الجمهورية الجزائرية الديمقراطية الشعبية
République Algérienne Démocratique et Populaire
وزارة التعليم العالي و البحث العلمي

Université Mohamed Khider - Biskra
Faculté des sciences et de la technologie
Département de génie électrique
Ref: ....

جامعة محمد خيضر بسكرة
كلية العلوم و التكنولوجيا
قسم الهندسة الكهربائية
المرجع : ....

Thèse présentée en vue de l'obtention

du diplôme de

Doctorat 3ème cycle LMD en: Communication Signaux et Réseaux

Segmentation des Images au Sens du Mouvement

(Image Segmentation Based on Motion)

Présentée par:

OURCHANI AMINA

Soutenue publiquement le : 2018

Devant le jury composé de :

Pr. Abdelkarim OUAFI	**Professeur**	**Président**	**Université de Biskra**
Pr. Zine-Eddine BAARIR	**Professeur**	**Rapporteur**	**Université de Biskra**
Pr. Abdelmalik TALEB-AHMED	**Professeur**	**Examinateur**	**Université de Valenciennes**
Pr. Mahmoud TAIBI	**Professeur**	**Examinateur**	**Université de Annaba**
Dr. Abida TOUMI	**MCA**	**Examinateur**	**Université de Biskra**
Dr. Athmane ZITOUNI	**MCA**	**Examinateur**	**Université de Biskra**

Dedication

I dedicate this modest work to:

My beloved parents

My beloved husband

My brothers and sisters: **Moulida**, **Abla**, **Hichem**, **Loubna**, **Malek**.

My daughter and son: **Rihem**, **Yacine**.

My supervisors: **Zine-Eddine Baarir** and **Abdelmalik Taleb-Ahmed**.

May **Allah** grant your health, happiness and long life and make sure that I never disappoint you.

All my friends and colleagues.

All my teachers of the **Department of Electrical Engineering of Biskra**.

Finally, I dedicate this modest work to all those I love and appreciate.

Acknowledgments

*First of all, I wish to express my sincere gratitude to my supervisor, **Mr. Baarir Zine-eddine**, professor of University of Mohamed Khider. for his guideline, advice and support during my thesis work.*

A special thanks also go to professor **Abdelmalik Taleb-Ahmed** and Professor **Mezerdi Brahim** .

I am very indebted to doctor **Azzedine. Benlamoudi** and doctor **Salahuddin. Bekhouche**, for their continuous encouragement and support. They were always ready to help with a smile.

I am also thankful to all the members of the jury who honor me by evaluating this work.

My sincere thanks to all the teachers that had provided me with kind words, a welcome ear, new ideas, useful criticism, or their invtime.

I am really thankful to my all friends and Colleagues in **LESIA** research laboratory.

To each and every one of you – Thank you.

Abstract

Detection of moving objects from background in video sequences is hard, but essential task in a large number of applications in computer vision. Most of the existing methods had given accepted results only in the case where both object and background are rigid, because of the serious occlusions and the complex computation, which presents limitations in case of occlusions and shadows.

In this thesis ,we developed three new approches to detect moving foregrounds from complex backgrounds . In the first approach called new method for the motion estimation and segmentation of moving objects ”NMES”, we focus on the combination of motion, color and texture features.

Firstly, we have used the block-matching method to compute the motion vectorand we have taken into our consideration the result of the frame difference technique, to improve the quality of the optical flow. Moreover, we have used the k-means clustering algorithm owing to group the pixels, having similar motion, color and texture features.

lastly, the results of grouping pixels are used as an input in Chan-Vese model, in order to attract the evolving contour of moving object contours.

In the second approach called combination between motion and shape features ”MSFs”, we applied a logical comparison between the results of the optical flow (motion feature) and the color space segmentation (shape feature) of each pixel.

In the third approach called hybridization between motion and texture features ”BMFS-LBP”, we concentrated on a combination between local binary pattern (texture feature) and block matching full Search algorithm (motion feature).

Both of MSFs and BMFS-LBP approaches are suitable to discriminate moving objects from both static and dynamic background.

Finally, to evaluate the performance of our proposed approaches, we experimentd them on challenging sequences. Ws have shown that our BMFS-LBP approach provided improved segmentation results compared with MSFs method and state of art methods.

Keywords

Segmentation, moving object, background, foreground, color feature, texture feature, shape feature, motion feature, NMES, MSFs, BMFS-LBP.

Résumé

La détection d'objets en mouvement à partir du fond dans des séquences vidéo est difficile, mais est une tâche essentielle dans un grand nombre d'applications en vision par ordinateur.

La plupart des méthodes existantes donnent des résultats acceptés uniquement dans le cas où les objets et l'arrière-plan sont rigides, à cause d'occlusions et de calculs complexes, ce qui présente des limitations en cas d'occlusions et d'ombres.

Dans cette thèse, nous avons développé trois nouvelles techniques pour détecter le premier plan à partir d'un arrière-plan complexe.

Dans la première approche appelée nouvelle méthode pour l'estimation du mouvement et la segmentation des objets en mouvement "NMES", nous nous sommes concentrés sur la combinaison des caractéristiques de mouvement, de couleur et de texture.

Tout d'abord, nous avons utilisé la méthode de mise en correspondance des blocks pour calculer le vecteur de mouvement et nous avons également pris en compte le résultat de la différence des trame, pour améliorer la qualité du flux optique. De plus, nous avons utilisé l'algorithme de regroupement k-means pour regrouper les pixels, ayant des caractéristiques de mouvement, de couleur et de texture similaires.

Enfin, le résultat du regroupement des pixels est utilisé comme entrée dans le modèle de Chan-Vese, afin d'attirer le contour évolutif des contours d'objets en mouvement.

Dans la deuxième approche appelée combinaison entre les caractéristiques de mouvement et de forme "MSF", nous avons appliqué une comparaison logique entre les résultats du flux optique (caractéristique de mouvement) et la segmentation de l'espace de couleur (caractéristique de forme) de chaque pixel.

Dans la troisième approche appelée hybridation entre les fonctions de mouvement et de texture "BMFS-LBP", nous nous sommes concentrés sur la combinaison entre le modèle binaire local (fonction de texture) et l'algorithme complet de recherche de bloc (fonction de mouvement).

Les approches MSF et BMFS-LBP sont toutes les deux adaptées pour distinguer les objets mobiles à partir des arrière-plans statiques et dynamiques.

Pour évaluer les performances de nos approches proposées, nous les avons testées sur des séquences complexes.

les résultats expérimentaux ont montré que, notre approche BMFS-LBP permet de trouver de meilleurs résultats que la segmentation avec la méthode MSF et les méthodes éxistants.

Mote Clé

Segmentation, objet dynamique, arrière-plan, premier-plan, fonction de couleur, fonction de texture, fonction de forme, fonction de mouvement, "NMES", "MSFs", "BMFS-LBP".

الملخص:

يعتبر استخلاص الأجسام المتحركة من خلفية الصورة في الفيديو أمرًا صعبًا ، ولكنه ضروري في عدد كبير من تطبيقات رؤية الكمبيوتر. تعطي معظم الطرق الحالية نتائج مقبولة فقط في حالة كون المتحرك والخلفية صلبين ، بسبب التداخل بين الأشياء المتحركة ،الظلال وكذالك الحسابات المعقدة ، والتي تمثل عوائق للقيام بالاستخلاص الفعال للأجسام المتحركة.

في هذه الأطروحة ، نقوم بتطوير ثلاث تقنيات جديدة لاستخلاص الأشياء المتحركة من خلفية فيديو معقدة ، نركز على آ پيض ١) في التقنية الأولى ، والتي تسمى طريقة جديدة لتقدير الحركة وتقطيع الأجسام الجمع بين خصائص الحركة واللون والملمس.

أولاً ، استخدمنا خوارزمية كتلة المطابقة لحساب متجه الحركة ، كما أخذنا في الاعتبار نتيجة فرق الإطار، لتحسين جودة التدفق البصري. بالإضافة إلى ذالك استخدمنا خوارزمية التجميع لتجميع البكسلات التي لها نفس خصائص الحركة و اللون و الملمس.

و أخيرا يتم استخدام نتيجة تجميع البيكسل كمدخل في نموذج شان و فيز، من اجل جذب الكفاف المتحرك نحو الأجسام المتحركة

٢) في التقنية الثانية التي تسمى الجمع بين خصائص الحركة و الشكل،حيث طبقنا مقارنة منطقية بين نتائج التدفق الضوئي (خاصية الحركة) و تجزئة مساحة اللون (خاصية الشكل) لكل بكسل في الصورة.

٣) في التقنية الثالثة التي تدعى الدمج بين خصائص الحركة وخصائص النسيج ، نركز على الجمع بين النموذج الثنائي المحلي و خوارزمية البحث الشامل .

تعتبر كل من التقنيتان الثانية و الثالثة مناسبة لتمييز الأجسام المتحركة من خلفيات الصورة سواء كانت ثابتة أو ديناميكية.

لتقييم أداء تقنياتنا المقترحة، سنقوم بتجربتها على مجموعة من الفيديوهات المعقدة.

اطهرت النتائج التجريبية أن تقنيتنا الثانية المقترحة تسمح بإيجاد نتائج أفضل من تقنيتنا الثالثة المقترحة و الخوارزميات الموجودة في النهج العلمي

الكلمات المفتاحية:

تجزئة، الأجسام المتحركة، الخلفية، المقدمة، خصائص اللون، خصائص الحركة، خصائص الشكل، خصائص النسيج.

Scientific Productions

Publications in journals

- **Amina Ourchani**, Zine-Eddine Baarir and Abdelmalik Taleb-Ahmed, “A new method for the optical flow estimation and segmentation of moving objects ‘NMES’,” Int. J. Intelligent Systems Technologies and Applications, Vol. 17, No. 1/2, 2018.

Publications in international conferences

- **Amina Ourchani**, Zine-Eddine Baarir and Abdelmalik Taleb-Ahmed, “Object Detection from Dynamic Background” International Conference on Automatic control, Telecommunication and Signals (ICATS’17), December 11th -12th,2017.
- **Amina Ourchani**, Zine-Eddine Baarir and Abdelmalik Taleb-Ahmed, “Forground detection using Block-Matching Full Search-LBP algorithm for dynamic video”. Fifth International Conference on Image and Signal Processing and their Applications (ISPA), December 2017.

Contents

List of Figures

List of Tables

Acronyms

BDM Block Distortion Measure

BMFS-LBP Hybridization between Motion and Texture Features

CAVIAR Image-based Active Recognition database

CCV Color Coherence Vector

CDNET change detection.net dataset

CIE Commission International de l'Eclairage

CIE-LAB L*a*b Color space

CMYK cyan, magenta, yellow and black

DS Diamond Search Algorithm

EM Expectation Maximization algorithm

ES Exhaustive Search

FN False negatives

FP False positives

FSS Four-Step Search

GLCM Grey Level Co-occurrence Matrix

HCI Human Computer Interaction

HEXBS Hexagon Based Search Algorithm

HSV Hue Saturation and Value

IEEE Institute of Electrical and Electronics Engineers

KDE Kernel Density Estimation

LBP Local Binary Patterns

MAD Error Mean Absolute Difference

ME Motion Estimation

MGM	Mixture of Gaussians Model
MSE	Mean Squared Error
MSFs	Combination Between Motion and Shape Features
NTSS	New Three-Step
OFC	Optical flow constraint
OMC	Optical Motion Capture
pdf	Gaussian probability density function
PETS	Performance evaluation of tracking and surveillance
PSNR	Peak to Signal Noise Ratio
RGA	Running Gaussian Average
RGB	Red-Green-Blue
SAD	Sum of Absolute Difference
TMF	Temporal median filter
TN	True negatives
TP	True positives
TSS	Three-Step

1

Introduction

Contents

1.1 Introduction

Image segmentation based on motion is an active research topic in computer vision, that tries to detect (partition) moving objects in a sequence of images and grouping the pixels belonging into significant meaningful region representation which simplifies the analysis of the image sequence. The moving objects are called foreground, and the rest of the image sequence is called background.

