SSI-DT000000

Doctoral Thesis

Co-occurrence Pixel-Block Background Model and its Application to Robust Event Detection

Wenjun ZHOU

June, 2019

Division of Systems Science and Informatics
Graduate School of Information Science and Technology
Hokkaido University

Doctoral Thesis
submitted to Graduate School of Information Science and Technology,
Hokkaido University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy.

Wenjun ZHOU

Thesis Committee: Shun'ichi KANEKO Professor
Satoshi KANAI Professor
Takayuki TANAKA Associate Professor

Co-occurrence Pixel-Block Background Model and its Application to Robust Event Detection*

Wenjun ZHOU

Abstract

As a basic approach utilized in many computer vision applications, foreground detection plays an important role in various tasks like video surveillance, traffic monitoring, scene background initialization and object tracking. One simple way to do background model is to acquire a background image without any moving objects. However, foreground detection is faced with many practical challenges, especially the background changes, not least of which is related to illumination changes, e.g. variable sunlight or lights being switched on and off indoors, and background motion, e.g. the swaying motion of the trees, fleeting cloud and moving waves on the water. To handle such challenges, previous statical methods have been proposed, in which the intensity of each pixel is independently analyzed in the temporal domain and then the current frame is subtracted, such as the Gaussian Mixture Model (GMM) to build a pixel-wise model for each pixel, however such kind of methods is difficult to solve illumination changes with the intensity varies rapidly and significantly. Recent many local feature based methods have been put forward for background modeling such as Barnich et al. proposed ViBe, a method that involves comparing each pixel with a set of previous values located the same or neighboring positions to evaluate whether a pixel belongs to the background. However, such local feature based background models are susceptible to be affected by the dynamic motion of the background, thus losing the robustness.

To overcome above problems, this thesis presents a novel background subtraction method called Co-occurrence Pixel-Block pairs (CPB) for detecting objects in dynamic scenes. CPB is a "pixel to block" structural model, which is evolved from the Co-occurrence Probability based Pixel Pairs (CP3) and it uses the correlation of multiple co-occurrence pixel block pairs to detect objects in dynamic scenes. It offers robust background subtraction against a dynamically changing background. We firstly propose a correlation measure for co-occurrence pixel-block pairs to realize a robust background model. We then introduce a novel evaluation strategy named correlation depended decision function for accurate object detection with the correlation of co-occurrence pixel-block pairs. Finally, CPB can estimate the foreground from a dynamic background with a sensitive criterion. Furthermore, a Hypothesis on Degradation Modification (HoD)

*Doctoral Thesis, Division of Systems Science and Informatics, Graduate School of Information Science and Technology, Hokkaido University, SSI-DT000000, June 6, 2019.

based on CPB is proposed to further resist background changes for foreground detection, such as illumination changes and background motion. HoD provides CPB with a model update strategy that can be used for a long time. HoD further improves the robustness of CPB, and stabilizes the efficiency of CPB over time.

Through the experimental comparisons with other existing foreground detection techniques based on challenging datasets, we demonstrated the good performance of our algorithms. In summary, CPB is sufficiently sensitive to detect foreground objects in dynamic scenes and CPB performs robust detection in outdoor or indoor environments with relatively low complexity. Furthermore, HoD provides a new and natural thought: the structure of background model can be updated by the designed correlation weigh, which is a new strategy can be utilized in the pixel-correlation based algorithms for the background model update.

This thesis is organized into the following chapters:

Chapter 1 introduces the related works in foreground detection. Some general problems are involved and discussed. Furthermore, the motivations and contributions of this study are described.

Chapter 2 introduces the Co-occurrence Pixel-Block Background Model (CPB) in detail, including the basic concept and essential mechanism of CPB. As an extension from the "pixel to pixel" structure that our previous work CP3, CPB proposes a "pixel-block" structure for the background model. In this chapter, we describe how to construct the "pixel-block" structure for background model and explain the process of model building in theory.

Chapter 3 discusses the application of CPB in the field of the foreground (event) detection. We also introduce a novel evaluation strategy named correlation depended decision function for accurate foreground detection and explain the theoretical meaning of the evaluation strategy. Moreover, we do a comparison to present the performance of CPB for foreground detection.

Chapter 4 focuses on the Hypothesis on Degradation Modification (HoD), which is proposed based on CPB to further improve the robustness of CPB and stabilize the efficiency of CPB over time. In this chapter, the basic knowledge and mechanism of HoD are discussed in detail. Finally, we verify the ability of HoD with adequate experiments.

Chapter 5 introduces the experimental setup in detail. In this chapter, the comparative experiments for CPB and CPB+HoD using several challenging datasets are designed and through these experiments we measure the robustness and efficiency of our methods, CPB and CPB+HoD in various indoor and outdoor challenges.

The final Chapter summarizes the main points of the study and discusses our algorithms. Finally, the plan and concept of future work are presented.

Keywords: Co-occurrence pixel-block pairs (CPB),Hypothesis on degradation modification (HoD), Foreground detection, Illumination changes, Background motion

Contents

List of Figures

List of Tables

Chapter 1. Introduction

1.1 Research background

Foreground detection is a critical lower-level vision task in many computer vision applications that separates moving objects, called the foreground, from static information, called the background, and classifies pixels as either foreground or background in the scene [1]. All foreground detection techniques are based on modeling the background of the image, by setting the background and detecting changes. Defining the background can be difficult when it contains shapes, shadows, and moving objects. In defining the background, it is assumed that stationary objects can vary in color and intensity over time. A common classification method is background subtraction, in which each frame in a video sequence is compared with a background model and pixels that deviate significantly from the background information are considered to be outliers and are identified as the foreground [2].

Background subtraction is generally used for images that are part of a video stream and it can provide an important cue for numerous applications in computer vision. The method is used for pre-processing in many computer vision tasks, including video surveillance [3, 4], traffic monitoring [5], understanding and describing human behavior [6, 7], object tracking [8, 9], and scene background initialization [10, 11]. Foreground is any change in a video sequence. Usually, the foreground comprises moving objects that are of interest, including pedestrians, vehicles, and animals, while the background consists of stationary objects that can vary in color and intensity under illumination changes over time [12]. One common and simple method of background subtraction is to acquire a background image with no moving objects and perform a difference operation between the current frame and background image. In general, foreground and background should be defined on the basis of the ground truth. However, foreground detection may prove challenging in complicated scenes and remains a difficult task to solve.

There are two particularly difficult challenges.

- **Illumination changes:** for example, changes in sunlight and lights being switched on and off indoors
- **Background motion:** scenes with strong (parasitic) background motion, such as trees moving in the wind, fleeting clouds, changing advertising boards, and waves on a body of water.

Typical examples of these challenges are shown in Fig. 1.1, and Fig. 1.2 shows two typical results of background subtraction in the presence of illumination changes and background motion.

To tackle these challenges, there are numerous background subtraction methods,most of which follow a simple processing procedure consisting of the following four main steps [2].

- **(1) Pre-processing:** the techniques like raw input frames denoising or gray-scale transformation to provide efficient data;
- **(2) Background modeling:** to achieve the background information;
- **(3) Foreground detection:** to extract the moving objects from background;
- **(4) Post-processing:** which is utilized to eliminate the error of foreground detection such as filtering.

Fig. 1.3 shows the flow diagram for background subtraction. Steps (2) and (3) are the main processes in background subtraction.

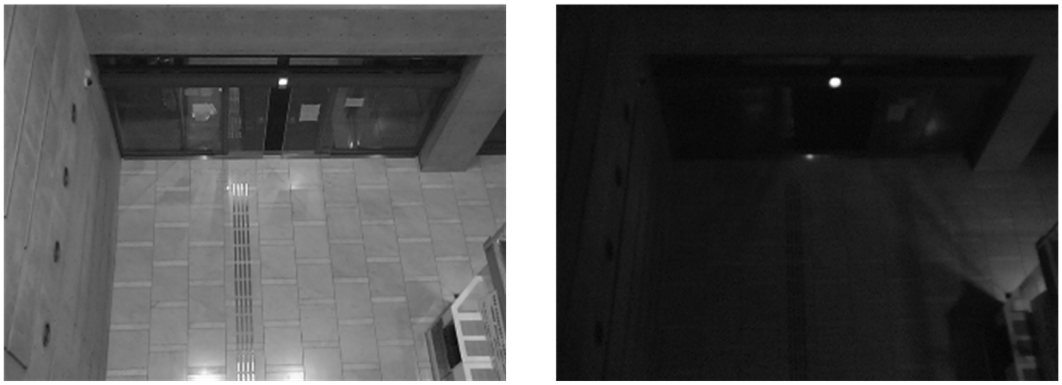
- **Background modeling:** builds a mathematical model to describe the background information and represent the background characteristics; this step determines the ability of the model to handle unimodal or multimodal backgrounds
- **Foreground detection:** classifies pixels as background or as foreground and labels them; this step is a classification task

In this thesis, we brief introduce different approaches to the following modeling methods.

- **Traditional background modeling:** (1) general basic models that were the first used in the field; (2) useful for solving specific cases; simple models; (3) underpin many improved algorithms and (4) may have reached their full potential
- **Recent background modeling:** (1) more complex and can handle more challenges; (2) most need to be improved to cope with real-time requirements

Table 1.1 shows some typical modeling methods in these two categories. In Sections 1.2 and 1.3, we introduce these approaches.

(a) variable sunlight outdoor



(b) lights being switched on and off outdoor

(c) waving water

(d) everchanging advertising board

Figure 1.1: Typical challenges of severe scenes.

1.2 Traditional background modeling

1.2.1 Basic model

McFarlane et al. [13] developed an algorithm with a median background and a Laplacian operator for segmenting and tracking piglets in surveillance video. In [14],

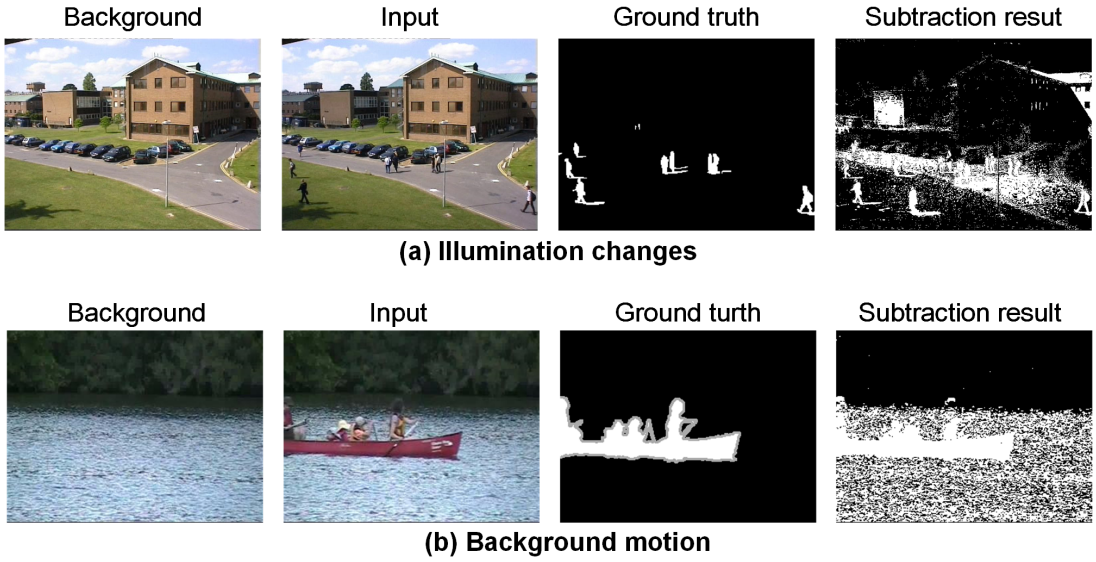


Figure 1.2: The typical example results of illumination challenges and background motions by using the static frame difference approach, (a) Illumination changes: one sequence with the light intensity typically varies during day. (b) Background motions: one sequence with the water rippling.

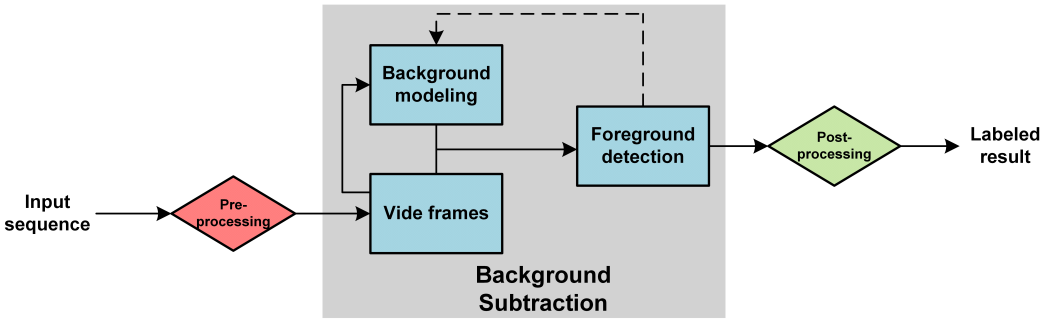


Figure 1.3: Flow diagram for background subtraction.

the background model is built through an average for extracting objects. Once the background is modeled, pixels of the current frame are classified as foreground by using a threshold to distinguish the difference between the background and foreground.

$$State = \begin{cases} d(I_t(x,y), B_{t-1}(x,y)) > T & foreground \\ otherwise & background \end{cases} \tag{1.1}$$

Here, T is a constant threshold and $I_t(x,y)$ and $B_{t-1}(x,y)$ are the current frame at time t and the background at time $t-1$, respectively. These models are easy to implement and can handle some simpler environments; however, they cannot adapt to complex situations, such as sudden illumination changes.

Table 1.1: Typical modelings of traditional and recent background modelings

Background modeling	Categories	Approaches
Traditional	Basic model	Median[13], Mean[14]
	Statistical model	GMM[15], KDE[16], PCA[17]
	Clusters model	Codebook[18], K-means[19]
Recent	Non parametric model	ViBe[20], PBAS[21] SuBSENSE[22]
	Co-occurrence model	SRF[23], GAP[24], CP3[25]
	Fuzzy model	T2FMOG-UM[26], T2FMRF-UV[27], FCM[28]

1.2.2 Statistical model

Gaussian mixture model (GMM)

To handle dynamic backgrounds, such as waving trees, or algae floating atop gently undulating water, a background model needs to cope with multiple background patterns. The Gaussian mixture model (GMM) (also called the mixture of Gaussian model) [15] has been used to adapt the dynamic background. Illumination in the real world can change gradually (e.g., sunlight from dawn to dusk) or suddenly (e.g., lights switching on or off indoors). Under these conditions, an object could be introduced to or removed from the scene. To adapt to changes, a mixture of K Gaussian models is introduced rather than relying on a single Gaussian model to cope with multiple types of background information. Numerous improvements [29, 30, 31, 32, 33, 34, 35, 36, 37, 38] based on GMM have been developed for robust object detection in the critical scene. For example, an efficient adaptive algorithm using Gaussian mixture probability density, which can select the required number of components per pixel automatically and is more adaptive to the scene, has been reported [31]. Whereas, to build a proper number of Gaussian distributions will not be accurate, GMM is sensitive to sudden changes in illumination. A wide background distribution will result in poor foreground detection in a large intensity range, which interferes with the real background information.

Kernel density estimation (KDE)

Elgammal et al. [16] used kernel density estimation (KDE), a non-parametric method that can be used to detect objects in dynamic scenes to estimate background probabilities at each pixel from many recent samples over time. In [16], the probability density function of the background is defined as

$$P(I_{s,t}) = \frac{1}{N} \sum_{i=t-N}^{t-1} K(I_{s,t} - I_{S,I}) \tag{1.2}$$

Here, K is the kernel and N is the number of previous frames. Several improvements on the basic KDE method are described in [39, 40]. For example, Seki [41] proposed

a method that involved estimating the co-occurrence correlation between neighboring blocks through the KDE estimation. However, this mechanism is time-consuming and has difficulty handling large, fast illumination changes because it is prone to interference due to varying illumination. Various improvements [42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53] have been developed for the background model, initialization, background maintenance, and foreground detection.

Principal component analysis (PCA)

Oliver et al. [17] used spatial eigenvalues to model each background pixel with principal component analysis (PCA), where N frames are used to build a background model; that is, the first p-significant feature vector extracted from the PCA is represented by the average frame and the projection matrix. Based on this method, foreground segmentation is performed by calculating the difference between the input frames and the reconstruction. To overcome the limitations of PCA, numerous improvements have been proposed in recent years [54, 55, 56, 57, 58, 59].

1.2.3 Clusters model

Codebook

A pixel-wise method using a codebook model for background modeling was proposed by Kim et al. [18]. The codebook, which consists of one or more codewords for each pixel, is computed. Then, each pixel is clustered into a set of codewords with the color distortion metrics and luminance boundaries. Due to the activity of pixels, the number of codewords for each pixel is different. After a cluster represented by the codeword is linked to a single Gaussian or other parameter distribution, a background model can be encoded on a pixel-by-pixel basis. By calculating the difference of the current frame from the background model with respect to color or brightness differences, the foreground can be extracted.

K-means

Butler et al. [19] proposed an algorithm based on K-means that represents each pixel in a frame with a group of clusters. These clusters are ordered by likelihood, and then the background is modeled. This method can be adapted to handle background and lighting variations. Incoming pixels are matched against the corresponding cluster group, and are classified according to whether the matching cluster belongs to part of the background or not.

1.3 Recent background modeling

1.3.1 Non parametric models

Visual background extractor(ViBe)

The visual background extractor, ViBe, proposed by Barnich et al. [20], is a method that compares each pixel with a set of previous values at the same or neighboring positions to evaluate whether a pixel belongs to the background. It is a simple algorithm with a classification model based on a small number of correspondences between a candidate set and the corresponding background model for each pixel. ViBe provides a random selection policy that supplies a smooth exponential decay that can be predicted easily to construct the pixel models. Moreover, ViBe uses a fast spatial information mechanism that randomly diffuses pixel values across neighboring pixels, which are set as the supporting candidates. Consequently, ViBe can handle challenging conditions, such as darker backgrounds, shadows, and frequent background changes and it also first proposed a novel design for the large family of background subtraction algorithms.

Pixel-based adaptive segmenter (PBAS) (PBAS)

The pixel-based adaptive segmenter (PBAS) [21] developed by Hofmann et al. models the background from recently observed pixel information. The foreground classification depends on a decision threshold and the background update is based on a learning parameter. PBAS introduced a feedback loop with two controllers for updating the decision threshold and the learning parameter to adapt to dynamic background changes.

SuBSENSE

SuBSENSE [22], a recent method following ViBe's strategy to build a nonparametric background model with the local binary similarity patterns [60] features. By using pixel-level feedback loops, SuBSENSE can adjust the internal parameters of the background model dynamically with no manual settings; thus, it can handle complex surveillance scenes and present many different challenges simultaneously.

However, all the above algorithms are local feature background models that are likely to be affected by dynamic background motion, and thus are not robust.

1.3.2 Co-occurrence models

Statistical reach feature (SRF)

The statistical reach feature (SRF) method [23] uses a set of statistical pair-wise features to build a model of the local texture for each pixel against various disturbances in the real world, including local or global variations of illumination, occlusions, and noises. Through sign coding, the SRF method acquires the local texture feature for each target pixel, P, comparing P with its neighboring pixels, Q_n, by

$$srf(P \succ Q_n; T_P) = \begin{cases} 1 & I(P) - I(Q_n) \geq T_P \\ 0 & I(P) - I(Q_n) \leq -T_P \\ \phi & otherwise \end{cases} \tag{1.3}$$

Here, T_P is the threshold of intensity difference. Multiple pixels Q_n are used and the resulting code string is used to calculate the local similarity of P. Fig. 1.4 shows an example of the local coding for P. Based on the sign coding, the similarity comparison between P and Q_n is used to judge the state of P and determine whether it belongs to the background.

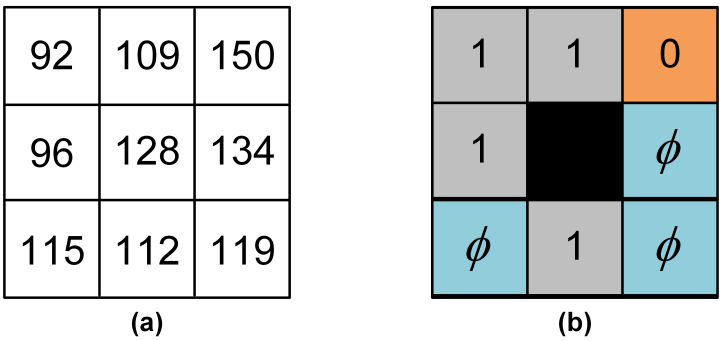


Figure 1.4: Local coding for target pixel P : (a) Example of intensities of 3×3 subregion centered at the target pixel P; (b) Values at each supporting Q_n, $T_P = 15$.

However, the SRF model is not sufficiently robust because its searching order is based on the neighboring pixels on the edge of an image; thus, each direction of P that satisfies the requirements is stored in the order, irrespective of its sign. In the case, to control the global features but not the individual Q_n with individual signs.

Gray-scale arranging pairs (GAP)

Gray-scale arranging pairs (GAP) [24] is a co-occurrence model based on the SRF model [61], and is defined as a set of statistical pair-wise features. First, GAP analyzes the stability of intensity between point pairs. The intensity difference shows better stability than the intensity, even in complex environments. Then, new selection rules are created to choose appropriate point pairs that maintain a stable intensity difference during changes in the global spatial domain [24]. For example, for each target pixel P, there are N positive supporting pixels that have higher intensity than P in the statistics, and N negative supporting pixels have lower intensity than P in the statistics. Based on this mechanism, a co-occurrence structure between target pixel P and its multiple supporting pixels can be constructed as the background model. For the current frame, P is calculated as the background when its intensity is steady; that is, lower than the positive supporting pixels and higher than the negative supporting pixels. In contrast, when the intensity of P is higher than the positive supporting pixels or lower than the negative supporting pixels, this situation is set as an unsteady state and P is classified as the foreground. GAP has been extended to object tracking [62], industrial defect detection [63], and facial recognition [64].

Co-occurrence probability-based pixel pairs(CP3)

In our previous work [25, 65, 66], we proposed the co-occurrence probability-based pixel pairs (CP3) method based on GAP [24] and SRF [23] for background modeling using a pixel-to-pixel structure. The method involves building a background model of a dynamic scene from pairs of pixels that have a high probability of spatial co-occurrence. Instead of using single pixels, CP3 uses the stable relationships between the pixels in co-occurring pairs; this is a more reliable method, especially for dramatically changing environments [25]. First, target pixel p is selected and one or more supporting pixels Q^p that have high normalized correlations with the target pixel are determined. Objects are detected by comparing the pixel pairs, $(p, \{Q_k^p\}_{k=1,2,\ldots,K})$, in turn.

Although CP3 is effective at dealing with unstable backgrounds, it suffers from the open problem of time consumption. For instance, for T sequence frames of size $U \times V$, we have $U \times V \times T$ intensity values within the range [0, 255] of gray-scale levels. To determine Q^p for one target pixel, CP3 must compare the target pixel with all the other pixels in the T sequence frames, and then acquire stable Q^p. This is a time-consuming approach; as the frame size increases, CP3 becomes computationally expensive. Practicality means that the computational cost must be weighed against the benefits of the approach and a more efficient strategy should be sought.

1.3.3 Fuzzy model

Type-2 fuzzy mixture of Gaussians

Fida El Baf et al. proposed the type-2 fuzzy mixture of Gaussians model (T2FMOG-UM) [26], which is used to build an uncertainties background model when there is a dynamic background. They introduced the uncertainty over the mean into the algorithm resulting in more robust adaptation than the simple GMM strategy [15]. Due to the related uncertainty, T2FMOG-UM can extract the foreground under challenging situations.

Type-2 fuzzy Markov random field model

Zhao et al. proposed a hybrid method for motion detection in dynamic backgrounds based on a T2FMOG-UM and a Markov random field. The model introduces spatio-temporal constraints into the T2FMOG-UM under a Bayesian framework to handle dynamic backgrounds and detect shadow easily.

Feature clustering model

W. Kim and C. Kim[28] designed an adaptive clustering-based feature using the fuzzy color histogram and such model can greatly attenuate color variations generated by background motions while still highlighting moving objects. This model is a simple way for background subtraction in dynamic scenes with sufficient texture and achieved a better results than the MOG[15] under dynamic background.

1.4 Other approaches

Hybrid models are also commonly used for background subtraction. For example, Ding et al. [67] proposed a mixture of a nonparametric regional model, KDE, with a parametric pixel-wise model, GMM, to estimate the color distribution of the background. The color distribution of the foreground can be extracted from neighboring pixels in the previous frame, and then the local distributions of the background and foreground are estimated by KDE. Independent multimodal background subtraction [68] is a real-time algorithm for creating a robust mechanism with the multimodal background model, which is sensitive to dynamic changes in background with good accuracy while maintaining real-time performance. The algorithm is based on separating the foreground and background in different domains. For example, Wren and Porikli [69] used fast Fourier transform to estimate the background information by capturing the spectral signatures of multimodal backgrounds, which are used to evaluate changes of scene when the background information differs from the signatures.

