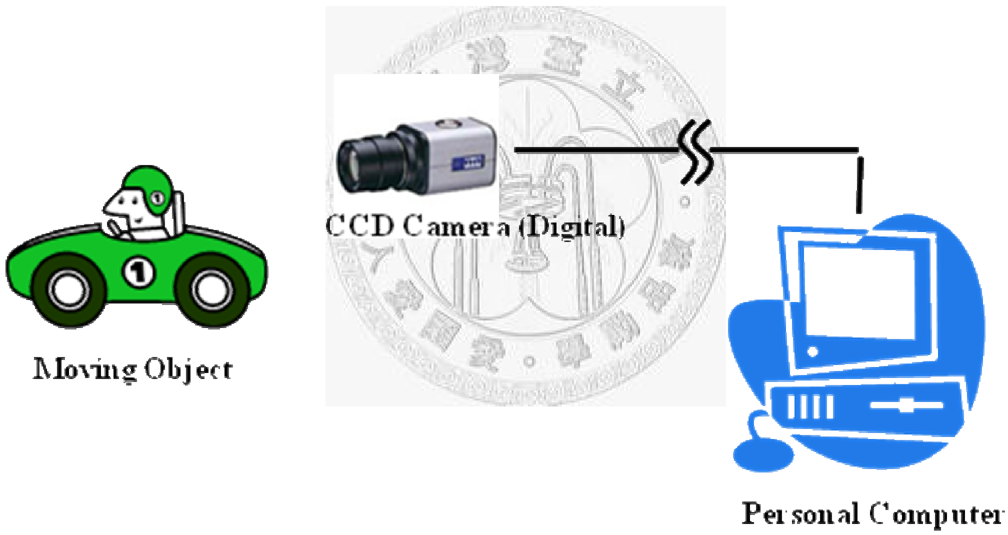


Figure 48 Frame work for system architecture

Table 7 Devices and its specifications which used in the experiments

Device	Specification
PC	Intel (R) Core (TM)2 Quad CPU Q6600 @ 2.40GHz RAM: 2.40GHz, 1.98GB OS: Microsoft Windows XP
CCD webcam	Frame rate: 30 fps Resolution: 320 * 240 pixels Interface: USB 2.0
Simulation software	BCB 6.0

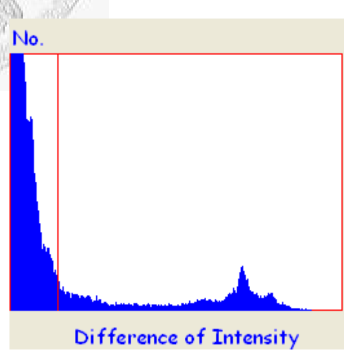
5.2 Experiment Results

The process of moving object detection using IRAB is shown in Figure 49. First is the captured frame (shown in Figure 49(a)) which is captured from CCD camera, which is one of frames in an image sequence. The background (shown in Figure 49(b)) is constructed by IRAB. The difference frame (shown in Figure 49(c)) is the difference of captured frame and background. By thresholding the histogram of difference frame using triangular auto-thresholding algorithm (shown in Figure 49(d)), the captured frame can be divided into two regions, foreground region and background region. The foreground region (shown in Figure 49(e)) includes moving object region, ghost region, shadow region and ghost shadow region. Moving object region is then detected by removing the shadow region and ghost shadow region which are detected by shadow detection. The ghost region is reduced by IRAB. The detection result in this stage is called as DMO region (shown in Figure 49(f)). There are a lot of false detections in the DMO region. To solve this problem, Two-StaBaS is suggested. First find the possible moving object region by dilating the DMO region, called as MOC region (shown in Figure 49(g)), then compute the normalization of the distance transform of MOC region to form weighting map (shown in Figure 49(h)). Finally, thresholding the background subtraction in MOC region with weighting map, 2nd-staged DMO region is obtained (shown in Figure 49(i)). It reduces the false detections. The final result of Two-StaBaS is shown in (shown in Figure 49(j)).

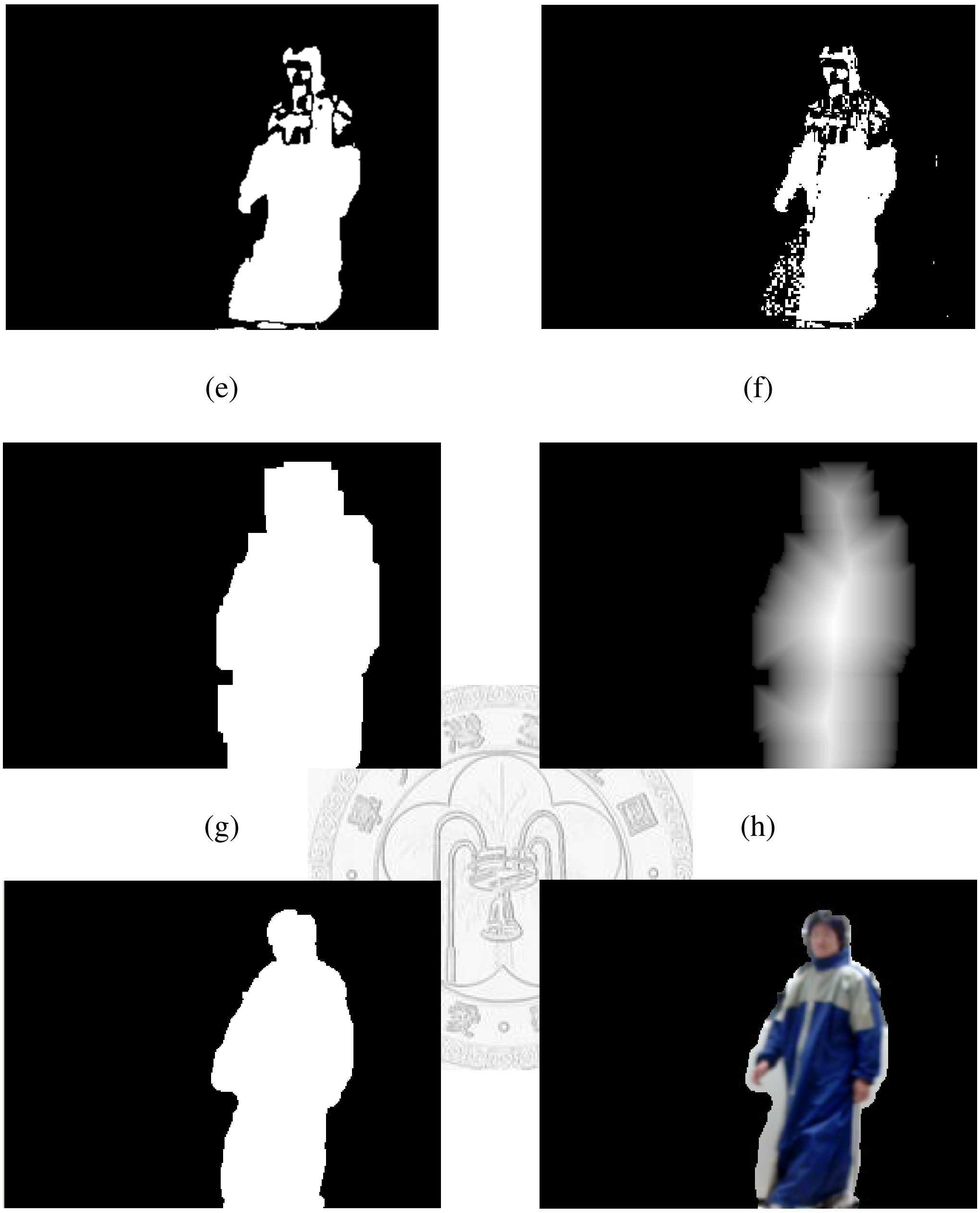
(a)