The resulted foreground regions should attain two fundamental characteristics which are homogeneity and connectivity.

The homogeneity concerns the pixels which belong to the same region with similar color, intensity, size, shape and textural features.

The connectivity concerns any two adjacent pixels with a connected path within the whole region.

Segmentation of moving objects in an image sequence plays an important role in image processing and analysis such as sports analyses, human computer interaction, optical motion capture, intelligent visual observation of animals and insects, intelligent visual surveillance and content based video coding.

However, perfect image segmentation based on motion is hardly to achieve since there are several problems that need to be solved, such as noisy video, camera jitter, illumination changes, camouflage, moved background objects, dynamic backgrounds, Shadows.......etc.

The objective of this thesis is to develop efficient full automatic detection techniques able to segment moving regions from real image sequences with critical situations like occlusions, illumination changes, shadows and dynamic backgrounds.

In our work, we choes to make a combination between both texture, color, shape and motion feature to extract moving region, then we use a clustering method in order to fuse them in homogeneous and connected regions.

This thesis is organized in six chapters as follows:

In Chapter 2, we present the different techniques related to moving object segmentation. We also present some challenges and issues in real image sequence, In addition, we review a number of relevant works in literature. Furthermore, we introduce available datasets developed in the last years, and we explain the meaning of the ground truth. Finally, performance measure evaluation methodology and resource of codes are discussed.

In chapter 3, we describe the details of various classical methods in computer vision which we use it in our work such as motion, color, texture, and shape feature methods.

In chapter 4, Our proposed approaches of image segment based on motion from static background are presented and discussed.

In chapter 5, We focus our study on two new techniques for discriminating moving objects from both static and dynamic backgrounds.

In chapter 6, A general conclusion of our work and proposed eventual future works.

2

Chapter II. State of art of segmentation related to moving objects

Contents

2.1 Introduction

Segmentation of moving objects in an image sequence draws a crucial role in the processing and analysis of image sequences. It is the process of separating moving objects called ”foreground” from the static information called ”background”. It is very important in many multimedia applications such as surveillance, mobile robot navigation, video object detection, tracking, etc......

Initially, the segmentation of moving objects includes the detection, tracking and extraction of moving objects. Furthermore, the detection, or the correspondence is the process of taking into account the course, and properties of moving objects.

Extraction is a meaningful segmentation of the moved objects from the scene [3].

Object detection can be classified into two types. The first one is fast moving object detection. The second one is slow moving object detection. The slow moving object detection is difficult in comparison to the fast moving object detection [4].

2.2 Importance of moving objects segmentation

The segmentation of moving foreground objects from video stream is the fundamental step in many computer vision applications such as:

2.2.1 Intelligent visual surveillance

In [5] There is a strong need to put cameras in the place of human eyes, to make the whole surveillance task automatic as much as possible, which make the intelligent visual surveillance a more active research area. The intelligent visual surveillance is needed and used in the whole world, not only by private companies, but also by governments and public institutions. Indeed, it is a key technology tool for public safety like in communities, governmental buildings, schools, hospitals, public facilities, traffic lights, railroad crossings, detection of military targets or forbidden places.

2.2.2 Intelligent visual observation of animals and insects

In the literature, there have been many researches on the analysis of video data on humans, but only few studies on visual analysis for zoological aspects like animal activities in protected areas, rivers oceans have been studied.

2.2.3 Optical Motion Capture (OMC)

The optical motion capture is one of the methods for turning real-life movement into digital data. It uses up to 200 special cameras. These cameras monitor the movements precisely from different views. The captured views are then used in animation to give a character life and personality.

2.2.4 Human Computer Interaction (HCI)

Researches on human-computer-interaction (HCI) have become an increasingly important part of our daily lives. The use of HCI technology needs a video acquired by fixed cameras in real time.

2.2.5 Sports analyses

In sports analyses, the main aim is to improve the performance of players by segmenting the motion of the players in video sequences. The sports video analyses can be found in many applications like ball, player tracking, game highlight extraction, and computer-assisted refereeing.

2.2.6 Content based video coding

The goal is to generate efficient methods for segmenting the video from static and dynamic environments into foreground and background, and then encoding the registered background and foreground separately.

2.3 Challenges and Issues

There are three main ideal conditions that ensure a smooth operation of the motion segmentation method. There are three main ideal conditions which insure a good functioning of the motion detection method: the fixed camera, the illumination and the background are static.

In a real image, there are critical situations need to be solved. In [6],the methodes identified thirteen difficult situations in the context of video surveillance:

2.3.1 Noisy Video

The video signal is usually superimposed on the noise. It is due to a poor quality video source such as sensor noise or compression artifacts in images.

2.3.2 Camera jitter

In some cases, the camera rocks due to wind causing false detection in the foreground mask.

2.3.3 Camera automatic adjustments

Many modern cameras have standard features. They automatically determine the focus, gain control and brightness control. These adjustments modify dynamically the color levels between different frames in the sequence.

2.3.4 Illumination changes

The illuminations can change gradually in the appearance of the environment such as those in a day in outdoor scene, or changes suddenly by switching on / off the light in an indoor environment. The changes of illumination lead to false positive detections in several parts of foreground mask.

2.3.5 Bootstrapping

In some environments, the initial data which is free from foreground objects is not available. So, it is impossible to compute the representative background image. At this point, the bootstrapping strategy can be use to solve this problem.

2.3.6 Camouflage

In some stances, we can not distinguish between the pixels of background and foreground objects. Camouflage is particularly a problem of temporal differentiating.

2.3.7 Foreground aperture

In certain situations, the foreground objects have unified colored regions that make the detection of the change inside these regions difficult. This problem generates foreground masks which contain false negative detection.

2.3.8 Moved background objects

Some parts of background objects may contain movement. These objects should not be considered as a part of the foreground.

2.3.9 Inserted background objects

In some environments a new background object may be inserted. This inserted background object should not be considered part of the foreground.

2.3.10 Dynamic backgrounds

Backgrounds can represent sets of disjointed pixel values. This disjunction caused by the movement of the background such as a fountain, the movements, swaying of tree branches, and water wave. these of movements may be periodic or irregular such as traffic lights and waving trees.

2.3.11 Beginning moving object

Detection of moving objects becomes a very difficult job when background objects are moving.

2.3.12 Sleeping foreground object

In certain circumstances, the foreground objects become stationery objects. This foreground object will be incorporated in the background.

2.3.13 Shadows

Shadows can result from background objects or moving objects. Researchers have proposed different methods for detection of shadows.

2.4 Traditional Motion Segmentation Techniques

A large number of approaches focusing on motion segmentation have been proposed in literature. Researchers use multiple techniques and in some cases they combine them with different methodologies.

Several surveys that focused on motion segmentation methods are available in literature [2, 3,7,8] and [9].The researchers give valid and reliable views on motion segmentation techniques in this topic and provide further support for novice researchers. For example, Wang et al. [7] provide a review of important segmentation techniques.

They classified motion segmentation into two categories. The first category is motion-based segmentation methods which employ only the motion information. These methods are the most appropriate with rigid scene. The second category is spatio -temporal segmentation methods which employ both spatial and temporal information embedded in the sequence. These methods aim to overcome the sensitive noise and inaccuracy problems in motion-based segmentation.

We deduce from this survey that the spatio-temporal methods are most appropriate then the motion-based segmentation methods. In [9], Inigo et al. presented categories of segmentation based on motion algorithms; they relied on their motion representations and their clustering criteria. They presented a significant survey of some common methods for the moving objects segmentation. In this survey, they classified object segmentation methods as background subtraction, contour and threshold, temporal-spatial, edge detection methods and optical flow.

In [3],Antony et al.classified the segmentation of moving objects into three categories: regions, boundaries, and combination of regions and boundaries. In region-based methods, the image is segmented into homogeneous regions by classifying neighboring pixels of similar intensity levels.

Region-based methods use various methods such as traditional background subtraction, optical flow methods, change detection mask, markov random field, grouping, and modified statistical methods.

Boundary-based methods are another category used to detect the edges of moving objects. This approach deals with discontinuities in images. The boundary-based methods used various methods such as active contour methods, edge-based methods, and optical flow methods.

The last category introduced in [3] is a combination of region and boundary-based methods. These methods use both concepts and the advantages of region and boundary based approaches equally. These methods use the region-based techniques as background subtraction and the boundary-based techniques as edge detection methods.

From this survey, we can say that the region based methods are more complex then the boundary based methods. The region based methods are more efficient in the presence of occluded objects. The combination of region and boundary-based methods are more efficient for complex background. In general, in literature, there are many surveys, but each survey has a specific classification. As stated previously, we review the most important techniques for the segmentation of moving objects in image sequences.

2.4.1 Background subtraction algorithm

Background subtraction algorithm is a popular method for detecting the moving objects in a sequence of video frames from static cameras. The idea is to subtract the moving regions in a sequence image by taking the difference between the current image and the reference background image in a pixel-by-pixel fashion.

Frequently, the background image is called background model. It is a representation of the scene without moving objects and must be regularly updated to adapt to the varying luminaries conditions and geometry settings. This method is useful for detecting the moving objects in a surveillance camera. By using this method, we can track or recognize the objects perfectly.

It is extremely sensitive to changes in dynamic scenes derived from lighting and extraneous events. In [10] and [11], the background subtraction algorithms pass through four major steps. These steps are pre-processing, background modeling, foreground detection and data validation. Figure (2.1) Shows the block diagram of the background subtraction algorithm steps.

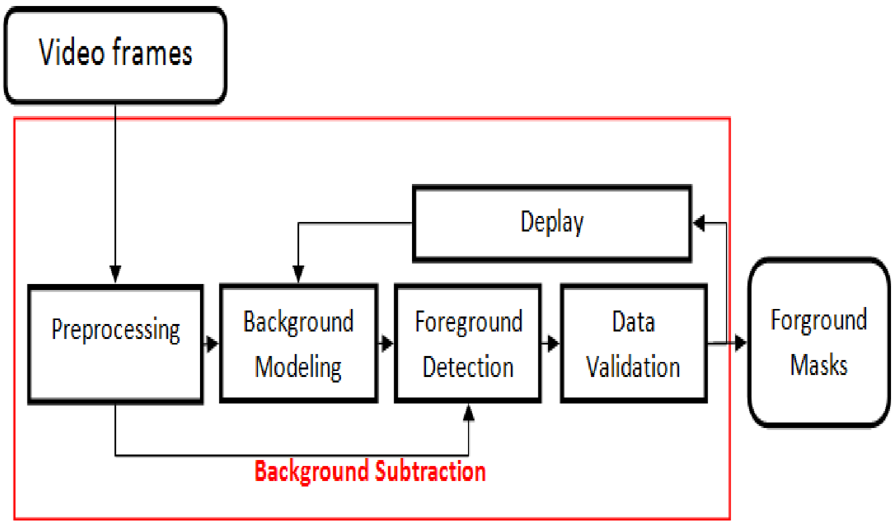


Figure 2.1: Block diagram of the background subtraction algorithm steps

2.4.1.A Pre-processing

The pre-processing step is a simple temporal and/or spatial smoothing used initially to filter out common types of unnecessary changes before making the object detection decision. This step is used to reduce camera noise and remove transient environmental noise (rain and snow captured by an outdoor camera).