Pattern classification based on convolutional neural networks (ConvNets) [70] is be-

coming widespread in background subtraction. For example, Braham et al. [71] developed an algorithm based on spatial features learned with ConvNets for scene-specific backgrounds. Zhang et al. [72] used deep learning techniques [70, 73] for the intrinsic high-level structural learning of scene information based on a huge amount of generic natural image patches, and then constructed a binary scene model for background subtraction. Sultana et al. [74] proposed a method called deep context prediction by using a generative adversarial network and image in-painting, which is a context prediction algorithm for foreground segmentation.

1.5 Introduction of the study

1.5.1 Challenges of background model

Background subtraction algorithms must overcome several challenges in real applications. According to the work of Brutzer et al. [75] and the classical definition by Toyama et al. [76], the challenges can be classified as follows.

- **Gradual illumination changes:** the light intensity typically varies during the day; for example, for outdoor images, the background model must adapt to the gradual changes in the appearance of the environment
- **Sudden illumination changes:** suddenly switching a light on affects the observation of an object, leading to errors in detection; one-off changes cannot be included in the background model
- **Dynamic background (background motion):** background motion includes regular movement (e.g., rippling water) and sudden movement (e.g., changing advertising boards), which must be classified correctly as background
- **Camouflage:** some objects may appear similar to the background and may be difficult to extract; for example, video surveillance in fog is particularly difficult for background subtraction
- **Shadows:** in videos with many hard and soft shadows and intermittent light, shadows may interfere with foreground extraction; in particular, overlapping shadows in foreground regions hinder separation and classification
- **Bootstrapping:** if data free from foreground objects is not available, a bootstrapping strategy should be used for background model initialization.

- **Video noise:** generally, the video signal is superimposed by noise; background subtraction methods must overcome such signals in degradation affected by noise (e,g., the noise of sensor or video compression)
- **Intermittent object motion:** object detection is difficult for background objects moving away, abandoned objects, and objects stopping for a short time, and then moving away; background subtraction has difficulty distinguishing changes in objects from the background accurately
- **Camera jitter:** in video surveillance, camera jitter is a problem that must be solved for background subtraction

This thesis focuses on illumination changes and dynamic backgrounds (background motion), which present major challenges, such as implementing background subtraction for real scenes with changeable backgrounds.

1.5.2 Motivation of the study

We discussed previous methods for overcoming the challenges in the background model in Sections 1.2 and 1.3, and these methods can be classified as follows.

- **Pixel-wise model:** the intensity of each pixel is analyzed independently in the temporal domain, and then the current frame is subtracted. For example, Pfinder [15] reported a real-time method for analyzing the color information (Y/U/V) of each pixel and building a pixel-wise model from the GMM [77]. Friedman et al. [78] built a pixel-wise model for traffic surveillance by using three mixed Gaussian functions. Elgammal et al. [16] used KDE, a non-parametric method that can be used to detect objects in dynamic scenes. Despite being efficient, these various methods have fallen out of use because they cannot effectively handle illumination changes in the absence of contextual spatial information, and because they are prone to interference due to varying illumination
- **Spatial model:** a background model is built by determining the spatial correlations between pixels or blocks. Seki [41] proposed a method that involved estimating the co-occurrence correlation between neighboring blocks. Matsuyama [79] proposed a block-matching method called the normalized vector distance for detecting objects in non-stationary scenes. Oliver et al. [17] used spatial eigenvalues to model each background pixel with PCA. Sheikh et al. [80] used range information and the correlation between proximal pixels to build both foreground and background KDE models for object detection. ViBe [20] is a method that involves comparing each pixel with a set of previous values located at the same or neighboring positions to evaluate whether a pixel belongs to the background. SuBSENSE [19], a method that follows ViBe's strategy to build a nonparametric background model with local binary similarity pattern features [60]. However,

most of these methods focus on the local spatial information of neighboring pixels or blocks and ignore global spatial information. Background models based on local features are affected by the dynamic background motion, decreasing their robustness.

In this thesis, a robust foreground detection method called the co-occurrence pixel-block pairs (CPB) background model [81] for handling large background changes is proposed to overcome these problems. To prevent background changes degrading foreground detection, we introduce the Hypothesis on Degradation Modification (HoD; described briefly in [82, 83, 84, 85]) into CPB. In our previous work [25], we proposed a pixel-to-pixel structure strategy to estimate target pixel p with other pixels one by one and select suitable supporting pixels for p. This strategy deals with background changes effectively; however, it suffers from the open problem of time consumption. To overcome this problem, we must find a strategy to avoid the problem with CP3 [25]; therefore, approaches using super-pixels for cost reduction in follow-up processing and for image matting [86, 87, 88] may be helpful. To reduce the processing cost, we propose the CPB model, which is a background subtraction method for robust event detection in dynamic scenes.

1.5.3 Contributions

The contributions of the CPB model described here are as follows.

- **Robustness:** we propose the CPB background model, which uses a pixel-block pair structure for the background model in contrast with the methods discussed above. CPB extracts the spatio-temporal information of each target pixel p through multiple supporting blocks Q_k^B, not just focusing on the local or global features in the scene, as in other approaches. The pixel-block structure of CPB is more stable than the parameterized/non-parametrized model based on a single pixel and is not sensitive to changes in the scene. This is because the structure of CPB maintains a statistical correlation that is more stable for each target pixel and abandons the prior assumption of local correlation. Furthermore, we propose the HoD based on the CBP model to handle background changes. The HoD provides the CPB model with a model update strategy that can be used for a long time. The HoD improves the robustness and stabilizes the efficiency of the CPB model over time.

- **High precision:** we introduce an evaluation strategy called the correlation-dependent decision function for accurate object detection with the correlation of CBP pairs. In contrast to the binary function with a global threshold for foreground detection, the correlation-dependent decision function uses a weighting factor based on the correlation between target pixel p and supporting blocks Q_k^B, which allows a calculable threshold for detecting the foreground to be set instead

of an empirical global threshold. This method can determine the state of the pixel with high precision and apply different data more generally.

- **Low training cost:** ConvNets approaches [70, 71] require a substantial amount of labeled data with teaching signals or ground truths for training, which is costly and may be unavailable [89]. In contrast, collecting the training data for CPB is cheap and requires no teaching signals. A limited selection of training samples according to the type of data is made, and then the background is modeled directly during the training stage.

- **Strategy for updating background models:** the new HoD strategy for updating the background model is applied to the CPB model and can also be applied to other pixel-correlation algorithms, such as ViBe [20] based on random neighboring pixels, SuBSENSE [22] based on local binary similarity patterns features, or CP3 [25]. HoD is intuitive because the structure of the background model can be updated by the designed correlation weight.

1.5.4 Overview

In Chapter 1, we gave an overview of foreground detection and stressed the challenges in background subtraction in real applications. We discussed previous studies in the field of background subtraction, including traditional and recent background modeling. We also introduced the CPB model for robust foreground detection, which can cope with the two main challenges of illumination changes and background motion. Furthermore, we also highlighted the advantages of the CBP model as being a robust, high-precision model with low training cost that provides a new strategy for updating background models.

The remainder of the thesis is organized as follows:

- In Chapter 2, the CPB model is introduced in detail, including the concept and essential mechanism. As an extension of the pixel-to-pixel structure in our previous CP3 method, the CPB model has a pixel-block structure for the background model. In this chapter, the construction of the pixel-block structure for the background model is described and the processing of the model building is explained.

- In Chapter 3, the application of CPB in the field of foreground (event) detection is discussed. An evaluation method called the correlation-dependent decision function for accurate foreground detection is introduced and the theoretical meaning of the evaluation strategy is explained. A comparison of the performance of CPB for foreground detection is presented.

- In Chapter 4, the HoD method, which is based on CPB to improve the robustness of CPB further and stabilize the efficiency of CPB over time, is described. The basic

concept and mechanism of HoD are discussed in detail. Finally, the performance of HoD is verified experimentally.

- In Chapter 5, the experimental setup is introduced in detail. Comparative experiments for CPB and CPB+HoD using several challenging datasets ($PETS2001 - camera1$, $AIST-Indoor$, $SceneBackgroundModeling.NET$ and $Change-detection\ dataset$) are designed and used to measure the robustness and efficiency of the CPB and CPB+HoD methods for various indoor and outdoor challenges. The measure metrics $Precision$, $Recall$, $F-measure$, and $PSNR$ (peak signal-to-noise ratio) for quantitative evaluation are described in detail.

- In Chapter 6, the main points of the study are summarized and the algorithms are discussed. Finally, plans for future work are presented.

Chapter 2. Co-occurrence Pixel-Block Background Model (CPB)

2.1 Basic concept of CPB

In our previous work[25], we proposed one "pixel to pixel" structure strategy to estimate the target pixel p with other pixels one by one and then to select the suitable supporting pixels for the target pixel p, and this strategy is quite effective at dealing with background changes, however it suffers from the open problem of time consumption. In order to handle this problem, we need to find a strategy to avoid the defect of CP3[25], therefore the approaches using super-pixels for cost reduction in follow-up processing and for image matting[86, 87, 88] could be helpful. As an extension of "pixel to pixel" structure, in CPB we design a "pixel to block" structure to reduce the time computing and the proposed CPB includes two processes: training process and detecting process as shown in Fig. 2.1. As discussed in chapter1.3, we seek to build on our previous work and develop a method that can reduce the computational time required for model building. To solve this problem at the training stage, we decided to group the reference pixels as a whole and then compute the correlation between the target pixels and the new reference pixels. Therefore, we propose comparing target pixel p with reference *block* Q^B, defining ref $\{Q_k^B\}_{k=1,2,...,K} = \{Q_1^B, Q_2^B, ..., Q_K^B\}$ as a reference block set for target pixel p. Fig. 2.1, which illustrates the two processes of CPB: training process and detecting process. In this work, we compare the target pixel p with the Q^B as block, and define $\{Q_k^B\}_{k=1,2,...,K} = \{Q_1^B, Q_2^B, ..., Q_K^B\}$ to denote a supporting block set for the target pixel p. We divide each frame (of size $U \times V$) into blocks Q^B of size $m \times n$. Hence, each block Q^B consists of $m \times n$ pixels:

$$Q^B = \begin{Bmatrix} Q_{11} & Q_{12} & \dots & Q_{1n} \\ Q_{21} & Q_{22} & \dots & Q_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ Q_{m1} & Q_{m2} & \dots & Q_{mn} \end{Bmatrix}. \tag{2.1}$$

As shown in Fig. 2.2, the number of blocks is $M \times N$ ($\frac{U}{m} = M, \frac{V}{n} = N$), where the size of each block Q^B is $m \times n$. Thus, the training time can be reduced further than is theoretically possible for CP3.

As with CP3, a reference pixel Q^p in CPB maintains an intensity difference relative

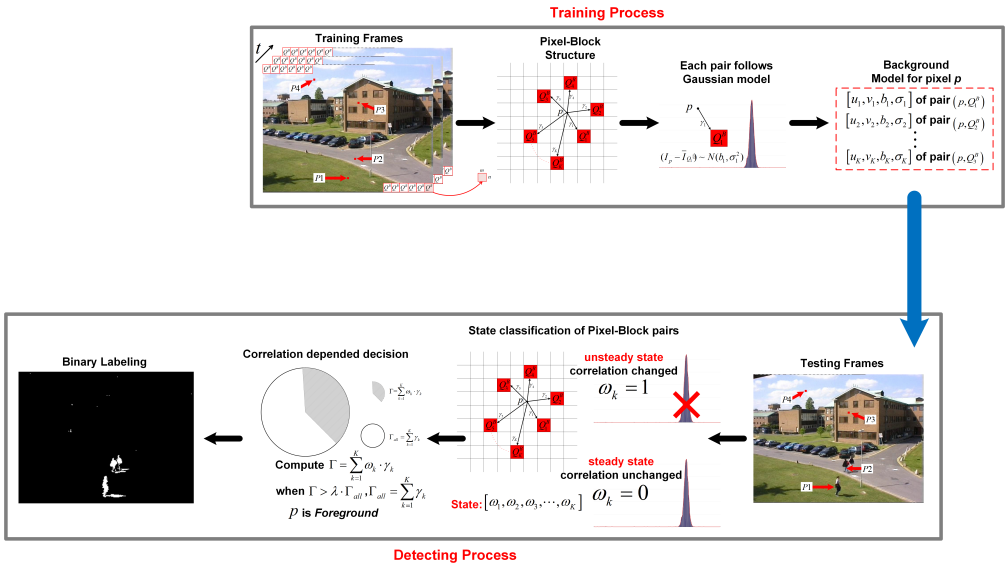


Figure 2.1: Overview of working mechanism of CPB.

to target pixel p. Hence, it is natural to assume that each pixel Q_{mn} that belongs to reference block Q^B is correlated with target pixel p. As a result, we expect one or more blocks Q^B to possess a stable intensity difference $I_p - \bar{I}_Q$ throughout the whole training frames ($\bar{I}_Q$ is the average intensity of block Q^B), even though pixel p and block Q^B can be at quite dissimilar positions, as shown in Fig. 2.3, in which the size of Q^B is set to 5×5. In theory, since a large part of computation cost can be reduced in the training process , CPB is expected mn times faster in the training than CP3[25]. When such relation maintains steady as time goes by, the deviation between the target pixel p and its supporting block Q^B would be follow a single Gaussian distribution. This relation is called as "Co-occurrence between intensity" as shown in Fig.2.3 (b) and we can utilize this knowledge to design the background model for the characteristics in background pixels.

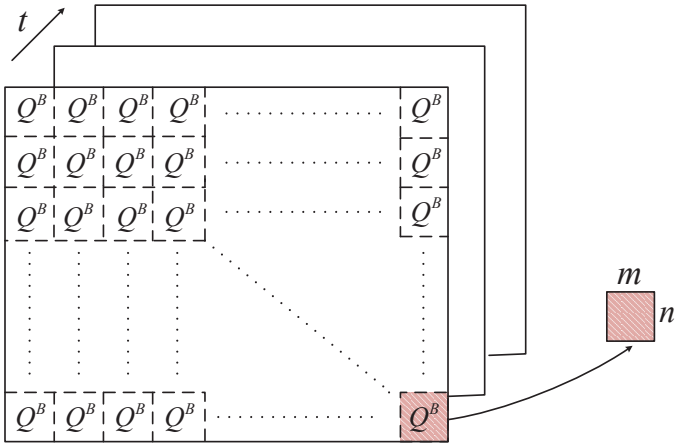
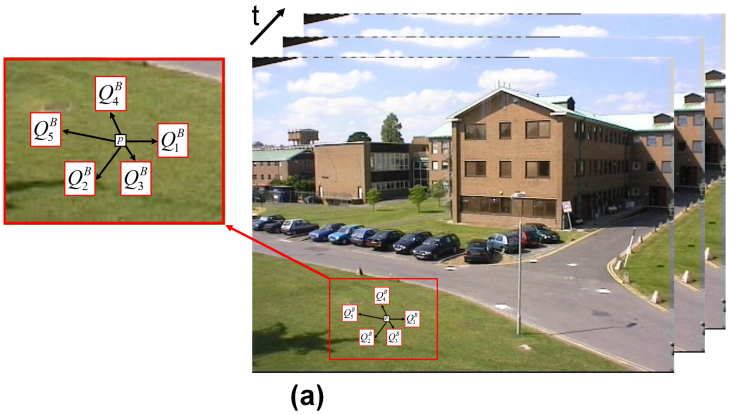


Figure 2.2: Dividing each frame into blocks.

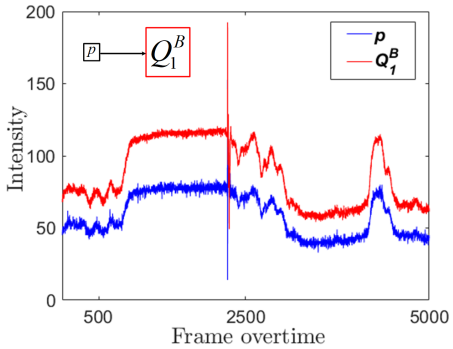
2.2 Advantages of CPB

The two factors of object occupation and illumination change can influence the current intensity of target pixel p when we detect objects in a dynamic scene. Hence, it is quite natural that block Q^B, being strongly correlated with target pixel p, can be used to determine the state of the latter. Target pixel p can be introduced as a reference to estimate its current intensity in the detecting frame, that is, a correlation between pixel p and block Q^B: $I_p = \bar{I}_Q + \Delta_k$ ($\bar{I}_Q$ is the average intensity of block Q^B in the current detecting frame). When the illumination intensity changes at target pixel p and these is no occlusion by objects, I_p changes concurrently with $\bar{I}_Q$, and consequently target pixel p is deemed to be a background pixel. Otherwise, if target pixel p is occupied by an object, the current intensity I_p ceases to be correlated with the current intensity $\bar{I}_Q$, and target pixel p is classified as a foreground pixel. At the same time, because the reference for target pixel p is the block Q^B, the following situation may arise: block Q^B contains pixel Q_{mn} that comes from two or more regions in the frame (e.g., one part of the pixel is a road and another part is grass). In this case, however, by comparing the intensity I_p of target pixel p with the average intensity $\bar{I}_Q$ of block Q^B, the foreground detection is not affected appreciably.

As Fig.2.4 indicates, target pixel p will maintain a relevant association with blocks $\{Q_k^B\}$ under the illumination changes over time, among them, target pixel p shares a significant correlation with some blocks. Based on this knowledge, we can evaluate the supporting blocks for target pixel p, that will be discussed in Section2.3.1 in detail.



(a)



(b)

(c)

Figure 2.3: Basic structure of co-occurrence pixel to block pair. (a) Co-occurrence pixel-block pair structure. (b) Correlation of pixel-block pair (p, Q_1^B). (c) Statistical model of pixel-block pair (p, Q_1^B).

In practice, CPB possesses two strengths compared with our previous CP3. One is its computational complexity: CPB can accelerate the training stage when computing the correlation between target and reference. In the CP3 method, we need to compute

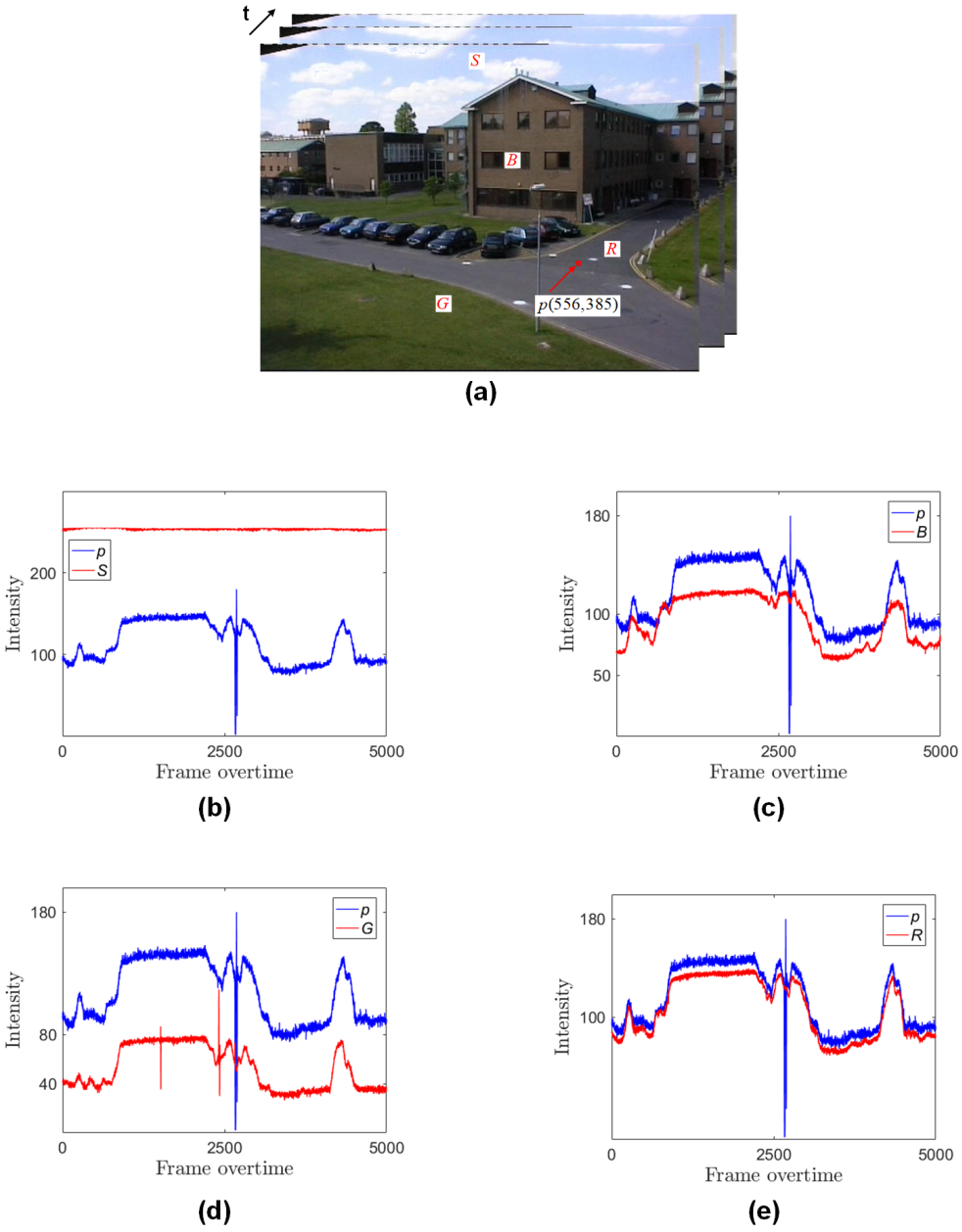


Figure 2.4: The intensity change of target pixel p and four random blocks. (a) Location of the target pixel p and four random blocks S, B, G, R at the sky, building, grass and road. (b) The intensity change of p and S. (c) The intensity change of p and B. (d) The intensity change of p and G. (e) The intensity change of p and R.

the correlation pixel by pixel and then select reference pixels for the target pixel; the computational complexity of CP3 can be defined as $O_1(UV) + O_2$, where O_1 is the training time and O_2 is the detecting time. Because $O_1(UV) >> O_2$, the computational complexity of CP3 can be taken as $O_1(UV)$. In the CPB method, we select the reference

block for each target pixel block by block; the computational complexity is based on the size of block: $O_1(\frac{UV}{mn})+O_2(mn) = \frac{1}{mn}O(UV)+mnO_2$. Given that $\frac{1}{mn}O(UV) >> mnO_2$, we can take the computational complexity of CPB to be $\frac{1}{mn}O(UV)$. Hence, it is clear in theory that CPB offers a greatly improved computational complexity compared to CP3, a fact that we demonstrate in the end of the Chapter 2.

Another advantage of CPB is its superiority over CP3 in actual object-detection tests. The fact that CP3 selects pixels with strong co-occurrence probability with the target pixel as the supporting references may lead to an concentrated distribution of supporting pixels . This could lead in turn to the target pixel being exposed in the frame whereas the majority of supporting pixels are occluded by objects. This type of situation can easily cause the state of the target pixel to be estimated erroneously. By using blocks as the supporting references of the target pixel in CPB, their distribution over a large and dispersed region avoids the occlusion by objects. In this way, CPB can reduce the likelihood of erroneous detection and lead to more reliable and robust object detection, which is why we adopt this new improved "pixel to block" structure. The results of doing so are discussed in next chapter.

In the CPB method, Δ_k is modeled at the training stage as a single Gaussian model whose unique mean and variance are then evaluated. We define $\{Q_k^B\}_{k=1,2,...,K} = \{Q_1^B, Q_2^B, ..., Q_K^B\}$ as the supporting blocks for target pixel p. We design two steps to distinguish between foreground and background pixels at the detecting stage: (1) identify whether a co-occurrence pair (p, Q_k^B) can be well described by the single Gaussian model at the detecting frame; (2) from a total of K pairs, record the unsteady state of pairs $(p, \{Q_k^B\})$ that do not accord with the single Gaussian model. If the probability of an unsteady state is high, the target pixel p should be a foreground pixel. We explain the steps of object detection in Chapter 3.

2.3 Training stage in CPB

2.3.1 Supporting blocks selection

In this section, we present an original measurement for selecting the supporting blocks $\{Q_k^B\}$ for each target pixel p. The procedures are summarized as follows.

1. Define the location of block Q^B. We divide each frame into blocks Q^B, the location of which is $Q^B(u', v')$.

2. Select the supporting blocks $\{Q_k^B\}$. We utilize Pearson's product-moment correlation coefficient to estimate the linear correlation between target pixel p and block Q_k^B:

$$\gamma(p, Q_k^B) = \frac{C_{p,\bar{Q}_k}}{\sigma_p \cdot \sigma_{\bar{Q}_k}} \tag{2.2}$$

and

$$-1 \leq \gamma(p, Q_k^B) \leq 1. \tag{2.3}$$

The covariance of pair (p, Q_k^B) can be defined as

$$C_p^Q = \frac{1}{T}\sum_{t=1}^{T}[I_p(t) - \varepsilon(I_p(t))][\bar{I}_Q(t) - \varepsilon(\bar{I}_Q(t))], \tag{2.4}$$

and the variances as

$$\sigma_p^2 = \frac{1}{T}\sum_{t=1}^{T}[I_p(t) - \varepsilon(I_p(t))]^2 \tag{2.5}$$

and

$$\sigma_{\bar{Q}_k}^2 = \frac{1}{T}\sum_{t=1}^{T}[\bar{I}_Q(t) - \varepsilon(\bar{I}_Q(t))]^2. \tag{2.6}$$

The expectation values of p and Q_k^B are

$$\varepsilon(I_p(t)) = \frac{1}{T}\sum_{t=1}^{T} I_p(t), \tag{2.7}$$

$$\varepsilon(\bar{I}_Q(t)) = \frac{1}{T}\sum_{t=1}^{T} \bar{I}_Q(t), \tag{2.8}$$

where

- T is the sequence of frames, the size of each is $U \times V$;
- the size of one block Q_k^B is $m \times n (m < M, n < N)$;
- each block consists of $m \times n$ pixels;
- $I_p(t)$ is the intensity of target pixel p in frame t; $\bar{I}_Q(t)$ is the average intensity of block Q_k^B in frame t.