(b)

(c)



(d)



(e) (f) (g) (h) (i) (j)

Figure 49 (a) Captured frame (b) Background modeling using IRAB (c) Difference frame (d) Foreground detection using thresholding the histogram of difference frame (e) Foreground region (f) DMO region (g) MOC region (h) Weighting map (i) 2^{nd}-staged DMO region (j) Final result using Two-StaBaS approach

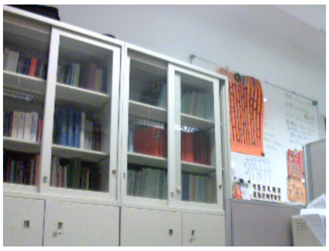
There are three main issues in this subsection. They are (1) reduction of noise and illumination variation, (2) background updating and false positive/ false negative analysis, and (3) the problem of near gray intensity in corresponding pixel of moving object and background. The experiments for three issues are shown in the following.

5.2.1 Reduction of noise and illumination variation

The first experiment emphasizes on the testing of the reduction of noise and various illumination. Noise comes from environment or CCD camera. Variation of illumination comes from sun shine outdoors or fluorescent lamp indoors. The second background updating parameter in IRAB, α_2, is used for reducing above-mentioned problems. Here, design four experiments for verification, there are 300 frames in every image sequence. There are different situations/environments, shown in Table 8. The environments in every experiment are shown in Figure 50.

Table 8 Four different cases for verification the reduction of the effect of noise and illumination

Experiment	Background	Illumination variation
(A)	Simple environment	No
(B)	Complex environment	No
(C)	Simple environment	Yes
(D)	Complex environment	Yes



Experiment (A)

Experiment (B)

Experiment (C)

Experiment (D)

Figure 50 The environments in every experiment

The illumination variation in experiments is produced by someone who turns on or turns off a fluorescent lamp. The experiment (A) and (B) are without illumination variation, only noise is included. The analysis on experiment (A) and (B) focuses on the effect of noise, and whether the effect of noise depends on environment. The experiment (C) and (D) have various illumination, the analysis focuses on the effect of illumination variation.

First define that the average error per frame is the average of difference of captured frame and background in every frame. The four experiments results (the average error per frame versus number of frame) are shown in from Figure 51 – Figure 54, respectively. By observation, the comparison of simple environment and complex environment (Figure 51 v.s. Figure 52 and Figure 53 v.s. Figure 54), the effect of noise and variation of illumination is independent of environment.

In the illumination variation case (see Figure 51 and Figure 54), it is obviously that IRAB suppresses various illumination. The illumination variation makes error pulse in the curve; nevertheless, IRAB reduces the error in few frames.