The use of frame size and frame rate reduction is very efficient in reducing data processing rate in real time systems .In the case of moving cameras, researchers use the image registration between successive frames before background modeling.

In addition, most algorithms handle the color image, or the Red-Green-Blue (RGB), Hue Saturation and Value (HSV) color space in order to identify objects in low-contrast areas and remove shadows from moving objects.

The addition of spatial and features temporal features to color is increasingly used in the background subtraction literature.

2.4.1.B Background modeling

Also called background maintenance, this is an essential step of any background subtraction algorithm.

In the following subsections, we will focuse only on the most popular techniques used in background subtraction.

1. Simple background subtraction [12]

   In simple background subtraction (figure (2.2)), the absolute intensity difference is taken between each current image I_t (x,y)and the background model $B(x,y)$to find the motion detection mask $D(x,y)$.

   The background model is a static image (supposed to have no present object), it is usually the first frame of a video.

$$D(x,y) = \begin{cases} 1 & \text{if } |I_t(x,y) - B(x,y)| \geq \tau \\ 0 & otherwise. \end{cases} \tag{2.1}$$

   Where: τ is a threshold that determines if the pixel is the foreground or the background. If the absolute difference is greater than or equal to τ, then the pixel is classified as foreground (1), otherwise the pixel is classified as background (0).

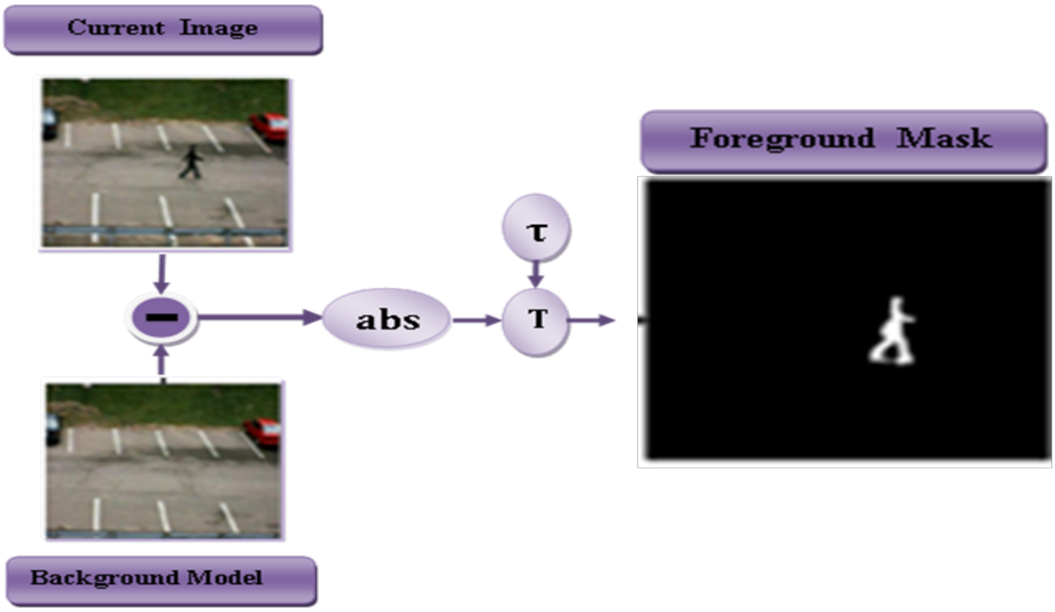


Figure 2.2: Simple background subtraction technique

2. Running Gaussian Average (RGA)

   Wren et al.([12, 13]) proposed Running Gaussian Average (RGA.) to model the background independently at each (x,y) pixel location to eliminate the ghost in the difference images. The RGA algorithm comprises two steps: differentiation step and background modeling step([13]). The differentiating step extracts motion pixels by computing the difference between current frame and background model by.

$$D(x,y) = |I_t(x,y) - B(x,y)| \tag{2.2}$$

   The background modeling step is based on fit a Gaussian probability density function (pdf).

$$B_t(x,y) = MB_{(t-1)}(x,y) + (1+M)(\propto I_t(x,y) + (1-\propto)B_{(t-1)} \tag{2.3}$$

   Where: α is an empirical weight often chosen as a trade-off between stability and quick update, and the binary value M is 1 in correspondence of a foreground value, and 0 otherwise. The binary motion detection mask $D(x, y)$ is calculated as follows

$$D(x, y) = \begin{cases} 1 & \text{if } |I_t(x, y) - B(x, y)| \geq \tau \\ 0 & otherwise. \end{cases} \tag{2.4}$$

3. Temporal median filter

   Temporal median filter is one of the simplest non-recursive background modeling techniques. In the Temporal median filter (TMF) technique, the median value of the last n frames is used as a background model at each pixel. We use this technique to increase the stability of the background model, but with recent pixel values, a calculation requires a buffer.

4. Mixture of Gaussians Model (MGM) [9]

   sometimes the changes in the background object are not permanent and appear at a rate faster than that of the background update, A typical example is that of an outdoor scene with snowing, raining, or waving trees. In these cases a multi-valued background model is obtained. The mixture of gaussian model is capable to raise the problem of handling multi-modal background distributions. In [14], They made a survey from 170 original papers covered background subtraction based on MGM.

   The Principle of the mixture of Gaussian algorithm is resumed as follows:

   (a) Assume that the RGB color components are independent and have the same variances.

   (b) Suppose that a background pixel corresponds to a high weight with a weak variance due to the fact that the background is more present than moving objects and that its value is practically constant.

   (c) Calculate the probability of observing the current pixel value.

   (d) Define the background model

(e) Use an Expectation Maximization algorithm Expectation Maximization algorithm (EM) algorithm to initialize the different parameters of the mixture of Gaussians (weight, mean and covariance).

(f) First foreground detection can be made. Two cases may occur:

   i. Case 1: A match is found with one of the K Gaussians. In this case, if the Gaussian distribution is identified as a background one, the pixel is classified as background, otherwise the pixel is classified as foreground.

   ii. Case 2: No match is found with any of the K Gaussians. In this case, the pixel is classified as foreground.

(g) Obtain a binary mask.

(h) The parameters are updated to make the next foreground detection.

More details can be found in [14, 15].

5. Kernel Density Estimation (KDE) [16, 17]

   In literature, there are different techniques that use a unique background to estimate the location of each pixel. In [16], they proposed a multimodal and original model based on the density estimation of observation of value pixel intensity. The KDE model can handle situations where the backdrop is cluttered and not completely static but contains small movements such as tree branches and bushes.

   KDE presents several advantages such as the ability to suppress detection of shadow using the color information. The KDE model can run in real-time for both gray level and color images. To perform this model, the researchers first estimate the probability density function using the Normal Gaussian function for each pixel. Second, they use the probabilities and a threshold to detect the foreground. Finally, they update the parameters to perform the next foreground detection by taking the intersection of the results of two background models: a short term one and a long term one, and extra false positives detection that occur in the long term model results. The only false positives detection that will remain will be rare events not represented in either model [17].

The combinations eliminate the persistence of false positives detection from the short term model and extra false positives detection that occur in the long term model results. The only false positives detection that will remain will be rare events not represented in either model 16.

6. Frame differencing

Frame differencing technique (figure (2.3)) is considered as the simplest background modeling techniques. It is based on two stages: The first one is to calculate the absolute difference between two or three consecutive frames in a sequence of images. The second one is to apply a thresholding function to determine the changes. The thresholding function depends on the speed of object motion. In some cases, the speed of the object changes notably that makes the quality of segmentation change. The differentiation of images consists very often of the development of holes in moving entities.

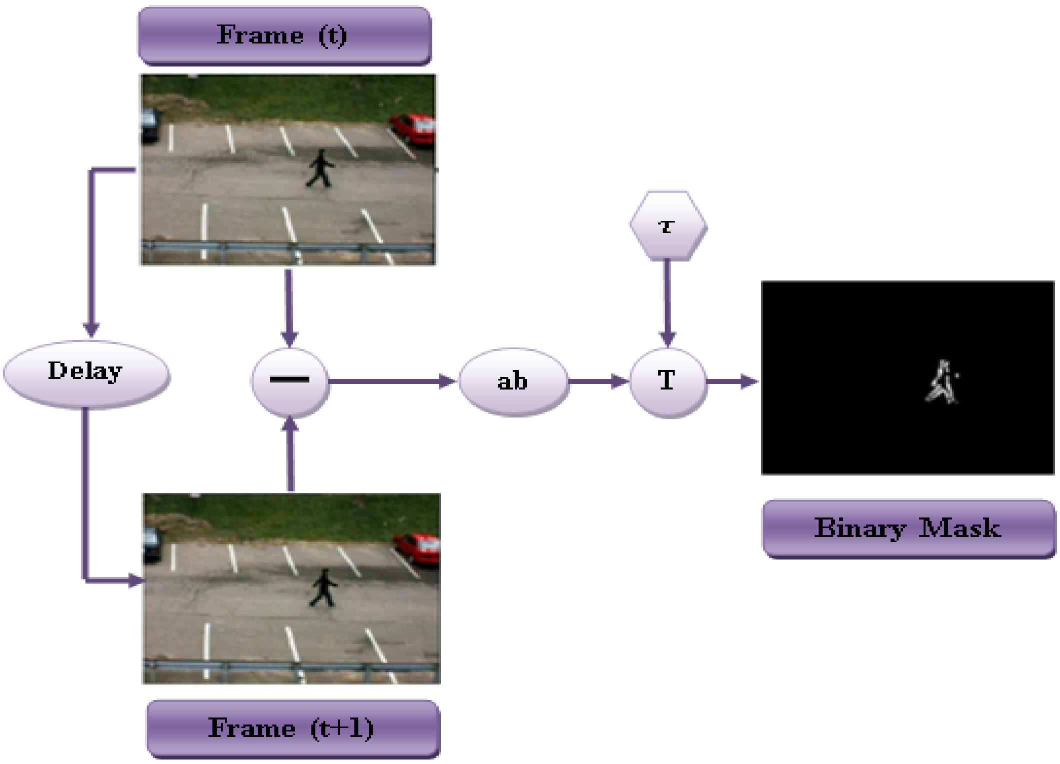


Figure 2.3: Frame differencing technique

2.4.1.C FOREGROUND DETECTION

Foreground detection is a classification task. This task labels the pixels of the scene as background or as foreground pixels by comparing the input video frame with the background model. Usually, to get this classification, the researchers have binarized the difference map by thresholding. The correct value of the threshold depends on the scene, the camera noise, and the illumination conditions [10]. There are various techniques used for foreground detection in literature such as absolute difference edge-Based, kernel density estimation, clustering-based, and single gaussian-based.