For each target pixel $p(u, v)$ with $M \times N - 1$ values of $\gamma(p(u, v), Q_k^B(u', v'))$ need to be calculated. Block $Q_k^B(u', v')$, having a high value of $\gamma(p, Q_k^B)$ with target pixel $p(u, v)$, can be selected as the reference block Q_n^B. Collectively, the reference blocks that satisfy the requirement of having the highest K components in the array $\{\gamma(p, Q^B)\}$ can be nominated as the candidate blocks. Taking the candidates $Q_1^B, Q_2^B \in Q_n^B$ by way of example, we order them so that $Q_1^B > Q_2^B$ if $\gamma(p, Q_1^B) > \gamma(p, Q_2^B)$.

The supporting block set $\{Q_k^B\}_{k=1,2,\ldots,K}$ is defined as follows:

$$\{Q_k^B\}_{k=1,2,\ldots,K} = \{Q^B | \gamma(p, Q_k^B) \text{is the highest K components}\} \text{and} \gamma(p, Q_k^B) > 0. \tag{2.9}$$

In general, we can expect that if the pixel-block pair (p, Q_k^B) keeps a high correlation coefficient, then the supporting block Q_k can provide some reliability to estimate the current state of the target pixel p. Fig. 2.5 shows example layouts of the supporting blocks using $PETS2001 - dataset3 - camera1$ and the target pixels are selected from the four representative regions: "Grass," "Road," "Building," "Sky," respectively. Here, K=10, and the size of each block is 5×5.

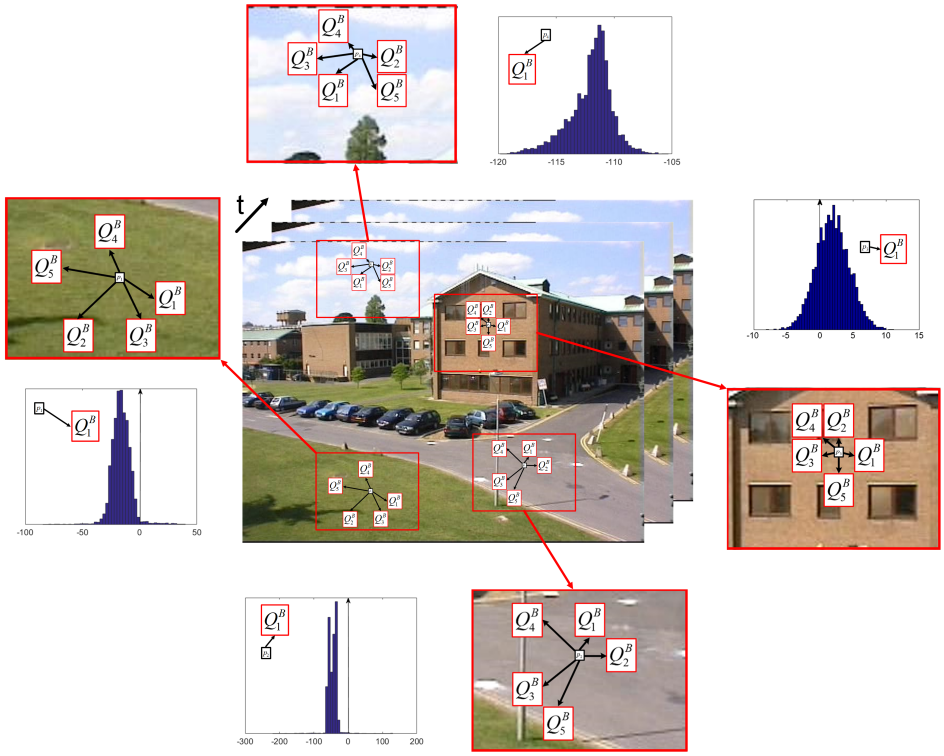


Figure 2.5: Example layouts of pixel-block pairs for different position pixels $p_1(256, 483)$, $p_2(551, 432)$, $p_3(481, 168)$ and $p_4(250, 41)$, respectively, where $K = 5$ and the size of each block is 5×5, and examples of the correlation of pairs at different position, respectively.

2.3.2 Background modeling

In this section, we discuss how to build the background model by using a single Gaussian model.

In Section 2.3.1, we introduce how to select supporting blocks $\{Q_k^B\}_{k=1,2,...,K}$ for each target pixel p. To ensure robustness and computational efficiency, we set the number of supporting blocks to $K = 20$. Each supporting block Q_k^B retains a differential increment with target pixel p:

$$\Delta_k \sim N(b_k, \sigma_k^2) \quad \Delta_k = I_p - \bar{I}_{Q_k}, \tag{2.10}$$

where I_p is the intensity of the pixel p at t frame, $\bar{I}_{Q_k}$ is the average intensity of the block Q_k^B at t frame and σ_k^2 be considered as the noise ε, which follows a normalized distribution $k \sim N(0, \sigma_\varepsilon^2)$. From our previous work [25], we know that a single Gaussian model works well with the selected pixel block pairs, which tolerate noise by maintaining a favorably steady relationship. In CPB, each pixel-block pair(p, Q_k^B) owns an unique Gaussian and we record two parameters that the differential increment b_k and the standard deviation σ_k as model as Fig.2.1 shows.

Algorithm 1 Background model building

Input:

T training frames;

Output:

a list of $\{Q_k^B\}_{k=1,2,...,K}$ consisting of $[u_k, v_k, \sigma_k, b_k]$;

for each pixel p **do**

1. Compute $\gamma(p, Q_k^B)$ based on Eq. 2.2.
2. Select$\{Q_k^B\} = \{\gamma(p, Q_k^B)$ is the highest K components, $\gamma(p, Q_k^B) > 0\}$.
3. Compute and record $[u_k, v_k, \sigma_k, b_k]$.

return $[u_k, v_k, \sigma_k, b_k]$

end for

Where, b_k is defined as the following expression:

$$b_k = \frac{1}{T} \sum_{t=1}^{T} \Delta_k \tag{2.11}$$

and the variance estimation is defined as follows:

$$\sigma_k^2 = \frac{1}{T} \sum_{t=1}^{T} (\Delta_k - b_k)^2, \tag{2.12}$$

where T is the sequence of frames. Through the training process, the parameters σ_k and b_k are recorded as a model description for the next detecting stage and then the background model is built as a list consisting of $[u_k, v_k, b_k, \sigma_k]$ for supporting block set $\{Q_k^B\}_{k=1,2,...,K}$, where (u_k, v_k) is the coordinate of supporting block.

In summary, the pseudo-code of background model building of CPB method is shown in Algorithm 1.

2.3.3 Robust against complex scenes

The CPB model can deal well with complex scenes such as illumination changes or background motion. It is also robust to some extreme scenes, including camera jitter, intermittent object motion, and dynamic background (sudden background motion). These extreme scenes are explained in detail in this chapter and the experimental verification is discussed in Chapter 3.

- **Camera jitter:** in video surveillance, camera jitter is a problem that must be solved for background subtraction. Camera jitter, which is movement of the camera, may introduce motion blur, and thus limit resolution quality [90]. Consequently, accurate object detection in the presence of camera jitter is challenging [91, 92]. However, CPB still achieves good performance in the presence of camera jitter by using the co-occurrence background model. This is because camera jitter

is a regular movement and this kind of motion information can be learned by CPB during training, thereby avoiding interference in the detection of the current frame.

- **Intermittent object motion:** is a difficult challenge for object detection and includes background objects moving away, abandoned objects, and objects stopping for a short while before moving away. It is difficult for background subtraction to detect the actual changes in objects from the background. CPB acquires the feature information from the initial background, which does not suffer from interference by object motion, and thus obtains the object motion information accurately.

- **Dynamic background (sudden background motion):** includes the following kinds of changes in background: regular movement (e.g., rippling water) and sudden movement (e.g., changing advertising boards). These movements in the scene are regarded as background motion. Here, sudden background motion is discussed, which is a different challenge than distinguishing foreground from background. CPB learns background information, and this information can also be the spatio-temporal context [93], which is stable against the interference of object movement. By obtaining the background spatio-temporal context, CPB is robust for object detection with dynamic backgrounds.

2.3.4 Computation time in training

We compare the computation times between CPB and our previous CP3. Here, we use *AIST-Indoor* for the test, which is facilitated by the 300 frames in the dataset that are provided specifically for training. In CP3, the author also offers an accelerated method for CP3, which we include in the experiment by way of competition. Table 2.1 gives the average processing speeds on the same platform (2.3 GHz Core i7CPU; 8 GB of RAM; MATLAB).

From the results, CPB can greatly reduce the computation time for model building in comparison with our previous work, CP3. This may contribute to the real applications, especially for the real-time application. In addition, the computation time of CPB in training is based on the setting of supporting block size, In theory, the larger supporting block will lead less training time, however, the size of supporting block is limited by the original size of the input data, and we also have to consider the balance between supporting block size and the accuracy foreground detection. In this thesis, because the data size of all experiments is 320×240, we set the block size as 8×8, which is also an experimental-based empirical value that satisfies the balance between training time and detection accuracy.

Table 2.1: Comparison of computation times

Method	Average computing time (sec.)
CP3	126.33
Accelerated CP3	5.02
CPB	**2.15**

2.3.5 Training data selection

In this section, we want give some comments on how to select the training data as an important step in our CPB's mechanism. We need an enough set of suitable data for training, and then CPB may train itself properly to detect excepted foreground pixels. It has been a common and important problem in the algorithms that need any training data, such like IMBS[68] or SuBSENSE[22] as Fig. 4.7 in Section 4.2.2 to do this preparation. In this paper, since we use many databases that have their own ground truth frames and therefore we can see some types of the expected foreground pixels, such as walking peoples or vehicles, it is possible to select some frames as the training data, which do not include any excessive foreground pixels. But in any real tasks in which it is not reality to take high cost for making effective ground truth data, one may have to make the training data or frames through implicit definition of foreground pixels and selecting the proper frames. Therefore, in our experimental demonstration, we prudently select the frames for training to avoid the emergence of adversarial data.

2.4 Summary

In Chapter 2, we presented a detailed introduction of the Co-occurrence Pixel-Block Background Model (CPB), which is a novel "pixel to block" structure for background model based on our previous work, CP3. Then, we explained the procedure of CPB in training to build the background model. Compared with CP3, CPB significantly reduces the computational complexity in training, which is proved in Table 2.1. Moreover, with the dispersed distribution in structure CPB can reduce the likelihood of erroneous detection and lead to more reliable and robust results in foreground detection, this will be verified in next chapter.

Chapter 3. Foreground Detection

The proposed methods establish a competitive binary classification mode for object detection [2]. We encompass this mode in our CPB by calculating the pixel block pairs $(p, \{Q_k^B\}_{k=1,2,...,K})$ in turn. This includes two procedures: (1) estimating the steady or unsteady state of each co-occurrence pixel block pair (p, Q_k^B); (2) estimating whether pixel p is a foreground or background pixel.

3.1 Pixel-block pair state

Based on the co-occurrence background model built above, CPB can acquire the spatial-temporal information of target pixel p and then compare the difference between target pixel p and supporting block Q_k^B to judge the state of target pixel p as shown in Fig.3.1. Once the co-occurrence relation appears an outlier, such situation would be regarded as an unsteady state of pixel-block pair(p, Q_k^B) at current frame, thereby we could estimate target pixel p as foreground. This knowledge can be realized as the following: the state F (unsteady) means p may be occluded by any foreground object, while the state B (steady) means that p may be exposed to the camera as it has been in the statistical training frames. In order to obtain any difference between these two states, for each pixel p, we introduce an index value as a "penalty" for violating the relationships authorized at the statistical training process. In other words, if the state F is associated with pixel p and the pixel value may also be changed, therefore we can utilize statistical tests in which the difference may belong to the registered distribution or be rejected as a value outside of the distribution.

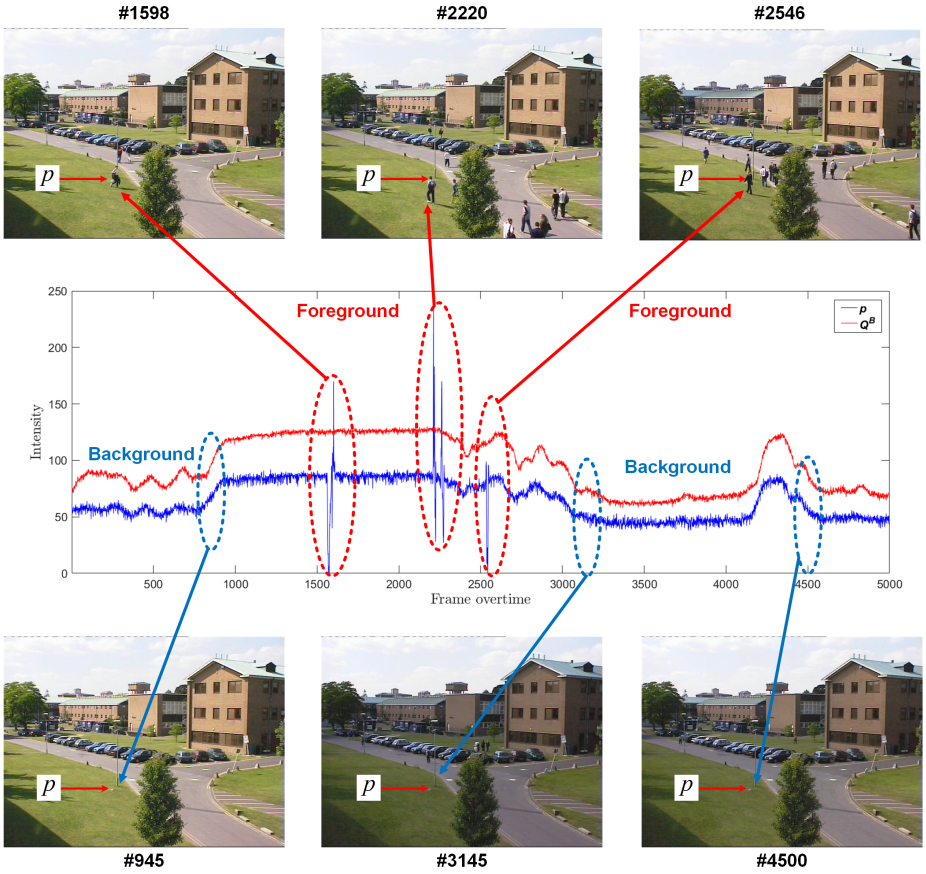


Figure 3.1: Co-occurrence intensity changes between target pixel p and supporting block Q^B overtime.

For each pair (p, Q_k^B), a binary function for identifying its steady or unsteady state can be defined as follows:

$$\omega_k = \begin{cases} 1 & \text{if } |(I_p - I_{Q_{mn}}) - b_k| \geq \eta \cdot \sigma_k \\ 0 & \text{otherwise} \end{cases}, \tag{3.1}$$

where I_p is the intensity value of pixel p at the current frame, $I_{Q_{mn}}$ is the intensity value of pixel Q_{mn} (as in equation 2.1), block Q^B consists of $m \times n$ pixels, Q_{mn} is a pixel that belongs to block Q^B, and η is a constant. If all the pixels $Q_{mn} : |(I_p - I_{Q_{mn}}) - b_k| \geq \eta \cdot \sigma_k$, then $\omega_k = 1$ and the pair (p, Q_k^B) is an unsteady state.

It is important to note that in our CPB method, we use $|(I_p - I_{Q_{mn}}) - b_k| \geq \eta \cdot \sigma_k$ to estimate the unsteady state of each pair (p, Q_k^B) and we record the unsteady state as $\omega_k = 1$. This is different to our previous CP3 method, in which we record the steady state of a pixel pair, for which the difference Δ_k still follows the Gaussian distribution at the current frame. In contrast, in the CPB method we record the unsteady state of each pair, for which the difference Δ_k fails to accord with the Gaussian model, meaning the correlation between pixel p and block Q_k^B is changed at the current frame. This is a

very important transformation in CPB, which will be confirmed in the next subsection.

3.2 Correlation dependent decision

In CP3, the decision of the state of the target pixel was based on the number of changes of co-occurrence pairs: when the number of changes reached a certain value, we determined the target pixel as being in the foreground. However, there is a problem with such a judgment: we need to set a global threshold for it, and usually this global threshold is an empirical value learned from experiments based on datasets. Note that this selection policy may be prone to an error and lacks generality for other datasets

To deal with this problem, we introduce a novel evaluation strategy named correlation depended decision function for accurate object detection based on the correlation of co-occurrence pixel-block pairs as follows. In Section 3.3, we present a comparison to verify our evaluation strategy.

At the statistical training stage of CPB, for each pixel p of the whole frame, a set of supporting blocks $\{Q_k^B\}_{k=1,2,\ldots,K}$ may be expected to have high co-occurrence probabilities and high values of γ_k ($\gamma(p, Q_k^B)$ is aliased as γ_k). As an instance, Fig. 3.2 illustrates a set of distributions that shows multiple pairwise relationships between the target pixel and each supporting block.

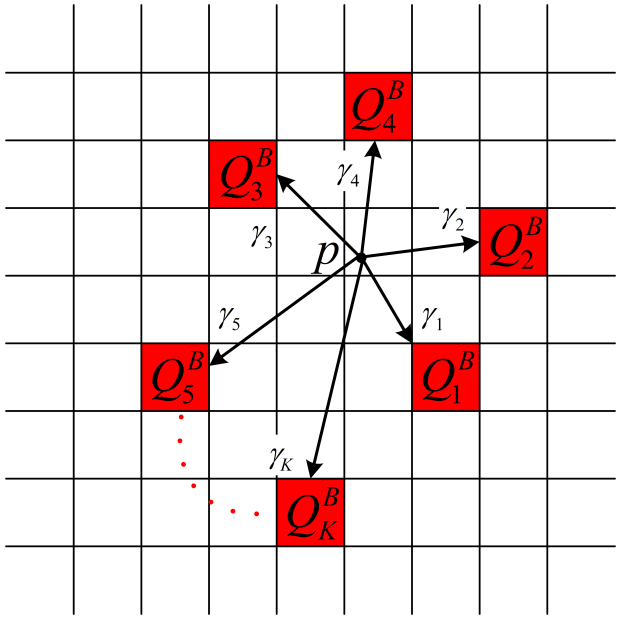


Figure 3.2: Correlation coefficient of pixel block pair (p, Q_k^B).

After identifying the steady or unsteady state of each pair (p, Q_k^B), K components of ω_k are produced for the following decision of each pixel p. To record the state of pixel p, we introduce another weighting factor into the above definition described in section 3.1. Obviously, each pixel p has multiple supporting blocks $\{Q_k^B\}_{k=1,2,...,K}$ with corresponding γ_k. We propose that a higher γ_k may have more of an effect on the state of pixel p, which is then realized by using γ_k values as weighting factors as follows:

$$\Gamma = \sum_{k=1}^{K} \omega_k \cdot \gamma_k. \tag{3.2}$$

with two following significances: first, Γ could count up the unsteady pairs; second, the maximum value of Γ could be possibly obtained in the case that all of K elemental pairs are in the unsteady state and it is also a relative value with respect to the target pixel. Furthermore, Γ would not miss to count any high γ_k in the summation to lead a wrong decision. To realize relative decision making on Γ, we can have the following possible maximum value of it.

According to Eq. 3.2, Γ records the state of pixel p from K pixel-block pairs $(p, \{Q_k^B\})$. In our CPB method, Γ has two roles. Firstly, Γ calculates the number of pairs $(p, \{Q_k^B\})$ that are in the unsteady state. Secondly, Γ indicates the unsteady state of pairs $(p, \{Q_k^B\})$ that have high co-occurrence probabilities. In other words, when the value of Γ is high, the majority of pairs are in unsteady state, and a large number of pairs that previously had high co-occurrence probabilities have been converted to the unsteady state at the current frame. Therefore, it is highly likely that pixel p is a foreground pixel.

To classify whether pixel p is a foreground or background pixel, the evaluation criterion is defined as follows:

if $\Gamma > \lambda \cdot \Gamma_{all}$, **then**
  p is *foreground*
else
  p is *background*, where

$$\Gamma_{all} = \sum_{k=1}^{K} \gamma_k \tag{3.3}$$

and Γ_{all} records the values $\{\gamma_k\}_{k=1,2,...,K}$ from K pixel-block pairs $(p, \{Q_k^B\}_{k=1,2,...,K})$. Contacting Eq. 3.2 with Eq. 3.3, it is natural to estimate the state of pixel p through a comparative analysis between the values of Γ and Γ_{all}.

As shown in Fig. 3.3, if the value of Γ is high, it is highly likely that pixel p is a foreground pixel. In our CPB, we set the threshold for Γ as $\lambda \cdot \Gamma_{all}$, $0 \leq \lambda \leq 1$, that is, when $\Gamma > \lambda \cdot \Gamma_{all}$, then pixel p should be in the foreground. The pseudo-code of the algorithm for detecting objects is presented in Algorithm 2.

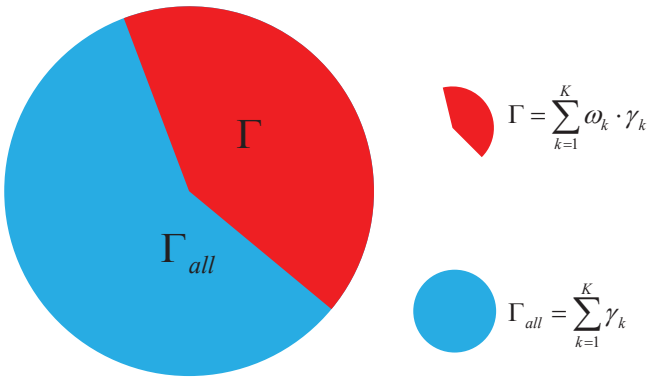


Figure 3.3: Relationship between Γ and Γ_{all}.

3.3 Performance verification

In this section, we select four challenging datasets including indoor an outdoor scenes to verify the performance of CPB, that come from the *Change Detection Challenge* [94], the *PETS2001* public dataset (dataset 3, camera 2, testing; the ground-truth comes from the Department of Advanced Information Technology, Kyushu University) and the *AIST-Indoor* dataset provided by the National Institute of Advanced Industrial Science and Technology in Japan. The datasets: (1) *PETS2001*-camera 2 with gradual illumination changes; (2) *pedestrians*, outdoor in high contrast; (3) *PETS2006*, indoor in low contrast; and (4) *AIST-Indoor* with sudden illumination changes.

Here, we analyze our method by comparing its results with those of four other approaches: GMM [15], KDE [16], Oliver's SL-PCA [17], the state of the art technique IMBS[68], and our previous algorithm CP3 [25].

Figures 3.5–3.8 show examples of foreground detection for a typical frame from each dataset sequence, the results of all four scenes from *PETS2001*, *baseline-pedestrians*, *baseline-PETS2006*, and *AIST-Indoor*. The sequences are challenging in relation to high or low contrast, moving trees, sunlight, and sudden changes in illumination conditions, respectively. Table 3.1 lists the results of our CPB compared with the other four methods for all four challenging datasets. Figure 3.9 shows an example of the performance between CPB and the other four methods over time in an outdoor scene with high brightness.

3.3.1 Performance of correlation depended decision function

In order to verify the performance of our novel evaluation strategy, correlation depended decision function for foreground detection, we designed an experimental competition. We compare the performance between two evaluation strategies: our correlation depended decision function introduced in Section 3.2, and the previous binary classification strategy [25]. We set a threshold of $T = 10$ for the previous binary classification

Algorithm 2 Correlation dependent decision

Input:

Testing frame; Parameters: η and λ ;

Output:

The state of pixel p;

for each pixel p **do**

1. Load $[\mu_k, v_k, b_k, \sigma_k]$ for each pair (p, Q_k^B);
2. Estimate the state of each pair;

for $k = 1, 2, \ldots, K$ **do**

if all the pixels $Q_{mn} : |(I_p - I_{Q_{mn}}) - b_k| \geq \eta \cdot \sigma_k$ **then**

$\omega_k = 1$

else

$\omega_k = 0$

end if

end for

3. Foreground detection

Compute Γ

if $\Gamma > \lambda \cdot \Gamma_{all}$, where $\Gamma_{all} = \sum_{k=1}^{K} \gamma_k$ **then**

p is *foreground*

else

p is *background*

end if

return the state of pixel p

end for

strategy, which means the target pixel should be in the foreground if more than half the pairs are in the unsteady state. In this competition, we select different number of frames from the *baseline-pedestrians* dataset for training (from 100 frames to 750 frames, with an interval of 50 frames) in order to compare the efficiency of both strategies.