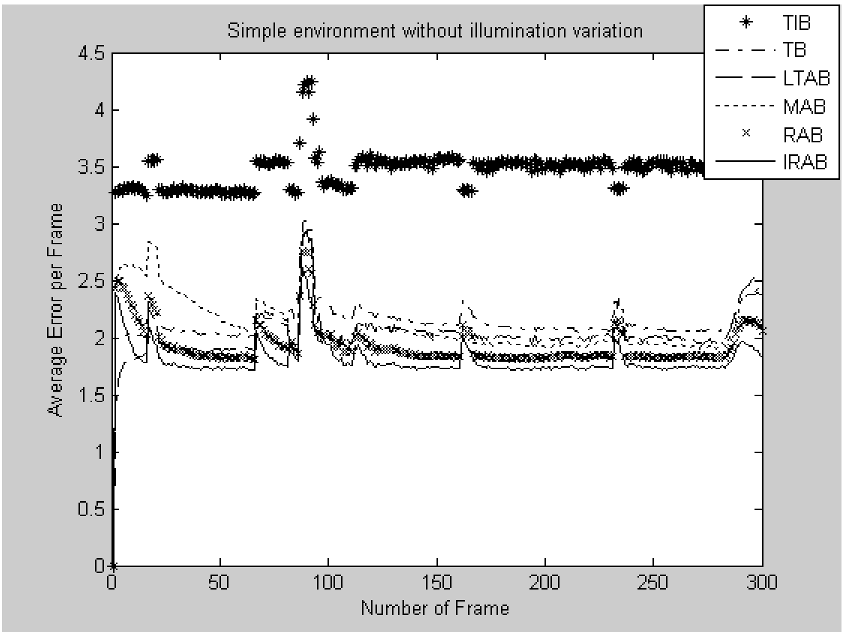


Figure 51 Simple environment without illumination variation

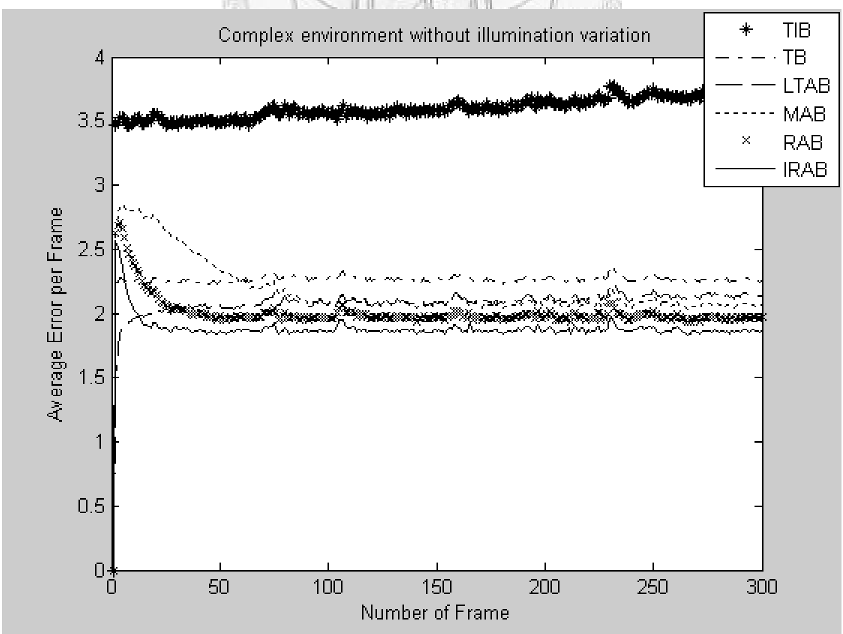


Figure 52 Complex environment without illumination variation

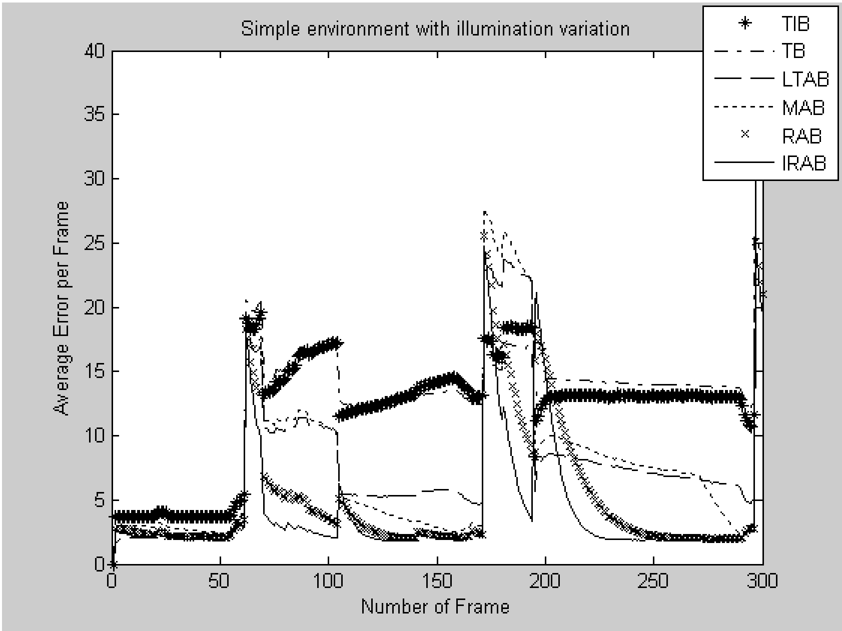


Figure 53 Simple environment with illumination variation

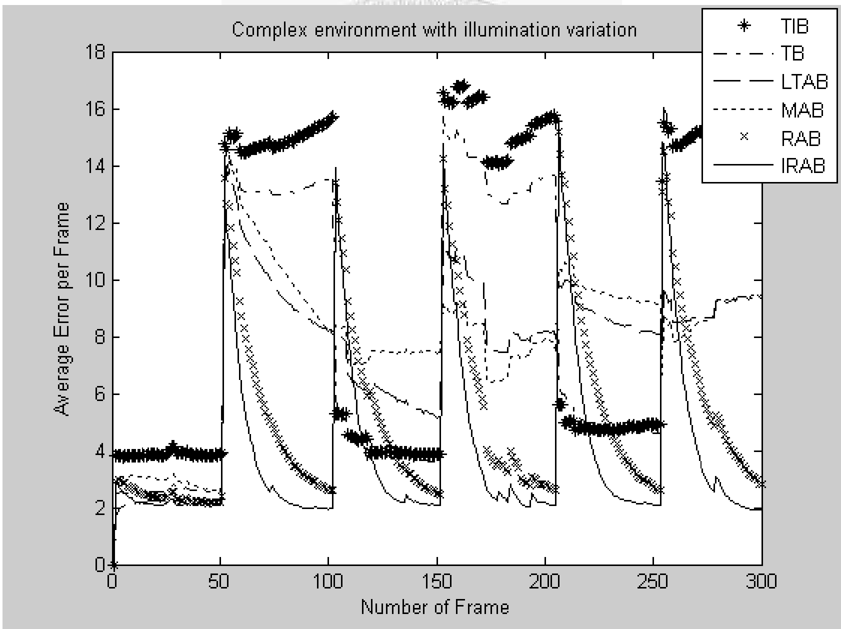


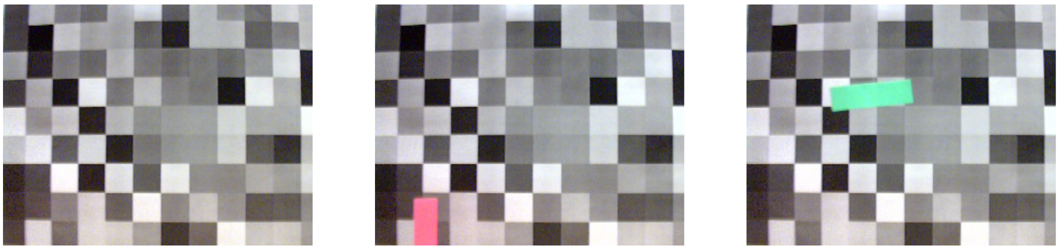
Figure 54 Complex environment with illumination variation

1. the effect of noise and variation of illumination is independent of environment
2. IRAB reduces the illumination faster than others

5.2.2 Background updating and false positive/false negative analysis

Design an image sequence which background has complex gray levels. The background is produced by uniform distribution probability density function, namely, the gray level in background is uniformly distributed (Figure 55(a)). There is a new static object (pink tag) in the background from frame number 320 to frame number 600 (Figure 55(b)), and a moving object (green tag) moves around in the scene that sometimes stops in few frames, sometimes moves fast, and sometimes disappears in the scene (Figure 55(c)). The noise and illumination variation (caused by fluorescent lamp) also be included. Namely, in the designed image sequence, there are noise, various illumination, new static object and moving object. In the next paragraph shows the results in moving object detection.

In the image sequence, the background updates the new static object (pink tag), and in the final, the pink tag appears in the background, shown as Figure 56. Define false positive and false negative are “The pixel belongs to background in the detection result but actually belongs to moving object region” and “The pixel belongs to moving object region in the detection result but actually belongs to background”, respectively. In the designed image sequence, the green tag (moving object) is assigned to be moving object region in the captured frame. Assume the method that detecting green color in captured frame is correct for detecting green tag. Compare the moving object detection using IRAB and green detector method, the false positive and false negative is then obtained in Figure 57 and Figure 58, respectively.