2.4.1.D DATA VALIDATION

In [11], Ahmed et al. defined data validation as the process of improving the candidate's foreground mask based on information from outside the base model.

This task aims to reduce the misclassifications that may occur in segmentation results. For instance, the misclassifications occurs when background regions are incorrectly classified as foreground (false positives) and the misclassifications occur when foreground regions are classified as background (false negatives).

2.4.2 Active contour methods

The active contour algorithms are useful methods discussed in literature for the segmentation of fast-moving objects. The active contour model is a deformable curve. This curve moves to the location of object edges. There are two difficulties to use active contour algorithms: The first is the starting problem and the second is how to catch the edge if the initial contour is far away from it [18].

The researchers proposed an automatic segmentation algorithm of moving objects based on information fusion. They used the frame difference method to get the initial contour of moving objects. Then, the fusion is made by active contour model with three types of information (motion information, region information (color), and edge-based image force). In [19], they proposed a method for detecting and tracking of moving objects. A gradient-based optical flow and an edge detector are used to create systems for line detecting and tracking object restoration; moreover, a line-based background subtraction of the previous frame is performed. Finally, the nearest-neighbor is used for clusting lines and contours of the clustered lines are extracted by using snakes. This approach satisfies requirement as a real-time system because the area of edge-based features is less than region based features. In [20] Zhengping et al. proposed a new technique that allows to segment automatically in the foreground using dynamic programming (deterministic decision-making process).

This technique consists of three phases. In the first one, they identified motion region by difference of frames. Motion edge of the current frame is extracted through employing Laplacian filter is done in the second phase. Finally, in third phase they used the thinning algorithm which emphasizes on the continuity to segment the contour of foreground region and used the thinning algorithm which emphasizes on color, intensity and gradient to segment the contour of background region.

In [21], a combination of optical flow and a modified active contour model are used for two dimension + time (2D + t) image segmentation. This method needs specialized hardware for real-time applications.

2.4.3 Optical Flow

Horn & Schunck defined the optical flow in [22] as the observed apparent motion of brightness patterns when a camera moves relative to captured objects. The optical flow is modelised as a vector motion field that describes the apparent velocity distribution of the brightness patterns in the image plan.

The optical flow is an ancient concept that was first computed for image sequences by Horn & Schunck in the 1980s [23]. It uses the flow vectors of moving objects .It is widely used to segment moving objects in video sequences. The area with motion field is considered object rest the other area is considered background. However, the main limitations of optical flow methods are the larg time of computation and high sensitivity to noise. Moreover, most of them cannot be executed in real time without specific equipment [8]. Nowadays, optical flow techniques are widely used [12] due to the availability of high speed computer.

The optical flow by it self is not enough to solve occlusions and sequential stops. In addition, these methods are very sensitive to illumination changes, shadows and noise [9].

2.4.4 Statistical techniques

Statistical techniques are widely used in the field of motion-based segmentation. They are the process of building a more advanced background subtraction. This can be done by comparing the statistical behavior of a small neighborhood for each pixel position in the difference image sequences. The comparison process is based on a meaningful test [24]. This technique is considered as a classification problem where each pixel must be classified in the background or in the foreground.

The statistical techniques work well with multiple objects and are able to handle noise, shadow, changes in lighting conditions and occlusions [25]. Many techniques are available in literature. Some of the statistical tests have been used such as the sum of absolute differences, the estimation of nucleus density and the simple Gaussian basis.

2.5 Datasets

Several datasets have been developed in the last years to evaluate and compare motion segmentation algorithms. These datasets cover specific challenges.

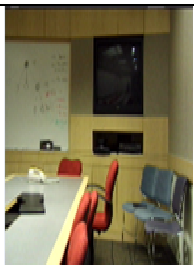
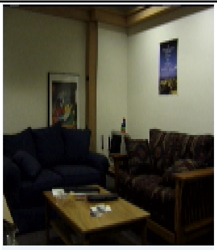
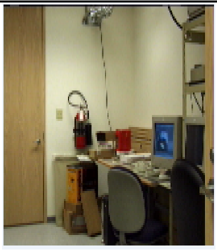
In this section, we mention the most recognized datasets as below:

2.5.1 Wallflower Dataset

It is available in [26]. The dataset of Wallflower was provided by Toyama et al (figure (2.4)). [27]. Wallflower is a test image sequences that represent the real-life problematic scenarios for developing and evaluating background subtraction.

Wallflower contains seven image sequences. Each one represent a specific challenge like illumination changes, background motion and shadow. Each image resolution is 160*120 pixels.

In the field of computer vision, Wallflower dataset is one of the most used datasets, but its main disadvantage is that there is only one ground- truth image per sequence [28].

Sequence	Moved Object	Time of Day	Light Switch	Waiving Trees
Test Frame				

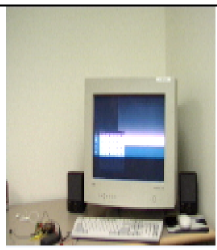
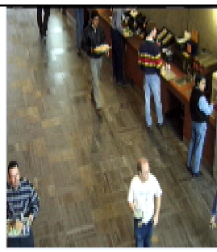
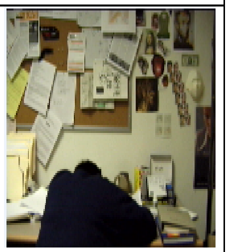
Sequence	Camouflage	Bootstrap	Foreground Aperture
Test Frame			

Figure 2.4: Example frames from Wallflower Dataset

2.5.1.A Moved Object

This video sequence represents a person walking in a conference room, performing different positions with a phone and a chair.

2.5.1.B Time of Day

This video sequence shows a person walking in a dark room and sitting on a couch for several minutes.

2.5.1.C Light Switch

This video sequence shows a room with the lights both on and off, and walking person.

2.5.1.D Waving Trees

This video sequence contains a person walking beside swaying tree. This scene presents multiple problems such as different lighting, shadows and a non-stationary background.

2.5.1.E Camouflage

This video sequence presents a person walking in the scene and occludes the monitor sits on a desk with rolling interference bars.

2.5.1.F Bootstrapping

This video sequence shows of walking people in different directions.

2.5.1.G Foreground Aperture

This video sequence shows a person sleeping at his desk from the back .

2.5.2 I2R Dataset

The dataset I2R24 provided by Lin and Huang [29] consists of nine video sequences (figure (2.5)). Each sequence presents multiple problems such as different illuminations, shadows, and non-stationary background. Each image resolution is 176 * 144 pixels.

This dataset has a twenty ground-truth masks taken when a critical situation occurs [30].

Figure 2.5: Example frames from I2R Dataset

2.5.2.A Office Environments

The video sequences shows laboratories, offices, meeting rooms, lobbies, corridors, and entrances. It contains two test sequences. It is composed of multiple problems that can be caused by shadows, changes of illumination conditions, camouflage foreground objects, waving curtains, running fans and flickering screens.

2.5.2.B Campus Environments

This video sequence shows campuses or parks containing vehicles of different colors. The dataset contains multiple Challenges such as changes in the background caused by motion of tree branches and their shadows on the ground surface, or changes in the weather and the tree shadows.

2.5.2.C Shopping Malls

This video sequence shows shopping centers, hotels, museums, airports, and restaurants. It contains three test sequences. In three video sequences, the lighting is distributed from the ceilings. In such cases, if multiple persons move in the scene, the shadows on the ground surface vary significantly in the image sequences [29].

2.5.2.D Subway Stations

This video sequence displays human crowds in subway stations with moving escalators and trains. In this sequence, the motion of background objects may make the detection of the background difficult.

2.5.2.E Sidewalks

In this sequence, the camera catches the images of the pedestrians all the day with a range of weather conditions. The sidewalks contains two test sequences.

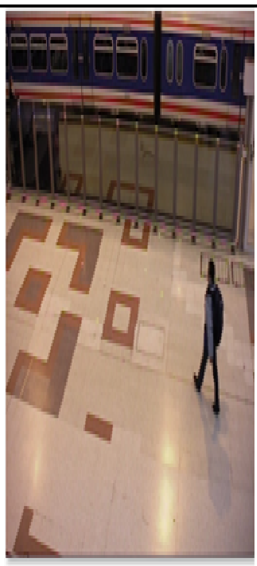
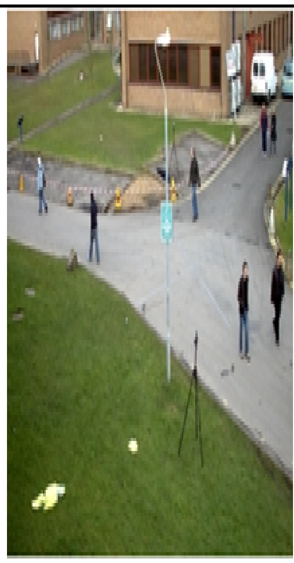
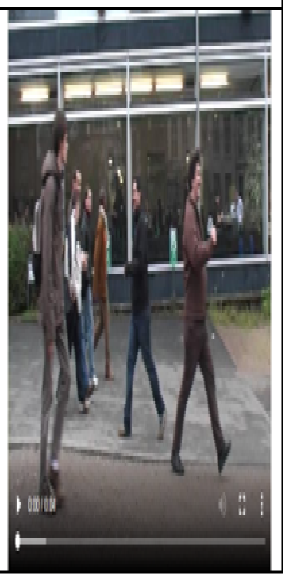
Sequece	PETS 2006	PETS 2009	PETS 2017
Test Frame			

Figure 2.6: Example frames from PETS Dataset

2.5.3 (Performance evaluation of tracking and surveillance (PETS) Dataset

The PETS program aims to solve the problem of evaluating visual tracking and surveillance algorithms. The program was held since 2000, in conjunction with the Institute of Electrical and Electronics Engineers (IEEE) Face and Gesture Recognition conference [31]. It was held for over than fifteen years (FG 00, CVPR'01, ECCV '02, ICVS '03, ICCV '03, ECCV'04, ..., AVSS'12, WVM'13, AVSS'14, AVSS'15, pets2016)(figure (2.6)).

2.5.4 CAVIAR2 Dataset

This dataset is available in [32]. It contains more than eighty staged indoor videos representing human activity such as walking, browsing, shopping, fighting, resting, slumping (figure (2.7)). The ground-truth for these sequences was done by hand-labeling the images.

Figure 2.7: Example frames from CAVIAR2 Dataset

2.5.5 ChangeDetection.NET Dataset (CD net)

This dataset is available in [33]. It contains a representation of the typical internal and external visual data captured in surveillance, smart environment, and video database scenarios. It is divided into two generations, the first generation is CD net of 2012 Dataset, and the second generation is CD net of 2014 Dataset (figure (2.8)).