From the results in Fig. 3.4, we note that our CPB with the correlation depended decision function performs better than CPB with the previous binary-classification strategy. Meanwhile, CPB with the novel evaluation strategy also leads to high precision with a small number of training data. Hence, the correlation depended decision function is associated with good performance and may reduce the required the training data. This means that this novel evaluation strategy could be applied more effectively in real-world situations.

Table 3.1: Comparison of CPB with other methods

Methods	**Accuracy**	**Precision**	**Recall**	**F-measure**
GMM	0.9891	0.4819	0.8886	0.6249
KDE	0.9034	0.2740	0.8209	0.4109
SL-PCA	0.9178	0.2970	0.9257	0.4497
IMBS	0.9835	0.4828	0.8632	0.6192
CP3	0.9850	0.3733	0.7611	0.5009
Our CPB	**0.9938**	**0.8441**	**0.5157**	**0.6402**

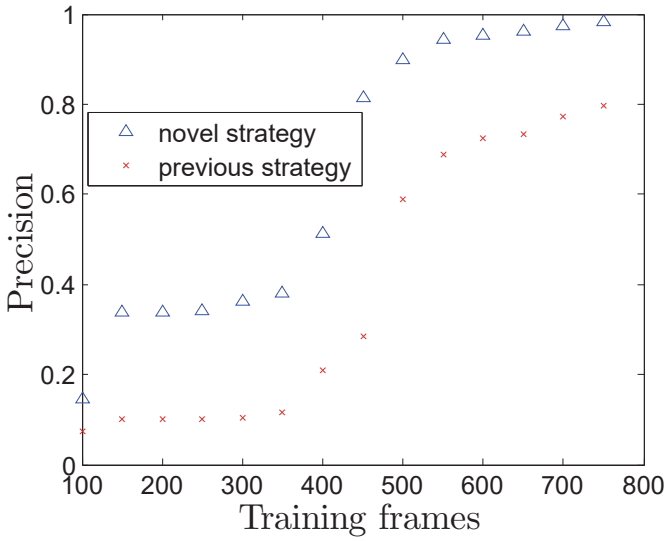


Figure 3.4: Comparison between novel evaluation strategy and previous binary classification strategy.

3.3.2 Discussion

Figs. 3.5–3.8 show that the performance of CPB is good compared with other methods. In particular, in Fig. 3.8, CPB is robust under sudden indoor illumination changes and the target pedestrian is extracted accurately. Moreover, CPB effectively restrains noise extraction at low or high contrast (Figs. 3.5 and 3.6), achieving an accurate result. This is also demonstrated in Figure 3.9(a), where CPB retains high precision under high brightness outdoors conditions over time. The high precision confirms that CPB resists noise.

CPB can achieve the first $F-measure$ level in the comprehensive comparison, which highlights the effectiveness of CPB for foreground detection in challenging scenes (3.1). In addition, the *Recall* of CPB is relatively low because, in real testing, CPB may cause some errors in the results called holes, which are incorrect pixels in the foreground. This is a particular problem if a foreground object's color is similar to that of the background, which may cause holes in the detected foreground [95]. After converting an image to a gray-scale equivalent, dark colors in the foreground and background may have similar intensities [25], which may cause holes in the foreground.

Generally, high precision is more important to us , especially because high-precision detection can greatly reduce errors in object tracking or traffic monitoring. Although real-world applications require high-precision detection, low *Recall* can be tolerated under certain circumstances. This problem is addressed by modifying CPB with a new strategy called the HoD, which will be discussed in the next section.

3.4 Summary

Chapter 3 proposed how to do the foreground detection by using CPB feature. First, we introduced a novel evaluation strategy named correlation depended decision function for accurate foreground detection, which is a weight based method using the correlation coefficient γ_k. And then, we also proved the performance of CPB in different scenes with high or low contrast, moving trees, sunlight, and sudden changes in illumination conditions, respectively. The resuls from Fig. 3.5–3.8 and 3.1 shows the interest of our CPB.

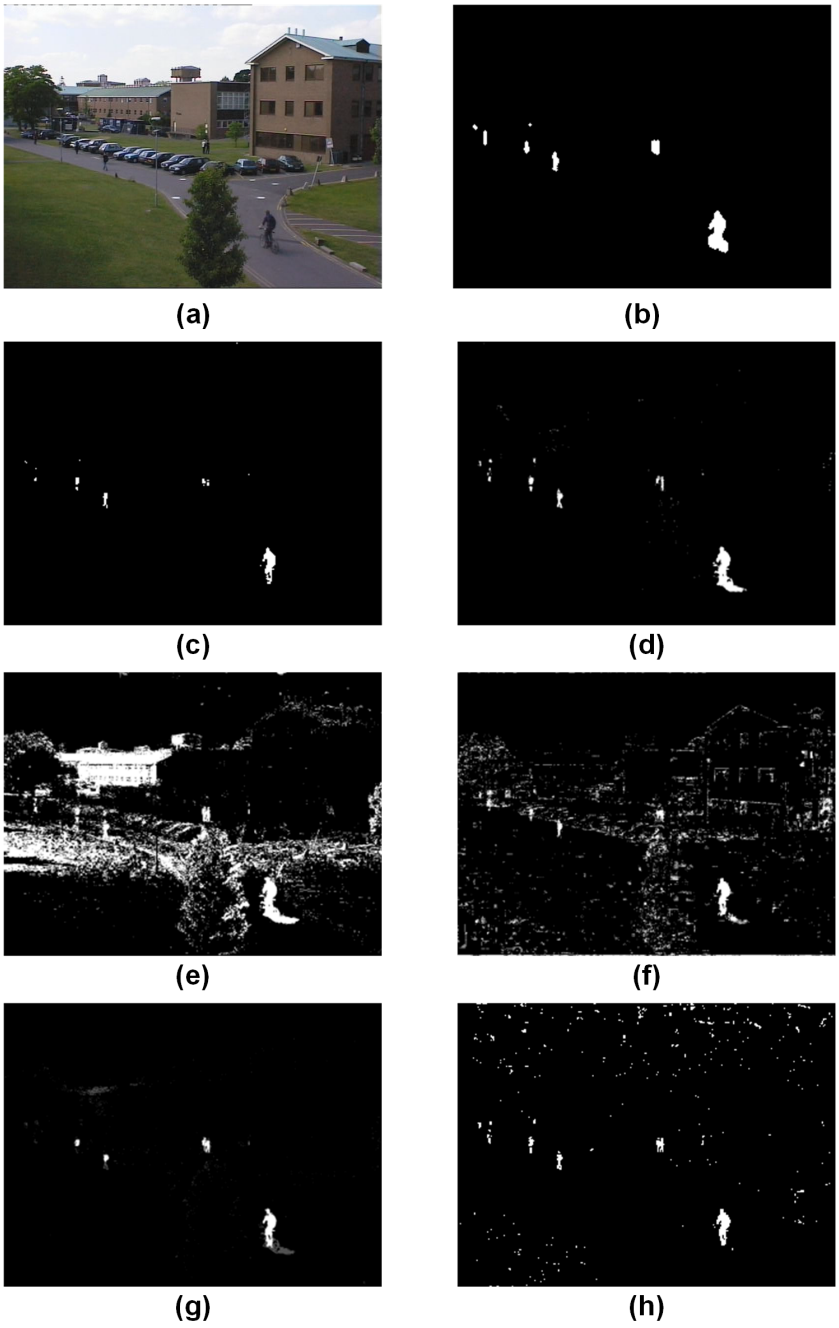


Figure 3.5: Comparative maps of five different background-subtraction methods for a frame taken from *PETS2001*: (a) Input frame; (b) Ground truth; (c) CPB; (d) GMM; (e) KDE; (f) SL-PCA; (g) IMBS; (h) CP3.

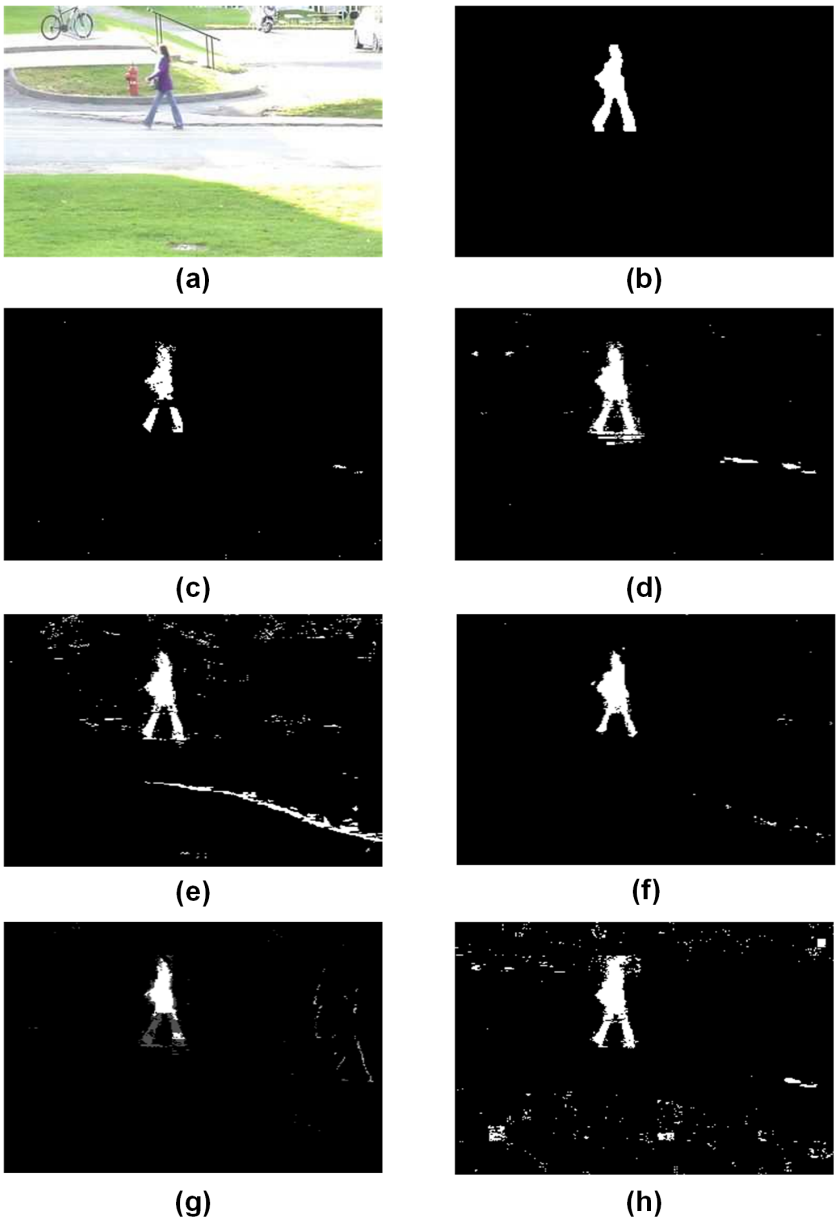


Figure 3.6: Comparative maps of five different background-subtraction for a frame taken from *baseline-pedestrians*: (a) Input frame; (b) Ground truth; (c) CPB; (d) GMM; (e) KDE; (f) SL-PCA; (g) IMBS; (h) CP3.

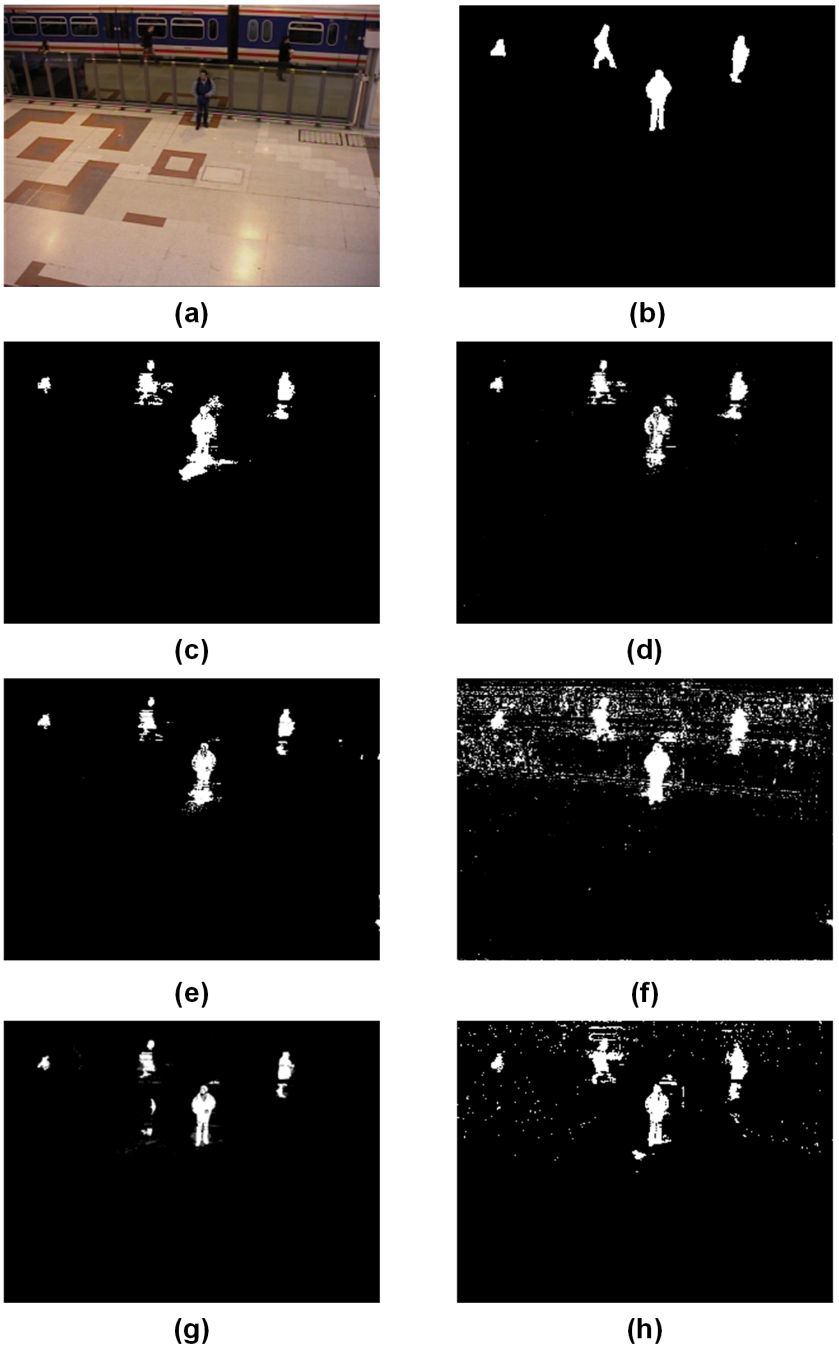


Figure 3.7: Comparative maps of five different background-subtraction methods for a frame taken from *baseline-PETS2006*: (a) Input frame; (b) Ground truth; (c) CPB; (d) GMM; (e) KDE; (f) SL-PCA; (g) IMBS; (h) CP3.

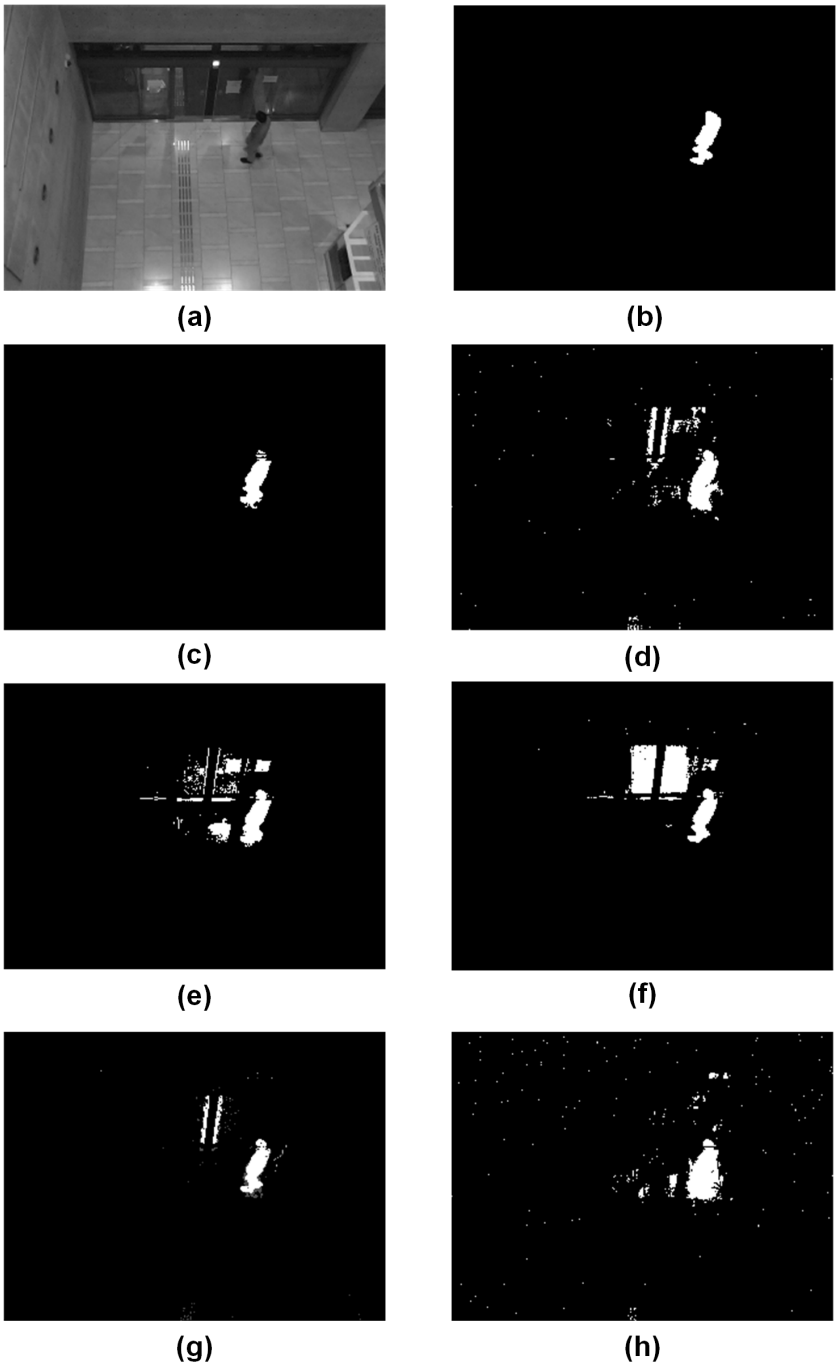


Figure 3.8: Comparative maps of five different background-subtraction methods for a frame taken from *AIST-Indoor*: (a) Input frame; (b) Ground truth; (c) CPB; (d) GMM; (e) KDE; (f) SL-PCA; (g) IMBS; (h) CP3.

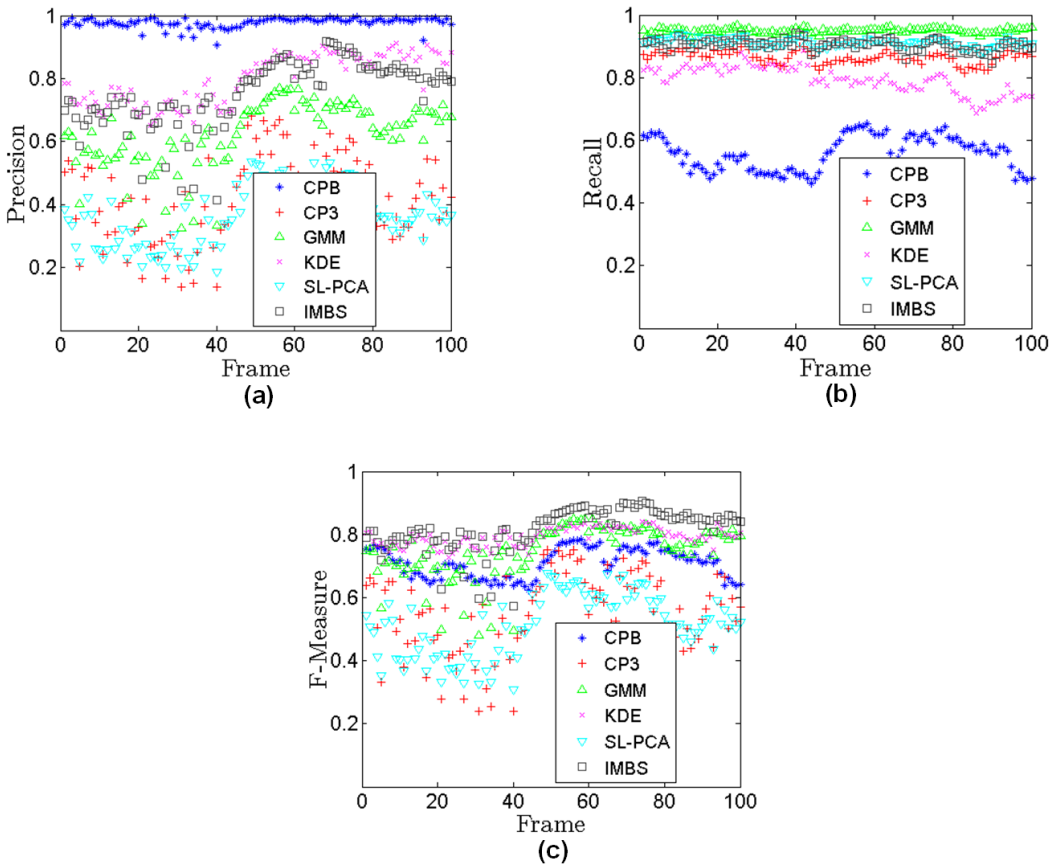


Figure 3.9: Comparison in outdoor scene for CPB, CP3, GMM, KDE, SL-PCA and IMBS for a frame over time taken from *pedestrians*: (a) precision; (b) recall; (c) F-measure.

Chapter 4. Hypothesis on Degradation Modification (HoD)

In Chapter 2 and 3, we have introduced the basic CPB algorithm for robust background subtraction, however, where the data for training and detecting are prepared in advance of the operations. In this work, we define two types of problems *open-set* (generalization problem), which is shown in Fig. 4.1 (a), where the data in training are known in advance but the data for detecting are unknown, continuable, and different from the training data. *Open-set* is different from the type *closed-set* (classification problem), where the detecting data can be selected from the same set of the training data and the also can be known in advance as shown in Fig. 4.1 (b). We may have some mechanism to modify the model to fix some errors which may be observed in *open-set* condition. In general, hypothesis of this paper follows two significances: (1) we assume that some "noise" may arise in detecting process due to a long time usage of initial CPB background model and we can not confirm such "noise" is true or not without any ground truth for verification in real applications; (2) second, we assume that after a prolonged using, the initial "Pixel to Block"structure can no longer adapt to the current, then resulting in errors. Then, based on the above assumptions, in this section, we intend to introduce a simple mechanism named Hypothesis on Degradation Modification (HoD) extended from CPB to adapt the background changes and reinforce the robustness of CPB to resist the "noise" in real applications.

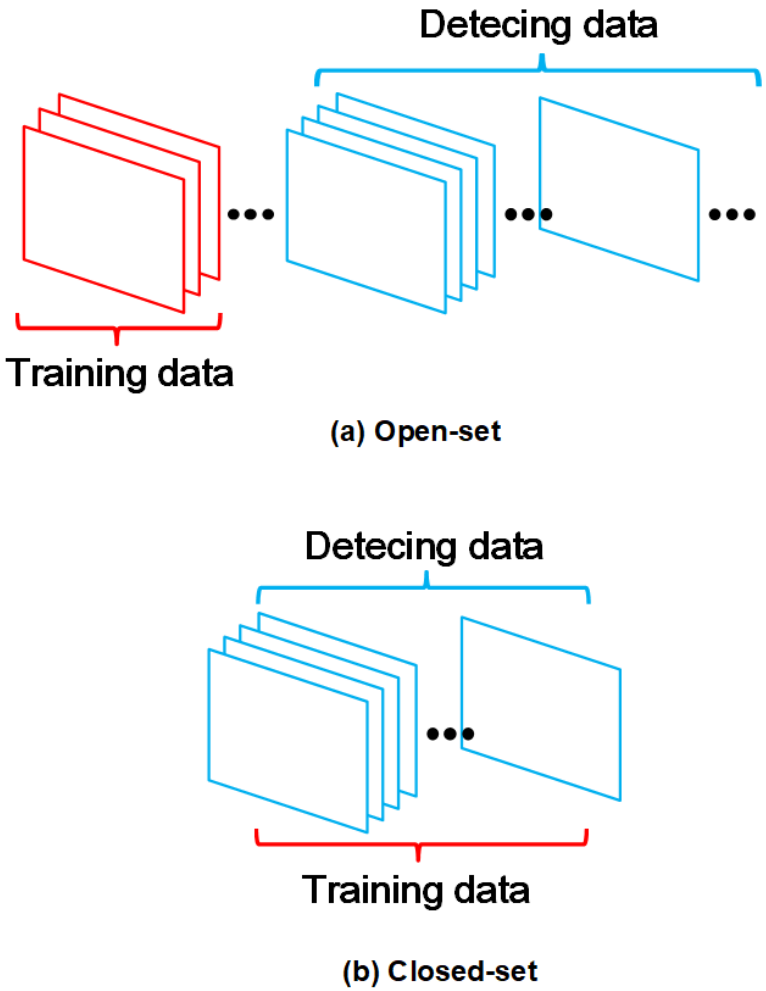


Figure 4.1: Descriptions of open-set and closed-set conditions.