(a) (b) (c)

Figure 55 (a) The gray level in background is uniformly distributed (b) There is a new static object in the background from frame number 320 to frame number 600 (c) A moving object moves randomly

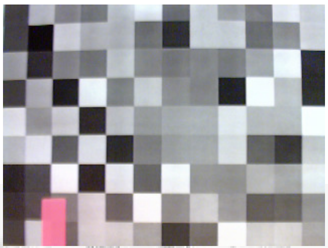


Figure 56 The background updates the new static object (pink tag)

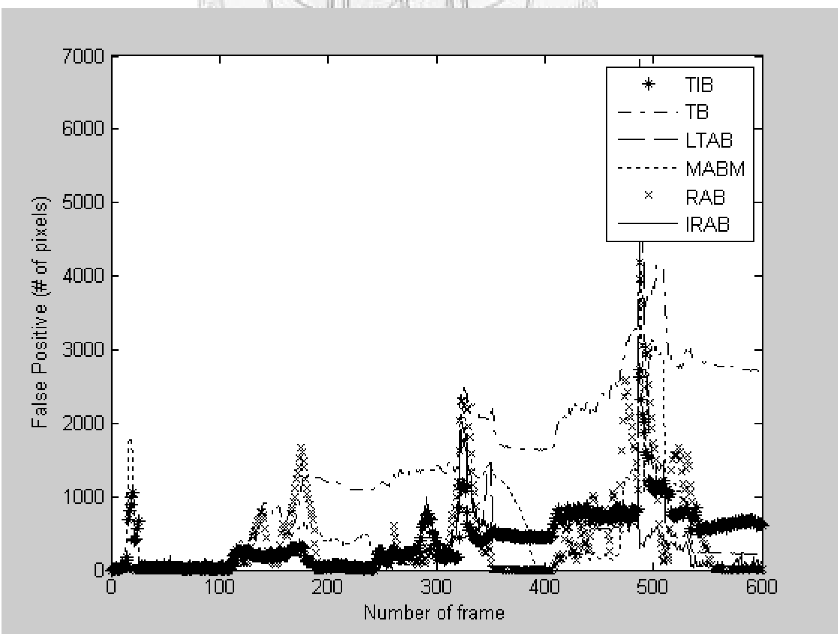


Figure 57 False positive

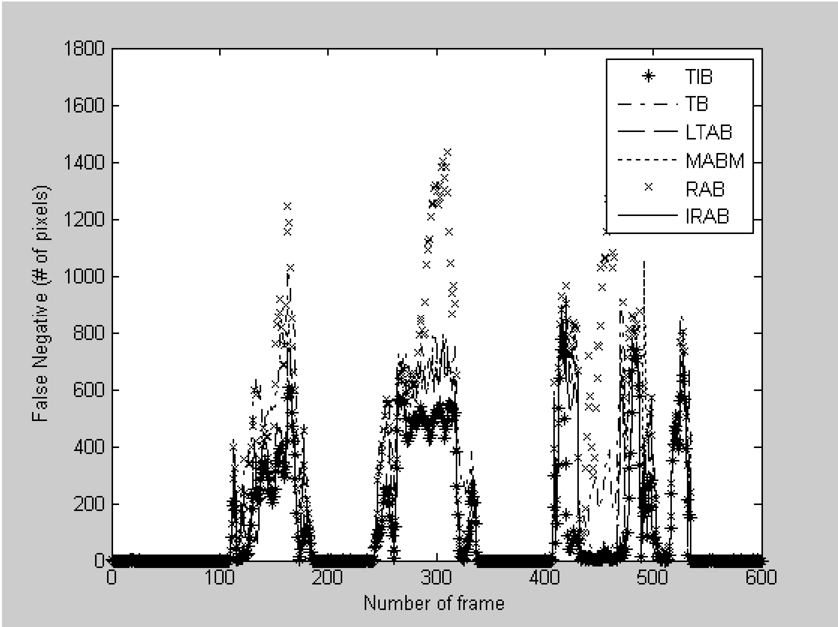


Figure 58 False negative

As the figures shows, the false positive and false negative are smaller than other backgrounds. It not only has good accuracy but also updates new static object which in the background. In Figure 59 compares six backgrounds, some background cannot suppress illumination variation such that its detected moving region has many false positive.

