國立臺灣大學電機資訊學院資訊工程研究所

碩士論文

Department of Computer Science and Information Engineering

College of Electrical Engineering and Computer Science

National Taiwan University

Master Thesis

# 結合時空背景模型與隨機森林分類器之前景分割及影子去除

# Combining Spatiotemporal Background Modeling and Random Forest Classifier for Foreground Segmentation and Shadow Removal

廖偉傑

Wei-Jie Liao

指導教授：洪一平 博士

陳祝嵩 博士

Advisor: Yi-Ping Hung, Ph.D.

Chu-Song Chen, Ph.D.

中華民國 103 年 7 月

July, 2014

# 國立臺灣大學碩士學位論文
# 口試委員會審定書

結合時空背景模型與隨機森林分類器之前景分割及影子去除

Combining Spatiotemporal Background Modeling and Random Forest Classifier for Foregroung Segmentation and Shadow Removal

本論文係廖偉傑君（學號 R01922106）在國立臺灣大學資訊工程學系完成之碩士學位論文，於民國 103 年 7 月 31 日承下列考試委員審查通過及口試及格，特此證明

口試委員： 洪一平 陳祝嵩

（指導教授）

黃彥男 陳祝嵩

簡韶逸 徐繼聖

系主任 許永真

# 誌謝

能完成這篇論文，我要特別感謝我的指導教授洪一平老師和陳祝嵩老師，在這兩年中不斷在學業和研究上指引我往正確的方向前進，並且不斷的提供我更多的知識和訊息，讓本身在影像處理這個領域沒有太多了解的我可以更順利的完成我的研究。也要特別感謝兩年來一直從旁指導我的林志瑋學長，每當在研究遇到瓶頸或是有不確定的想法時，學長總是非常熱心且不厭其煩的和我討論，讓我的研究可以更加完整且有系統。另外也要感謝陳宣輯學長和林可昀同學在學業和實驗室的計畫總是協助我和我一起完成不同階段的考驗，讓我可以有更多的時間專心在自己的研究上。還有臺大資工所的同學和學弟們在兩年中互相扶持，除了課業上互相討論之外,總是在我挫折失意的時候鼓勵我讓我重新拾起信心努力下去，和我一起打球一起玩樂。當然最重要的感謝一路上家人總是扮演我身後最偉大的後盾，讓我可以專心的在我的學業並且有個溫暖的家庭讓我在結束辛苦的一天後能擁有最好的環境放鬆自己。

因為自己本身不是資訊工程背景出身，在學習和研究的過程中常常感到力不從心，真的很感謝一路上大家的幫忙和支持，每當我覺得疲累的時候讓我能夠重新打起精神繼續努力下去，由衷的感謝大家一路上的幫助，在這邊致上最深最深最誠懇的感謝。

# 中文摘要

在眾多視訊監控的應用中，前景偵測和影子的消除一直是一個很重要的議題，我們對於前景偵測和影子消除提出了一個新的架構，我們的架構包含兩個五要的部份分別是前景偵測和影子偵測。我們提出時空背景截取器來進行前景偵測，時空背景截取器主要包含背景截取和背景邊緣截取兩部份，時空背景截取器在動態背景和瞬間光線變化有不錯的表現。在影子消除的部份我們選擇色度、物理性質和紋理三種特徵來當作判斷的資訊，我們利用隨機森林分類器和所選擇的特徵學習出適合各個場景的影子偵測模型。我們使用建出來的影子偵測模型針對時空背景截取器的前景結果進行影子消除。除此之外我們和目前較常使用的前景偵測和影子消除的方法進行比較。

關鍵字：前景物偵測、影子消除、支持向量機、隨機森林

# ABSTRACT

Cast shadows detection and removal is indispensable in the object detection to many surveillance applications. In this paper, we present a novel framework for removing cast shadow of moving objects. Two main components, moving objects detector and redundant shadow remover, are integrated. For moving objects, we adopt the spatiotemporal background extractor (SBE) to detect the moving objects which is comprised of the background extractor (BE) and the background gradient extractor (BGE). SBE features the object detection in the dynamic background and the sudden lighting changes environment. For shadow removal, we use the classifier, Random Forest, to learn the shadow features, which are chromaticity, physical properties, and texture. Then, we remove the shadow from the result of SBE with the shadow classifier. The proposed method can effectively detect the moving objects and remove the shadow effect. Furthermore, we demonstrate the performance of our method compared with some state-of-the-art techniques of object detection and shadow removal.

Key word：Object detection, Shadow removal, Support vector machine, Random forest.

# CONTENTS

# LIST OF FIGURES

# LIST OF TABLES

# Chapter 1 Introduction

## 1.1 Motivation

In recent years, security issues have been noticed because of terrorist attacks occur, such as 911 in 2001, London bombing in 2005 and local violent events. In the past we consume lots of people and money to solve these problems. Due to the advancement of technology, we can reduce the resource using visual surveillance. There are many application s in visual surveillance, such as human detection, people counting, object left, and object stolen, in which object detection is the most important issue. The moving object can be extracted by using object detection for various scenes. After doing objects detection, some useful information can be obtained, such as the location, shape, size and other details of the objects, and that can be applied to different applications.

## 1.2 Background Knowledge

In the method of the present, object detect can rough class to two category, temporal differencing and spatial differencing. The famous method of temporal method is Gaussian Mixture Model (GMM) [1], using the Gaussian model to learn the background model and to compute the probability of each pixel which is used to identify that it belongs to background of foreground. Another method of temporal is Codebook, proposed by Kim *et al*. [2, 3]. Kim *et al*. [2, 3] train the background model by using the occur frequency of each pixel data, the information of frequency is used to decide the data should be added to the background model or be eliminated. In the spatial differencing, learning the background model refers to not only time information but also the spatial information. Recently, ViBe proposed by O. Barnich and M. Van

Droogenbroeck [4] is using the spatial information. ViBe learne the background model by referring the information of each pixel and it's neighbors. Because of the performance of visual surveillance in related to the accuracy of object detection. There have some challenges of object detection, such as the dynamic background, the ghost effect and shadow. Shadow detection is the most challenge of visual surveillance. The shadow of moving object should not be classified into the part of foreground. If the shadow is classified into the part of moving object, the information of object will become inaccurate. The shadow may cause the wrong information of object. For example, distorted shape, larger size and error location are the wrong information of object that will reduce the performance of visual surveillance application. How to remove the shadow is the importance issue of visual surveillance.

There are lots of works to solve the problem of shadow removal. Wu *et al*. [5] classified the architecture of shadow removal into two categories, feature-based and model-based. The method of feature-based, use the features of shadow to classify the foreground and shadow. In this method, the performance is sensitive to parameter and threshold of each feature. Different feature is suitable for different scene. So, how to choose the threshold is the important issue of feature-based method. In the method of model-based, use the value of feature to train the model of classifier to divide the foreground and shadow. Because model-based can adjust the classifier according to the condition of each scene, we can get the better performance for major scene instead of specific one. The difficult of model-based is that it needs the prior information, ground truth or object's class. In order to get the better performance of major scene, we chose architecture of model-based in our work.

# Chapter 2 Related Work

In this chapter, we introduce the method about background modeling and shadow removal that are associated with our method. In Section 2.1, we explain the method of background modeling using the temporal and spatial differencing. In Section 2.2, we introduce two major categories of shadow removal, feature-based and model-based. In Section 2.3, describe four widely used features of shadow removal and the advantage and disadvantage of each feature. In Section 2.4, we introduce the widely used classifier of shadow removal and the reason of why we chose the Random Forest to be our method's classifier.

## 2.1 Background Modeling

In previous chapter, the method of background modeling can be classified into two categories, temporal differencing and spatial differencing. Codebook is the one of the famous methods in the temporal differencing. The architecture of Codebook can be divided into two steps: training and detecting steps. Because Codebook combines all features of each pixel in the training step, each pixel in background model has lots of information. After training, Codebook compares the pixel value of current frame with background model. If it can find the similar enough value in the background model, this pixel is classified into the background. Conversely, the pixel is classified into foreground. Codebook records the data of background includes color, intensity, occur frequency, the time of the background model begin and last when it occurs. Because Codebook separates the color and intensity, the capable of overcoming light changes is well. According to the time information and the frequency, Codebook adds a new background model if it is often in the scene and eliminates the model when it is seldom appear in the scene. This mechanism can handles the problem of dynamic background. In spatial differencing, ViBe is a current method of background modeling using spatial information. There are two phase in ViBe, training and detecting phases. Because ViBe considers the background is associate with not only the value of each pixel but also the neighbor of each pixel. However, this work considers that a consistent threshold of time used to perfectly classify the background and foreground is not exist. Therefore, ViBe proposes a novel method to eliminate and add the background model using the random probability. When the object often appears in the scene will have higher probability add to the background model. Instead of directly use the time to choose the background model, ViBe uses the properties of random distribution to present the time information. ViBe considers that the spatial information of neighbors affects the background model of the pixel, ViBe randomly chooses a neighbor to update the background information from the pixel and update the information from the neighbor to each pixel. According to this architecture, ViBe learns the background from the refer pixel and it’s neighbor. In addition, ViBe only uses one frame to build the background model in the training phase. In the training step, ViBe combine all neighbor data and pixel’s information to initiate background model. Because of no matter how long the object stay in the scene, this object has same probability leave to the scene as any object stays in scene or not. ViBe considers that the update policy of the background model is more important than training step. Thus, ViBe uses one frame to build the background model and uses the update mechanism to adjust the model to suit for the background’s change. Base on ViBe, M. Van Droogenbroeck propose the improve work of ViBe named ViBe plus [6]. ViBe Plus add the information of intensity to be a condition of background and adjust the probability of random chose. In addition to adjust the parameter of ViBe, ViBe plus

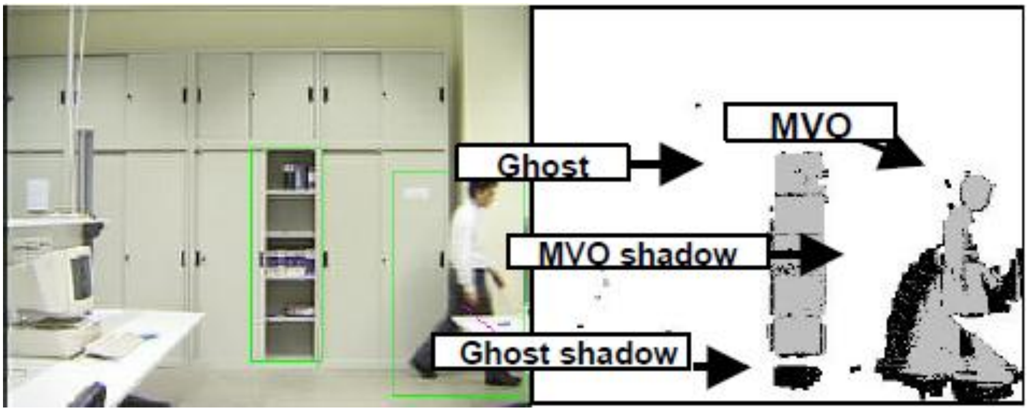



Fig. 1 The object classification result of [6]

add the post process to the architecture to improve the performance. We propose a novel algorithm combine these two properties named Spatiotemporal Background Extractor (SBE).