4.1 Hypothesis on degradation modification

We have introduced the basic CPB algorithm for robust background subtraction, however, where the data for training and detecting are prepared in advance of the operations. In this work, we define two types of problems *open-set* (generalization problem), which is shown in Fig. 4.1 (a), where the data in training are known in advance but the data for detecting are unknown, continuable, and different from the training data. *Open-set* is different from the type *closed-set* (classification problem), where the detecting data can be selected from the same set of the training data and the also can be known in advance as shown in Fig. 4.1 (b). We may have some mechanism to modify the model to fix some errors which may be observed in *open-set* condition. In general, hypothesis of this paper follows two significances: (1) we assume that some "noise" may arise in detecting process due to a long time usage of initial CPB background model and we can not confirm such "noise" is true or not without any ground truth for ver-

ification in real applications; (2) second, we assume that after a prolonged using, the initial "Pixel to Block"structure can no longer adapt to the current, then resulting in errors. Then, based on the above assumptions, in this section, we intend to introduce a simple mechanism named Hypothesis on Degradation Modification (HoD) extended from CPB to adapt the background changes and reinforce the robustness of CPB to resist the "noise" in real applications.

4.1.1 Hypothesis on degradation

In practice, after a long time utilization of initial CPB background model in an unlearned sequence, the expected relative relation of the pixel-block pair might be broken. In other words, initial CPB model might generate a degradation with the passage of time, then some "noise" might arise in detecting process. Here, we define such assumption as "Hypothesis on Degradation" and name the "noise " in detecting process as "hypothetical noise": (1) the hole surrounded by the detected foreground pixels, which is estimated as the background and we named it 'NaB'; (2) the dot surrounded by the non-detected pixels, which is estimated as the event and we named it 'NaE'. Fig. 4.2 shows an example of the hypothetical noise using *copyMachine* from CDW-2012 dataset[94]. To reinforce the merits of CPB background model, we introduce a tactic named Hypothesis on degradation modification (HoD) into the CPB structure to remove the hypothetical noise.

Fig. 4.3 describes an overview of the proposed HoD. Note that HoD is not one post-processing technique, in this study, HoD is an update approach of model structure to reinforce the robustness of CPB, and it is also a feasible on-line mode of CPB structure in future. Moreover, we also can clearly notice that HoD is a self-checking mode, which is completely different from the retraining mode. In HoD mode, it costs less time and consumes less data cost over a period of usage, and is more efficient than the retraining mode.

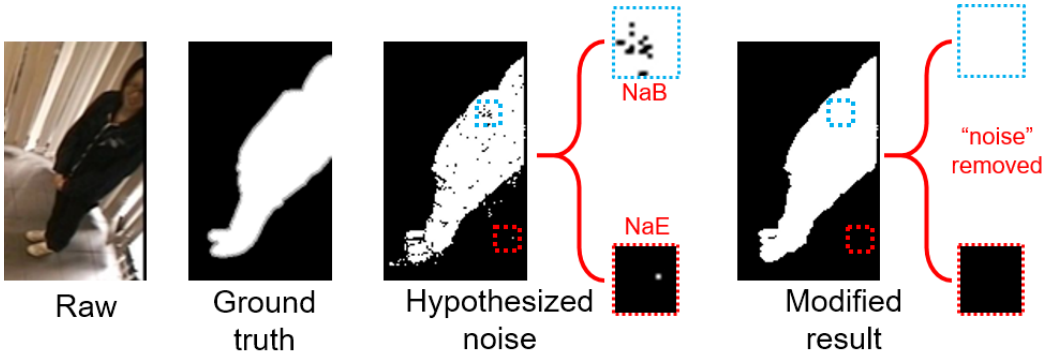


Figure 4.2: Description of hypothesized noise.

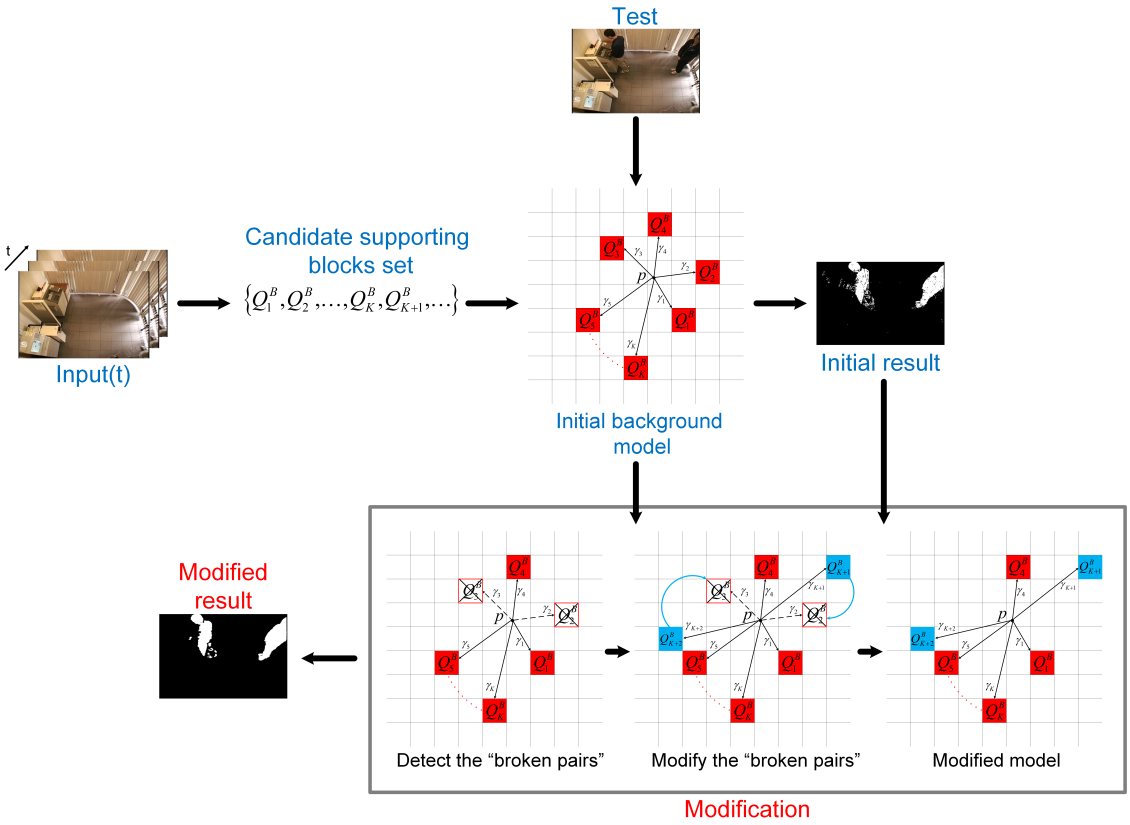


Figure 4.3: Overview of HoD Modification.

4.1.2 Broken pixel-block pairs detection

As shown in Fig. 4.3, first we need to detect the broken elemental pairs in pixel-block structure of the hypothetical noise. In this study, we assume that the larger γ (mentioned in Section 3.2) could hold a higher weight in the trained pixel-block structure and such pair would be more likely to affect the state of pixels. Thus it is obvious that the pairs with large γ in unsteady state might cause a decision on NaE, whereas the pairs with large γ in steady state might cause a decision on NaB. With the above assumption, we propose a weight-based decision function to detect the broken pair:

$$if\ \gamma_m \geq \bar{\gamma},\ \ then\ (p, Q_m^B)\ is\ broken \tag{4.1}$$

where (p, Q_m^B) is the pair, which is in unsteady state of NaE or steady state of NaB.

Depending on the noise is NaE or NaB, the threshold $\bar{\gamma}$ owns different definition. In the case of NaE, it is defined by use of the total number of unsteady pairs $M = \sum_{k=1}^{K} \omega_k$ as follows:

$$\bar{\gamma} = \frac{1}{M} \sum_{k=1}^{K} \gamma_k \cdot \omega_k = \frac{1}{M} \Gamma. \tag{4.2}$$

In the other hand, for NaB case, it is defined as follows:

$$\bar{\gamma} = \frac{1}{K-M}\sum_{k=1}^{K}\gamma_k \cdot (1-\omega_k) = \frac{1}{K-M}\left(\Gamma_{all} - \Gamma\right). \tag{4.3}$$

There is a slight difference in the above definitions, and then we record these broken pairs for the next process. This process is shown in pseudo-code in Algorithm 2.

Algorithm 3 Broken pixel-block pairs detection

Input: Initial model structure of the hypothetical noise p;
Output: Broken pixel-block pairs

```
Detect the broken pixel-block pairs;
for each pair (p, Q_m^B) do
  if γ_m ≥ γ̄ then
    (p, Q_m^B) is broken pair
  else
    (p, Q_m^B) is stable pair
  end if
  return the state of pair (p, Q_m^B)
end for
Record the broken pixel-block pairs.
```

4.1.3 Structure modification

Then, we try to exchange the broken pair by new one which is kept as a spare pair in the training process and remove the hypothesized noise by using the modified pixel-block structure as shown in Fig. 4.3. First, we remove the broken pixel-block pairs and replace by the new pairs from candidate supporting blocks set $\{Q_1^B, Q_2^B, ..., Q_K^B, Q_{K+1}^B, ...\}$ (update in order). At last, a modified structure can be constructed and then we use the modified structure to modify the "hypothetical noise" to get a modified result.

As we discussed at first, HoD is a new strategy for the background model update. It is not only applied on the top of CPB, but also can be applicable to the other pixel-correlation based algorithms (such as ViBe[20] based on random neighboring pixels, SuBSENSE[22] based on local binary similarity patterns features or our previous work CP3[25]), HoD provides a new and natural thought: the structure of background model can be updated by the designed correlation weight, which is discussed in details in Section 3.2 and the validity of HoD is proved in Section 4.2 and 4.3.

4.2 Ability of HoD

4.2.1 Verification of HoD's performance

In this section, in order to verify the HoD's performance under *open-set* condition, we compare the results of CPB and CPB+HoD in the sequence *canoe*, which is a typical scene with rippling water [94]. In this experiment, we select the first 300 frames for training and then at detecting process, the frame #845 to #930 with the continuous movement of the canoe, a total of 86 frames are selected as the testing frames. Fig.4.4 shows the typical results of CPB and CPB+HoD and Fig. 4.5 illustrates the *F-measure* and *False Positives* comparison between CPB and CPB+HoD in the sequence *canoe* overtime. From Fig. 4.4 and Fig. 4.5, it is clear that with the help of HoD, CPB+HoD has a significant improvement over CPB and further restrained the noise in scene. Table.5.2 illustrates a change in False Positives of ♯860 and ♯900 between CPB and CPB+HoD, from the table, we can note that HoD greatly restrains the noise in dynamic scene, in which as one example, the $FalsePositives$ in #860 drastically dropped from 320 to 1. These results suggest that HoD can effectively suppress the degradation in CPB with the passage of time.

Table 4.1: A change in False Positives of ♯860 and ♯900

Methods	The number of False Positives	
	♯860	♯900
CPB	320	350
CPB+HoD	**1**	**33**

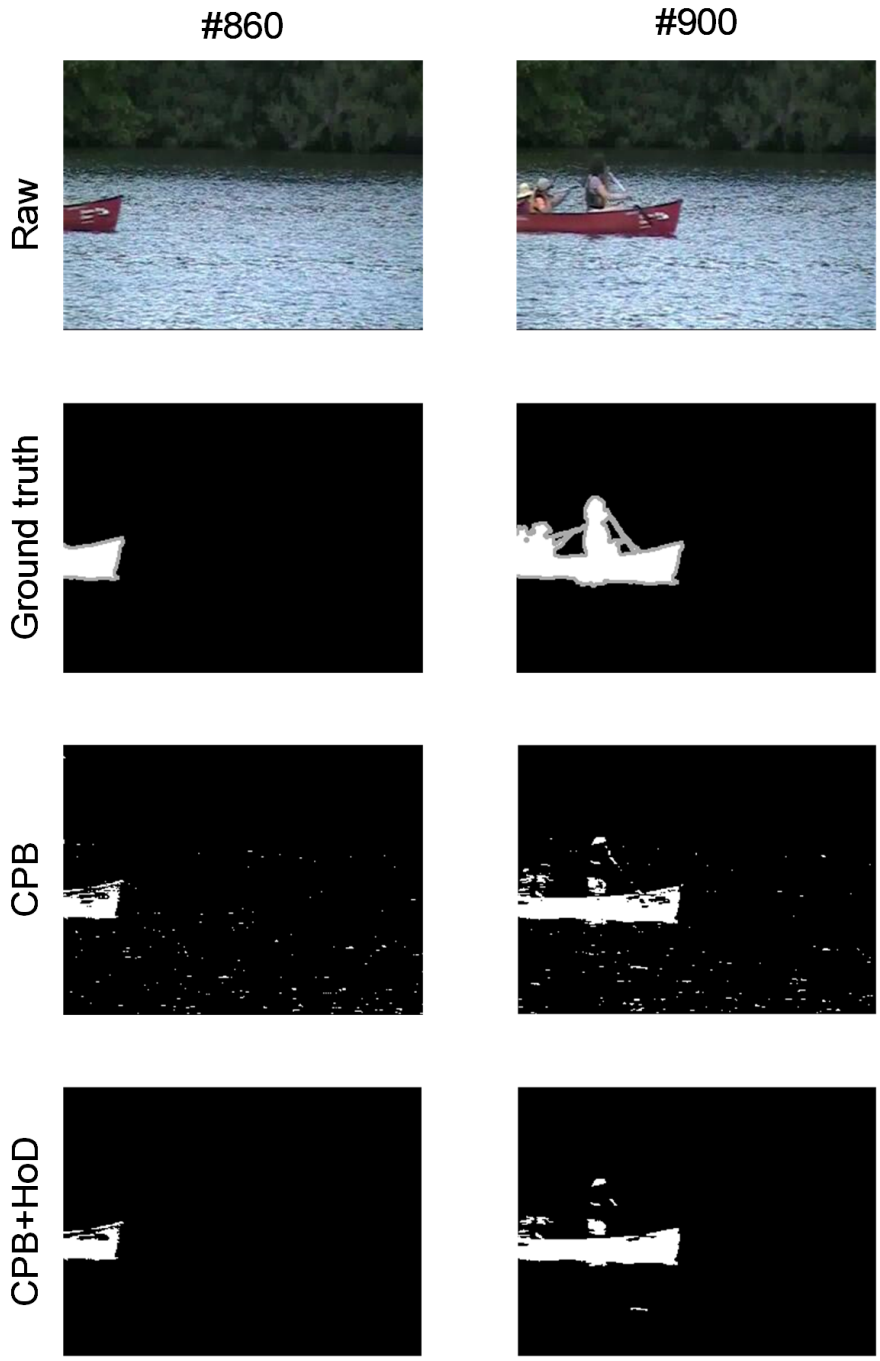
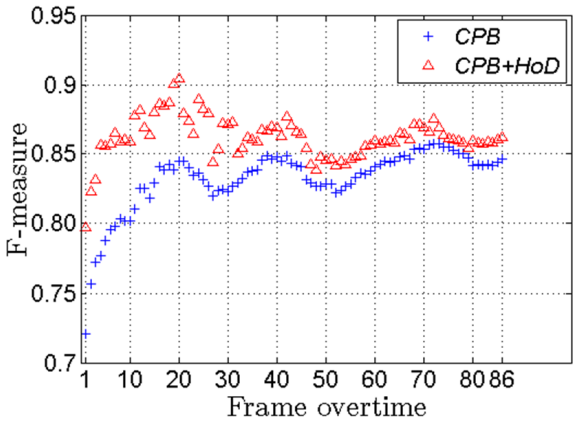
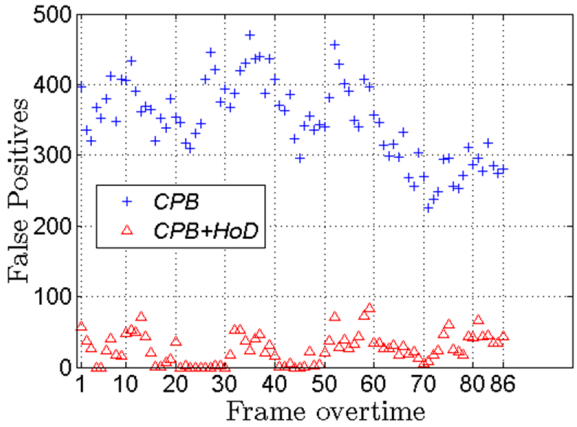


Figure 4.4: Typical results for CPB and CPB+HoD in the sequence *canoe*.



(a) F-measure comparison



(b) False Positives comparison

Figure 4.5: Comparison of CPB and CPB+HoD in the sequence *canoe* overtime.

As we introduced, with the help of HoD, CPB can update the pixel-block structure to further restrain the noise in background and adapt the changes overtime. This is the reason that CPB+HoD can lead a better result than CPB.

4.2.2 Ability of HoD under adversarial data

To get an idea of what adversarial looks like, consider one demo from "Explaining and Harnessing Adversarial Examples"[96]: inputting a panda image, and adding some perturbation that has been evaluated to make the image be recognized as a gibbon with high confidence. Similarly, we can define the following training data as the adversarial data:

- training data includes a high-density crowd or large-scale object;

- foreground information is mixed with or even covers background.

Fig. 4.6 shows a typical example of adversarial data, which is from the sequence *fall* with swaying branches in the CDW-2012 dataset[94]. The giant truck passes the background and covers half of the background information. The typical results under adversarial training are shown in Fig. 4.7. In this case, the training data includes 150 frames (#2460-#2609, selected from the sequence *fall*): (a) 120 frames without any large-scale objects (#2460-#2579); (b) 30 frames with a giant truck (#2580-#2609) and the interference rate is 20 %, we define the interference rate as the percentage of adversarial frames to total frames.

As mentioned in Section 2.3.5, CPB is not good at the adversarial data. However, HoD can help CPB to resist the interference from the adversarial data and we design the experiments to verify the ability of HoD and the details are presented in Table. 4.2. Fig. 4.9 shows the typical results of CPB and CPB+HoD and Fig. 4.8 shows the *F-measure* comparison between CPB and CPB+HoD in the six different cases.

Figure 4.6: A typical example of adversarial data: giant truck passes the background.

Table 4.2: Experimental design under the adversarial training data from the sequence *fall*

Case	Interference rate	Frames without any large-scale objects	Frames with a giant truck	Total number of frames
1	5%	#2010-#2579 (570 frames)	#2580-#2609 (30 frames)	600 frames
2	10%	#2310-#2579 (270 frames)		300 frames
3	15%	#2410-#2579 (170 frames)		200 frames
4	20%	#2460-#2579 (120 frames)		150 frames
5	25%	#2490-#2579 (90 frames)		120 frames
6	30%	#2510-#2579 (70 frames)		100 frames

We note that CPB will lose the efficiency as the interference increases. However, HoD can help CPB to resist the interference, which is demonstrated in Fig. 4.9 and Fig. 4.8, because HoD can repair and stabilize the initial model structure of CPB by selecting new pixel-block pairs from the candidate supporting block set as described in Section 4.1. The results demonstrate the ability of HoD under the adversarial data.

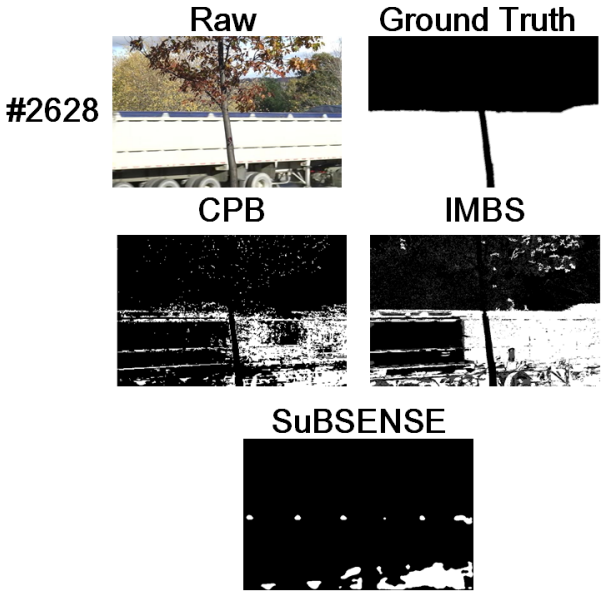


Figure 4.7: Influence of adversarial data on the detection.

4.2.3 Discussion

The HoD prevents the degradation of the CPB model over time (Figs. 4.4 and 4.5). The HoD allows the CPB model to adapt to background changes to better resist noise generation and maintain its robustness. In addition, the HoD has better adaptability to adversarial data (Figs. 4.9 and 4.8). Under certain circumstances, CPB+HoD can achieve better results without being affected by the adversarial data.

Because the HoD can provide a new update strategy for the CPB model with the modified background model structure to replace the broken pairs, this strategy stabilizes the robust structure of the model and enables the CPB model to maintain efficiency over long-term use.

In general, the HoD provides a new, intuitive strategy for the pixel-correlation algorithms of the background model to update to maintain the robustness of algorithms to background changes after prolonged use.

4.3 Summary

Chapter 4.1 proposed the Hypothesis on degradation modification (HoD) . First, we introduced the mechanism of HoD in detail including three steps: (1) 'broken pairs' detection; (2) 'broken pairs' modification; (3) update with modified model. Then, we verified the performance of HoD with the reliable experiments. Moreover, we proved the ability of HoD under the adversarial data.

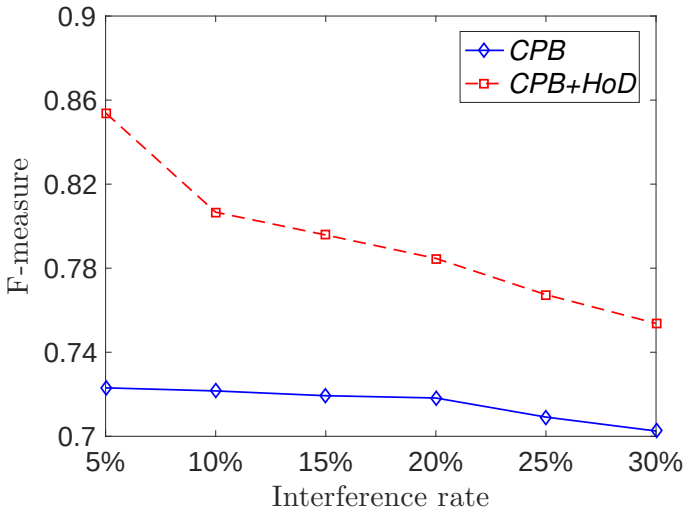


Figure 4.8: Comparison of CPB and CPB+HoD in six interference cases.

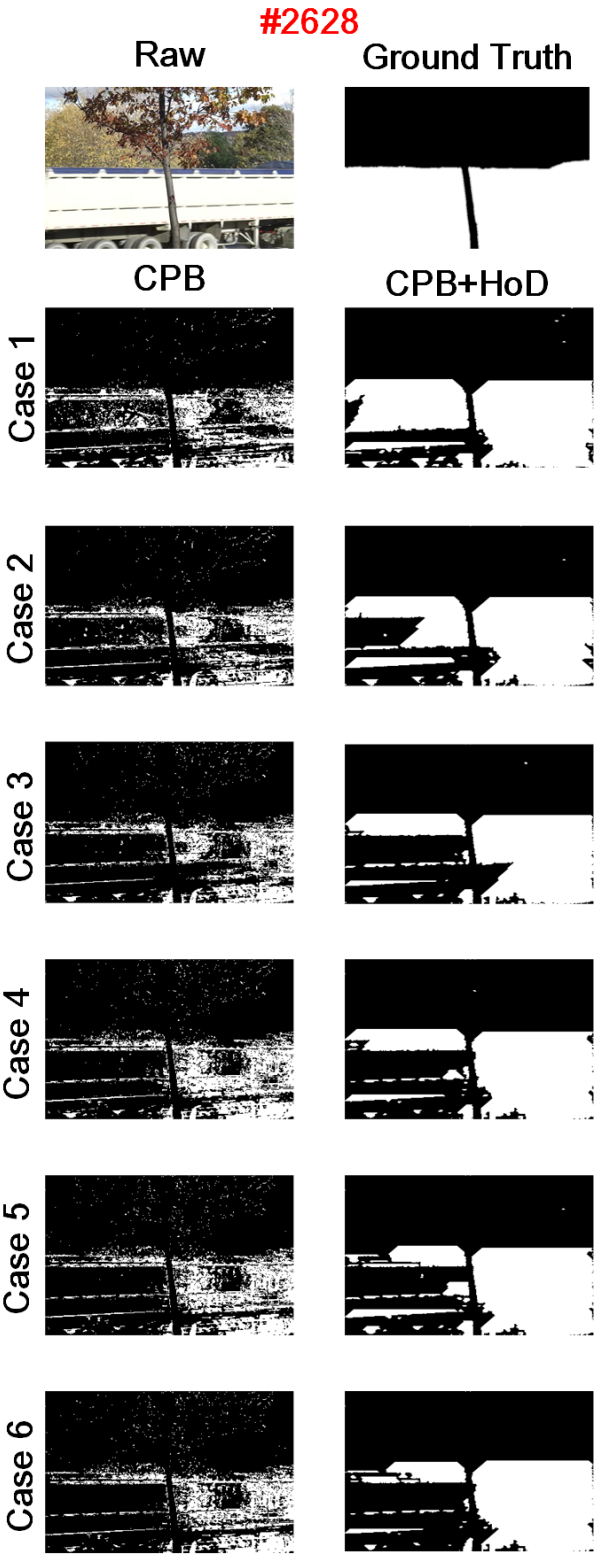


Figure 4.9: Typical results of CPB and CPB+HoD in different interference cases.