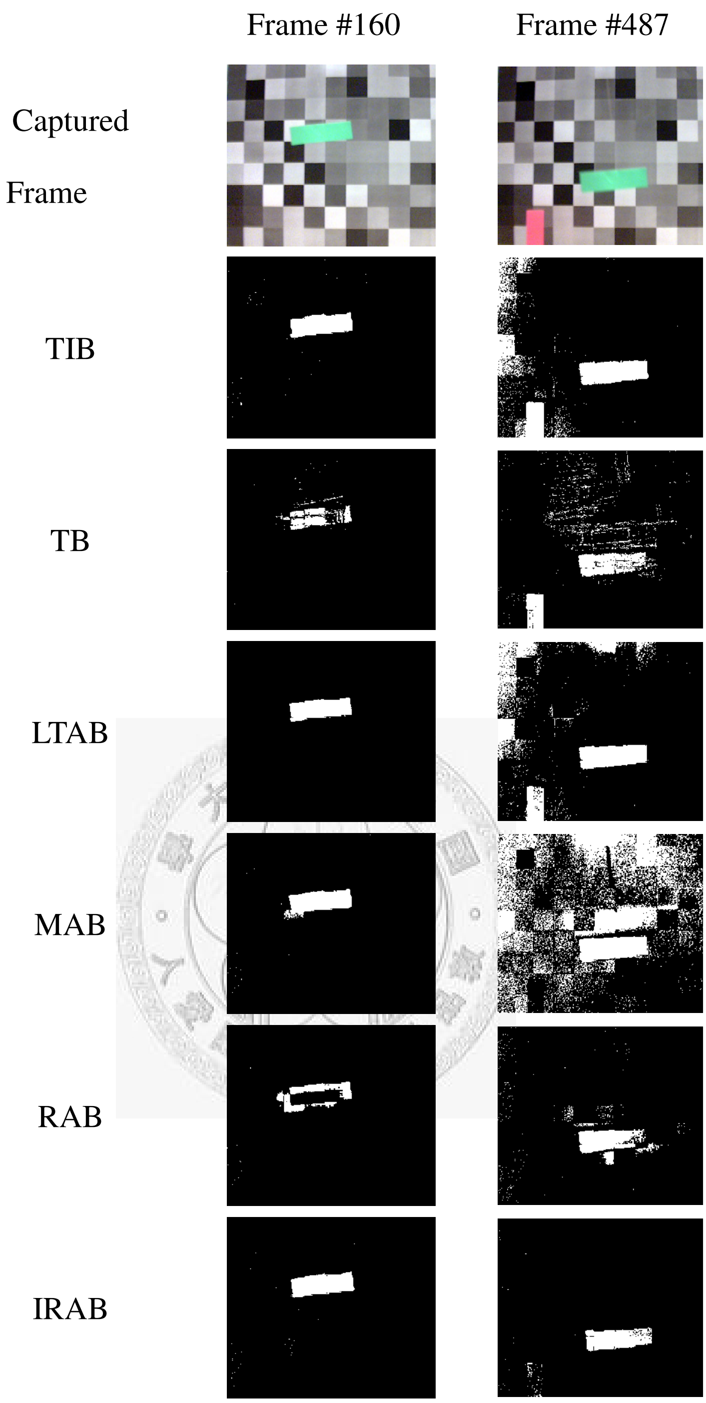


Figure 59 Moving object detection in noisy and various illumination situation

5.2.3 The problem of near gray intensity in corresponding pixel of moving object and background

The near gray intensity in corresponding pixel of moving object and background usually causes false negative. Two-StaBaS reduces this problem. Capture three image sequences for testing.

I. Image sequence 1: Moving Bear Model

Bear model is moved in uniform distributed gray level background. In the image sequence, there are some pixels has near gray level value between bear model and background. The moving object detection result shows in Figure 60(C). The near gray level value causes incomplete DMO region. However, the result in Two-StaBaS is shown in Figure 60(D). The 2nd-staged DMO region has complete contour of bear model.

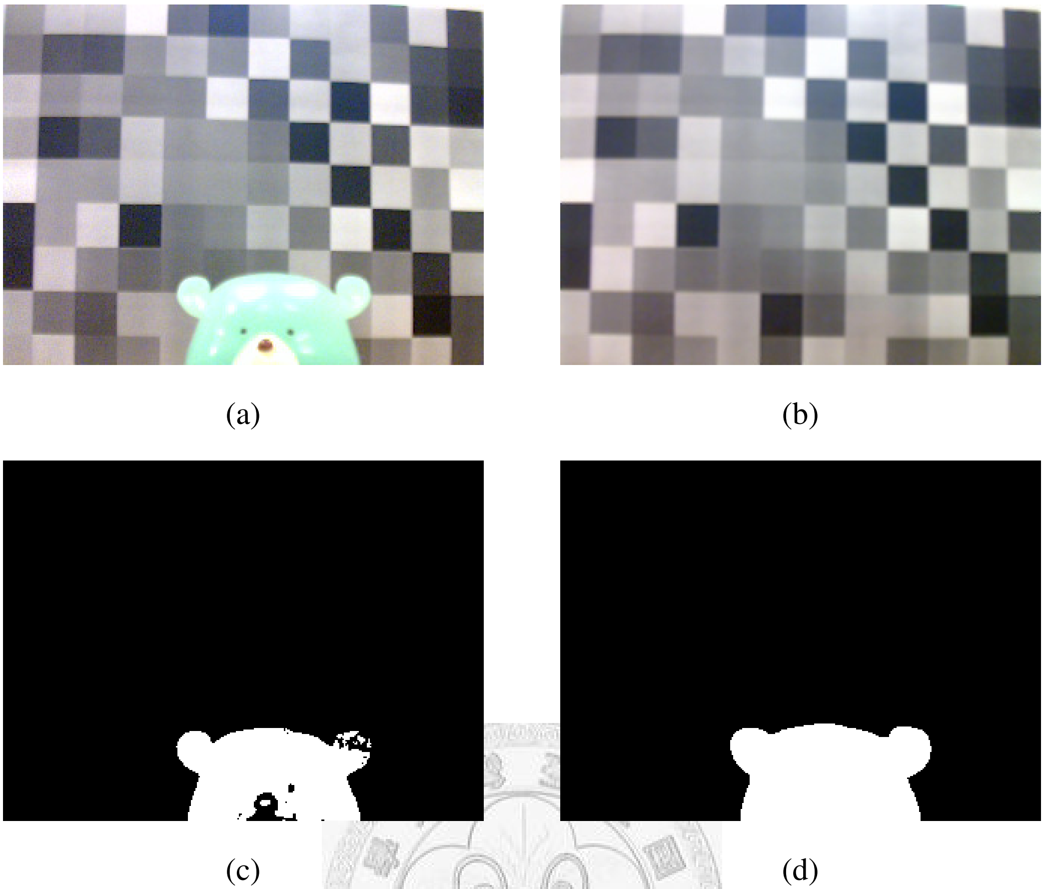


Figure 60 (a) Captured frame (b) Constructed background using IRAB (c) Detecting moving object using IRAB background subtraction method (d) Detecting moving object using Two-StaBaS method

II. Image sequence 2: People walk in the hallway

People walk in the hallway case, the camera does not calibrate and has high brightness. The person is occluded by railings. In the DMO region shown in Figure 61 (C), the region in the railings is not detected and the near gray level value in the cloth and window frame does not detected as DMO region. However, the Two-StaBaS approach reduces the problem.