## 2.2 Shadow Removal

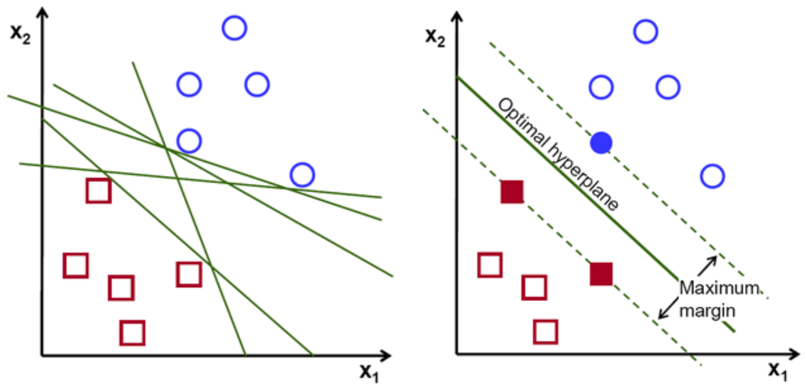
According to the algorithm of shadow removal, the paper [5] divides the method to two major architectures. First architecture is based on the result of background modeling and the threshold of parameter. In the method proposed by Cucchiara *et al*. [6], it divide the result of object detection to fore components, moving object, moving object shadow, ghost and ghost shadow in Fig. 1. Using the optical flow to eliminate the ghost and ghost shadow. This method considers the region of moving object shadow has the darker brightness and similar chromaticity. Using this property, the region with lower brightness and similar chromaticity in the foreground will define to the moving object shadow. In order to get more accurate moving object, this method has to remove the moving object shadow. The performance of remove the shadow of moving object is based on the threshold of the brightness. This method of shadow removal is refer the chromaticity and brightness. The other method use the property of environment proposed by Huang and Chen [7]. The light source can divide to two component, direct light and ambient light. Direct light is white like sun light and ambient light is more bluer than direct light. The cause of shadow is the direct light was blocked but still has ambient light. If the direct light was totally blocked, the region of shadow only has ambient light. So this method considers the change of color in the shadow region will between the full direct light and without direct light. In the training step, this method uses Gaussian model to build the shadow model base on the ground truth that label the background, foreground and shadow in whole scene. In addition to build the model with using color information, this method also build the model with the gradient in each pixel. In the detect step. This method uses the weak detector to detect the foreground without the impossible shadow sample (e.g. the region with brighter color than background). Using these Gaussian models to compute the probability of these pixel belonging to background, foreground or shadow. If the probability of these pixel belonging to foreground is larger than background and shadow, these pixel will be classify to foreground. This method uses the property of light source to detect the shadow by build the model to learn the change of light. Some method using this property will compute the relation between blue and other color information. The other method uses the information of texture was proposed by Sanin *et al*. [8]. This method considers the region of shadow has similar texture with background. Base on the result of object detection, this method use the chromaticity to choice the region is likely to shadow. After select the candidate region of shadow, this method computes the gradient direction distance of the region pixel which has significant gradient magnitude. If the average gradient direction distance of region is lower than the threshold, the region will classified to the shadow. The performance of this method is based on the object detection and threshold of gradient direction distance. The method uses texture is consider that has better performance than the method use color information. This architecture always attempt to adjust the parameter and threshold in the feature or find the different combined ratio of each feature to get the best performance in general scene. There have a problem occur when the scene is very extreme. The performance will become poor than our expectation. In the previous paragraph, even the general scene may have great difference between each other. For example when the combination is suitable for the simple foreground scene, it may not suitable for the case in outdoor. It is hard to find the combination has best performance in every scene.

In order to find the good combination of each scene, second architecture proposed the method by using the classifier to build the model for each scene. The method proposed by Wang *et al*. [9] uses this architecture. This method uses LPGMM to detect the moving object first. LPGMM is the method of background modeling different from GMM with refer the information not only current pixel but also the neighbor pixel. This foreground includes some region of shadow was misclassified to foreground. So this method uses the SVM to learn the feature of shadow that can classify the foreground and shadow. Combining the object detection and classifier can get the more accurate foreground without the shadow. In this architecture, we need the ground truth data of the scene in order to label the feature of each pixel with it's category (foreground, shadow or background) for the training stage. Because of these feature is label by the ground truth of each scene, we can build the model was suitable for each scene to classifier foreground and shadow.

In the propose method, we choose the feature those are suitable for different scene and use classifier to build the model in order to find the best performance of each scene. Because we chose the feature are suitable for different scene, there has some feature has good performance and other has poor result in the same time. Base on the architecture which using the classifier, the feature which has greater performance will get higher weight in the model by training use the classifier. So we can build the great shadow removal model for each scene.

## 2.3 Feature of Shadow

In the first architecture, how to choose the feature with the property of feature is very important. According to [10, 11], there have a lot of features can rough classify to four category such as chromaticity, physical properties, geometric and texture. In the chromaticity [6], we convert color space from RGB to HSV. HSV include three component by hue, saturation and value. Hue is mean the color like red, green, blue…etc. In RGB color space use combination of red, green and blue to represent the color. In HSV just use value of hue to represent it. Saturation mean the purity of color. Color have higher value of saturation will become brighter and lower value of saturation will become like gray. Value is the light value of color. For example, light red and dark red may have the same value of hue but different value of light. This classification propose that the region of shadow have the similar hue and saturation. But due to the light was blocked so have darker lighter value than background. When the region have similar color and darker light value will be seen to the shadow region. Moreover, the region with different color or extreme light change will seem to the foreground. The advantage of this feature is simple to implement and computationally. However, this method use the color space. So when the foreground has similar color with background. This feature will consider the foreground is part of background. Physical properties [7, 12] consider there have two major illumination sources in outdoor are ambient light (blue light) and the sun (white light). The cause of shadow is the sun blocked by moving object. Though the object block the light source of sun, ambient light is always irradiate on the region. According to this situation, the color of shadow region should trend to blue color because the sun has been blocked so the ratio of ambient light has increase. Using the ratio of blue and light value can calculate the difference of physical properties. The advantage of physical properties is more accuracy in outdoor scene than other color feature. However, this method will misclassified the foreground with similar color because it use the color information too. It has common performance in indoor scene. Geometry against the specific foreground by use it's orientation, size and shape. In the method proposed by Hsieh *et al*. [13], we use the foreground's gravity to calculate the angle between foreground and shadow. Then use the gradient of object's contour to find out the cut point in order to divide foreground and shadow region. This method do not rely on an accurate detect of foreground so can detect the foreground with similar color with background. But they have lots of limitation of scene. Because different object have different size and shape, we only can detect specific object type (ie. vehicles or standing people). In order to divide the foreground and shadow by the angle between them. We requiring object and shadow to have different orientation and the scene have only one light source. These limitations make geometry's feature become hard to used. The method base on texture which proposed by Sanin *et al*. [14], consider the region of shadow will become darker than original but retain the texture of this region. This method always follow two steps: (1) rough foreground detection select the candidate region of shadow include foreground and shadow and (2) classify the candidate region to foreground or shadow by using texture feature. Texture method calculate the difference of gradient and angle between pixel and it's neighborhood. Because the region of shadow will retain the texture, the difference value will small. If the difference is large mean there have some object across the region. Because texture feature do not




(a) (b)

Fig. 2 SVM uses the hyper-plane to divide training data. (a) SVM may have multiple hyper-planes that all can divide the training data. (b) only one hyper-plane can divide the data with largest maximum margin.

use color space and few limitation. It has great performance in majority case. But if the object and background have similar texture will cause the misclassified (ie. Smooth object on the floor). Each feature is suitable for different scene. So, we always combine more than one feature to raise the removal performance.

When the scene change, the combination between every feature have to change too. For example, in the outdoor, physical properties may has the best performance make it has more ratio than other feature. When the scene has simple foreground (ie. only working people in the hall), geometry may substitute physical properties become the most suitable feature in this scene. There have not any new feature or perfect combination of different feature can suitable for every scene. We should learn the property of every scene then train the exclusive classifier for each scene.。

## 2.4 Classifier

In the second architecture by using classifier to build the model, in large part method chose the classifier Support Vector Machine (SVM) to train the model. SVM is a supervised learning method can widely used in many area likes statistical analysis, regression analysis..etc. Using the SVM, we label the training data according to the category of each pixel before training stage. Then SVM will map these data to the space has greater than or equal to the dimensions of training data. In the space, SVM try to find a hyperplane to divide training data with it's category. In addition to divided the data as possible as clear. Hyperplane has to find the largest minimum distance between training data. We use the figure reference to OpenCV to explain the policy of SVM. See the Fig. 2 (a) We have two class data, blue and red then SVM map these training data to 2-dimensions space. We can see the SVM find multiple line divide each class data clear. The Fig. 2 (b) indicate the optimal hyperplane with largest minimum distance (maximum margin). This optimal hyperplane can divide these data most clear than other hyperplane. After SVM get the hyperplane by training data, we can use this model to class other data to the each category. We get the non-classified data than mapping this data into the training space. Substituting non-classified data into the hyperplane function can get a value. According to this value is belonging to which side of hyperplane can decide the classify result. Besides using the standard method of SVM, we try to combine the weight vector which was represent the importance of each feature of training data.