Chapter 5. Experimental Evaluation

In this chapter, we would like to introduce the evaluation of experiments for our methods including five main contents: (1) Analysis measurements; (2) Experimental setup; (3) Experimental comparison: (4) Parameter discussion and (5) Computational cost. In the end, the discussion will be presented to have a summary of the experimental results for our methods CPB and CPB+HoD.

5.1 Analysis measurements

In our CPB, we use the correlation between co-occurrence pixel-block pairs to detect objects. This can lead to more sensitive pixel detection than the coarse block detection of co-occurrence-based block-correlation methods [41, 97]. To analyze the quality of our method, we utilize four common analysis measurements: *Accuracy*, *Precision*, *Recall*, and *F-measure*. These metrics are widely used to estimate performance in pattern recognition and binary classifiers[98, 99, 100], as well being used to evaluate the quality of background subtraction methods [75, 101]. Here, $Acc(i)$, $Prec(i)$, and $Rec(i)$ for test frame i are defined as

$$Acc(i) = \frac{TP(i) + TN(i)}{TP(i) + TN(i) + FP(i) + FN(i)}, \tag{5.1}$$

$$Prec(i) = \frac{TP(i)}{TP(i) + FP(i)}, \tag{5.2}$$

$$Rec(i) = \frac{TP(i)}{TP(i) + FN(i)}, \tag{5.3}$$

where

- $TP(i)$ is the number of true positive pixels in frame i;
- $FP(i)$ is the number of false positive pixels in frame i;
- $TN(i)$ is the number of true negative pixels in frame i;
- $FN(i)$ is the number of false negative pixels in frame i.

True positive, *false positive*, *true negative* and *false negative* are four possible outcomes in an instance. If the instance is positive and it will be classified as positive, it is counted as a true positive; if it is classified as negative, it is counted as a false negative. If the instance is negative and it will be classified as negative, it is counted as a true

negative; if it is classified as positive, it is counted as a false positive. Fig. 5.1 shows an abridged general view of the analysis measurements: *Accuracy*, *Precision* and *Recall* that can be calculated from it. Through Fig. 5.1, *Accuracy* can be seen as the proximity of measurement results to the ground truth, *Precision* can be seen as a measure of exactness for foreground extraction and *Recall* can be seen as a measure of completeness of detection.

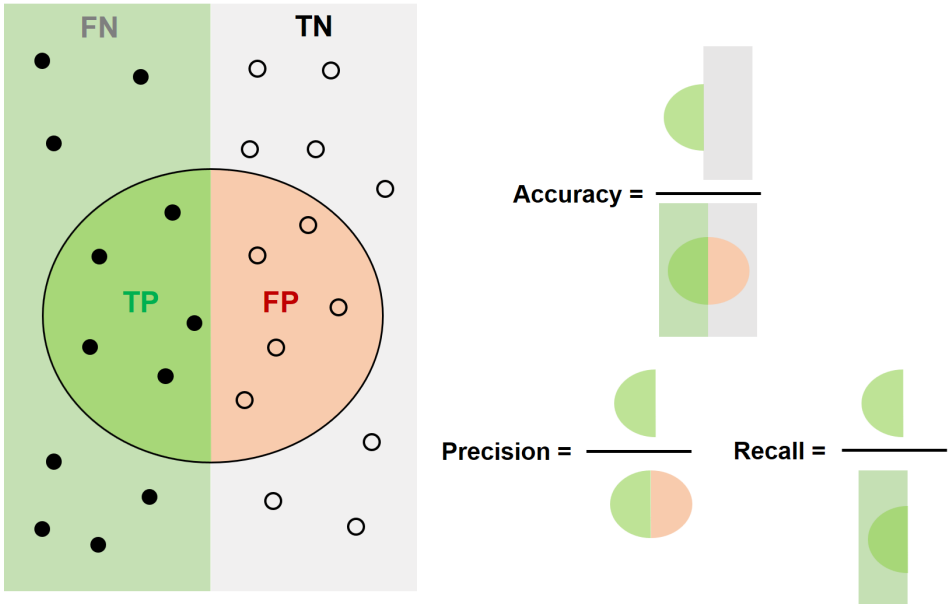


Figure 5.1: Abridged general view of the analysis measurements: *Accuracy*, *Precision* and *Recall*.

The *Accuracy* is defined by:

$$Accuracy = \frac{1}{I}\sum_{i=1}^{I} Acc(i). \tag{5.4}$$

The *Precision* is defined as:

$$Precision = \frac{1}{I}\sum_{i=1}^{I} Prec(i). \tag{5.5}$$

The *Recall* is defined as:

$$Recall = \frac{1}{I}\sum_{i=1}^{I} Rec(i). \tag{5.6}$$

Accuracy, *Precision*, and *Recall* are calculated as average value for the test frames. The $F - measure$ is defined as

$$F - measure = \frac{2Precision \cdot Recall}{Precision + Recall}. \tag{5.7}$$

$F-measure$ is calculated as an average value for the test frames, which is a computing score of $Precision$ and $Recall$ as a weighted harmonic mean value; I is the test frame sequence.

For further evaluating our CPB and CPB+HoD, we introduce the peak signal-to-noise ratio (PSNR) as our metric[102, 103], which can be used o measure the quality of the estimated resulted compared with the background truth[104]. The definition of $PSNR$ is calculated as follows:

$$PSNR = 10 \cdot \log_{10} \left(\frac{255^2}{MSE} \right), \tag{5.8}$$

where MSE is the mean square error.

5.2 Experimental setup

Considering the several challenges of video surveillance for background subtraction algorithm[75]. We consider the following datasets to evaluate the proposed methods:

- **PETS2001 dataset[105]:** one typical sequence of gradual illumination changes.
- **AIST-Indoor dataset:** the sequence with sudden illumination change, which contains the strong sudden light changes when the auto-door opening, in such moment it is difficult to detect true foreground from the scene. AIST-Indoor dataset is provided by the National Institute of Advanced Industrial Science and Technology in Japan.
- **SBMnet dataset[11]:** one sequence *advertisementBoard* with strong background motion is selected from SBMnet dataset for testing, and this sequence contains an ever-changing advertising board in the scene.
- **CDW-2012 dataset[94]:** one typical sequence *canoe* with water rippling is selected from the CDW-2012 dataset, the *sofa* sequence with objects stopping for a short while and then moving away and the *sidewalk* sequence under camera jitter condition.

Above these datasets, we use six difficult challenges which are the issues that need to be solved for background subtraction for the evaluation of our methods:

- **Gradual Illumination Changes:** the light intensity typically varies during day.
- **Sudden Illumination Changes:** for example the sudden switch of light, strongly affects the observation of object to lead a fault for detection.
- **Sudden Background Motion:** sudden changes in background.
- **Regular Background Motion:** regular movement e.g. swaying tree, waving water.

- **Intermittent Object Motion:** abandoned objects and objects stopping for a short while and then moving away.

- **Camera Jitter:** scene is unstable due to the camera is not fixed.

Experimental comparisons are discussed in Section 5.3. In the thesis, we mainly focus on gradual illumination changes, sudden illumination changes, regular background motion and sudden background motion. Major experiments will be presented around these four challenges as discussed in Section 5.3 in detail.

We compare the proposed CPB and CPB+HoD with six different foreground detection techniques: GMM[15] and KDE[16], which are two well-known traditional algorithms, and four state of the art techniques IMBS[68], T2FMRF-UV[27],ViBe[20] and SuBSENSE[22].

At first, GMM[15] and KDE[16] are two main basic standard techniques that are often used to make the basic comparison[106, 20, 107, 108]. Second, the state of the art techniques IMBS[68] and T2FMRF-UV[27] are the foreground extraction techniques specifically for dynamic background. And then, ViBe[20] and SuBSENSE[22], which are two of the leading unsupervised techniques for foreground detection, especially SuBSENSE[22] is one of the top-ranked techniques in CDW-2012 dataset at present. Based on the above reasons, we select these six different techniques for comparative experiments.

In contrast to the methods with complex strategies[68, 27, 22], CPB is a low-complexity algorithm that is more easily realized. The parameters for GMM, KDE, IMBS, T2FMRF-UV, ViBe and SuBSENSE were set by using the tool bgslibrary[109]. In experiments, we set each block as 8×8 pixels for CPB, the parameters are shown in Table. 5.1, here we set η as 2.5 based on our previpus work[25] and another important threshold λ has been discussed in 5.3 how to decide.

Table 5.1: Parameters setting of CPB

Number of supporting blocks K	20
Gaussian model threshold η	2.5
Correlation dependent decision threshold λ	0.5

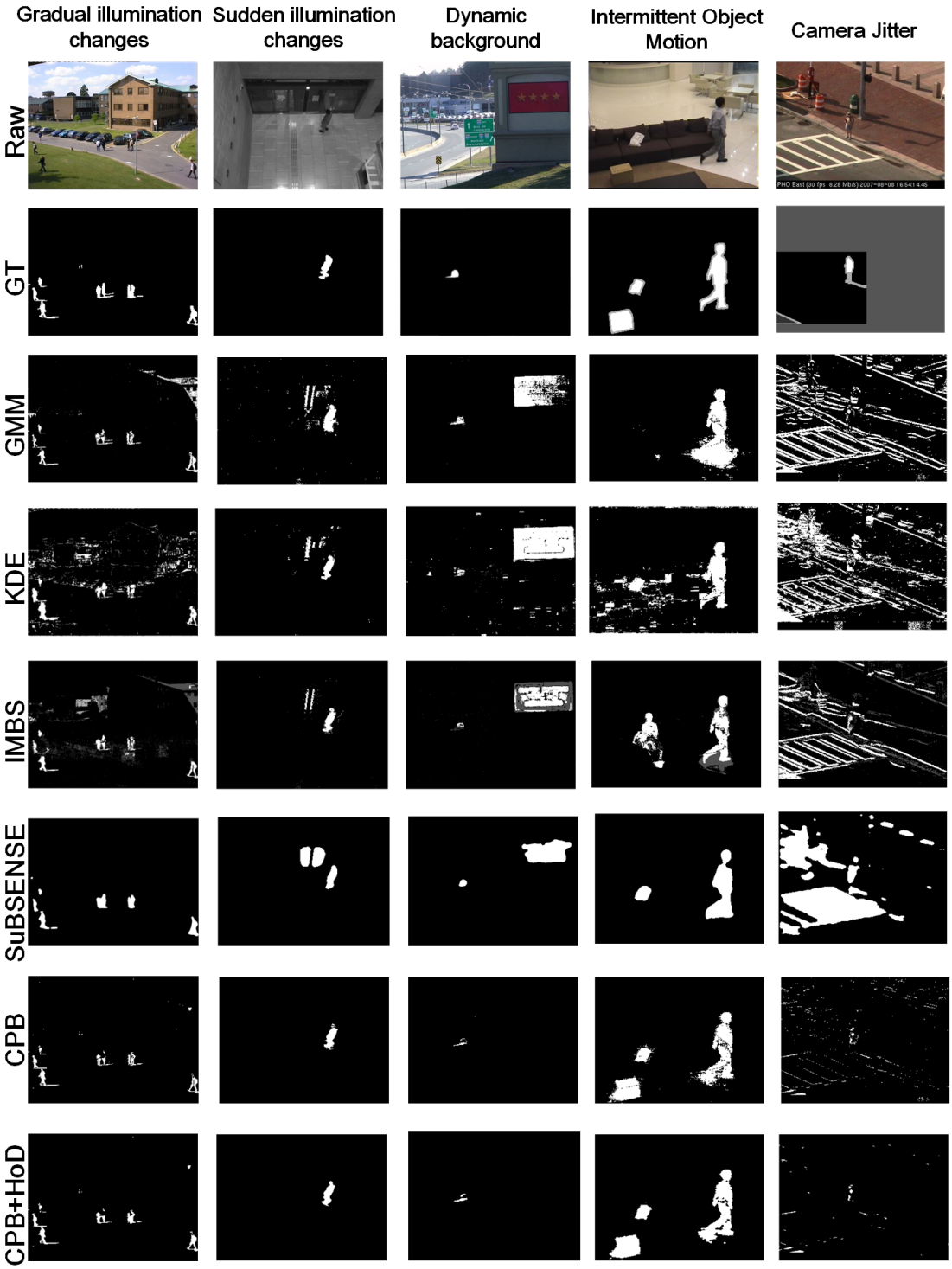


Figure 5.2: Foreground detection results in different challenging sequences.

Table 5.2: Comparison in different challenging categories

Method	Measure	Category				
		Gradual illumination changes	Sudden illumination changes	Dynamic background	Intermittent object motion	Camera jitter
GMM	Precision	0.6465	0.6523	0.5151	0.8462	0.4150
	Recall	0.9508	0.9207	0.5196	0.6876	0.4811
	F-measure	0.7697	0.7636	0.5174	0.7587	0.4456
	PSNR	39.46	40.57	26.92	36.11	31.79
KDE	Precision	0.5181	0.5896	0.4962	0.7361	0.5640
	Recall	0.8836	0.6944	0.4856	0.7820	0.5167
	F-measure	0.6532	0.6377	0.4909	0.7583	0.5393
	PSNR	17.77	38.16	21.67	25.99	33.68
IMBS	Precision	0.5162	0.5760	0.5095	0.8353	0.4457
	Recall	0.8841	0.6923	0.5118	0.7298	0.4879
	F-measure	0.6518	0.6288	0.5107	0.7790	0.4658
	PSNR	16.20	36.36	30.09	28.21	32.10
SuBSENSE	Precision	0.9008	0.5864	0.5018	0.9556	0.5966
	Recall	0.8840	0.7047	0.5033	0.7803	0.5079
	F-measure	0.8923	0.6401	0.5025	0.8591	0.5487
	PSNR	54.11	37.14	27.62	32.29	34.92
CPB	Precision	0.9566	0.8651	0.7653	0.8928	0.6365
	Recall	0.7517	0.8181	0.5118	0.8691	0.5051
	F-measure	0.8418	0.8409	0.6133	0.8808	0.5633
	PSNR	56.05	53.14	36.64	32.04	34.22
CPB+HoD	Precision	0.9652	0.8668	0.7973	0.9079	0.6384
	Recall	0.7562	0.8227	0.5214	0.8750	0.5055
	F-measure	0.8480	0.8442	0.6305	0.8912	0.5642
	PSNR	56.39	53.31	37.39	32.69	34.28

* Note that red entries indicate the best in $F-measure$, and blue entries indicate the second best.

5.3 Experimental comparison

Fig. 5.2 shows examples of foreground detection for a typical frame from each dataset sequence. Table 5.2 lists the results of the performance measurements of CPB and CPB+HoD with other methods from all the categories, respectively. Compared with above foreground detection results, the proposed algorithms outperform the methods GMM, KDE , IMBS and SuBSENSE in most testing sequences. Meanwhile, CPB+HoD is quite efficient in extracting foreground from sequences that suffers from sudden illumination changes and dynamic background. Furthermore, it is should be noted that CPB and CPB+HoD can lead high $Precision$ and $PSNR$ in most testing sequences as the results shown in Table 5.2, that means our algorithm is robust against noise for detecting foreground in severe scenes.

Based on co-occurrence pixel-block pairs, CPB can build one prospective background model from a scene, such background model contains spatial and temporal information of each pixel in sequence, and then CPB can analyze the current sate of each pixel effectively with these information. In other words, at training process, CPB can learn the information of scene, whether the scene is dynamic or static, our model can acquire the regularity of scene. Then, at detecting process, when any object enters into the scene and the information of this object is out of range of our model, so we can extract

the object from the scene efficiently.

For that reason, CPB does well in above scenes. On the basis of this, we introduce a HoD into CPB to adapt dynamic changes in scenes and reinforce robustness in real conditions. Through the results of above experiments, CPB+HoD leads a good performance in various scenes.

Table 5.3: Performance evaluation for foreground detection during illumination changes

Datasets	**PETS 2001**				**AIST**			
Methods	Precision	Recall	F-measure	PSNR	Precision	Recall	F-measure	PSNR
GMM[15]	0.6465	0.9508	0.7697	39.46	0.6523	0.9207	0.7636	40.57
KDE[16]	0.5181	0.8836	0.6531	17.77	0.5896	0.6944	0.6377	38.16
IMBS[68]	0.5162	0.8841	0.6518	16.20	0.5760	0.6923	0.6288	36.36
T2FMRF-UV[27]	0.5818	0.8365	0.6863	34.94	0.6382	0.5818	0.6087	45.65
ViBe[20]	0.7059	0.8821	0.7842	43.42	0.5005	0.5146	0.5074	9.11
SuBSENSE[22]	0.9008	0.8840	0.8923	54.11	0.5864	0.7047	0.6401	37.14
CPB	0.9566	0.7517	0.8418	56.05	0.8651	0.8181	0.8409	53.14
CPB+HoD	0.9652	0.7562	0.8480	56.39	0.8668	0.8227	0.8442	53.31

* Note that red entries indicate the best in measurement, and blue entries indicate the second best.

Table 5.4: Performance evaluation for foreground detection during background motion

Datasets	**SBMnet**				**CDW-2012**			
Methods	Precision	Recall	F-measure	PSNR	Precision	Recall	F-measure	PSNR
GMM[15]	0.5151	0.5196	0.5174	26.92	0.6748	0.7024	0.6883	21.71
KDE[16]	0.4962	0.4856	0.4909	21.67	0.6584	0.8630	0.7468	17.22
IMBS[68]	0.5095	0.5118	0.5107	30.09	0.7315	0.8911	0.8035	21.60
T2FMRF-UV[27]	0.5508	0.5179	0.5338	35.38	0.6797	0.6114	0.6438	23.49
ViBe[20]	0.6427	0.5368	0.5850	35.16	0.8114	0.7821	0.7965	28.02
SuBSENSE[22]	0.5018	0.5033	0.5025	27.62	0.9766	0.7649	0.8573	30.80
CPB	0.7653	0.5118	0.6133	36.64	0.9283	0.7730	0.8436	32.61
CPB+HoD	0.7973	0.5214	0.6350	37.39	0.9809	0.7830	0.8708	34.16

* Note that red entries indicate the best in measurement, and blue entries indicate the second best.

We also do more experiments in the main challenges: **Illumination changes** and **Background motion**. Fig. 5.3 shows the typical results on the two challenges and more experimental results of the foreground detection are presented in Fig. 5.4 and 5.5. Table 5.3 and Table 5.4 list the evaluation of these approaches in pixel level. *F-measure* is the comprehensive evaluation for foreground detection in pixel level and it should be as large as possible.

- **Illumination changes:** Fig. 5.4 shows the illumination change challenges, which are gradual illumination changes (global illumination changes) and sudden illumination changes (local illumination changes). The results demonstrate that our methods work well during illumination changes, especially sudden illumination changes. Here, we explain the difference in our model. For example, ViBe[20] is based on an assumption that the correlation of pixels, that is depended on the distance in spatial between them (e.g. the LBP feature in SuBSENSE[22],

where the target pixel has a high correlation with its neighboring pixels). However, this mechanism ignores the localized relation between each pixel, and the detection is insensitive and cannot adapt to local illumination changes as shown in Fig. 4.5. In CPB, due to the multiple supporting blocks for each target pixel, the co-occurrence pixel-block pairs build a multiple and spatial structure; thus, this structure maintains a stable statistical correlation more steadily for each target pixel and abandons the prior assumption of local correlation. This is why, CPB can extract the foreground sensitively under both global and local illumination changes as shown in Table 5.3.

- **Background motion:** Fig. 5.5 also shows two background motion challenges, which are sudden changes in background like a continuously changing advertising board in the scene and regular movement like rippling water. Video sequences contain the temporal context information and our CPB model can learn this information from the training data to avoid interference from background information such as background motion, during the detection process, and then accurately extract the current foreground information (*object*). This is different from the approaches based on local features (e.g., SuBSENSE [22] or ViBe[20]), which cannot adapt in non-ideal cases, for example, where textures are missing or there is a dynamic background. Based on this knowledge, our model can handle both of the changes well, and outperforms other methods significantly for sudden background motion as shown in Table 5.4.

Based on the results of Fig. 5.4 and Fig. 5.5, Table 5.5 groups the ability of our methods and others to deal with **Illumination changes** and **Background motion**. The suitable type of secen for each method is also indicated at Table 5.5 in third column.

Table 5.5: Performance evaluation on illumination changes and background motion

Methods	Illumination changes	Background motion	Outdoor/ Indoor
GMM[15]	Slow changes	-	Outdoor
KDE[16]	-	Slow movement	Outdoor
IMBS[68]	Slow changes	-	Outdoor
T2FMRF-UV[27]	Slow changes	Slow movement	Outdoor
ViBe[20]	Slow changes	Slow movement	Outdoor
SuBSENSE[22]	Yes	Yes	Outdoor & Indoor
CPB	Yes	Yes	Outdoor & Indoor
CPB+HoD	Yes	Yes	Outdoor & Indoor

5.4 Parameter discussion

The proposed method involves a significant parameter λ. Next, we discuss how to select a suitable value for λ. As discussed in Section 3.2, λ is a way to identify the

state of pixel p. In this case, we utilize the *AIST-Indoor* for the test and select three random frames #1653, #2184, and #2945 in order to estimate the results. We test the distribution of pixels for which $\frac{\Gamma}{\Gamma_{all}} > 0$ in the frame, and the distribution of pixels in the foreground. Figure 4.4 illustrates an example of the distribution graphs.

Through analyzing the results shown in Fig. 5.6, we find that $\frac{\Gamma}{\Gamma_{all}}$ follows the normalized distribution $\frac{\Gamma}{\Gamma_{all}} \sim N(\mu, \sigma^2)$. In other words, when $\frac{\Gamma}{\Gamma_{all}} > \mu - \sigma$, pixel p is likely to be in the foreground. Consequently, we can estimate the value of λ through the value of $\mu - \sigma$. We compute the values of μ as 0.7457, 0.6116, and 0.8089 from test frames #1653, #2184, and #2945, respectively; the corresponding values of σ are 0.1140, 0.1849 and 0.2753, respectively. From the calculation, the value of $\mu - \sigma$ is approximately 0.5307. From the analysis of pixel distribution based on most of our databases, we define an approximate value of 0.5 as a suitable value for $\mu - \sigma$. As a result, the value of λ is 0.5 here.

In order to assess the performance of $\lambda = 0.5$, we set an interval of [0.05,1] for λ for reference as a comparison. In this case, we use the sequences from the *AIST-Indoor* dataset for the test. We select 10 random test frames based on the ground truth to calculate the average metrics *Precision*, *Recall*, and *F-measure*. Figure 5.7 illustrates the results.

As shown in Fig. 5.7, when $\lambda = 0.5$, the *F-measure* has the highest value, and leads to the best performance considering both the *Precision* and *Recall*. Thus, we select $\lambda = 0.5$ as a suitable value for our CPB.

5.5 Computational cost

This section, we compare the processing time of our proposed methods with others in terms of fps. We evaluate the time required in foreground detection with the tool in the MATLAB platform (Intel E3 3.5GHZ and 16G) and utilize the testing frames from *canoe*[94] (frame size: 320×240). From the above results in Table. 5.6, observing that our methods lead an intermediate level in the detection process. CPB does not dominate in detecting time as illustrated in the Table. 3.1. Because of the multiple "pixel to block" structure of CPB, it takes some time to estimate the current state of the target pixel p as discussed in [81]. For each pair (p, Q^B), we define all the pixels Q_{mn} of block Q^B follow the equation 3.1, then we can estimate the current state of the pair (p, Q^B) as described in Section 3.2. Based on this mechanism, it takes time on detection. To solve this problem, on the one hand, we can appropriately reduce the size of block to achieve the reduction in detecting time. On the other hand, we would like to introduce the parallel processing implement into our detection, which is also employed in[110]. For example, we can divide the current input frame into N non-overlapping regions based on the number of available CPU cores. Then, detecting the foreground pixels of each individual region instead of the full scene detection. We would like to optimize the program to further reduce the processing time on detection in the future.

Table 5.6: Processing time comparison in FPS

Methods	Processing speed
GMM	81
KDE	69
IMBS	33
T2FMRF-UV	60
ViBe	149
SuBSENSE	14
CPB	**30**
CPB+HoD	**27**

5.6 Discussion

Compared with the classification algorithms based on ConvNets, training data preparation for the CPB model is simple. CPB is a statistical model based on the extraction of background information to distinguish the outliers (i.e., foreground) from the background. In contrast to ConvNets algorithms, CPB does not need labeled data (separate background and foreground data) for training. Hence, the cost of training data preparation for CPB is lower. However, CPB has its own disadvantages for training data selection. For instance, CPB cannot handle adversarial data well if the training data includes a high-density crowd or large-scale object, as described in Section 4.2.2. This is because CPB learns the background information to build the initial background model, which is a single Gaussian model based on the training data. If the foreground information severely interferes with the background, this produces a faulty background model; therefore, the training data for CPB must be selected carefully.