The CD net of 2012 dataset was composed of nearly 90,000 frames in 31 video sequences.representing various challenges divided into 6 challenges as stated below:

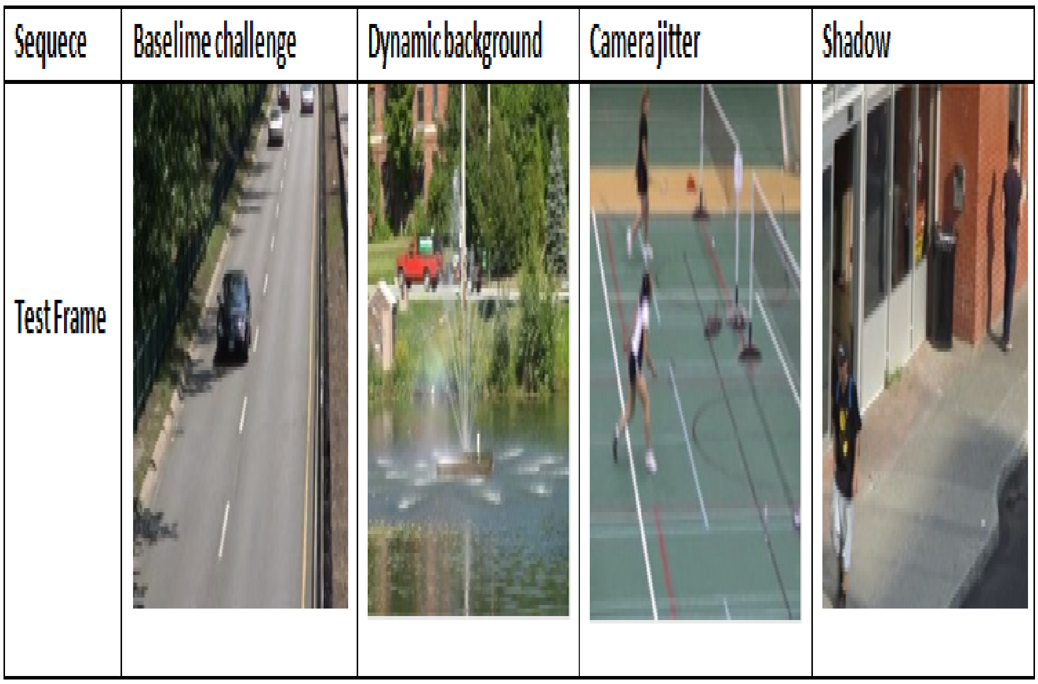


Figure 2.8: Example frames from ChangeDetection.NET Dataset

2.5.5.A Baseline challenge:

It contains four videos (two indoors and two outdoors videos).

2.5.5.B Dynamic background challenge:

It contains six videos depicting outdoors videos.

2.5.5.C Camera Jitter challenge:

It contains four videos captured with unstable cameras (one indoor and three outdoor videos).

2.5.5.D Shadow

It represents six videos (two indoors and four outdoor videos).

2.5.5.E Intermittent object motion challenge :

It contains six videos.

2.5.5.F Thermal challenge:

It represents five videos (three outdoor and two indoor videos). The CD net of (2014) dataset includes all the CD net of (2012) videos in addition to the following challenges: challenging weather, low frame-rate, acquisition at night, PTZ capture and air turbulence [28]. Each dataset includs accurate ground-truth segmentation.

2.6 Ground-truth ”gold standard”

The ground-truth can be defined as the correct answer for what precisely the motion segmentation algorithms is expected to produce [34]. It is designed manually to validate motion detection methods. Thus, without any doubt, the production of binary ground-truth images for camera captured videos is very difficult especially near moving object boundaries, partially-opaque objects, and in semi-transparent areas (dirty windows). In this case, the difficulty is due to the pixels in these areas which may be moving objects and background.

In the ground-truth mask, the background pixels assigned grayscale value of 0, and the foreground pixels are assigned grayscale value of 255. In the ground-truth mask, the shadow pixels should be considered as background pixels (figure (2.9)).

Figure 2.9: Examples of ground truth masks

2.7 Performance Measure Evaluation Methodology

In order to evaluate quantitatively the ability of the moving object segmentation algorithms, a region matching procedure is usually adopted between the results of the moving object segmentation and the ground truth mask [34].

Each comparison determines if we have:

2.7.1 True positives (TP):

Number of foreground pixels (pixels′ change) correctly detected.

2.7.2 False positives (FP):

Number of background pixels (no-change pixels) incorrectly detected as foreground known as false alarms.

2.7.3 True negatives (TN):

Number of background pixels (no-change pixels) correctly detected;

2.7.4 False negatives (FN):

Number of foreground pixels (change pixels) incorrectly detected as background as misses. Based on the above mentioned quantities, we mention three methods to evaluate quantitatively the moving object segmentation algorithms as below:

$$Precision = \frac{TP}{TP + FP} \tag{2.5}$$

$$Recall = \frac{TP}{(TP + FN} \tag{2.6}$$

$$F - measure = \frac{2 \times precision \times recall}{precision + recall} \tag{2.7}$$

By definition, for good detection the number of false positives and false negatives should be small, and the values of Precision, Recall, and F-measure should be high.

2.8 conclusion

In this chapter, we presented a review on the existing motion segmentation approaches with the aim of pointing out their advantages and drawbacks.

Background subtractions algorithms are based on dense representation of objects. It combines simplicity and good overall results able to deal with occlusions, multiple objects and non-rigid objects. The drawbacks of these algorithms are the difficulty to deal with sleeping foreground objects and with moving cameras. Furthermore, these algorithms are still very sensitive to illumination changes and noisy video.

The optical flow is theoretically a good technique used in order to segment motion. However, the OF by itself is not enough since it cannot help to solve occlusions and temporal stopping. Moreover, these methods are sensitive to noise and light changes.

Statistical techniques work well with multiple objects and are able to deal with occlusions and foreground aperture. The weaknesses are the difficulty to deal with illumination changes and moving background objects. Furthermore, most of the statistic approaches require some prior knowledge.

Considering the aspects emerged in this chapter, we conclude that all the approaches are able to segment the moving object but they are not so efficient in presence of challenge situations. For this risen we propose to combines color, texture and motion features to segment moving objects.

3

Chapter III. Review of different techniques related to feature extraction and segmentation

Contents

3.1 Introduction

As we have seen in Chapter II, we choose to use a combination of spatial and temporal features to segment moving objects. Thus, in the first section of this chapter, we will study the effective techniques of image features extraction .In the second section, we will review the clustering algorithms.

3.2 Feature extraction techniques

Feature extraction is a crucial step for image processing. As we known, the most image annotation and retrieval systems have been constructed based on visual features include color, texture, shape and motion [35]. It is employed to separate the visual information in images. Then, it stores visual information in the form of feature vectors to find the image information. The feature extraction technique is used to extract the appropriate features for object classification or detection .

There are various methods in the field that have been used to extract features such as color, texture, shape and motion as feature vector. The methods for feature extractions are classified as shown in Figure (3.1).

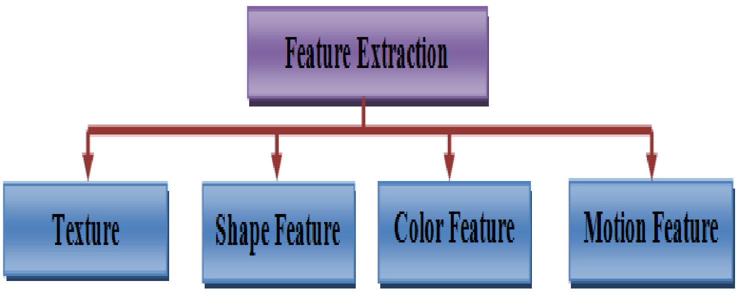


Figure 3.1: Classification of Feature Extraction Methods

These features are also classified as spectral features (color features), spatial features (shape features, texture features) and temporal features (motion features). Moreover, the feature extraction methods can be also classified as the low level feature and high level feature extraction.

In the low level features, we can extract from the automatically image small details of the image (point, line, edge, corner) without knowing the shape.

High-level features are based on low-level features to detect objects and larger shapes in the image [36].

3.2.1 Texture feature

Texture f0eature can only be measured from a group of pixels. However, it is an important way of detecting the features of images. Hence, they can be used for recognition and interpretation [37].

Texture feature can be classified into two categories. The first is spatial texture category. It is easy to use and can be extracted information from any shape; however, it is very sensitive to noise and distortions. The second category is spectral texture; it is robust and requires less computation. This category lacks the square of image regions [36].

In ancient texture techniques, features are extracted by calculating the statistical pixels or by finding the local pixel structures in the original image domain [38].

Studies on texture feature extraction shows a variety of techniques such as fractals [39–43], Grey Level Co-occurrence Matrix (GLCM) [44–46], and Autocorrelation based texture features [47]. Here, we focuse only on well known ones.

3.2.1.A Local Binary Patterns (LBP)

Ojala et al. [48] proposed a new version of the LBP method of Wang et al. [49]. They suggested using two level versions in order to represent the texture unit. The aim of the proposed method is to reduce the number of possible texture units to $2^8 = 256$ instead of 6561.

The LBP characterizes the spatial structure of a texture and presents the characteristics of being invariant to monotonic transformations of the gray levels [50].

So, the LBP operator is a gray scale invariant texture primitive statistic. It labels the pixels of an image region by comparing the neighborhood of each pixel with the center value and considering the result as a binary number (binary pattern). It creates an order based feature for each by comparing the intensity value of each pixel with that of its neighboring pixels.

The LBP operation at the location (x_c, y_c) can be defined from the following formulation:

$$LBP_{P,R}(x_c, y_c) = \sum_{p=0}^{7} S(g_p - g_c)2^p \tag{3.1}$$

Whereg_c corresponds to the grey value of the center pixel (x_c, y_c) of a local neighborhood and g_p to the gray values of spaced pixels on a circle P, and R corresponds to the radius. The function $S(g_p - g_c)$ is defined as follows:

$$s(g_p - g_c) = \begin{cases} 1 & \text{if } (g_p - g_c) \geq 0 \\ 0 & \text{if } (g_p - g_c) < 0 \end{cases} \tag{3.2}$$

Figure (3.2) shows framework of the LBP operator. Since correlation between pixels decreases with the distance, a lot of the texture information can be obtained from local neighborhoods. Thus, the radius R is usually kept small [50].

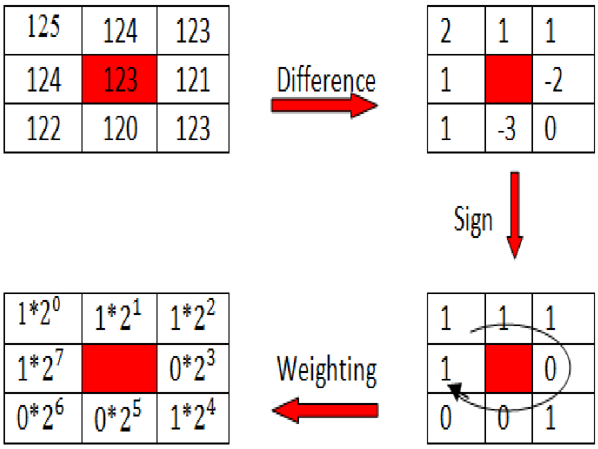


Figure 3.2: Illustration of the LBP operator

3.2.1.B Shannon entropy

Shannon's Entropy is one of the widely known texture feature. It is used to build a robust descriptor in order to classify occluded pixels in image sequences. Shannon was the first scientist who introduced the notion of entropy in the quantification of information [51].

The theory was based on the probabilistic modeling of the information containing messages [52]. Therefore, the entropy measures the randomness of intensity distribution from the neighborhood of each pixel $P(x, y)$ in the image.