The other newer method of training model using classifier is named Random Forest. Random Forest includes multiple decision tree to classify the data. Just the same as SVM, Random Forest has to label the training data according to the category of each pixel. In the training stage, Random Forest random chose part of training data several time. Because the data is random chose, each part of data will have some is same as other part of data and some is different. Random Forest will use every part of data to build a decision tree. In the decision tree, each node is a minor classifier to divide the

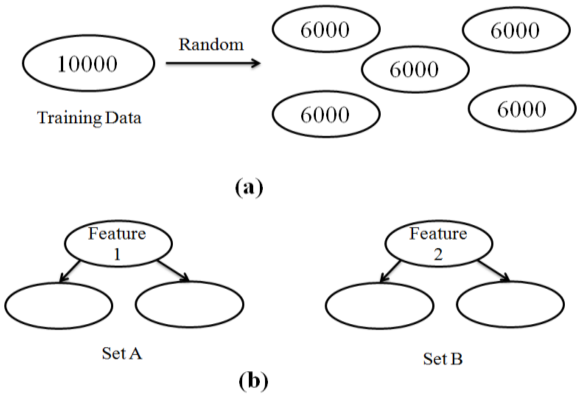



Fig. 3 Random forest use multiple division tree to classify the data. (a) Random forest randomly chose 6000 data from training data five time. (b) set A of data use the feature 1 be the rot of decision tree and set B chose the feature 2 be the root.

input data by data's label and distribution to left subtree and right subtree. When the number of data has been divided to less than the threshold, this node is become to the leaf represent the result of classifier. See the Fig. 3 (a). Initially, we have 10000 training data. We random chose 6000 data to five sets. Assume each data has three feature. Each set build the decision tree according to feature's distribution. In the Fig. 3 (b), setA (decision tree A), it chose feature 1 be a first node (root) of this decision tree. But in the setB, it chose the feature 2 be a first node. It is because the data is chosen by random, every set's data distribution will different from other set. After build several decision, we use these decision tree to classify testing data. We put testing data into every decision tree. After classify the testing data by each decision tree, we get a several result of classification and calculate the plural of these result. The plural is the final result of Random Forest classifier. Because the Random Forest will reference the importance of each feature of input data, it like the concept that each feature of shadow removal has different weight of each scene. So, we think Random Forest is more suitable than SVM in the shadow removal by using the training model with classifier.

# Chapter 3 Methodology

We propose a method combine object detection and shadow removal by using the spatiotemporal background extractor to detect the moving object, extract the feature of shadow and the training the classifier. In Section 3.1, we introduce the architecture of our method. In Section 3.2, we propose a method combine the temporal information and spatial information to detect the moving object. In Section 3.3, we explain property of each feature and why we chose these features or not. After extract the feature, we use the SVM and Random Forest to train the classifier of shadow removal.

## 3.1 System Architecture

Our architecture can rough divide into two parts, object detect using SBE and shadow removal in Fig. 4. Initially, we use the classifier to training the model can remove the shadow for each scene. Then we use the SBE to detect the moving object. Because the result of SBE include some region of shadow is classified to the moving object, we use the classifier model training in former step to remove the shadow in the result of SBE. In the architecture of shadow removal we have two steps, feature extraction and classifier training in Fig. 4. In the first step, we extract the feature in each frame of ground truth. In the previous chapter we introduce four kinds of feature, chromaticity, physical properties, geometry and texture. According to the category of ground truth, we extract three features by every pixel of foreground and shadow then label the data of it's category for latter step. We use the information of previous step to training the classifier in final step. In addition to the basic method of SVM, we try to combine the weight vector which represent the importance of each feature to the data.

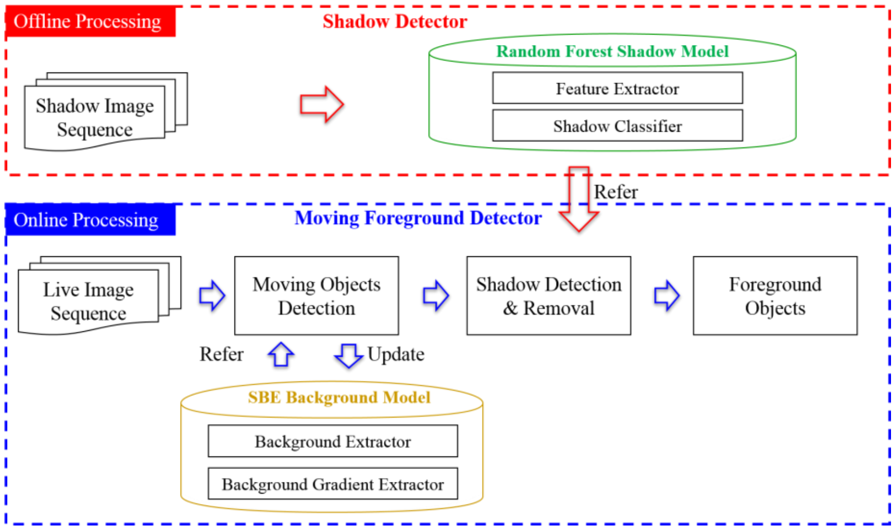



Fig. 4 The work flow of the proposed method

According to the properties of feature that have different performance in each scene is similar to the concept of Random Forest, we decide to use the Random Forest as the classifier of our architecture. In the architecture of SBE have two components, Background Extractor (BE) and Background Gradient Exactor (BGE) in Fig. 4. The component of BG is used to train the background model and detect the moving object. In order to get the more accurate performance of object's edge, we propose the component of BGE that can train the gradient model of background. We get the result of SBE by execute these two components simultaneously.

## 3.2 Spatiotemporal Background Extraction

In this section we will explain the algorithm of SBE include the two components BE and BGE. The architecture of SBE that has two major components is inspired by [2, 3]. BE combines temporal and spatial information to train the background model and eliminate the noise from background. BGE can ensure the shape completeness of the moving object. The temporal information is contributed from the video stream can train the information of static background. The spatial information is gotten from the

**Algorithm 1.** The contraction of BE

Input: The first frame of image sequence
Output: The background extractor
$\mathcal{C} \leftarrow \emptyset$ (empty set)
for each pixel do
    To construct the codebook $\mathcal{C} = \{c_i | 1 \le \mathrm{i} \le \mathrm{n}\}$
    according to the adjacent neighborhoods and each
    codeword $c_{\mathrm{i}}$ is composed of $v_i$ and $aux_i$
    for neighbor $j$ = 1 to n do
        $I_i \leftarrow \sqrt{{R_j}^2 + {G_j}^2 + {B_j}^2}$
        $v_i \leftarrow (R_j, G_j, B_j)$
        $aux_i \leftarrow (\check{I}_i = I_j, \hat{I}_i = I_j, f_i = 1, \lambda_i = 1,\ p_i =$
        $1, q_i = 1)$
    end
end

**Algorithm 2.** The update policies of BE

**Input**: The incoming pixel $x_t$, the position of current
    pixel $X$, and the positions of randomly chosen
    neighbors $C$ and $H$.
**Output**: The updated background extractor
**Regular update policy:**
    **if** $x_t$ is classified into background **then**
        The matched codeword $c_m$ is updated.
    **end if**
**Two-way propagation policy:**
    First-propagation direction:
        To propagate the color information from the
        current pixel $X$ to a randomly chosen neighbor $C$.
    Second-propagation direction:
        To propagate the color information from a
        Randomly chosen neighbor $H$ to current pixel $X$.
end

neighbors of each pixel includes the color information and intensity that can handle with the dynamic background.

### 3.2.1 Background Extractor (BE)

BE is constructed as a single-layer Codebook model in [2, 3]. But, different form [2, 3] that need a long period to construct the background, we only use one frame to construct the background of BE. The algorithm of BE is designed for color imagery and each pixel of BE is modeled as a single-layer Codebook $\mathcal{C} = \{c_i | 1 \le i \le n\}$ consist of n codewords. Each codeword $c_{\mathrm{i}}$ is includes two elements, a RGB vector $v_i = (\bar{R}_i, \bar{G}_i, \bar{B}_i)$ and a six-tuple $aux_i = \langle \check{I}_i, \hat{I}_i, f_i, \lambda_i, p_i, q_i \rangle$. Where $\check{I}_i$ and $\hat{I}_i$ are the minimum and maximum brightness of each codeword $c_i$, respectively, $f_i$ is the occurred frequency, $\lambda_i$ is the longest time interval that the codeword $c_i$ is **NOT** recurred, and $p_i$ and $q_i$ are the first and last access times, that the codeword has occurred, respectively. The algorithm of BE construction is shown in Algorithm 1. We use the first frame of video sequence to construct the BE. In the construction procedure, each pixel is constructed as a Codebook based on the region that include a set of pixels in a neighborhood around the current pixel, as shown in Fig. 5, where the current pixel is referred to as ***X***. The length of Codebook would be dependent on the pixel location. In

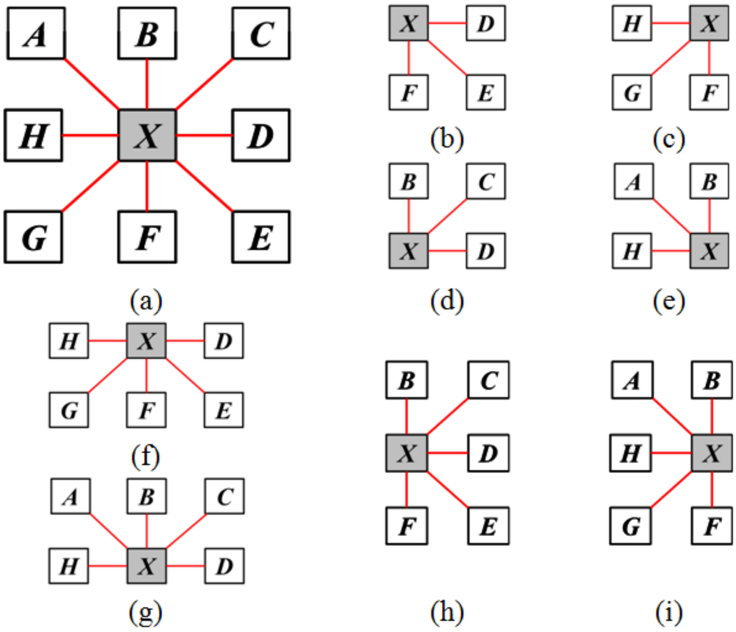



Fig. 5 The chart of connected neighborhoods around the current pixel ***X*****.** (a) the current pixel is inside the image. (b)-(e) the current pixel is at the corner of the image.(f)-(i) the current pixel is at the border of the image.