To reinforce the robustness of CPB, a HoD is proposed here, and the experimental results demonstrate the performance of the HoD. The HoD can be introduced at any time to repair or fix broken models. In general, the timing can be predicted by checking the amount of detected noise in the frames and comparing it with the normal detection frequency. However, the HoD may lose its effectiveness in some specific applications, such as small object detection. For example, in detecting and tracking honeybees [111], using the HoD to modify the initial result may remove the true target (honeybees) and lead to an erroneous modification. Because small objects are too similar to the hypothesized noise, as defined in Section 4.1, the HoD may detect the foreground as the hypothesized noise and lead to an erroneous modification. Thus, the HoD may not be suitable for this type of application.

5.7 Summary

In the chapter, we first introduced the design of experiments in detail and then we proved the performance of CPB and CPB+HoD on illumination changes and background

motion. It can be concluded that CPB and CPB+HoD can deal well with these two challenges, in particular, our methods perform well at sudden changes challenges (either sudden illumination changes or sudden background motion). Moreover, we also discussed the parameter setting in the CPB and compared the time consumption with other methods for foreground detection.

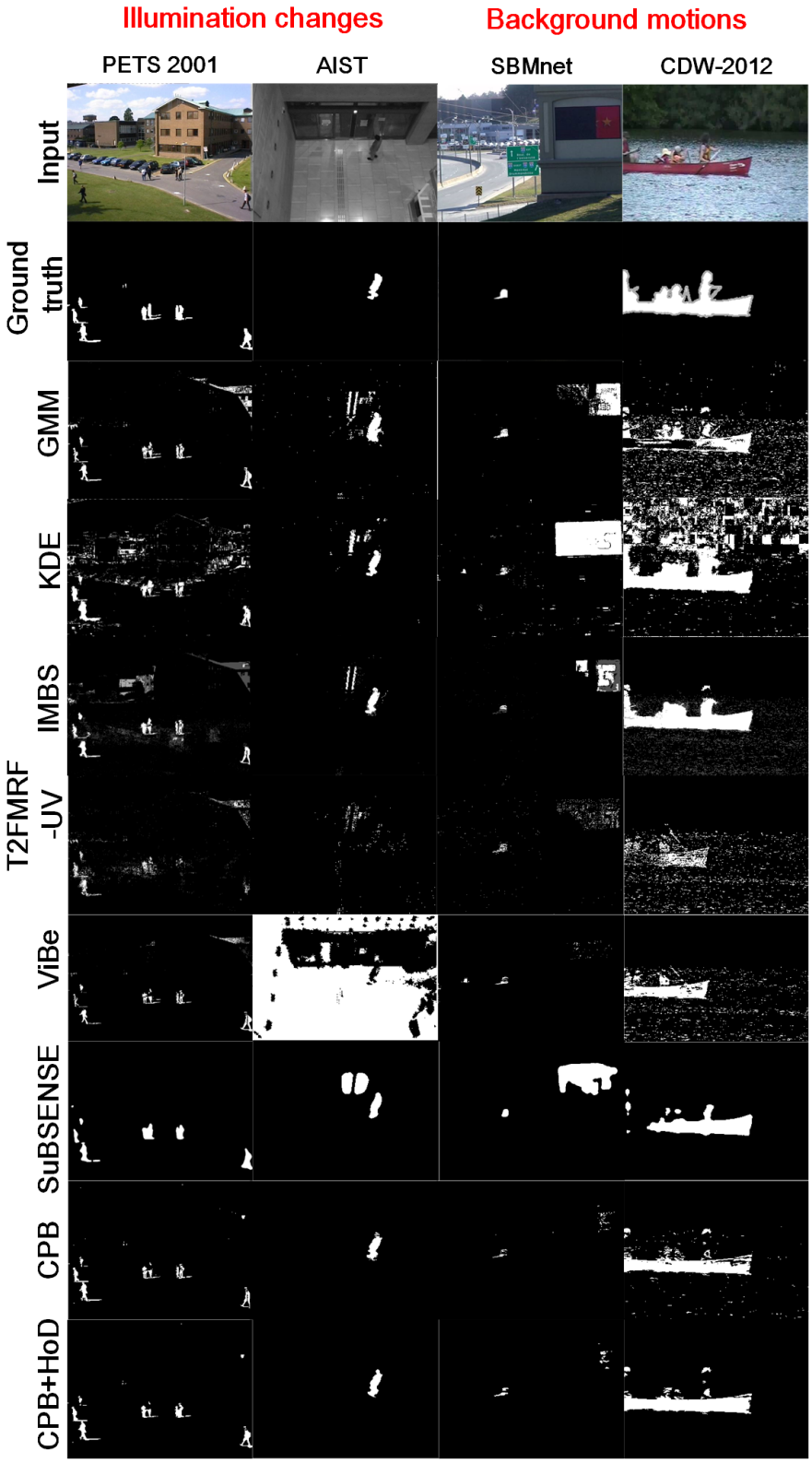


Figure 5.3: Typical results of each method on illumination changes and background motion.

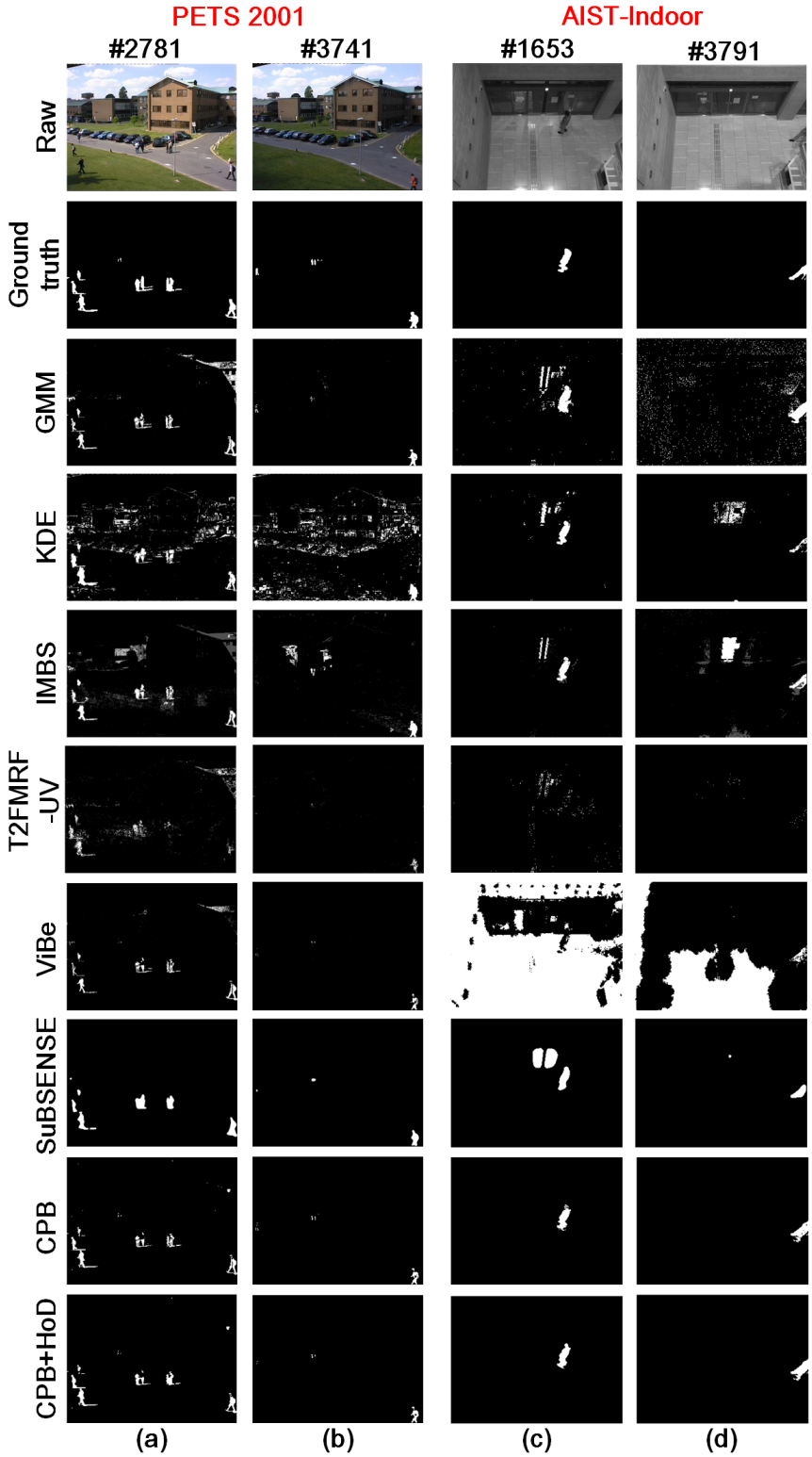


Figure 5.4: Representative results from the illumination change challenges: (a) illumination becomes stronger in daylight; (b) illumination becomes lower in daylight; (c) automatic door suddenly opens and the light changes; (d) person suddenly enters the scene, and the light switches on automatically.

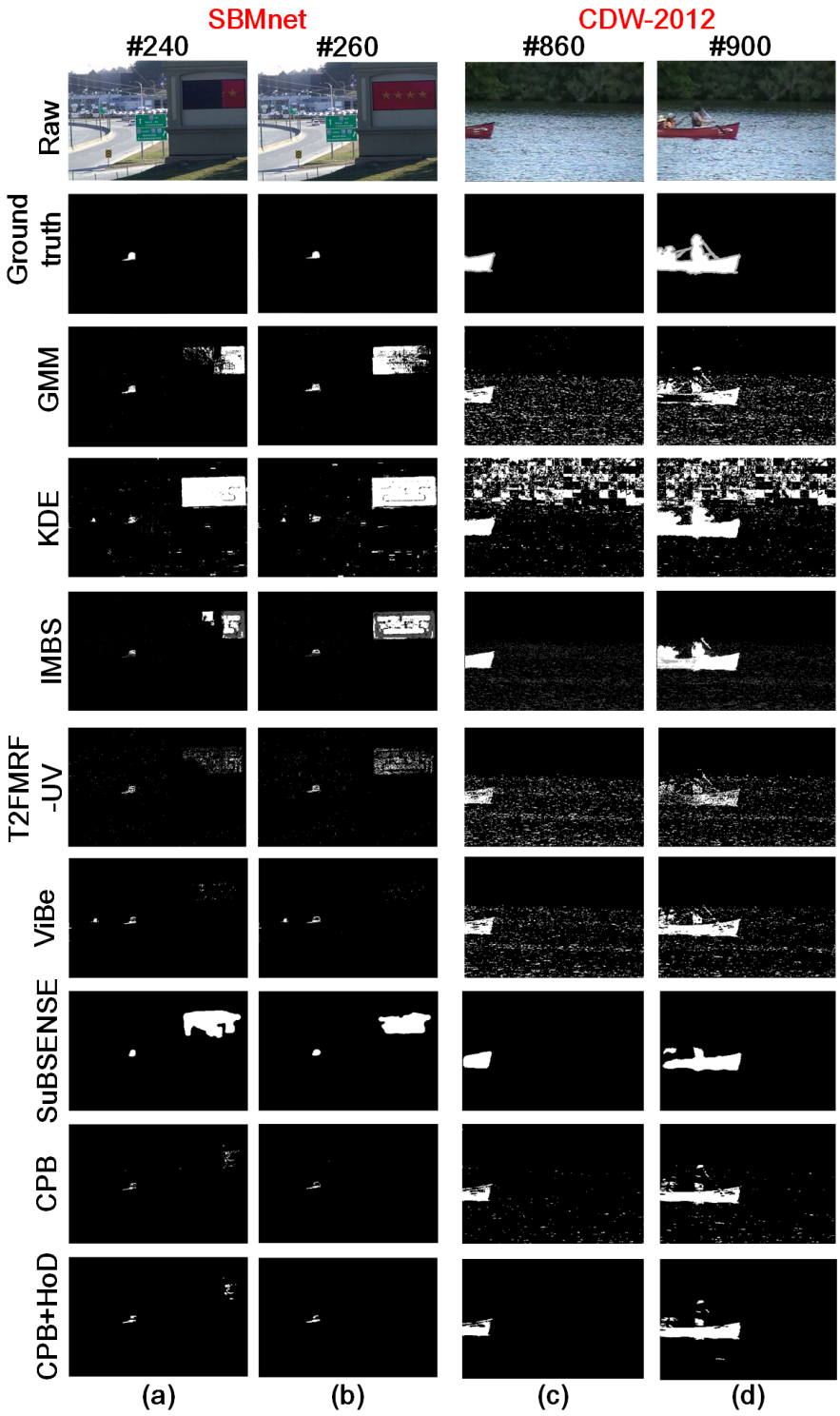


Figure 5.5: Representative results from background motion challenges: (a) advertisement board starts to change; (b) advertisement board stops changing; (c) canoe enters the scene; (d) canoe continues to move.

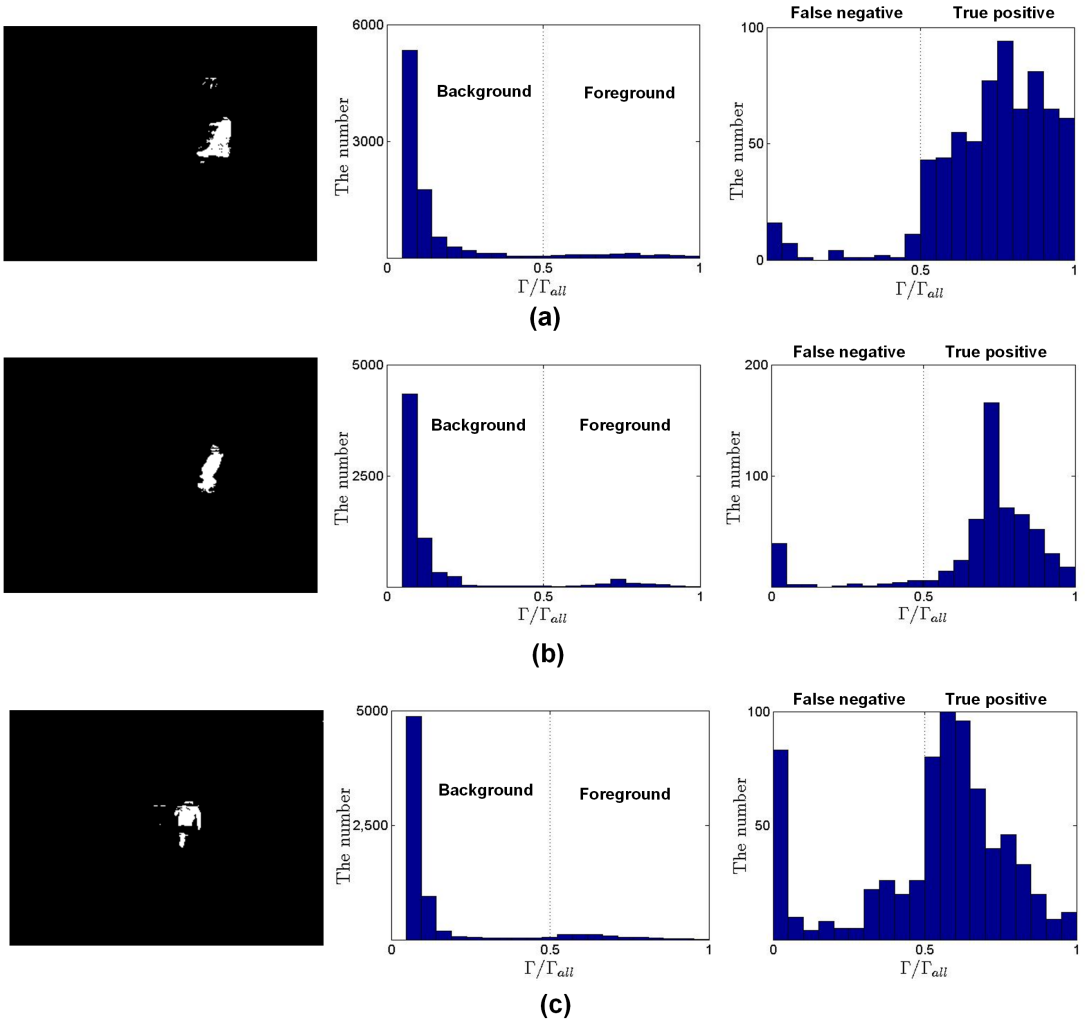


Figure 5.6: (a) distribution of pixels on #1653; (b) distribution of pixels on #2184; (c) distribution of pixels on #2945.

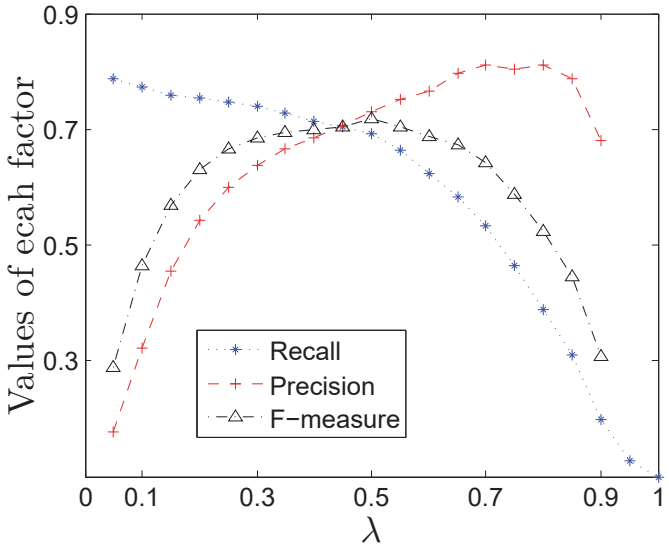


Figure 5.7: Efficiency of different value of λ.

Chapter 6. Conclusions and future works

6.1 Conclusions

A prospective background model based on the CPB model for foreground detection against dynamic scenes was developed. Because the CPB model is a pixel-block structure model with a correlation-dependent decision function, CPB is more reliable for foreground detection with background changes. By using a correlation-dependent decision mechanism, we created a non-parametrized model for CPB. As a spatio-temporal model, CPB consists of both local and global features for each pixel and achieves accurate detection, even under extreme environments, such as sudden illumination changes or sudden background motion. Moreover, compared with our previous model, CP3, CPB can greatly reduce the time consumed for training, making it suitable for real applications.

Furthermore, based on the CPB model, an HoD was proposed for foreground detection under dynamic scenes. The HoD was designed to handle the problem of strong background changes. The HoD improved the robustness of CPB and stabilized its effectiveness over long-term use. Furthermore, the HoD also helped CPB to resist interference under adversarial data. Experimental results from different challenges show the promise of proposed method.

In summary, the correlation between co-occurrence pixels and block pairs was verified. The correlation was more stable in intricate scenes than using the intensity of single pixels, and hence was used for background modeling. Secondly, a new selection strategy called the correlation-dependent decision function was introduced for accurate detection based on the correlation of the CPB model; this led to better performance than that of the previous binary classification. Finally, the CPB model allowed more efficient model building than the CP3 model. The HoD based on the CPB model was introduced to resist dynamic background changes, such as illumination changes and background motion. The HoD provides a new, intuitive concept: the structure of the background model can be updated by the designed correlation weight, which can be used in pixel-correlation algorithms for updating the background model. Experimental comparisons with other background-subtraction techniques based on challenging datasets demonstrated that the performance of the method was good.

6.2 future works

The HoD provided a new strategy for updating models, and in future work we would like to develop an online mode of the CPB structure by using the HoD to adapt to practical applications. At present, the block size is determined experimentally, although a general form should be found for this.

Moreover, CPB should be integrated into other fields, such as scene background initialization, object tracking, or defect detection.

Scene background initialization

Scene background initialization, which provides an initial background model that describes the scene with no foreground objects, is a prerequisite for many applications, including video surveillance, video segmentation, privacy protection for videos, and computational photography. The CPB model is a robust foreground detection method with a spatio-temporal background model. This background model can effectively learn the background information and features from training data without the effect of foreground objects; thus, the CPB model can effectively initialize the scene with the background model. We would like to apply the CPB model to scene background initialization.

Object tracking

Object tracking has great practical importance, and it is a difficult problem in various application scenarios, such as traffic monitoring, security monitoring of public scenes, and pedestrian volume analysis in complex scenes. Because existing algorithms are based on the single features, such as color, edges, or contours, in complex scenes, such algorithms lose their robustness and fail. Due to the merits of the CPB model in background modeling and foreground detection, we would like to consider using the CPB model for object tracking, and consequently to integrate the robust features of CPB to overcome the challenges in complex scenes for object tracking. The work presented in this thesis demonstrates that CPB and CPB+HoD can be effective for foreground detection in extreme scenes, such as those with illumination changes or background motion.

Defect detection

Defect detection is critical for industrial workpiece inspection. Using an image algorithm for detection is limited by insufficient numbers of samples and the effect of the workpiece material on the robustness of the algorithm. The CPB model provides a new strategy for defect detection. Defects can be detected as the foreground, whereas defect-free areas can be assigned as the background, similar to the foreground detection problem. The CPB model is resistant to changes in illumination and dramatic changes

in scene, and it does not require a large number of samples and can build a model from a small number of samples without learning the defects. These advantages can be adapted to actual production environments. In the future, we would like to apply the CPB model to defect detection applications to expand its practicality and applications.

Acknowledgements

First of all, I would like to thank my supervisor, Professor Shun'ichi Kaneko. In 2015, I came to Japan to start my study journey. A the beginning, I was very upset due to coming to an unfamiliar environment and different culture, but fortunately I met the best supervisor in my life, Professor Kaneko. Prof. Kaneko guided me very patiently, and also helped me in my study and abroad life. More importantly, he also told me the philosophy of life and taught me how to behave in real world. During the four years, under the careful guidance of Professor Kaneko, I understood what is the scientific research, what is the scientific thinking, what is the training of scientific research, and finally I can step into the real scientific research. Thank you Professor Kaneko for bringing me to the door of science research, thank you very much. I will never forget those four years, all of these will be remarkable in my whole life.

Thank you, Professor Satoshi Kanai, for reviewing my thesis and providing me so many key comments and suggestions.

Thank you, Associate Professor Takayuki Tanaka. Associate Professor Tanaka treats every student kindly and is very enthusiastic about every international student. Thank you very much for being my thesis supervisor. Your suggestion has helped me a lot in my research and papers, and I also thank you for your daily help. Every time in general meeting, your valuable experience and advice have benefited me a lot.

Thank you, Professor Manabu Hashimoto from Chukyo University, Japan, Professor Yutaka Satoh from National Institute of Advanced Industrial Science and Technology (AIST), Japan, Dr. Dong Liang from Nanjing University of Aeronautics and Astronautics, China. All of them are my cooperators , This study is based on their valuable suggestions and ideas. At the same time, they also gave me great help. Professor Hashimoto provided great information and suggestions for me and recommended me to attend the conference of VISAPP. Professor Satoh, during my internship at AIST, he gave me great support and help, which made me gain a lot. Dr. Liang, at the beginning of the study, his advice, experience and encouragement helped me to complete the initial implementation of CPB. Without their help, this study could not be completed, sincerely thank you all.

I would like to thank all the staffs and members in the Human-centric Engineering Laboratory, and I have been very happy with each other for the four years. Your friendship and help have accompanied me for four years of study and life, and have witnessed my growth. Especially Dr. Yoshio Tsuchiya, thank you for taking care of me during my initial entry into the lab, so that I can quickly integrate into this big HCE

family. Thank you, my good friend, Mr. Sheng Xiang, in daily life, not only in learning, but also helping me in life, I am grateful for his friendship in the four years. Thank you all for sharing me so many wonderful and unforgettable days and nights together in the Human-centric Engineering Laboratory.

Finally, I would like to thank my family in particular. On my way in study during the four years, my parents have unconditionally supported me and supported me to finish the road of my doctoral study. Thanks so much to them, to support my ideality, when my most difficult and upset , the encouragement and comfort from my parents, inspiring me to go on. Thanks for the unselfish support and understanding to me over the years. Without my parents' support, I can't finish my doctoral study. I love them!

References

[1] T. Bouwmans, "Traditional and recent approaches in background modeling for foreground detection: An overview," *Computer science review*, vol. 11, pp. 31–66, 2014.

[2] S. Y. Elhabian, K. M. El-Sayed, and S. H. Ahmed, "Moving object detection in spatial domain using background removal techniques-state-of-art," *Recent patents on computer science*, vol. 1, no. 1, pp. 32–54, 2008.

[3] W. Hu, T. Tan, L. Wang, and S. Maybank, "A survey on visual surveillance of object motion and behaviors," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 34, no. 3, pp. 334–352, 2004.

[4] A. Iosifidis, S. G. Mouroutsos, and A. Gasteratos, "A hybrid static/active video surveillance system," *International Journal of Optomechatronics*, vol. 5, no. 1, pp. 80–95, 2011.

[5] S.-C. S. Cheung and C. Kamath, "Robust background subtraction with foreground validation for urban traffic video," *EURASIP Journal on Advances in Signal Processing*, vol. 2005, no. 14, p. 726261, 2005.

[6] T. B. Moeslund, A. Hilton, and 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.