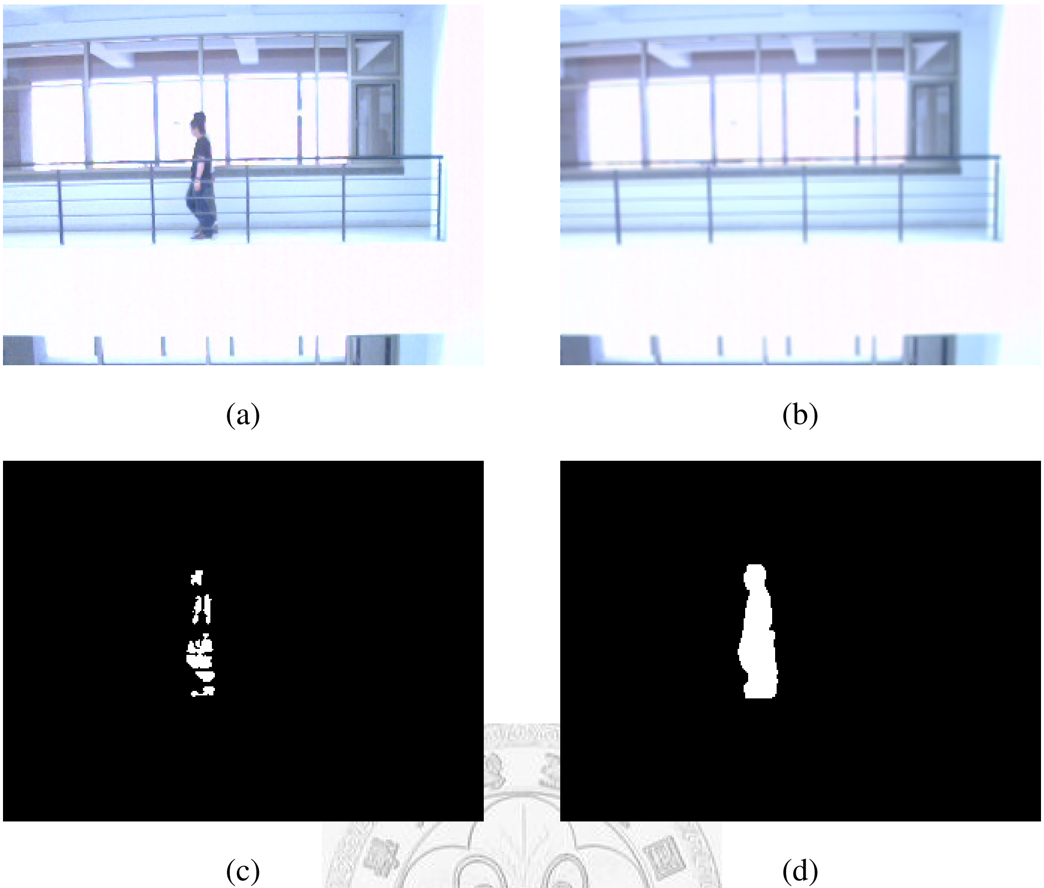


Figure 61 (a) Captured frame (b) Constructed background using IRAB (c) Detecting moving object using IRAB background subtraction method (d) Detecting moving object using Two-StaBaS method

III. Image Sequence 3: Road surveillance

For the road surveillance, there are many objects in the scene, such as human, bicycles, motorcycles, and cars. The speeds of different objects are diverse, and the background updating parameters depends on the slowest object – human. In the scene, some moving objects are occluded by trees and buildings. In the image sequence, the people who walk on the sidewalk and vehicles which drive on the road are detected by Two-StaBaS.

(a)

(b)

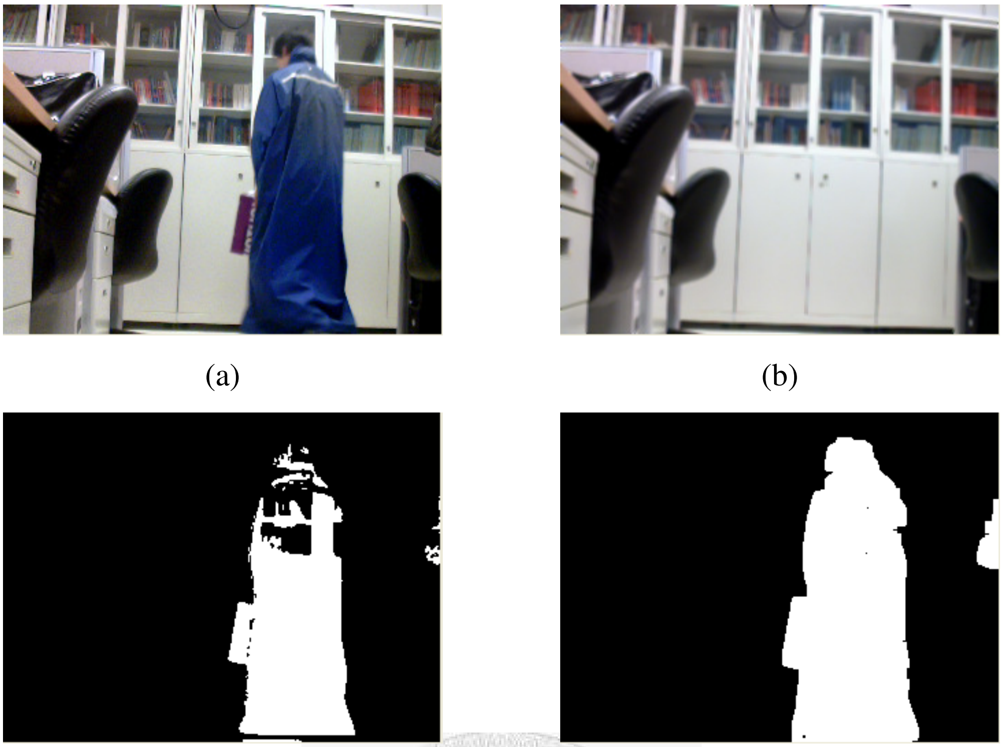
(c)

(d)

Figure 62 (a) Captured frame (b) Constructed background using IRAB (c) Detecting moving object using IRAB background subtraction method (d) Detecting moving object using Two-StaBaS method

IV. Image Sequence 4: Office Surveillance

In the official surveillance, there are two people in the scenario. A person moves in the scene and puts a box on the ground as a new static background. The IRAB updates the box into background. In the scenario, the gray intensity of people and the bookshelf which is in the background are near. In the Figure 63(c), the DMO region is not ideal, however, Two-StaBaS modifies the false detections.



(a) (b)

(c) (d)

Figure 63 (a) Captured frame (b) Constructed background using IRAB (c) Detecting moving object using IRAB background subtraction method (d) Detecting moving object using Two-StaBaS method

5.3 Summary

The specifications of RAB are shown in Subsection 5.1. RAB reduces the variation of noise and illumination. It not only updates backgrounds, but also has high computational speed and low memory requirement. The results of moving object detection based on Two-StaBaS are shown in Subsection 5.2. The first stage is moving object detection using IRAB gets good DMO region. The second stage modifies the false detections in the DMO region such that the 2^{nd}-staged DMO region and MO region are almost the same.

Table 9 The advantages and drawbacks of Two-StaBaS

Advantage	Drawbacks
● Good Accuracy ● Reduce noise, illumination variation ● Backgrounds updating ● Reduce false detections if near color in moving object and background ● Real time operation ● Low memory requirement	● 2^{nd}-staged DMO region is fatter than MO region ● Large DMO region false detections causes false detections in 2^{nd}-staged DMO region ● Connection phenomenon occurs if moving objects near

Chapter 6 Conclusions and Discussions

In this thesis, Two-StaBaS is proposed for moving object detection. There are two stages in Two-StaBaS. The first stage includes background modeling using IRAB and moving object detection using background subtraction approach. IRAB is a fast algorithm and low memory requirement for moving object detection. It not only reduces noise and illumination variation, but also updating backgrounds and removing shadows. The result in the first stage is called DMO region. There are a lot of false detections in the DMO region due to the near color between moving object and background. In the second stage, the process focuses on the reduction of first stage false detections. The method is background subtraction in MOC region with weighting map. The detection result, called 2nd-staged DMO region, actually reduces these false detections. However, the 2nd-staged DMO region is fatter than MO region. If the detection error in the first stage is terribly bad, then the result in the second stage is also not good.