Fig. 5 (a), as the current pixel ***X*** is inside the image and the neighborhoods ***A-H*** are adopted. Fig. 5 (b)-(e) and (f)-(i) represent the other cases of current pixel ***X's*** position when it is located at the corners or the border of an image. Because we use single frame to construct the BE, the minimum and maximum brightness of each codeword $c_i$, $\check{I}_i$ and $\hat{I}_i$, are set to be the same. Because BE only use a single frame to construct model, BE is faster than [2, 3]. The spatial information contributed by neighborhoods of current pixel, which is reflected in the simplified single-layer Codebook model and useful to overcome the problem of dynamic background.

After the construction of BE, we classify the incoming pixel to either background or foreground in the pixel classification stage. The incoming pixel is classified into background if it satisfies two conditions, the color distortion and the range of brightness. In the condition of color distortion, the matched codeword $c_m$ with RGB vector $v_m$ is found from its Codebook $\mathcal{C}$ base on the following color distortion measure:

$$p^2 = \|x_t\|^2 \cos\theta = \frac{(x_t, v_t)^2}{\|v_i\|} = \frac{\left(\overline{R}_i R + \overline{G}_i G + \overline{B}_i B\right)^2}{\left\|\overline{R}_i + \overline{G}_i + \overline{B}_i\right\|}, \text{ and}$$

$$colordis(x_t, v_i) = \delta = \sqrt{\|x_t\|^2 - p^2} \leq \varepsilon, \tag{1}$$

where $x_t = (R, G, B)$ is the current pixel at time t with a RGB vector. In order to solve the problem with changes of illumination, the brightness of the current pixel is considered:

$$brightness\left(I_{x_t}, \langle \check{I}, \hat{I} \rangle\right) = \begin{matrix} true & if\ I_{low} \leq I_{x_t} \leq I_{hi} \\ false & otherwise \end{matrix} \tag{2}$$

where $I_{low} = \alpha\hat{I}$ and $I_{hi} = min\left\{\beta\hat{I}, \frac{\check{I}}{\alpha}\right\}$ are the lower and upper bounds of illumination, and $I_{x_t}$ is the illumination of current pixel $x_t$. The parameters, $\alpha$ and $\beta$, are set as that in [2, 3] which are used to allow large brightness bound and limiting $I_{hi}$, respectively. If the current pixel simultaneously satisfies the color distortion and brightness, it is classified as background. The current pixel will be updated into BE if it is classified as background in the pixel classification stage. We propose two policy of updating: regular update and two-way propagation, which in order to combine the temporal information and maintaining the spatial consistency for BE construction. The algorithm of the update policies of BE is shown in Algorithm 2.

Information of temporal is added into the regular update policy as that in [2, 3]. When the current pixel is classified as the background, the matched codeword $c_m$ is updated base on the following equations:

$$v_m = \left(\frac{f_m\bar{R}_m + R}{f_m + 1}, \frac{f_m\bar{G}_m + G}{f_m + 1}, \frac{f_m\bar{B}_m + B}{f_m + 1}\right), \quad \text{and}$$

$$aux_m = \begin{Bmatrix} min\{I_{x_t}, \check{I}_m\}, max\{I_{x_t}, \hat{I}_m\}, f_m = f_m + 1, \\ max\ \{\lambda_m = t - q_m\}, p_m, q_m = t \end{Bmatrix} \tag{3}$$

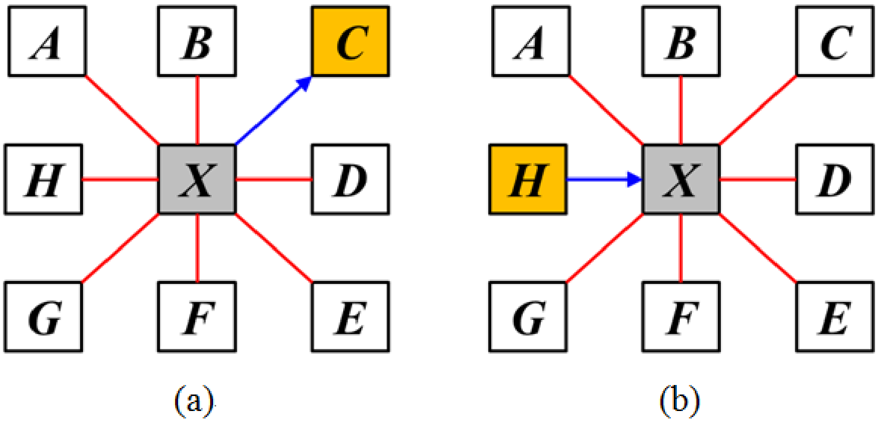



Fig. 6 The two-way propagation policy. (a) the first-propagation direction. (b) the second-propagation direction.

The RGB vector $v_m$ is computed by averaging the value of red, green, and blue colors with the current pixel $x_t$ and the parameters in $aux_m$ are all updated. In addition, the two-way propagation policy is used to combine the color information around the current pixel means that the current pixel propagates the color information to a neighbor and a neighbor propagates the color information to the current pixel mutually. In the first steps of the two-way propagation, we randomly choice a neighbor from 8 connected neighborhoods around the current pixel. In order to find the matched codeword from that neighbor, we base on two criteria: the color distortion and the range of brightness that described in previous section. If we found the matched codeword, we update the color information into the matched codeword according to the equation (3). Otherwise, we create a new codeword $c_L$ add to the chosen neighbor and then we propagate the color information of the current pixel $x_t$ which is classified as background to the codeword $c_L$ by the following setting:

$$v_L = (R, G, B),$$

$$aux_L = \{\check{I}_L = I_{x_t}, \hat{I}_L = I_{x_t}, f_L = 1, \lambda_L = 1, p_L = 1, q_L = 1\} \quad (4)$$

On the other step, we randomly choice a codeword from the 8 connected neighborhoods around the current pixel. If there is a matched codeword, we updated the matched codeword base on the equation (3). Otherwise, we create a new codeword $c_L$ for the current pixel and assign the value according to the equation (4). An example of two-way propagation is shown in Fig. 6. Fig. 6 (a) shows the first step which propagates the color information from the current pixel $\boldsymbol{X}$ to a randomly chosen neighbor pixel $\boldsymbol{C}$, as indicated by the blue line. Fig. 6 (b) shows the second step of the two-way propagation. The regular update policy is useful to maintain BE and capture the foreground accurately. Furthermore, the two-way propagation policy considers the space information which can efficiently solve the problems of dynamic background. Because endless increase the codewords, makes the size of Codebook became too large and that increased the time of matching. Therefore, we rejected the unsatisfied codewords which have the longest interval $\lambda$ according to the following equation:

$$\lambda > T_{\lambda}, \tag{5}$$

where $T_{\lambda}$ is the number of frames which is set to be 500 in our method. The filter policy would make the Codebook became empty when the threshold $T_{\lambda}$ is too small or the codeword is not recurred for a longest interval. Therefore, we proposed the reconstruction policy to solve the problem, if the Codebook became empty, the reconstruction policy is triggered according to the construction of BE that described in Section 3.2.2.

### 3.2.2 Background Gradient Extractor (BGE)

BE is useful to handle with the dynamic background and sudden changes of illumination. However, the intensities of foreground become blur because the propagation policy passes the background information to a neighbor. Hence, we propose the forbidden propagation policy that maintains the completeness of foreground. The forbidden propagation policy is based on the BGE introduced below.

BGE is used to construct the background gradient image $\nabla I_{BGE_i}$ for each frame. There are two phases of BGE. In phase 1, we constructed $\nabla I_{BGE_i}$ by using the first N frames according to the following equation:

$\nabla I_i = \frac{\partial I_i}{\partial x} + \frac{\partial I_i}{\partial y}$, and

$$\nabla I_{BGE_i} = \nabla I_{BGE_{i-1}}\ AND\ \nabla I_i, \quad (6)$$

where $\nabla I_i$ is the gradient results of the $i^{th}$ frame after binarization, $\frac{\partial I_i}{\partial x}$ and $\frac{\partial I_i}{\partial y}$ are the gradient in the x and y directions, respectively, and $\nabla I_{BGE_1} = \nabla I_1$. In order to aid the accuracy of $\nabla I_{BGE_i}$, we built the BGE cache model to be the phase 2. Each pixel in the BGE cache model is modeled as a vector $v$ which records the gradient of the captured image $\nabla I_i$ with the latest N frames as that in the phase 1. After the first N frames, $\nabla I_{BGE_i}$ is updated base on the following equation:

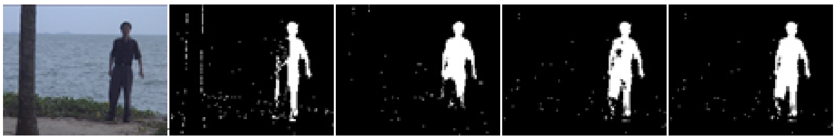
$$\nabla I_{BGE_i}(x,y) = \begin{cases} 255 & if\ \frac{\sum_{k=1}^{N} g_k/_{255}}{N} \geq T_{BGE}, \\ 0 & otherwise \end{cases} \quad (7)$$

where $g_k$ is the gradient value which is 0 or 255, and $T_{BGE}$ is the threshold. In our experiment, the value of $T_{BGE}$ is equal or larger than 0.6 which has sufficient amount of reliable samples and that shows the higher performance of forbidden propagation.