[7] I. Haritaoglu, D. Harwood, and L. S. Davis, "W/sup 4/: real-time surveillance of people and their activities," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 809–830, 2000.

[8] A. Yilmaz, O. Javed, and M. Shah, "Object tracking: A survey," *Acm computing surveys (CSUR)*, vol. 38, no. 4, p. 13, 2006.

[9] V. Mahadevan, W. Li, V. Bhalodia, and N. Vasconcelos, "Anomaly detection in crowded scenes," IEEE. Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on, 2010, pp. 1975–1981.

[10] L. Maddalena and A. Petrosino, "Towards benchmarking scene background initialization," in *International Conference on Image Analysis and Processing*. Springer, 2015, pp. 469–476.

[11] P.-M. Jodoin, L. Maddalena, A. Petrosino, and Y. Wang, "Extensive benchmark and survey of modeling methods for scene background initialization," *IEEE Transactions on Image Processing*, vol. 26, no. 11, pp. 5244–5256, 2017.

[12] M. Piccardi, "Background subtraction techniques: a review," vol. 4, IEEE. Systems, man and cybernetics, 2004 IEEE international conference on, 2004, pp. 3099–3104.

[13] N. J. McFarlane and C. P. Schofield, "Segmentation and tracking of piglets in images," *Machine vision and applications*, vol. 8, no. 3, pp. 187–193, 1995.

[14] M. H. B. Lee, "Background estimation for video surveillance," in *Image and Vision Computing New Zealand, IVCNZ, 2002*, pp. 315–320.

[15] C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," vol. 2, IEEE. Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on., 1999, pp. 246–252.

[16] A. Elgammal, R. Duraiswami, D. Harwood, and L. S. Davis, "Background and foreground modeling using nonparametric kernel density estimation for visual surveillance," *Proceedings of the IEEE*, vol. 90, no. 7, pp. 1151–1163, 2002.

[17] N. M. Oliver, B. Rosario, and A. P. Pentland, "A bayesian computer vision system for modeling human interactions," *IEEE transactions on pattern analysis and machine intelligence*, vol. 22, no. 8, pp. 831–843, 2000.

[18] K. Kim, T. H. Chalidabhongse, D. Harwood, and L. Davis, "Background modeling and subtraction by codebook construction," in *2004 International Conference on Image Processing, 2004. ICIP'04.*, vol. 5. IEEE, 2004, pp. 3061–3064.

[19] D. E. Butler, V. M. Bove, and S. Sridharan, "Real-time adaptive foreground/background segmentation," *EURASIP Journal on Advances in Signal Processing*, vol. 2005, no. 14, p. 841926, 2005.

[20] O. Barnich and M. Van Droogenbroeck, "Vibe: A universal background subtraction algorithm for video sequences," *IEEE Transactions on Image processing*, vol. 20, no. 6, pp. 1709–1724, 2011.

[21] M. Hofmann, P. Tiefenbacher, and G. Rigoll, "Background segmentation with feedback: The pixel-based adaptive segmenter," in *2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, 2012, pp. 38–43.

[22] P.-L. St-Charles, G.-A. Bilodeau, and R. Bergevin, "Subsense: A universal change detection method with local adaptive sensitivity," *IEEE Transactions on Image Processing*, vol. 24, no. 1, pp. 359–373, 2015.

[23] R. Ozaki, Y. Satoh, K. Iwata, and K. Sakaue, "Statistical reach feature method and its application to robust image registration," in *TENCON 2009-2009 IEEE Region 10 Conference*. IEEE, 2009, pp. 1–6.

[24] X. Zhao, Y. Satoh, H. Takauji, S. Kaneko, K. Iwata, and R. Ozaki, "Object detection based on a robust and accurate statistical multi-point-pair model," *Pattern Recognition*, vol. 44, no. 6, pp. 1296–1311, 2011.

[25] D. Liang, S. Kaneko, M. Hashimoto, K. Iwata, and X. Zhao, "Co-occurrence probability-based pixel pairs background model for robust object detection in dynamic scenes," *Pattern Recognition*, vol. 48, no. 4, pp. 1374–1390, 2015.

[26] F. El Baf, T. Bouwmans, and B. Vachon, "Type-2 fuzzy mixture of gaussians model: application to background modeling," in *International Symposium on Visual Computing*. Springer, 2008, pp. 772–781.

[27] Z. Zhao, T. Bouwmans, X. Zhang, and Y. Fang, "A fuzzy background modeling approach for motion detection in dynamic backgrounds," *Multimedia and signal processing*, vol. 346, pp. 177–185, 2012.

[28] W. Kim and C. Kim, "Background subtraction for dynamic texture scenes using fuzzy color histograms," *IEEE Signal processing letters*, vol. 19, no. 3, pp. 127–130, 2012.

[29] V. Morellas, I. Pavlidis, and P. Tsiamyrtzis, "Deter: Detection of events for threat evaluation and recognition," *Machine Vision and Applications*, vol. 15, no. 1, pp. 29–45, 2003.

[30] D.-S. Lee, "Online adaptive gaussian mixture learning for video applications," in *International Workshop on Statistical Methods in Video Processing*. Springer, 2004, pp. 105–116.

[31] Z. Zivkovic *et al.*, "Improved adaptive gaussian mixture model for background subtraction." in *ICPR (2)*. Citeseer, 2004, pp. 28–31.

[32] Y. Zhang, Z. Liang, Z. Hou, H. Wang, and M. Tan, "An adaptive mixture gaussian background model with online background reconstruction and adjustable foreground mergence time for motion segmentation," in *2005 IEEE International Conference on Industrial Technology*. IEEE, 2005, pp. 23–27.

[33] J. Cheng, J. Yang, Y. Zhou, and Y. Cui, "Flexible background mixture models for foreground segmentation," *Image and Vision Computing*, vol. 24, no. 5, pp. 473–482, 2006.

[34] A. Shimada, D. Arita, and R.-i. Taniguchi, "Dynamic control of adaptive mixture-of-gaussians background model," in *2006 IEEE International Conference on Video and Signal Based Surveillance*. IEEE, 2006, pp. 5–5.

[35] M. Amintoosi, F. Farbiz, M. Fathy, M. Analoui, and N. Mozayani, "Qr decomposition-based algorithm for background subtraction," in *2007 IEEE International Conference on Acoustics, Speech and Signal Processing-ICASSP'07*, vol. 1. IEEE, 2007, pp. I–1093.

[36] C. Cuevas, N. García, and L. Salgado, "A new strategy based on adaptive mixture of gaussians for real-time moving objects segmentation," in *Real-Time Image Processing 2008*, vol. 6811. International Society for Optics and Photonics, 2008, p. 681111.

[37] Y. Wang, Y.-h. Tan, and J.-w. Tian, "Video segmentation algorithm with gaussian mixture model and shadow removal [j]," *Opto-electronic engineering*, vol. 35, no. 3, pp. 21–25, 2008.

[38] S. Varadarajan, P. Miller, and H. Zhou, "Spatial mixture of gaussians for dynamic background modelling," in *2013 10th IEEE International Conference on Advanced Video and Signal Based Surveillance*. IEEE, 2013, pp. 63–68.

[39] T. Bouwmans, F. El Baf, and B. Vachon, "Statistical background modeling for foreground detection: A survey," in *Handbook of pattern recognition and computer vision*. World Scientific, 2010, pp. 181–199.

[40] T. Bouwmans, "Recent advanced statistical background modeling for foreground detection-a systematic survey," *Recent Patents on Computer Science*, vol. 4, no. 3, pp. 147–176, 2011.

[41] M. Seki, T. Wada, H. Fujiwara, and K. Sumi, "Background subtraction based on cooccurrence of image variations," vol. 2, IEEE. Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on, 2003, pp. II–II.

[42] A. Tavakkoli, "Automatic video object plane extraction using non-parametric kernel density estimation," *Mathematical Methods in Computer Vision, University of Nevada, Reno, NV*, 2005.

[43] A. Tavakkoli, M. Nicolescu, and G. Bebis, "Automatic statistical object detection for visual surveillance," in *2006 IEEE Southwest Symposium on Image Analysis and Interpretation*. IEEE, 2006, pp. 144–148.

[44] C. Ianasi, V. Gui, C. I. Toma, and D. Pescaru, "A fast algorithm for background tracking in video surveillance, using nonparametric kernel density estimation," *Facta universitatis-series: Electronics and Energetics*, vol. 18, no. 1, pp. 127–144, 2005.

[45] T. Tanaka, A. Shimada, D. Arita, and R.-i. Taniguchi, "Non-parametric background and shadow modeling for object detection," in *Asian Conference on Computer Vision*. Springer, 2007, pp. 159–168.

[46] ——, "A fast algorithm for adaptive background model construction using parzen density estimation," in *2007 IEEE Conference on Advanced Video and Signal Based Surveillance*. IEEE, 2007, pp. 528–533.

[47] Z. Zivkovic and F. Van Der Heijden, "Efficient adaptive density estimation per image pixel for the task of background subtraction," *Pattern recognition letters*, vol. 27, no. 7, pp. 773–780, 2006.

[48] S. Cvetkovic, P. Bakker, J. Schirris, and P. H. De With, "Background estimation and adaptation model with light-change removal for heavily down-sampled video surveillance signals," in *2006 International Conference on Image Processing*. IEEE, 2006, pp. 1829–1832.

[49] S. D. Witherspoon and M. Zhang, "Negative coeffcient polynomial kernel density estimation for visualization," in *Proceedings of the 18th conference on Proceedings of the 18th IASTED International Conference: modelling and simulation*. ACTA Press, 2007, pp. 398–403.

[50] R. Ramezani, P. Angelov, and X. Zhou, "A fast approach to novelty detection in video streams using recursive density estimation," in *2008 4th International IEEE Conference Intelligent Systems*, vol. 2. IEEE, 2008, pp. 14–2.

[51] Y. Mao and P. Shi, "Multimodal background model with noise and shadow suppression for moving object detection," *Journal of Southeast University*, vol. 20, no. 4, pp. 423–426, 2004.

[52] P. Tang, L. Gao, and Z. Liu, "Salient moving object detection using stochastic approach filtering," in *ICIG 2007*, 2007, pp. 530–535.

[53] ATavakkoli, M. Nicolescu, and G. Bebis, "Robust recursive learning for foreground region detection in videos with quasi-stationary backgrounds," in *ICPR 2006*, 2006.

[54] J. Rymel, J. Renno, D. Greenhill, J. Orwell, and G. A. Jones, "Adaptive eigen-backgrounds for object detection," in *2004 International Conference on Image Processing, 2004. ICIP'04.*, vol. 3. IEEE, 2004, pp. 1847–1850.

[55] Y. Li, "On incremental and robust subspace learning," *Pattern recognition*, vol. 37, no. 7, pp. 1509–1518, 2004.

[56] D. Skocaj and A. Leonardis, *Weighted and robust incremental method for subspace learning*. IEEE, 2003.

[57] J. Zhang and Y. Zhuang, "Adaptive weight selection for incremental eigen-background modeling," in *2007 IEEE International Conference on Multimedia and Expo*. IEEE, 2007, pp. 851–854.

[58] L. Wang, L. Wang, Q. Zhuo, H. Xiao, and W. Wang, "Adaptive eigenbackground for dynamic background modeling," in *Intelligent Computing in Signal Processing and Pattern Recognition*. Springer, 2006, pp. 670–675.

[59] R. Li, C. Yu, and X. Zhang, "Fast robust eigen-background updating for foreground detection," in *2006 International Conference on Image Processing*. IEEE, 2006, pp. 1833–1836.

[60] G.-A. Bilodeau, J.-P. Jodoin, and N. Saunier, "Change detection in feature space using local binary similarity patterns," IEEE. Computer and Robot Vision (CRV), 2013 International Conference on, 2013, pp. 106–112.

[61] K. IWATA, Y. SATOH, R. OZAKI, and K. SAKAUE, "Robust background subtraction based on statistical reach feature method," *The IEICE transactions on information and systems*, vol. 92, no. 8, pp. 1251–1259, 2009.

[62] X. Zhao, Y. Satoh, H. Takauji, and S. Kaneko, "Robust tracking using particle filter with a hybrid feature," *IEICE transactions on information and systems*, vol. 95, no. 2, pp. 646–657, 2012.

[63] X. Zhao, Z. He, and S. Zhang, "Defect detection of castings in radiography images using a robust statistical feature," *JOSA A*, vol. 31, no. 1, pp. 196–205, 2014.

[64] X. Zhao, Z. He, S. Zhang, S. Kaneko, and Y. Satoh, "Robust face recognition using the gap feature," *Pattern Recognition*, vol. 46, no. 10, pp. 2647–2657, 2013.

[65] D. Liang, S. Kaneko, M. Hashimoto, K. Iwata, X. Zhao, and Y. Satoh, "Robust object detection in severe imaging conditions using co-occurrence background model," *International Journal of Optomechatronics*, vol. 8, no. 1, pp. 14–29, 2014.

[66] D. Liang, S. Kaneko, H. Sun, and B. Kang, "Adaptive local spatial modeling for online change detection under abrupt dynamic background," in *2017 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2017, pp. 2020–2024.

[67] J. Ding, M. Li, K. Huang, and T. Tan, "Modeling complex scenes for accurate moving objects segmentation," in *Asian Conference on Computer Vision*. Springer, 2010, pp. 82–94.

[68] D. Bloisi and L. Iocchi, "Independent multimodal background subtraction." *CompIMAGE*, pp. 39–44, 2012.

[69] C. R. Wren and F. Porikli, "Waviz: Spectral similarity for object detection," in *IEEE International Workshop on Performance Evaluation of Tracking and Surveillance*, 2005, pp. 55–61.

[70] I. Goodfellow, H. Lee, Q. V. Le, A. Saxe, and A. Y. Ng, "Measuring invariances in deep networks," in *Advances in neural information processing systems*, 2009, pp. 646–654.

[71] M. Braham and M. Van Droogenbroeck, "Deep background subtraction with scene-specific convolutional neural networks," IEEE. Systems, Signals and Image Processing (IWSSIP), 2016 International Conference on, 2016, pp. 1–4.

[72] G. Zhang, Z. Yuan, Q. Tong, M. Zheng, and J. Zhao, "A novel framework for background subtraction and foreground detection," *Pattern Recognition*, vol. 84, pp. 28–38, 2018.

[73] N. Wang and D.-Y. Yeung, "Learning a deep compact image representation for visual tracking," in *Advances in neural information processing systems*, 2013, pp. 809–817.

[74] M. Sultana, A. Mahmood, S. Javed, and S. K. Jung, "Unsupervised deep context prediction for background estimation and foreground segmentation," *Machine Vision and Applications*, vol. 30, no. 3, pp. 375–395, 2019.

[75] S. Brutzer, B. Höferlin, and G. Heidemann, "Evaluation of background subtraction techniques for video surveillance," in *CVPR 2011.* IEEE, 2011, pp. 1937–1944.

[76] K. Toyama, J. Krumm, B. Brumitt, and B. Meyers, "Wallflower: Principles and practice of background maintenance," in *Proceedings of the Seventh IEEE International Conference on Computer Vision*, vol. 1. IEEE, 1999, pp. 255–261.

[77] T. Bouwmans, F. El Baf, and B. Vachon, "Background modeling using mixture of gaussians for foreground detection-a survey," *Recent Patents on Computer Science*, vol. 1, no. 3, pp. 219–237, 2008.

[78] N. Friedman and S. Russell, "Image segmentation in video sequences: A probabilistic approach," in *Proceedings of the Thirteenth conference on Uncertainty in artificial intelligence.* Morgan Kaufmann Publishers Inc., 1997, pp. 175–181.

[79] T. Matsuyama, "Background subtraction for non-stationar scenes," in *Proc. ACCV*, 1999, pp. 662–667.

[80] Y. Sheikh and M. Shah, "Bayesian modeling of dynamic scenes for object detection," *IEEE transactions on pattern analysis and machine intelligence*, vol. 27, no. 11, pp. 1778–1792, 2005.

[81] W. Zhou, S. Kaneko, D. Liang, M. Hashimoto, and Y. Satoh, "Background subtraction based on co-occurrence pixel-block pairs for robust object detection in dynamic scenes," *IIEEJ transactions on image electronics and visual computing*, vol. 5, no. 2, pp. 146–159, 2017.

[82] W. Zhou, S. Kaneko, M. Hashimoto, Y. Satoh, and D. Liang, "Co-occurrence background model with hypothesis on degradation modification for robust object detection," SciTePress. International Conference on Computer Vision Theory and Applications (VISAPP), 2018, pp. 266–273.

[83] ——, "A co-occurrence background model with hypothesis on degradation modification for object detection in strong background changes," in *2018 24th International Conference on Pattern Recognition (ICPR)*. IEEE, 2018, pp. 1743–1748.

[84] ——, "Foreground detection based on co-occurrence background model with hypothesis on degradation modification in background changes," in *2018 12th France-Japan and 10th Europe-Asia Congress on Mechatronics*. IEEE, 2018, pp. 77–82.

[85] ——, "Foreground detection based on co-occurrence background model with hypothesis on degradation modification in dynamic scenes," *Signal Processing*, vol. 160, pp. 66–79, 2019.

[86] X. Li, K. Liu, and Y. Dong, "Superpixel-based foreground extraction with fast adaptive trimaps," *IEEE transactions on cybernetics*, vol. 48, no. 9, pp. 2609–2619, 2018.

[87] X. Li, K. Liu, Y. Dong, and D. Tao, "Patch alignment manifold matting," *IEEE transactions on neural networks and learning systems*, vol. 29, no. 7, pp. 3214–3226, 2018.

[88] D. Stutz, A. Hermans, and B. Leibe, "Superpixels: an evaluation of the state-of-the-art," *Computer Vision and Image Understanding*, vol. 166, pp. 1–27, 2018.

[89] S. Javed, A. Mahmood, T. Bouwmans, and S. K. Jung, "Background–foreground modeling based on spatiotemporal sparse subspace clustering," *IEEE Transactions on Image Processing*, vol. 26, no. 12, pp. 5840–5854, 2017.

[90] M. Ben-Ezra, A. Zomet, and S. K. Nayar, "Jitter camera: High resolution video from a low resolution detector," in *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.*, vol. 2. IEEE, 2004, pp. II–II.

[91] P.-M. Jodoin, J. Konrad, V. Saligrama, and V. Veilleux-Gaboury, "Motion detection with an unstable camera," in *2008 15th IEEE International Conference on Image Processing*. IEEE, 2008, pp. 229–232.

[92] H. Sajid and S.-C. S. Cheung, "Background subtraction for static & moving camera," in *2015 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2015, pp. 4530–4534.

[93] M. Wu and X. Peng, "Spatio-temporal context for codebook-based dynamic background subtraction," *AEU-International Journal of Electronics and Communications*, vol. 64, no. 8, pp. 739–747, 2010.

[94] N. Goyette, P.-M. Jodoin, F. Porikli, J. Konrad, and P. Ishwar, "Changedetection. net: A new change detection benchmark dataset," IEEE. Computer Vision and Pattern Recognition Workshops (CVPRW), 2012 IEEE Computer Society Conference on, 2012, pp. 1–8.

[95] H. Wang and D. Suter, "A consensus-based method for tracking: Modelling background scenario and foreground appearance," *Pattern Recognition*, vol. 40, no. 3, pp. 1091–1105, 2007.

[96] I. J. Goodfellow, J. Shlens, and C. Szegedy, "Explaining and harnessing adversarial examples." International Conference on Learning Representations(ICLR), 2015.

[97] T. Matsuyama, T. Ohya, and H. Habe, "Background subtraction for non-stationary scenes," in *2000 Asian Conf. on Computer Vision.*, 2000, pp. 662–667.

[98] T. Fawcett, "An introduction to roc analysis," *Pattern recognition letters*, vol. 27, no. 8, pp. 861–874, 2006.

[99] A. Sobral and A. Vacavant, "A comprehensive review of background subtraction algorithms evaluated with synthetic and real videos," *Computer Vision and Image Understanding*, vol. 122, pp. 4–21, 2014.

[100] N. Lazarevic-McManus, J. Renno, and G. Jones, "Performance evaluation in visual surveillance using the f-measure," in *Proceedings of the 4th ACM international workshop on Video surveillance and sensor networks*. ACM, 2006, pp. 45–52.

[101] S. Brutzer, B. Höferlin, and G. Heidemann, "Evaluation of background subtraction techniques for video surveillance," IEEE. Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on, 2011, pp. 1937–1944.

[102] Q. Huynh-Thu and M. Ghanbari, "Scope of validity of psnr in image/video quality assessment," *Electronics letters*, vol. 44, no. 13, pp. 800–801, 2008.

[103] S. Winkler and P. Mohandas, "The evolution of video quality measurement: from psnr to hybrid metrics," *IEEE Transactions on Broadcasting*, vol. 54, no. 3, pp. 660–668, 2008.

[104] T. Huynh-The, O. Banos, S. Lee, B. H. Kang, E.-S. Kim, and T. Le-Tien, "Nic: a robust background extraction algorithm for foreground detection in dynamic scenes," *IEEE transactions on circuits and systems for video technology*, vol. 27, no. 7, pp. 1478–1490, 2017.

[105] "Performance evaluation of tracking and surveillance dataset 2001," *http://ftp.pets.rdg.ac.uk/pub/PETS2001*.

[106] L. Maddalena and A. Petrosino, "A self-organizing approach to background subtraction for visual surveillance applications," *IEEE Transactions on Image Processing*, vol. 17, no. 7, pp. 1168–1177, 2008.

[107] M. Hofmann, P. Tiefenbacher, and G. Rigoll, "Background segmentation with feedback: The pixel-based adaptive segmenter," in *2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*. IEEE, 2012, pp. 38–43.

[108] T. Elguebaly and N. Bouguila, "Background subtraction using finite mixtures of asymmetric gaussian distributions and shadow detection," *Machine vision and applications*, vol. 25, no. 5, pp. 1145–1162, 2014.

[109] A. Sobral, "Bgslibrary: An opencv c++ background subtraction library," vol. 7. IX Workshop de Visao Computacional (WVC ' 2013), 2013.

[110] D. D. Bloisi, A. Pennisi, and L. Iocchi, "Parallel multi-modal background modeling," *Pattern Recognition Letters*, vol. 96, pp. 45–54, 2017.

[111] G. Chiron, P. Gomez-Krämer, and M. Ménard, "Detecting and tracking honeybees in 3d at the beehive entrance using stereo vision," *EURASIP Journal on Image and Video Processing*, vol. 2013, no. 1, p. 59, 2013.

Appendix A Publications lists

A.1 Journal Papers

1. W. Zhou, S. Kaneko, Y. Satoh, M. Hashimoto, and D. Liang, "Foreground Detection based on Co-occurrence Background Model with Hypothesis on Degradation Modification in Dynamic Scenes," Signal Processing, 160, pp. 66-79, 2019.

2. W. Zhou, S. Kaneko, D. Liang, M. Hashimoto, and Y. Satoh, "Background subtraction based on co-occurrence pixel-block pairs for robust object detection in dynamic scenes," IIEEJ transactions on image electronics and visual computing, vol. 5, no. 2, pp. 146-159, 2017.

A.2 International Conferences

1. W. Zhou, S. Kaneko, M. Hashimoto, Y. Satoh, and D. Liang, "Foreground Detection based on Co-occurrence Background Model with Hypothesis on Degradation Modification in Background Changes." 2018 12th France-Japan and 10th Europe-Asia Congress on Mechatronics. IEEE, 2018, pp. 77-82.

2. W. Zhou, S. Kaneko, M. Hashimoto, Y. Satoh, and D. Liang, "A Co-occurrence Background Model with Hypothesis on Degradation Modification for Object Detection in Strong Background Changes," in Proceedings of the 24th International Conference on Pattern Recognition (ICPR) on 2018, pp. 1743-1748.

3. W. Zhou, S. Kaneko, M. Hashimoto, Y. Satoh, and D. Liang, "Co-occurrence background model with hypothesis on degradation modification for robust object detection," in Proceedings of the 13th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications - Volume 5: VISAPP, 2018, pp. 266-273.

4. W. Zhou, S. Kaneko, D. Liang, M. Hashimoto, and Y. Satoh, "Co-occurrence Pixel Block Pairs Background Subtraction for Object Detection in Dynamic Scene," in Proceeding of the International Symposium on Optomechatronic Technology 2016, November 7-9, 2016, Tokyo, Japan.

A.3 Domestic Conferences

1. W. Zhou, S. Kaneko, M. Hashimoto, Y. Satoh, and D. Liang, "Co-occurrence based Foreground Detection with Hypothesis on Degradation Modification in Severe

Imaging Conditions," Precision Engineering Society Fall Conference 2018, September 5-7, 2018, Hakodate, Japan.

2. W. Zhou, S. Kaneko, M. Hashimoto, Y. Satoh, and D. Liang, "Co-occurrence Background Model with Model Modification based on Hypothesis on Degradation for Robust Object Detection," in Proceeding of Vision Engineering Workshop 2017, December 7-8, 2017, Yokohama, Japan.

3. W. Zhou, S. Kaneko, D. Liang, M. Hashimoto, and Y. Satoh, "Co-occurrence Pixel Block Pairs Background Subtraction for Object Detection by Introduction of Hypothesis-based Modification," in Proceeding of Vision Engineering Workshop 2016, December 8-9, 2016, Yokohama, Japan.