References

[1]. S.T. Su, Y.Y. Chen, "Moving Object Segmentation Using Improved Running Gaussian Average Background Model," *Digital Image Computing: Techniques and Applications*, pp. 24-31, 2008.

[2]. C.R. Wren, A. Azarbayejani, T. Darrell, A. P. Pentland, "Pfinder: Real-Time Tracking of the Human Body," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 19, No. 7, pp. 780-785, 1997.

[3]. C. Stauffer, W.E.L. Grimson, "Learning Patterns of Activity Using Real-Time Tracking," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 22, No. 8, pp. 747-757, 2000.

[4]. N.M. Oliver, B. Rosario, A.P. Pentland, "A Bayesian Computer Vision System for Modeling Human Interactions," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 22, No. 8, pp. 831-843, 2000.

[5]. K. Toyama, J. Krumm, B. Brumitt, B. Meyers, "Wallflower: Principles and Practice of Background Maintenance," *IEEE Inte'l Conf. on Computer Vision*, Vol. 1, pp. 255-261, 1999.

[6]. A. Elgammal, D. Harwood, L. Davis, "Non-Parametric Model for Background Subtraction," In *Proc. of the 6th European Conference on Computer Vision*, pp. 751-767, 2000.

[7]. A. Monnet, A. Mittal, N. Paragios, V. Ramesh, "Background Modeling and Subtraction of Dynamic Scenes," *IEEE Inter'l Conf. on Computer Vision*, Vol.2, pp. 1305-1312, 2003.

[8]. R. Poppe, "Vision-Based Human Motion Analysis: An Overview," *Computer Vision and Image Understanding*, Vol. 108, No.1-2, pp. 4-18, 2007.

[9]. T.B. Moeslund, A. Hilton, V. Krüger, “A survey of advances in vision-based human motion capture and analysis,” *Computer Vision and Image Understanding*, Vol. 104, No. 2-3, pp. 90-126, 2006.

[10]. H.P. Moravec, "Visual Mapping by A Robot Rover," In *Proc. of the 6th International Joint Conference on Artificial Intelligence*, pp. 598-600, 1979.

[11]. C. Harris, M. Stephens, "A Combined Corner and Edge Detector," In *Proc. of The Fourth Alvey Vision Conference*, pp. 147-151, 1988.

[12]. J. Shi, C. Tomasi, "Good Features to Track," *Computer Vision and Pattern Recognition*, pp. 593-600, 1994.

[13]. D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," *Int'l. J. of Computer Vision*, Vol. 60, No. 2, pp. 91-110, 2004.

[14]. K. Mikolajczyk, C. Schmid, "An Affine Invariant Interest Point Detector," *European Conference on Computer Vision*, Vol. 1, pp. 128-142, 2002.

[15]. D. Comaniciu, F. Meer, "Mean Shift Analysis and Applications," In *Proc. of IEEE Int'l Conf. on Computer Vision*, Vol. 2, pp. 1197-1203, 1999.

[16]. J. Shi, J. Malik, "Normalized Cuts and Image Segmentation," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 22, No. 8, pp. 888-905, 2000.

[17]. V. Caselles, R. Kimmel, G. Sapiro, "Geodesic Active Contours," *Inte'l Journal of Computer Vision*, Vol. 22, No. 1, pp. 61-79, 1997.

[18]. C.P. Papageorgiou, M. Oren, T. Poggio, "A General Framework for Object Detection," *IEEE Inte'l Conf. on Computer Vision*, pp. 555-562, 1998.

[19]. H.A. Rowley, S. Baluja, T. Kanade, "Neural Network-Based Face Detection," *Computer Vision and Pattern Recognition*, pp. 203-208, 1996.

[20]. P. Viola, M.J. Jones, D. Snow, "Detecting Pedestrians Using Patterns of Motion and Appearance," *IEEE Int'l Conf. on Computer Vision*, pp. 734-741, 2003.

[21]. M. Brand, "Shadow Puppetry," *IEEE Int'l Conf. on Computer Vision*, Vol. 2, pp.

1237-1244, 1999.

[22]. R. Rosales, S. Sclaroff, "Inferring Body Pose without Tracking Body Parts," *Computer Vision and Pattern Recognition*, Vol. 2, pp. 721-727, 2000.

[23]. N.R. Howe, "Silhouette Lookup for Automatic Pose Tracking," *Computer Vision and Pattern Recognition*, pp. 15-22, 2004.

[24]. S. Belongie, J. Malik, J. Puzicha, "Shape Matching and Object Recognition Using Shape Contexts," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 24, No. 4, pp. 509-522, 2002.

[25]. G. Mori, J. Malik, "Recovering 3D Human Body Configurations Using Shape Contexts," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 28, No. 7, pp. 1052-1062, 2006.

[26]. L. Kakadiaris, D. Metaxas, "Model-Based Estimation of 3D Human Motion," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 22, No. 12, pp. 1453-1459, 2000.

[27]. S. Wachter, H.H. Nagel, "Tracking of Persons in Monocular Image Sequences," *IEEE Proc. of Nonrigid and Articulated Motion Workshop*, pp. 2-9, 1997.

[28]. T. Drummond, R. Cipolla, "Real-Time Tracking of Highly Articulated Structures in the Presence of Noisy Measurements," *IEEE Int'l Conf. on Computer Vision*, Vol. 2, pp. 315-320, 2001.

[29]. K. Rohr, "Towards Model-Based Recognition of Human Movements in Image Sequences," *Image understanding*, Vol. 59, No. 1, pp. 94-115, 1994.

[30]. A. Bottino, A. Laurentini, "A Silhouette Based Technique for the Reconstruction of Human Movement," *Computer Vision and Image Understanding*, Vol. 83, No. 1, pp. 79-95, 2001.

[31]. K.M.G. Cheung, S. Baker, T. Kanade, "Shape-from-Silhouette of Articulated Objects and Its Use for Human Body Kinematics Estimation and Motion

Capture," *Computer Vision and Pattern Recognition*, Vol. 1, pp. 77-84, 2003.