3.2.2 Color features

Color feature extraction is one of the most important features technique. To extract color feature we should go through two essential steps. the first step is select an appropriate color space. the second step is to extract color feature from images or regions.

3.2.2.A Color space

There are several color spaces models discussed in the survey of [53], such as RGB, cyan, magenta, yellow and black (CMYK), Hue Saturation and Value (HSV), CIE- L*u*v*, and L*a*b Color space (CIE-LAB). Details of it can be found in [54].

1. RGB color space

   A normal grayscale image can only be defined by one matrix. In the RGB model, colored image consists of three different image plans (matrices), one in each of the primary colors, red, green and blue. The color is obtained by mixing the three primary colors.

   The RGB color space is very suitable space in image processing and signals for various purposes. However, RGB color space is not perceptually uniform and it not suitable in describing color in a way that is easily interpreted by human [55].

2. HSV color space

   The HSV has been developed to avoid illumination changes problems. So, color feature is suitable to handle problems with illumination changes. The HSV is the description of color space in cylindrical polar [56, 57], where Hue is the property of color that varies in passing from red to green. Saturation is the property of color that varies in passing from red to pink, and Value is the property that varies in passing from black to white [57].

$$H = arccos\frac{\frac{1}{2}((R-G)+(R-B))}{\sqrt{((R-G)^2+(R-B)(G-B))}} \tag{3.3}$$

$$S = 1 - 3\frac{\text{min (R,G,B)}}{(R+G+B)} \tag{3.4}$$

$$V = \frac{1}{3}(R+G+B) \tag{3.5}$$

Equations(3.3), (3.4), and (3.5) illustrate the transformation of the RGB color space to HSV color space.

3. CIE-LAB color space

The CIE-LAB is a color component space. It stands for Luminance, red to green, and blue to yellow. Commission International de l'Eclairage (CIE) stands for 'Commission International de l'Eclairage'. The vertical "L" axis represents Lightness and the horizontal "A" and "B" axes represent. The CIE-LAB color model encompasses the entire spectrum that includes colors outside of human vision [58, 59]. Equations (3.6), (3.7), (3.8) and (3.9) illustrate the transformation of the RGB color space to CIE-LAB color space [60].

$$L = 116f(\frac{Y}{Y_n}) - 16 \tag{3.6}$$

$$a = 500(f(\frac{X}{X_n}) - f(\frac{Y}{Y_n})) \tag{3.7}$$

$$b = 200(f(\frac{Y}{Y_n}) - f(\frac{Z}{Z_n})) \tag{3.8}$$

with:

$$f(t) = \begin{cases} \sqrt[n]{t}^3 & \text{if } t > \delta^3 \\ \frac{t}{3\delta^3} & otherwise \end{cases} \tag{3.9}$$

where: $X_n = 95.047$, $Y_n = 100.000$, $Z_n = 108.883$, $\delta = \frac{6}{29}$

3.2.2.B Extract color feature

There are different techniques used to extract the color features from images such as color histograms, color coherence vector, color moments based and color correlogram. More details concerning these techniques, are mentionned in [55, 61]. The well known color feature extraction techniques are presented as follows :

1. Color histograms

   Color histogram is successful and fast to describe color distribution features in any given image. A color histogram is a type of line graph which quantifies color space in different bins. Each color bin represents a particular color of the used color space. The number of bins depends on the number of colors in an image. In other words, the color histogram technique denotes how many pixels in an image of a particular color. This technique is effective in translational and rotational changes. Though a color histogram is simple to compute, it does not take into account spatial information of pixels. As a result, different images can have similar color histograms. In addition, the size of a histogram is generally very high.

2. Color Coherence Vector (CCV)

   Color Coherence vector is more sophisticated form of histogram because the color coherence vector incorporates spatial information into the basic color histogram [61].

   It classifies pixels as either coherent or incoherent. In this technique, the values of pixel are replaced by the average values in a small local neighborhood. Then, the classification of pixels as coherent or incoherent can be determined through computing connected components. The connected components are computed using four-connected neighbors within a given discredited color bucket. Consequently, the coherent pixels include those pixels which are spatially connected. The non- coherent pixels include those pixels that are isolated. However, the color coherence vector is more complex because of the high dimensionality and computation cost [62].

3. Color moments

   The color moment is the most simplest, compact, robust technique which can be used effectively to extract the color feature [55]. In addition, it offers minimum of storage. This technique is able to characterize the distribution of any color by different moments like mean (first order), variance (second order), skewness (third order), standard deviation and kurtosis. The different moments can usually be calculated separately for each color channel (components). The weakness of the different moments that they can not represent all the color information of an image, and are not able to use spatial information.

3.2.3 Shape features

Shape features are important visual features. They are one of the basic features used to identify ,recognize and describe the content of images. Shape features extraction play an important role in several domains like shape recognition, classification, approximation and simplification. The efficient shape features must respect some essential properties such as identify-ability, translation, rotation, scale invariance and affine invariance, occlusion invariance and noise resistance. More details are found in [63].

There are two main different classifications of shape extraction techniques in literature. The first one is Contour-based method, the second is region-based method.

The figure (3.3) gives an hierarchy of the classification of shape feature extraction approaches mentioned in [63].

3.2.4 Motion features (motion estimation)

The motion feature extraction techniques are based on patterns of motion. They are a representation of images by their location and motion direction. The choice of the picture's element is an important factor for motion estimation. This element may be a pixel, or a block. The size of the element determines the robustness to noise and precision. There are two basic approaches to motion estimation.

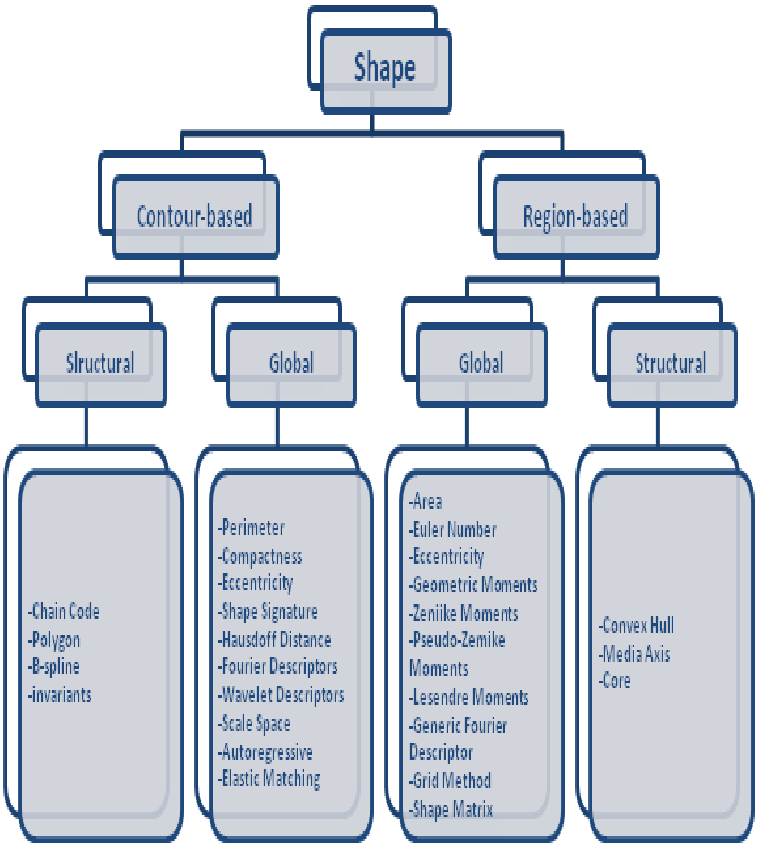


Figure 3.3: Classification of shape feature extraction approaches

3.2.4.A Pixel Based Motion Estimation (optical flow)

The optical flow is used to detect moving regions between two consecutive frames. It is a two dimensional motion dense field of displacement vectors, which describes the distribution of the apparent velocities of brightness patterns in image sequences.

The optical flow can also be defined as the 2D projection of the 3D motion of objects and surfaces in the image. The term dense indicates the flow vector. Most dense approaches are based on the linear formulation of Optical flow constraint (OFC), adding with another constraint called prior or regularization.

The formulation of OFC is given as follow

$$I(x, y, t) = I(x + u, y + v, t + 1) \tag{3.10}$$

Where $w = (u, v)$ are the two components of the optical flow vector; they refer to the horizontal and vertical displacement of pixel $p(x, y)$ from frame t to $t + 1$. The linearized version of OFC is expressed in equation (3.11)

$$I_x.u + I_y.v + I_t = 0 \tag{3.11}$$

Optical flow algorithms can be divided into three categories which are global, local and global-local techniques. Global techniques examine the full image and calculate the flow for all pixels at the same time [64]. The local techniques analyze each pixel individually and estimate how it has moved locally to calculate optical flow. The global-local techniques combine the above techniques.

1. Global techniques

   Horn & Schunk [23] developped a classical global optical flow algorithm. This algorithm relies on the brightness constancy assumption [64]. In this algorithm, they minimize the global energy function. This latter is composed from a quadratic norm of OFC and a smoothness constraint. The global energy function assumes that optical flow field varies almost everywhere in the image. The global energy functional for Horn & Schunk is written as:

$$HS(u, v) = \int ((\frac{\delta I}{\delta x}u + \frac{\delta I}{\delta y}v + \frac{\delta I}{\delta t})^2 + \alpha^2(\parallel \nabla u \parallel^2 + \parallel \nabla v \parallel^2))d_x d_y \tag{3.12}$$

   or

$$HS(u, v) = \int ((I_x u + I_y v + I_t)^2 + \alpha(\mid \nabla_u \mid^2 + \mid \nabla_v \mid^2))d_x d_y \tag{3.13}$$

   Where: I is the intensity of the image u and v is the flow in x and y direction from a pixel p at time t and α (usually $\alpha > 0$) is a weighting parameter that regularizes the balance between the OFC and the smoothness constraint. But we should note that a larger value of α can produce blurred contour. The solution of energy function is achieved by solving the corresponding Euler-Lagrange equation with an iterative implementation Gauss-Seiled.

$$D(x,y) = \begin{cases} \frac{\delta I}{\delta x}(\frac{\delta I}{\delta x}u + \frac{\delta I}{\delta y}v + \frac{\delta I}{\delta t}) - \alpha^2 \Delta u = 0 \\ \frac{\delta I}{\delta y}(\frac{\delta I}{\delta x}u + \frac{\delta I}{\delta y}v + \frac{\delta I}{\delta t}) - \alpha^2 \Delta v = 0 \end{cases} \tag{3.14}$$

The drawback of Horn-Schunk method is that discontinuities occur at the motion boundaries [24] because the smoothness of small movements are ignored.

3.2.4.B Block- Based Motion Estimation (Block matching methods)

These techniques are based on no linear formulation of optical flow constraint.

Unlike pixel-based Motion Estimation (ME) techniques, block-based ME techniques can be used alone.

Block matching methods can be classified in the category of matching primitive techniques.

The purpose of these methods is to estimate the motion vector between two successive frames. The algorithm divide each image frame into group of non overlapping macro-blocks of equal sizes. After that, each macro-block is compared with the corresponding block or its adjacent neighbors within a search window in the reference frame. This comparison is made to find the best matching block. The same principle is repeated for all macro-blocks in the whole frame.