The forbidden propagation policy is used to ensure the shape completeness of the foreground that is based on BGE. There are two phases in the forbidden propagation policy. First phase, the background-gradient Image $\nabla I_{BGE_i}$ is used to find the foreground gradient of the current frame by using the following equation:

$$\nabla I_{FGG_i} = \nabla I_i - \nabla I_{BGE_i}, \quad (8)$$

where $\nabla I_{FGG_i}$ is the foreground-gradient image. Second phase, the color



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

Fig. 7 The comparison of forbidden propagation. (a) the original image of water surface; (b) the result of codebook; (c) the result of ViBe; (d) Proposed method without forbidden propagation; (f) Proposed method with forbidden propagation

information of the current pixel is forbidden to propagate when the foreground-gradient value of itself or a randomly choice neighbor is 255. The performance of the forbidden propagation is demonstrated by using the data of water surface on the Perception datase. Fig. 7 (a) is the original scenery of water surface. Fig. 7 (b) and (c) are the results of Codebook and ViBe that have a lot of false positive because of tree and wave of the sea and has a huge broken on the human body. Fig. 7 (d) is the result of our proposed method without forbidden propagation which has better performance than Fig. 7 (b) and (c). The result of forbidden propagation policy is shown in Fig. 7 (e). The body is become more completeness and the false positive is not rising compared to the Fig. 7 (c) and (d).

The foreground gradient is a criterion that effectively keeps the completeness of foreground. The performance of the forbidden propagation policy is shown in the experiment.

## 3.3 Shadow Removal Classifier

In this section, we explain the algorithm of feature extraction and classifier training in the shadow removal. In Section 3.3.1, we explain how we extract the value of each feature. In Section 3.3.2, we explain SVM and Random Forest which we use in the shadow removal.

### 3.3.1 Feature Extraction

According to the survey paper [10], divide the spectral feature to four category, chromaticity, physical properties, geometry and texture. In this section we will explain why are the feature we chose.

In the chromaticity methods, the most important issue is chose the color space can separate the intensity and chromaticity. Several color spaces such HSV [6], c1c2c3 [18] and YUcUv[19] have great performance of shadow removal. We chose the method propose by Cucchiara et al. [6], since that color space is divide the chromaticity and intensity automatic and it is widely use in shadow removal. Because the region of shadow have similar hue and saturation but intensity is lower than background's intensity. We classify the feature of chromaticity according to the following equation:

$$\mathrm{Chormativity}\left(\mathrm{H}_{(\mathrm{p})}\right) = \left|\mathrm{F}^{\mathrm{H}}{}_{(\mathrm{p})} - \mathrm{B}^{\mathrm{H}}{}_{(\mathrm{p})}\right|, \tag{9}$$

$$\mathrm{Chormaticity}\left(\mathrm{S}_{(\mathrm{p})}\right) = \mathrm{F}^{\mathrm{S}}{}_{(\mathrm{p})} - \mathrm{B}^{\mathrm{S}}{}_{(\mathrm{p})}, \tag{10}$$

$$\mathrm{Chromaticity}\left(\mathrm{V}_{(\mathrm{p})}\right) = \mathrm{F}^{\mathrm{V}}{}_{(\mathrm{p})} / \mathrm{B}^{\mathrm{V}}{}_{(\mathrm{p})}, \tag{11}$$

We use (9) and (10) to calculate the difference between the frame and background and use (11) to calculate the proportion of intensity. The symbol $F^{H}{}_{(p)}$, $F^{S}{}_{(p)}$ and $F^{V}{}_{(p)}$ represent the hue , saturation and intensity at pixel p of current frame. The $B^{H}{}_{(p)}$, $B^{S}{}_{(p)}$ and $B^{V}{}_{(p)}$ represent the value of background.

There have two light source in the outdoor scene, ambient light and sun light. The physical properties method consider the ambient light is blue and the sun is white. Because the region of shadow have fewer source of sun light, it become blur than the background which was not blocked. We chose the method survey paper [10], since that use the RGB color space to calculate the physical properties. The method mapping the R, G and B component to the three-dimension space and each color is seem to the vector in this space. In the feature of physical properties, we classify feature of physical properties according to the following equation:

$$\mathrm{Physucal}(\emptyset_{(p)}) = \cos^{-1}\left(F^{B}{}_{(p)}/\left\|F_{(p)}\right\|\right), \tag{12}$$

$$\mathrm{Physucal}\left(\theta_{(p)}\right) = \tan^{-1}\left(F^{G}{}_{(p)}/F^{R}{}_{(p)}\right), \tag{13}$$

$$\mathrm{Physucal}(\alpha_{(p)}) = \left\|F_{(p)}\right\|/\left\|B_{(p)}\right\|, \tag{14}$$

In the three-dimension, this method use (12) to compute the angle of blue and color to represent the important part of physical properties. Then use the (13) to compute the angle of green and red to confirm the color is similar to the background and use the (14) to calculate the illumination attenuation of the current frame and background. The symbol $F^{R}{}_{(p)}$, $F^{G}{}_{(p)}$ and $F^{B}{}_{(p)}$ represent the component of red, green and blue of the pixel p in the current frame. The $F_{(p)}$ and $B_{(p)}$ represent the value of combine the RGB component of current frame and background at the pixel p.

Because we are not use the color information, we convert the color space to the gray space. There have lots of different formulation to compute the texture of the image. But have not any formulation is perfect to whole case because the scene is diverse. So, we chose the widely use method of [8], since it is representative and simple to implement that can reduce the computing time. We classify feature of physical properties according to the following equation:

$$\mathrm{Texture}\left(\lambda_{(p)}\right) = \left(F_{(p)}/B_{(p)}\right)/\left(F_{(p)} - B_{(p)}\right), \tag{15}$$

$$\mathrm{Texture}(\nabla_{(p)}) = \sqrt{\nabla_x^2 + \nabla_y^2}, \tag{16}$$

$$\mathrm{Texture}\left(\theta_{(p)}\right) = \tan^{-1}\left(\nabla_y/\nabla_x\right), \tag{17}$$

we use the (15) to compute the difference between current frame and background.

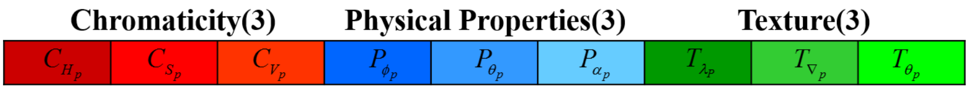



Fig. 8 A 9-dimensional vector of three features

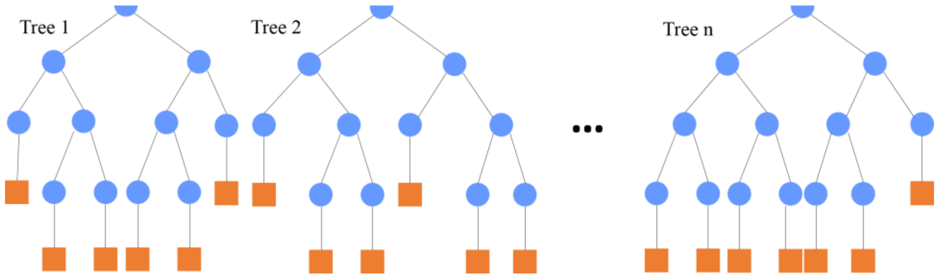



Fig. 9 The Random Forest Classifier

Then use the (16) and (17) to calculate the gradient and orientation of each pixel. The symbol ▽x and ▽y represent the gradient of horizontal and vertical.

We are not chose the feature of geometry because we extract this feature value pixel by pixel. Then use these feature of each pixel to training the classifier in latter step of our architecture. But the method of geometry extract the feature base on the region information. So we are not chose geometry to our method.

### 3.3.2 Classifier Training

In this paper, the Random Forest learning mechanism is adopted, because of the properties of feature that have different performance in each scene is similar to the concept of Random Forest.

Each feature described in Section 3.3.1 has three dimensions. However, it is difficult to choose a feature for a scene. Therefore, in this work, we merge three features to be one instance with nine dimensions, as shown in Fig. 8. An instance can be expressed as a vector, $f_i = \langle C_{H_p}, C_{S_p}, C_{V_p}, P_{\phi_p}, P_{\theta_p}, P_{\alpha_p}, T_{\lambda_p}, T_{\nabla_p}, T_{\theta_p} \rangle$, where $f_i$ is a combined instance for training data at pixel $p$.

The classifier is obtained according to the Random Forest algorithm as shown in

Fig. 9 For each tree of Random Forest, it randomly choose a subsample from $f$ to be the training data. For each node of tree, Random Forest chose the most suitable part of features according to the distribution of the selected training data, and train the best threshold for that node. The number of node of each tree is growing, until the training data is small enough.

# Chapter 4 Experiments

In this section, we analyze the result of our propose include SBE, shadow removal with classifier and combine these two components. We compare SBE with GMM, Codebook and ViBe by two popular surveillance video datasets, Wallflower [15] and Perception [16]. These datasets include major challenge of visual surveillance, dynamic background, sudden changes of illumination and crowded people. The detailed list of the experiment is shown in Table 1. In the experiments of shadow removal, we select six dataset which is often used in shadow removal. These dataset include indoor, outdoor, simple background and complex background. The detailed list of the experiment is shown in Table 2. Five indicative items: true positive (TP), false positive (FP),true negative (TN), false negative (FN) and total error (TE) are used to compute the performance of our method. The true positive refers to the number of pixels which are the background pixel then marked as background. The true negative refers to the

| | |
|---|---|
| Compared method | Gaussian Mixture Model(GMM) |
| | Codebook |
| | ViBe |
| Test Videos | Dataset |
| 1#:LightSwitch | Wallflower |
| 2#: TimeOfDay | |
| 3#: WavingTrees | |
| 4#: Campus | Perception |
| 5#: Curtain | |
| 6#: Escalator | |
| 7#: WaterSurface | |

Table 1 List of experiment items of object detect

| Dataset | Scene Type | Shadow Size | Shadow Strength |
|---|---|---|---|
| Campus | Outdoor | Large | Weak |
| Hallway | Indoor | Medium | Weak |
| Highway1 | Outdoor | Large | Strong |
| Highway3 | Outdoor | Small | Strong |
| Lab | Indoor | Medium | Weak |
| Room | Indoor | Large | Medium |

Table 2 List of experiment items of Shadow removal

number of pixels which are the background pixel but marked as foreground. The false positive and false negative are contrary to the true positive and true negative which is the foreground pixel and be marked as foreground and background. The total error is the sum of the false positive and false negative.