[32]. I. Mikic', M. Trivedi, E. Hunter, P. Cosman, "Human Body Model Acquisition and Tracking Using Voxel Data," *International Journal of Computer Vision*, Vol. 53 , No. 3, pp. 199-223, 2003.

[33]. R. Plankers, P. Fua, "Tracking and Modeling People in Video Sequences," *Computer Vision and Image Understanding*, Vol. 81, No. 3, pp. 285-302, 2001.

[34]. I. Haritaoglu, D. Harwood, L.S. Davis, "W4S: A Real-Time System Detecting and Tracking People in 2 1/2D," *European Conference on Computer Vision*, Vol. 1, pp. 877-892, 1998.

[35]. N. Jojic, J. Gu, H.C. Shen, T. Huang, "3-D Reconstruction of Multipart Self-Occluding Objects," *Asian Conference on Computer Vision*, pp. 455-462, 1998.

[36]. D. Ramanan, D.A. Forsyth, "Finding and Tracking People from The Bottom Up," *Computer Vision and Pattern Recognition*, Vol. 2, pp. 467-474, 2003.

[37]. T.J. Roberts, S.J. McKenna, I.W. Rickettsa, "Human Tracking Using 3D Surface Colour Distributions," *Image and Vision Computing*, Vol. 24, No. 12, pp. 1332-1342, 2006.

[38]. C. Barrón, I.A. Kakadiaris, "Monocular Human Motion Tracking," *Multimedia Systems*, pp. 118-130, Vol. 10, No. 2, 2004.

[39]. M.W. Lee, I. Cohen, "Proposal Maps Driven MCMC for Estimating Human Body Pose in Static Images," *Computer Vision and Pattern Recognition*, Vol. 2, pp. 334-341, 2004.

[40]. C. Cedras, M. Shah, "Motion-Based Recognition: A Survey," *Image and Vision Computing*, Vol. 13, No. 2, pp. 129-155, 1995.

[41]. B.K.P. Horn, B.G. Schunck, "Determining optical flow," *Artificial Intelligence*, pp. 185-203, 1981.

[42]. P. Viola, M. Jones, "Rapid Object Detection Using a Boosted Cascade of Simple Features," *Computer Vision and Pattern Recognition*, vol. 1, pp. 511-518, 2001.

[43]. P. Viola, M. Jones, "Robust Real-Time Object Detection," *International Journal of Computer Vision*, 2001.

[44]. A. Yilmaz, O. Javed, M. Shah, "Object Tracking: A Survey," *ACM Computing Surveys*, Vol. 38, No. 4, pp. 1-45, 2006.

[45]. Larry S. Davis, *Computer Vision Lecture Notes in Marry Land University*, 1999.

[46]. M. Piccardi, "Background Subtraction Techniques: A Review," *IEEE Int'l Conf. on Systems, Man and Cybernetics*, Vol. 4, pp. 3099-3104, 2004.

[47]. D. Koller, J. Weber, T. Huang, J. Malik, G. Ogasawara, B. Rao, and S. Russell, "Towards Robust Automatic Traffic Scene Analysis in Real-Time," *IAPR Int'l Conf. on Pattern Recognition*, Vol. 1, pp. 126-131, 1994.

[48]. N. Friedman, S. Russell, "Image Segmentation in Video Sequences: A Probabilistic Approach," *Uncertainty in Artificial Intelligence*, pp. 1-13, 1997.

[49]. I.T. Jolliffe, "Principal Component Analysis," *Springer*, 2002.

[50]. J. Fazenda, D. Santos, P. Correia, "Using Gait to Recognize People," *Int'l Conf. on Computer as a Tool*, Vol. 1, pp. 155-158, 2006.

[51]. L. Wang, T. Tan, H. Ning, W. Hu, "Silhouette Analysis-Based Gait Recognition for Human Identification," *Pattern Analysis and Machine Intelligence*, Vol. 25, No. 12, pp. 1505-1518, 2003.

[52]. C. Stauffer, W.E.L. Grimson, "Adaptive Background Mixture Models for Real-Time Tracking," *Computer Vision and Pattern Recognition*, Vol. 2, No. 23-25, pp. 246-252, 1999.

[53]. R. Cucchiara, C. Grana, M. Piccardi, A. Prati, "Detecting Moving Objects, Ghosts, and Shadows in Video Streams," *Pattern Analysis and Machine Intelligence*, Vol. 25, No. 10, pp.1337-1342, 2003.

[54]. S.Y. Chien, S.Y. Ma, L.G. Chen, "Efficient Moving Object Segmentation Algorithm Using Background Registration Technique," *IEEE Trans. on Circuits and Systems for Video Technology*, Vol.12, No. 7, pp. 577-586, 2002.

[55]. S.Y. Chien, Y.W. Huang, B.Y. Hsieh, S.Y. Ma, L.G. Chen, "Fast Video Segmentation Algorithm with Shadow Cancellation, Global Motion Compensation, and Adaptive Threshold Techniques," *IEEE Trans. on Multimedia*, Vol.6, No. 5, pp. 732-748, 2004.

[56]. R.C. Gonzalez, R.E. Woods, "Digital Image Processing 2/e," *Prentice-Hall Inc.*, 2002.

[57]. W. LIU, Y.N. LUO, N. SUN, "Mean Shift Tracking Algorithm Based on Background Optimization," *Journal of Computer Applications*, Vol. 29, No. 4, 2009.

[58]. R. Cucchiara, C. Grana, M. Piccardi, A. Prati, S. Sirotti, "Improving Shadow Suppression in Moving Object Detection with HSV Color Information," *IEEE Int'l Conference on Intelligent Transportation Systems*, pp. 334-339, 2001.

[59]. A. Prati, R. Cucchiara, I. Mikic, M.M. Trivedi, "Analysis and Detection of Shadows in Video Streams: A Comparative Evaluation," *IEEE Int'l Conf. Computer Vision and Pattern Recognition*, Vol. 2, No. 2, pp. 571-576, 2001.

[60]. P. Viola, M.J. Jones, "Robust Real-Time Face Detection," *Int'l Journal of Computer Vision*, Vol. 57, No. 2, pp. 137-154, 2004.

[61]. Y. Freund, R.E. Schapire, "A short introduction to boosting," *Journal of Japanese Society for Artificial Intelligence*, Vol. 14, No. 5, pp. 771-780, 1999.

[62]. R.E. Schapire, "The Boosting Approach to Machine Learning: An Overview," *Nonlinear Estimation and Classification, Springer*, 2003.

[63]. http://homepages.inf.ed.ac.uk/rbf/HIPR2/distance.htm