To find a motion vector, we can use the previous frame as a reference frame, it is called backward motion estimation technique. In the opposite case, if the future frame is used as a reference frame, then it is called backward prediction or forward motion estimation technique.

Block matching techniques differ in four essential ways. The first is the search window, the second is the matching criterion, the third is the block size, and the fourth is the search strategy.

1. Search window

   The search window is an appropriate set of candidate vectors used to be selected in the motion vector. It represents the area where the most similar block will be searched. A search window is chosen as a rectangular area centered in the macro block of the reference image [65]. The choice of a search window structure has a huge impact on both complexity of the motion estimation algorithm and precision.

2. Matching criteria

   The matching criteria or the distortion function are used to find the best match of the current macro block in the search window in the frame of reference. A number of matching criteria are used as a block distortion measure (Block Distortion Measure (BDM)) for block matching motion estimation. The most popular used matching criteria for block motion estimation are:Mean Squared Error (MSE), Error Mean Absolute Difference (MAD), Peak to Signal Noise Ratio (PSNR) and Sum of Absolute Difference (SAD)given by equations (3.15), (3.16), (3.17) and (3.18) respectively.

$$MSE(i,j) = \frac{1}{N^2}\sum_{i=0}^{n-1}\sum_{j=0}^{n-1}(C_{ij} - R_{ij})^2 \tag{3.15}$$

$$MAD(i,j) = \frac{1}{N^2}\sum_{i=0}^{n-1}\sum_{j=0}^{n-1} \mid C_{ij} - R_{ij} \mid \tag{3.16}$$

$$PSNR + 10log_{10}\frac{(peaktopeakvalueofdata)^2}{MSE} \tag{3.17}$$

$$SAD(i,j) = \sum_{i=0}^{n-1}\sum_{j=0}^{n-1} \mid C_{ij} - R_{ij} \mid \tag{3.18}$$

   Where N×N is the size of macro-block, C_ij and R_ij are pixel values compared to current and reference macro-block respectively. In literature, MAD, MSE mostly used. To finding image quality the PSNR are the used.

3. Macro-block size

   In motion estimation algorithm, the macro block size is an important parameter for motion vector computation. tIf the block size becomes smaller then accuracy effect problem is reduced. To minimize the possibility that the macro block contains different objects moving in different directions, researchers choose a smaller block size. As a result, a smaller macro-block size lead to more blocks per frame and more computing time.

4. Search strategy

   There are many search strategies. Each one aims to maximize the efficiency and the speed of this phase such as Exhaustive Search (ES), Three-Step (TSS), New Three-Step (NTSS), Four-Step Search (FSS), Diamond Search Algorithm (DS), and Hexagon Based Search Algorithm (HEXBS). In this sub-section, we concentrate on the exhaustive search.

   The exestive search algorithm is also known as full search. It is a simple block matching algorithm. Full search strategy compares the current block with all the candidate blocks of the reference frame within the search area. It consist of finding out the best matched block in the reference frame. The comparison is made by calculating the sum absolute difference (SAD) value MSE (Mean Squared Error) formula at each possible location in the search window. The full search is the most computationally expensive block matching algorithm. because we have to check all the candidates within the search windows and its computation complexity makes it difficult to be implemented in real time.

   The previous step extracted the set features. The following step combines the extracting features as a single feature vector. This step requires integrating those features into groups. Each group consists of objects that are similar between themselves and dissimilar to objects in other groups.

   For this reason, we will describe the clustering algorithms.

3.3 Clustering algorithms

Cluster analysis is one of the most important steps in data exploratory analysis. It consists of groups of data objects based only on information found in the data, whereas the data describes the objects and their relationships. The aim of clustering is to group similar objects, and separate them from different objects (or unrelated to). In some sources, the clustering can be considered as a form of classification. The most popular and useful clustering algorithms are: K-Means clustering, Expectation-Maximization (EM) clustering, graph cuts and mean-shift clustering.

3.3.1 K-Means Clustering

K-means clustering is the simplest and most used clustering algorithm. It is an iterative process that divides a given data set into K disjoints groups (clusters), where the value of K must be provided by the user [66]. Let $X = X_1, \ldots\ldots\ldots, X_N$ be a set of N image pixels, and let $V(X_i)$ denotes the feature vectors associated with a pixel X_i. The K-means algorithm has the following steps:

(a) Parameter Initialization:

The points are assigned to the initial centroids which are all in the larger group of points. In the classic K-means algorithm, the means of each of the K clusters are initialized as centroid.

(b) Hard Assignment of Pixels to Clusters:

At this step, each pixel X_i is associated with a single cluster. This is made by computing the mean of each clusters, then each pixel X_i is assigned to the centroid with the closest mean. The euclidean distance computes the distance between two vectors features.Then, the centeroids are updated.

(c) Parameter Recomputation

The means of the clusters are now computed again based on the feature vector values of all pixels in each cluster (group). The above steps ((b)and (c)) are repeated respectively until convergence occurs when no pixel moves from one cluster to another in a given iteration [67].

3.4 Conclusion

In this chapter, we presented the most common techniques for extracting image features. In addition, we explored the different clustering algorithms that we can be used in our proposed approaches.

4

Chapter IV. Moving object segmentation based on feature extraction with static background

Contents

4.1 Introduction

In this chapter, we will describe a new, and simple technique. This technique is able to segment the moving objects in a complex scene (shadow, partial occlusion).

Also, the obtained results of our experiments will be discussed in details. These experiments include challenge databases. Each experiment has specific discussions. Furthermore, we will compare the obtained results with the existing state of art.

The aim of our proposed technique is to extract groups of pixels from static background. This can be done by regrouping both motion and some color and texture features. The proposed algorithm is divided into two essential steps. The first step is feature extraction. The second step is object segmentation. A block diagram of the overall system is represented in Figure (4.1).

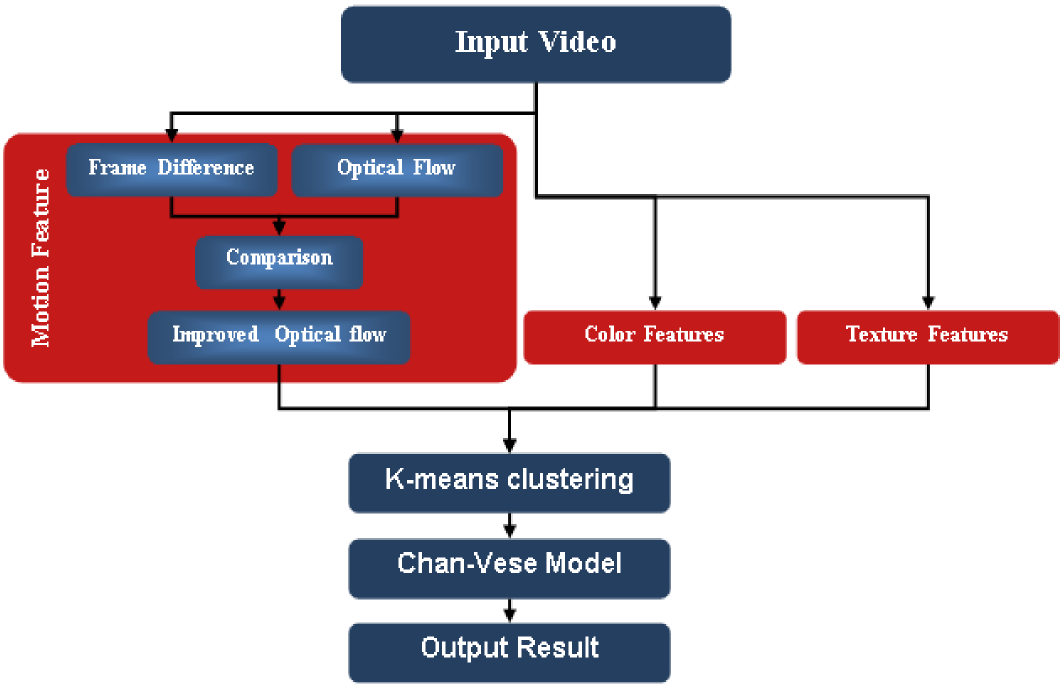


Figure 4.1: Proposed moving object segmentation framework

4.2 Features extraction

Extracting features of the image has become popular tools in the computer vision and image processing. They are applied widely in a large number of applications such as image classification, object recognition, robot localization, motion detection and tracking. We can define the features as a piece of information (key points) used to describe the image.

Generally, there are two methods of image features which are global features and local features [50]. The global features are based on the property interpretation of all pixels of image. This property can be color histograms, texture, or edge. The local features are based on the detection of salient regions (pixels) in image and describe them [50]. The goal of our proposed method is to group pixels with similar motion, color and texture features into clusters as illustrate in figure (4.8).

4.2.1 Motion features

In this subsection, we describe our essential approach to motion feature extraction. Motion Estimation (ME) algorithm plays a vital role in video processing. The different methods of ME are available in literature such as Lucas & Kanade [68], Horn & Schunk [23], and block-matching [69].

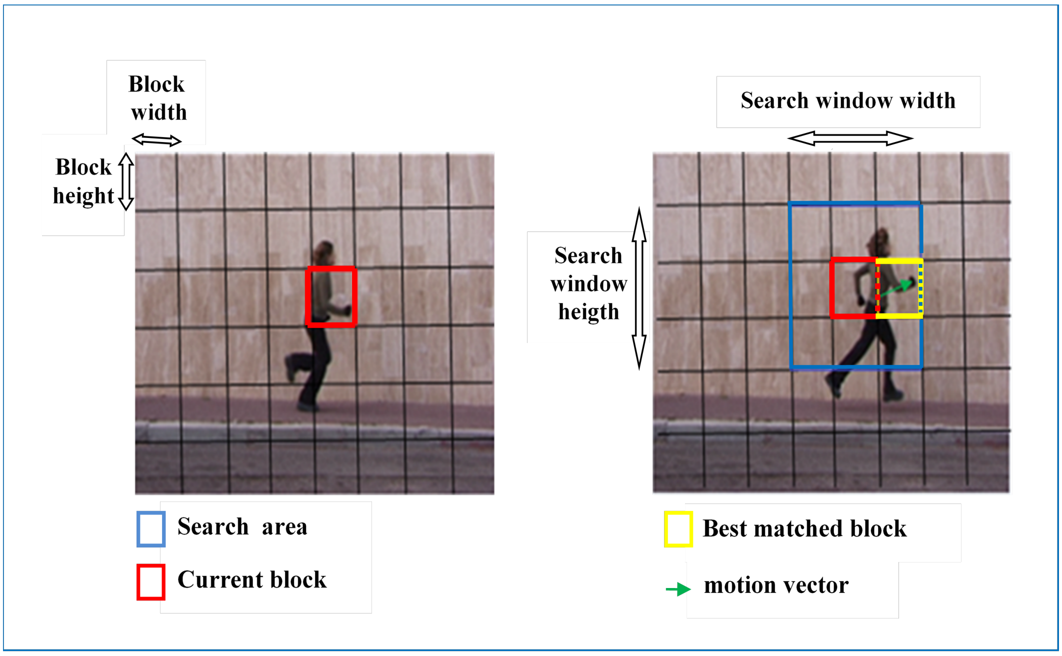


Figure 4.2: Block matching full search diagram

We have focused on the block-matching algorithm (figure (4.2)). However, these algorithms are more expensive in competition, but we aim to obtain a dense and very accurate estimation in all images.