## 4.1 Experimental Result Analysis of Spatiotemporal Background Extractor

We select the FP, FN and TE to represent the performance of SBE and other

| Total Error(TE) | | | | |
|---|---|---|---|---|
| Video | MoG | Codebook | ViBe | SBE |
| 1# | 15828 | 11887 | 15053 | 4221 |
| 2# | 1044 | 1081 | 1143 | 854 |
| 3# | 1807 | 1011 | 1172 | 378 |
| 4# | 1168 | 1535 | 605 | 317 |
| 5# | 511 | 1471 | 1768 | 993 |
| 6# | 362 | 1205 | 746 | 557 |
| 7# | 376 | 1091 | 1172 | 478 |
| Average | 3013 | 2754 | 3047 | 1114 |

Table 3 Experiment results of object detection


Table 4 Results on the dataset

method. The TEs of each video for SBE and other method are shown in Table 3. Although some result of GMM has least TEs, our method (SBE) has the least average error. That is mean our method is more suitable for different scene. Table 4 shows the results with images. In the video 1# which has sudden changes of illumination, GMM, Codebook and ViBe have poor performance than SBE because they need a period time to update the background model to solve the problem. SBE has the mechanism to detect the sudden change of illumination and reconstructing the background model. In the case with the dynamic background, SBE has the better performance in video3#, 4# and 7#. In all video dataset, SBE has the least FN and the most completeness foreground. Fig. 10 shows the performance with FP and FN for each method. The blue bar and red bar in Fig. 10 represents the value of FP and FN. We chose the video has sudden changes of illumination and dynamic background in Fig. 10 (a) to (c). The total error with of all

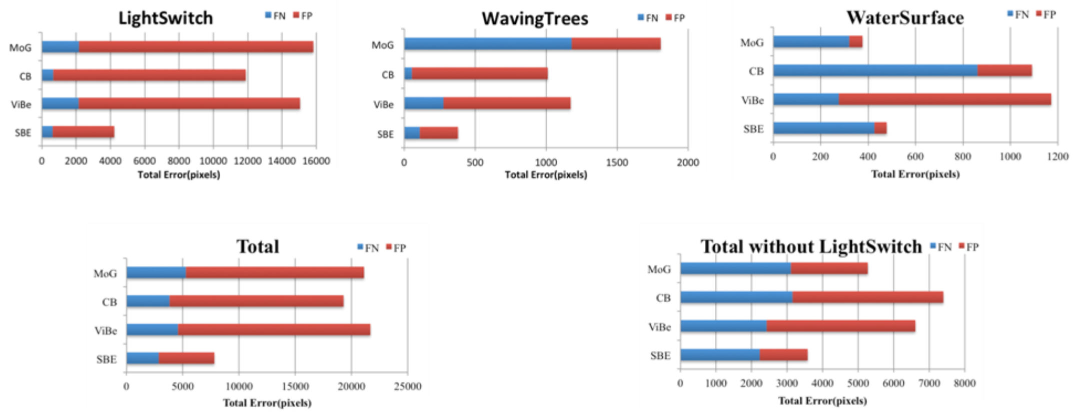
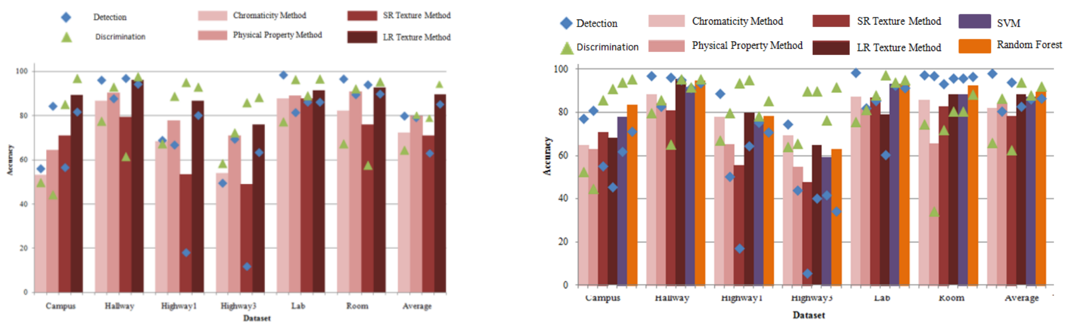



Fig. 10 Experiment results of object detection by FN and FP

video is shown in Fig. 10 (d) and the total error without the sudden change of illumination is shown in Fig. 10 (e). Although GMM has the least TE in Fig. 10 (c), the performance of GMM in other dataset has the worst , as shown in Fig. 10 (d) and (e). Comparing with Codebook, because of SBE consists the spatial information, the FP of SBE is batter than Codebook obvious and the total error of proposed method is the least. To compare with ViBe, the total error and the case of sudden change of illumination is the least.

## 4.2 Experimental Result Analysis of Shadow Removal with Single Image

In this section, we analyze the performance of shadow removal. First, we shown the results of each and combined features which is classified by SVM or Random Forest. Then, we demonstrate the performance of the proposed method with other methods which is described in [10]. Two metrics are used, the shadow detection rate ($\eta$) and the shadow discrimination ($\xi$), to evaluate the performance of shadow removal which are proposed by Prati *et al*. [17], as shown below:



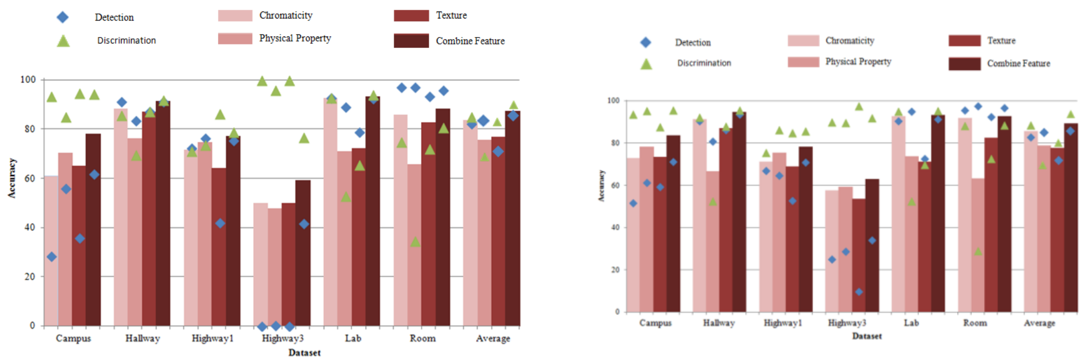

(a) (b)

Fig. 11 Comparison of shadow detection results on various sequences. (a) The results from [10] (b) The results of the program released from [10]

$$\eta = \frac{TP_S}{TP_S + FN_S}, \tag{9}$$

$$\xi = \frac{TP_F}{TP_F + FN_F}, \tag{10}$$

There have four indicative items in shadow: $TP_S$, $TP_F$, $FN_S$ and $FN_F$. In our method of shadow removal, we use the ground truth that has three class pixel, background, foreground and shadow to train the classifier model. The $TP_S$ refers to the number of pixels which are the shadow pixel then marked as shadow. The $TP_F$ represents to the number of pixels which are the foreground pixel then marked as foreground. The $FN_S$ refers to the number of pixel which are the foreground pixel then marked as shadow. The $FN_F$ represent to the number of pixel which are the shadow pixel then marked as foreground. The value of $\eta$ and $\xi$ show the performance of each method. If the




(a) (b)

Fig. 12 Experimental results of different classifiers. (a) SVM (b) Random Forest

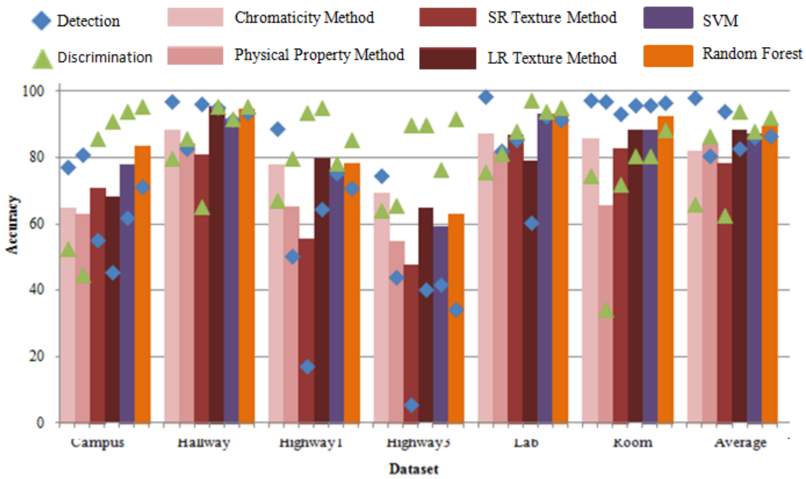



Fig. 13 Experimental result of different methods

classifier has poor performance that marked all testing data to foreground. That has one hundred percent accuracy of $\xi$ and zero accuracy of $\eta$. Thus, we use the average of $\eta$ and $\xi$ to represent the accuracy of performance.

[10] releases the programs for geometry method, chromacity method, physical method, small-region texture method, and large-region texture method. Fig. 11 shows the average shadow detection and discrimination rates on various sequences. Fig. 11 (a) and (b) are the results from the paper of [10] and that of released programs which has optimal parameters, respectively. From Fig. 11, there are some of difference between Fig. 11 (a) and (b), because the performance of those programs are all depend on the parameters tuning. In the following, the programs are used in the different experiments and the parameters are all optimal.

### 4.2.1 Combined Strategy between Different Features

Various features have different performance in different environments. Here, we first shown the performance of SVM classifier with individual features and the combined feature. Then, we compared the performance of SVM with Random Forest.

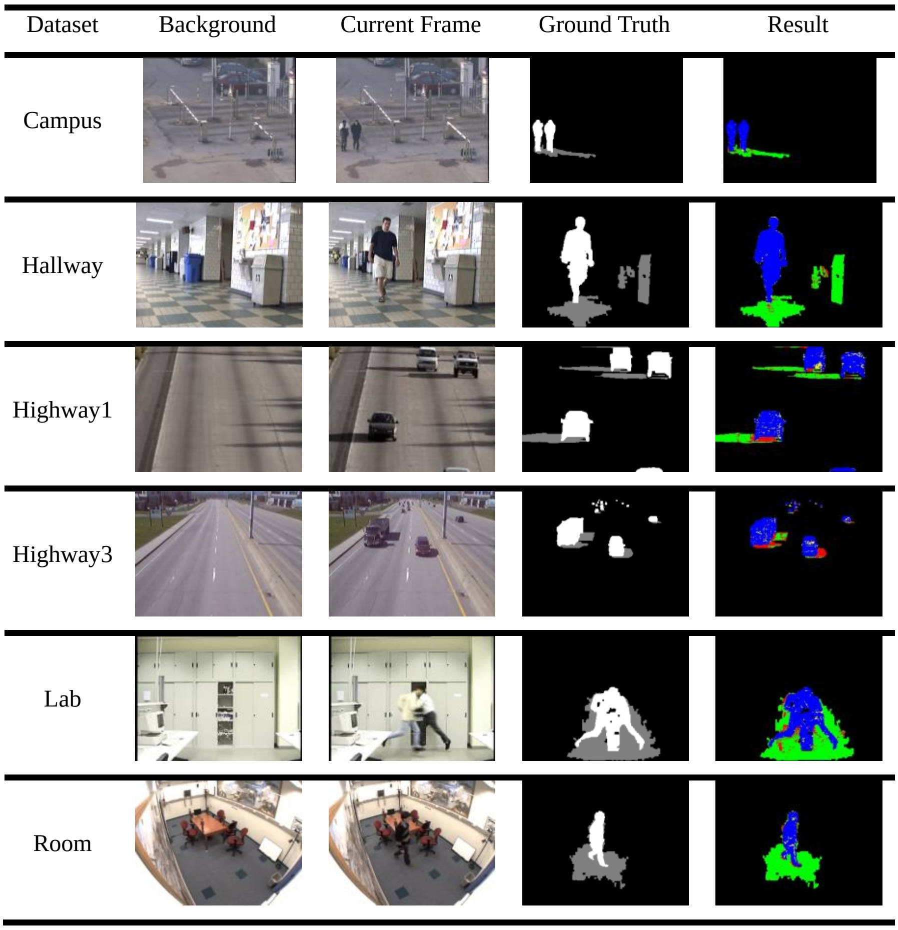


Fig. 14 Experimental result of Shadow with Random Forest

Fig. 12 (a) is the results using SVM classifier with individual features and the combined feature in different environments. From Fig. 12 (a), the combined feature has the best performance in SVM classifier. As the same as Fig. 12 (a), the combined feature has the best performance in Random Forest classifier as shown in Fig. 12 (b). From Fig. 12, the combined feature has the best performance in different environments. The performance of Random Forest classifier with combined feature is better than SVM classifier.

### 4.2.2 Experimental Result of Different Classifier

In this subsection, we shown the results of SVM, Random Forest and each method which is described in [10] in Fig. 13.

In Fig. 13, each method has the worst performance in the dataset of Highway3 in

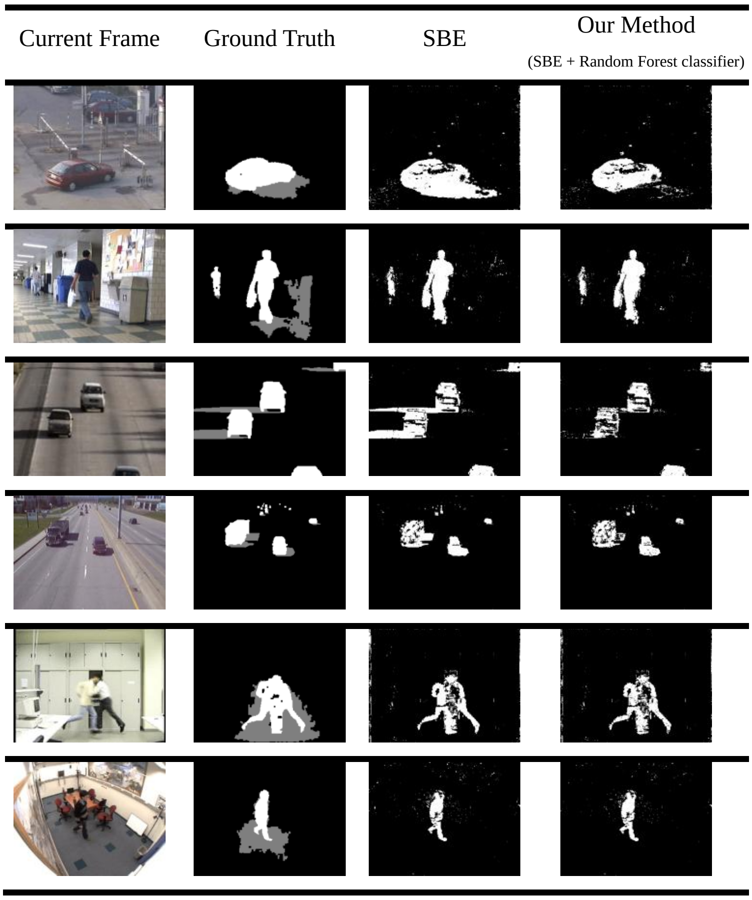



Fig. 15 Experimental results of the proposed method

which the chromaticity method has the best performance, and SVM and Random Forest also performed well. Although, the large-region texture method has the best performance in some datasets, SVM and Random Forest have the best performance in average. To compare SVM with Random Forest, Random Forest has the best performance in the most datasets and has the best performance in average. The performance of each method described in [10] is depend on the parameter tuning, the proposed method, Random Forest shadow classifier, is suitable for various scenes.

Fig. 14 shows the proposed method, Random Forest shadow classifier. The first column is the current frame, the second column is the ground truth in which the foreground pixels are indicated as white and the shadow pixels are marked in gray, and the third column is results of Random Forest classifier in which the shadow pixels

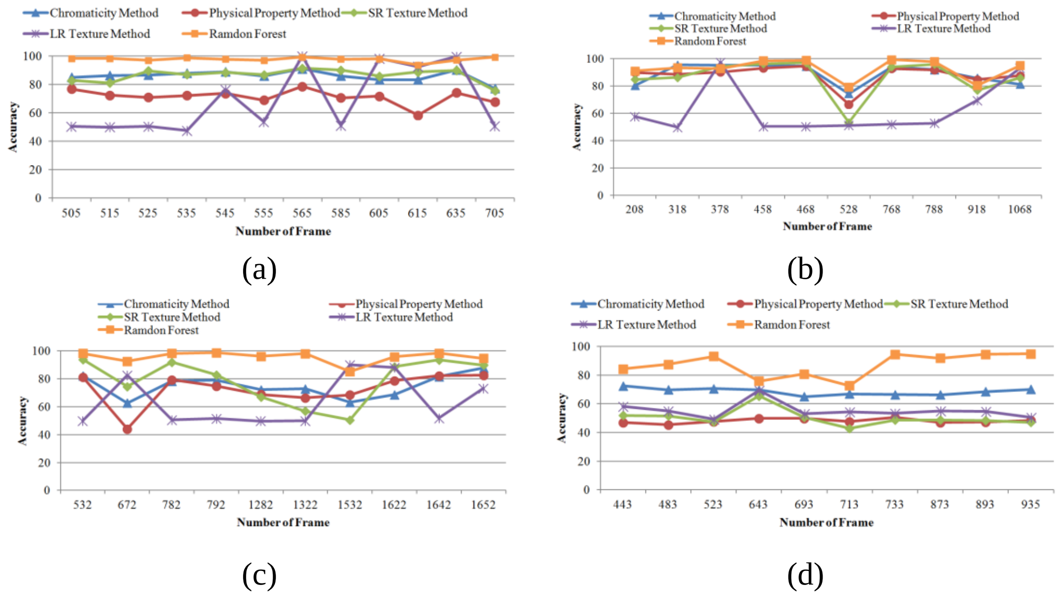



Fig. 16 Experimental results of the proposed method with the proposed (a)oceanwaves (b)NTUclassroom (c)NTUhallway (d)NTUoutdoor

which is classified as shadow are marked in green, the shadow pixels which are classified as foreground are marked in red, the foreground pixels which are classified as foreground marked as blue, and the foreground pixel which are classified as foreground are marked in yellow.

## 4.3 Experimental Result of the Proposed Method

We combine the SBE and Random Forest shadow detector to detect moving objects. The classifier is trained according to the ground truth of training data, and the shadow is removed from the results of SBE, the results are shown in Fig. 15. In Fig. 15, some results of SBE have great performance of shadow. We can see the results of Hallway, Lab and Room have the clear shadow without using shadow removal. Thus, the performance of shadow removal is not obvious. In the cases of Campus and Highway1, SBE did not perform well in cast shadow. After using the Random Forest classifier, the shadow is removed and the results of moving objects detection become

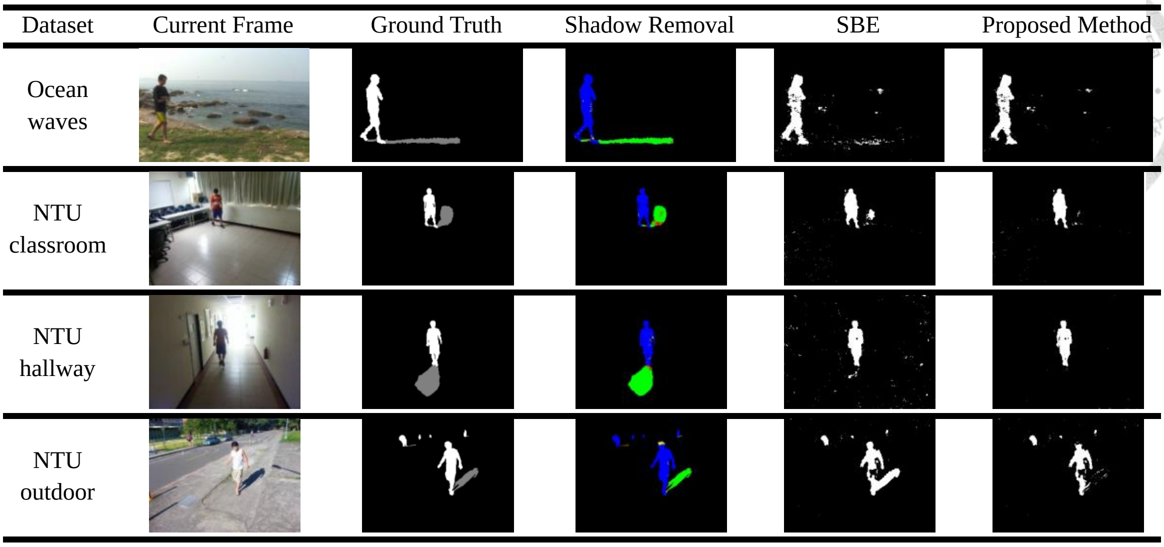
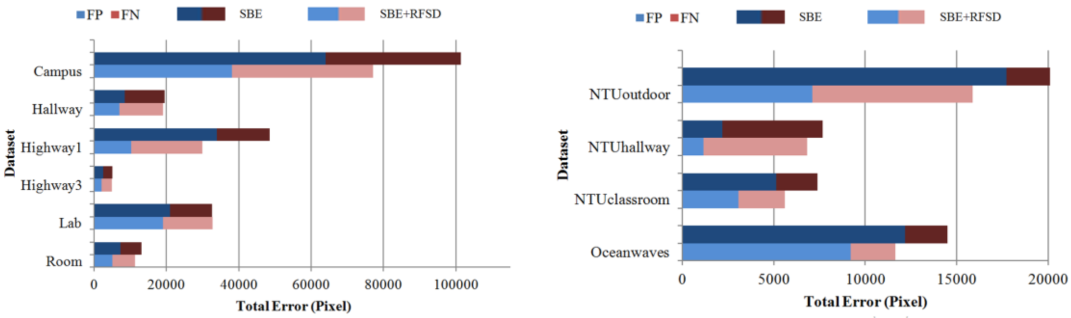



Fig. 17 Experimental results of the proposed method with the proposed dataset

accurately.

We also proposed a new datasets, oceanwaves, NTUclassroom, NTUhallway and NTUoutdoor, for testing. Fig. 16 shows the results of shadow removal using the proposed dataset. From Fig. 16, the proposed classifier, Random Forest classifier, has the best performance. Fig. 17 shows the results of proposed method, SBE + Random Forest shadow classifier, SBE is useful for dynamic background and Random Forest shadow classifier can effectively remove the shadow effect, as shown in Fig. 17 . In Fig. 18, we compute total error of dataset between result of SBE and SBE with shadow removal. The result of SBE with shadow is better than SBE because shadow removal can remove major shadow and remain complete foreground.




(a) (b)

Fig. 18 Experimental result with total error (a) previous dataset (b) new dataset

The performance of SBE with frame size 160*120 is 60~65 fps and frame size 320*240 is 20 fps. The performance of combine SBE and shadow removal with frame size 320*240 is 1.5 fps. The procedure of SBE and shadow include three part, background modeling, feature extract and shadow removal. The performance of feature extract and shadow removal are 6 fps and 2 fps. The most time-consuming part is shadow removal. In shadow removal part, random forest needs to test every pixel by all decision trees. We use 500 decision trees to build the random forest. Although reduce the number of decision tree can promote the speed of testing step, it may cause the poor performance of accuracy.

# Chapter 5 Conclusions and Future Works

We propose a method combine the object detect and shadow removal. Object detection includes two major part of the detect policy, temporal information and spatial information. We design a novel framework, the spatiotemporal background extractor (SBE) that combine the temporal information and spatial information. Two main components are proposed: the background extractor (BE) and the background gradient extractor (BGE). The BE detects the foreground and eliminates the noise in background. Only use one frame to construct the BE which is based on a single-layer Codebook model and the spatial information is propagated from the adjacent neighbors. The BE can efficiently solve the most challenges of object detection such as dynamic background and the sudden changes of illumination. However the policy of propagation makes the foreground more incompleteness. Therefore, the propagation is forbidden base on the result of BGE that keeps the completeness of foreground. The BGE and BE is constructed synchronously. To construct the BGE, the stable background gradient is used which is used to find foreground gradient of the current frame. In order to handle the problem of shadow removal in visual surveillance. We propose a method combine the advantage of different features of shadow that can remove the shadow base on the properties of each scene. We combine the features, chromaticity, physical properties and texture, from the ground truth. Using the combined data to train the classifier which can remove the shadow from the result of object detection. We choose Random Forest algorithm to train the shadow classifier model because the policy of Random Forest is more suitable for properties of the feature. After training the Random Forest classifier, we use this classifier to remove the shadow from the result of SBE.

In the experimental results of object detection, the result on a set of background modeling databases, Perception and Wallflower. We analyze the performance of the proposed method (SBE) to compare with GMM, Codebook, and ViBe. The total error of the proposed method is the least, and the performance of solving the dynamic background and the sudden changes of illumination are greater than others. In the experimental result of shadow removal, the performance of combine different feature is better than use the single feature. Comparing the method of feature-based and our method, the Random Forest classifier has the higher accuracy than SVM classifier. The result of the proposed method has the similar accuracy with feature-based using feature of texture and better performance than other features. Compare to the method of feature-based which is sensitive to the threshold of each feature, our method can train the model that suitable for each scene by it's properties.

The result of spatiotemporal background extractor can handle the dynamic background and sudden change of illumination, the foreground has some error region cause by two-way propagation. Although we proposed the forbidden propagation policy, it still has some region of foreground classified to background. We want to use color information about neighbor to solve this problem to get the more complete foreground. In our proposed work, we use three feature, chromaticity, physical property and texture to train the shadow model. We want to combine other feature do not use the color information and gradient information to improve the performance of shadow removal.

# REFERENCE

[1] L. Liyuan, H. Weimin, I. Y. H. Gu, and T. Qi. Statistical modeling of complex backgrounds for foreground object detection. IEEE Transactions on Image Processing, 13(11):1459-1472, 2004.

[2] K. Kim, T. H. Chalidabhongse, D. Harwood, and L. Davis. Real-time foreground–background segmentation using codebook model. Real-time imaging, 11(3):172-185, 2005.

[3] K. Kim, T. H. Chalidabhongse, D. Harwood, and L. Davis. Background modeling and subtraction by codebook construction. In Proceeding of the International Conference on Image Processing (ICIP'04), 3061-3064, 2004.

[4] O. Barnich and M. Van Droogenbroeck. ViBe: A universal background subtraction algorithm for video sequences. IEEE Transactions on Image Processing, 20(6):1709-1724, 2011.

[5] Wenping Wu, Jing Shao, and Wei Guo. Moving-object Detection Based on Shadow Removal and Prospect Reconstruction. Lecture Note in information Technology, vol. 12, 2012.

[6] R. Cucchiara, C. Grana, M. Piccardi, A. Prati, "Detecting moving objects, ghosts, and shadows in video streams," *IEEE Transactions on Pattern Analysis and Machine Intelligence* Vol. 25, No.10, pp. 1337-1342, 2003.

[7] J. B. Huang, C. S. Chen, "Moving cast shadow detection using physics-based features," in Proceeding of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'09), pp. 2310-2317, 2009.

[8] A. Sanin, C. Sanderson, B. Lovell, "Improved shadow removal for robust person

tracking in surveillance scenarios," in Proc. of ICPR'10, pp. 141-144, 2010.

[9] Shih-Chieh Wang, Te-Feng Su and Shang-Hong Lai"Detecting Moving Object From Dynamic Background with Shadow Removal"ICASSP2011 IEEE.

[10] A. Sanin , C. Sanderson , Brian C. Lovell, "Shadow detection: A survey and comparative evaluation of recent methods," *Pattern Recognition*, Vol. 45, No. 4, pp. 1684-1695, 2012.

[11] A. Amato, I. Huerta, Mikhail G. Mozerov, F. Xavier Roca ,Jordi Gonzàlez, "Moving Cast Shadows Detection Methods for Video Surveillance Applications," *Augmented Vision and Reality*, Vol 6, pp 23-47. 2014.

[12] Sohail Nadimi, Bir Bhanu, Physical Models for Moving Shadow and Object Detection in Video. IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 26, NO. 8, AUGUST, 2004.

[13] Chia-Jung Chang, Wen-Fong Hu, Hsieh, Jun-Wei, Yung-Sheng Chen, "Shadow elimination for effective moving object detection with Gaussian models", Pattern Recognition, 2002. Proceedings. 16th International Conference on, On page(s): 540 - 543 vol.2 Volume: 2, 2002

[14] A. Sanin, C. Sanderson, B. Lovell, "Improved shadow removal for robust person tracking in surveillance scenarios," in Proc. of ICPR'10, pp. 141-144, 2010.

[15] K. Toyama, J. Krumm, B. Brumitt, and B. Meyers. Wallflower: Principles and practice of background maintenance. In Proceedings of the 7th IEEE International Conference on Computer Vision (ICCV'99), 255-261, 1999.

[16] L. Liyuan, H. Weimin, I. Y. H. Gu, and T. Qi. Statistical modeling of complex backgrounds for foreground object detection. IEEE Transactions on Image Processing, 13(11):1459-1472, 2004.

[17] A. Prati, I. Mikic, M. Trivedi, R. Cucchiara, Detecting moving shadows:

algorithms and evaluation, IEEE Transactions on Pattern Analysis and Machine Intelligence 25 (7) (2003) 918–923.