The brightness constancy equation is only valid for small motions. In order to extract the large motions, we make a comparison between the various searching algorithms [70–74]. After comparison we found that the full search block matching algorithm gives more optimal solution, best performance, low control and the highest Peak-Signal-to-Noise-Ratio (PSNR) [75]. The full search block matching algorithm is described in subsection (3.2.4.B).

Let (MV_x, MV_y) is the motion vector of each pixel P(x,y) (figure (4.3)).

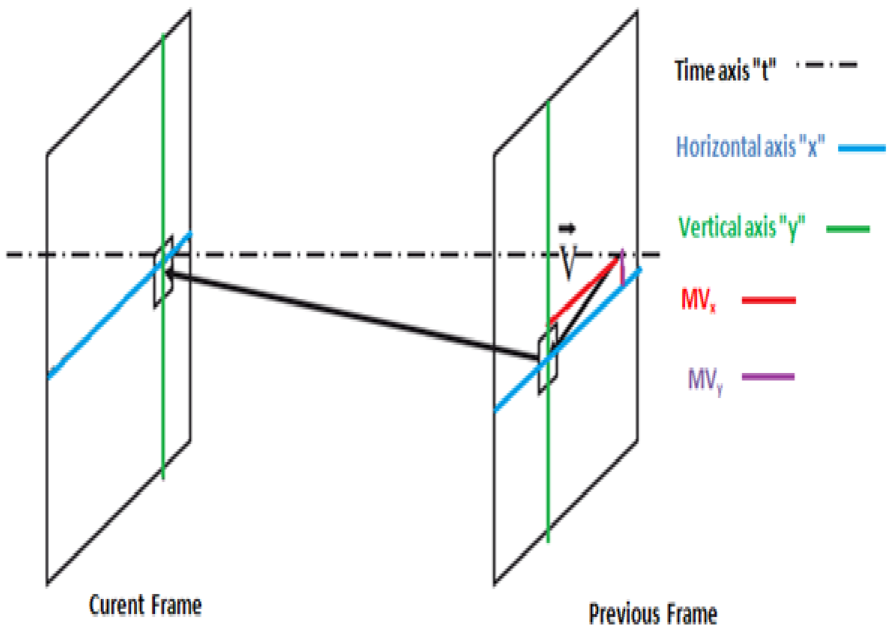


Figure 4.3: Complex presentation of the motion vector

The motion vector can also be defined by its magnitude (V) and direction (θ) as:

$$V(x, y) = \sqrt{(MVx)^2 + (MVy)^2} \tag{4.1}$$

$$\theta(x, y) = tan^{-1}(\frac{MVy}{MVx}) \tag{4.2}$$

The motion vector is smooth and reliable in moving objects. However, it is erratic and unreliable at the object boundaries [76]. Furthermore, it is inherently sensitive to illumination changes and shadows. To solve these problems, we have proposed an adaptive concept that combines the magnitude of the motion vector and frame difference images. The Frame difference image is given by:

$$D(x, y) = \mid f_{t+1}(x, y) - f_t(x, y) \mid \tag{4.3}$$

where $f_t(x, y)$ and $f_{t+1}(x, y)$ are the current and the previous frames, respectively. The result of this step is extracting moving objects in image with an open contour. To get binary mask and remove the holes, we use the binary threshold. The binary threshold image is defined as follows:

$$\begin{cases} 1 & \text{if } D(x, y) > T_D \\ 0 & otherwise \end{cases} \tag{4.4}$$

T_D denotes a pre-specified threshold [77], experimentally evaluated. To validate values of displacements,we made a comparison between the magnitude$V(x, y)$ of each pixel $p(x, y)$ and the frame difference image$D(x, y)$ as:

$$\begin{cases} V(x, y) & \text{if } D(x, y) = 1 \\ 0 & otherwise \end{cases} \tag{4.5}$$

A block diagram of the improved optical flow algorithm is shown in Figure (4.4).

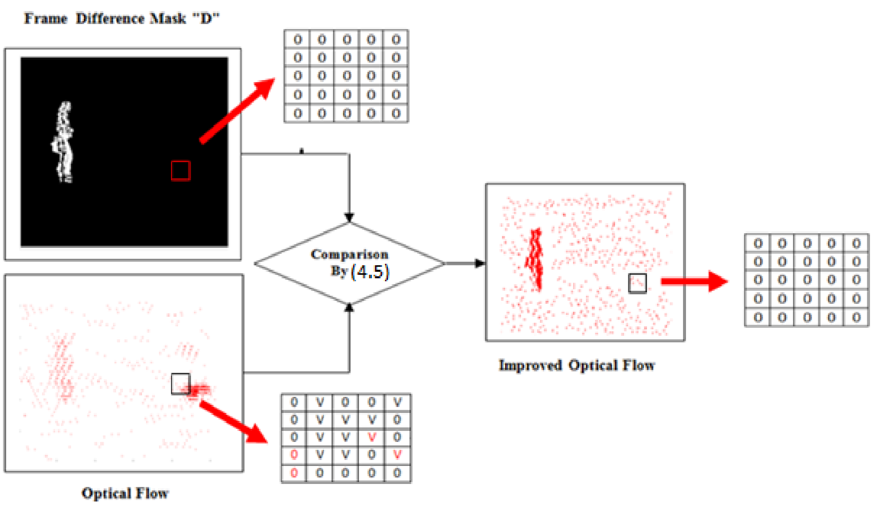


Figure 4.4: Concept of the improved optical flow algorithm combining frame difference image and optical flow

Then, we normalize the magnitude of optical flow value [78], as following: Firstly, we calculate the maximum (V_{max}) and the minimum (V_{mni}) of optical flow value using equations (4.6) and (4.7):

$$V_{max} = max(V(x, y)) \tag{4.6}$$

$$V_{min} = min(V(x, y)) \tag{4.7}$$

then we normalized the magnitude of optical flow value by $V_{nor}(x, y)$:

$$V_{nor}(x, y) = int(\frac{V(x, y)}{V_{max} - V_{min}} * 256) \tag{4.8}$$

As illustrated in figure (4.4), errors of the optical flow motion estimation are successfully removed.

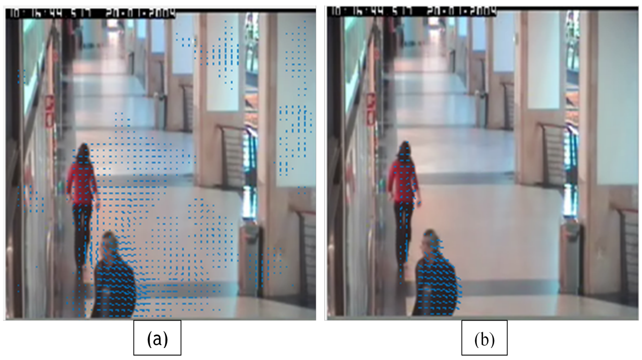


Figure 4.5: (a) Classical optical flow field and (b) Improved optical flow field by [2]

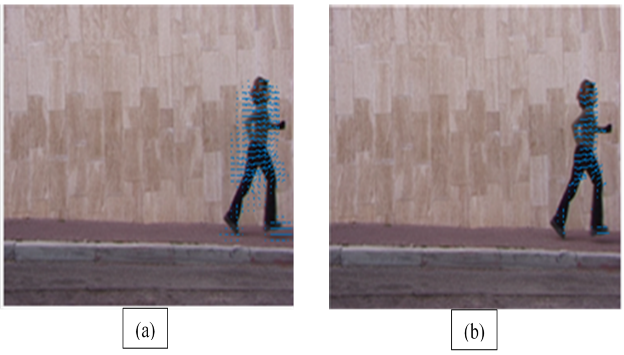


Figure 4.6: (a) Classical optical flow field and (b) Improved optical flow field by [2]

Each part of the human body has a specific direction during the movement, which causes an inconsistency in object motion. In order to correct this problem, we classified the motion directions into four orientations, instead of nine [77].
Specifically, let $\{\theta_1 = 0, \theta_2 = 90, \theta_3 = 180, \theta_4 = 270\}$ be the four existing orientations, and $\theta(x, y)$ the direction of each pixels, then the classified orientation is determined as:

$$\theta(x, y) = \theta_i, if \theta_i - \frac{n}{2} \leq \theta \leq \theta_i + \frac{n}{2} \tag{4.9}$$

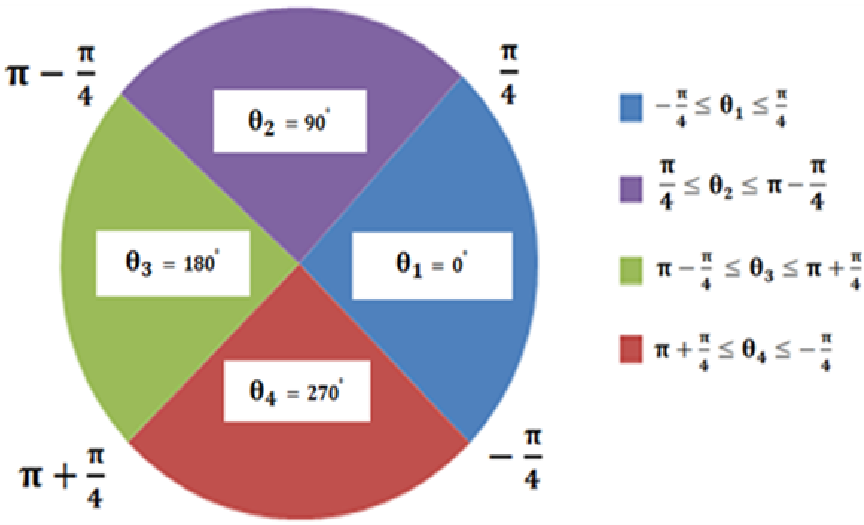


Figure 4.7: Direction classification of each pixel in four regions

Figure (4.7) explains how the classification algorithm works. When each pixel in the frame is processed in this way, we extract motion features presented by its improved magnitude and its classified direction. However, motion vector alone is not sufficient since it can not help to solve occlusions, shadows, and illumination changes problems. That is why, we have extracted the color and texture features to remedy the problem of partial occlusions and illumination changes.

4.2.2 Color features

Most of our test sequences include human skin. The use of the color space feature is suitable to describe the property of human skin. The RGB color space is the default color space for most available image formats. Any other color space can be obtained from a linear or non linear transformation from RGB [79]. RGB corresponds to the three primary colors: red, green and blue, respectively. Moreover, during our experiments, we observed that the RGB color space does not give the best representation of images. It should be pointed out that they are sensitive to illumination changes, and the mixing of chrominance and luminance data. So, we use the HSV color space to avoid illumination changes problems (subsection (3.2.2.A)). Thus, color feature is suitable to deal with illumination changes problems.

The color is insufficient because the contrast between an object and the background can be small. Therefore, we have used the texture feature with the color feature.

4.2.3 Texture features

In motion- based segmentation, a set of occluded pixels may remain unclassified. This is due to the fact that during occlusions, the optical flow is erratic and unreliable at object boundaries.

However, the texture features can solve this problem. Texture features represent important characteristics used in image analysis. It is useful to perform visual decision tasks such as detection and recognition.

Shannon's Entropy is one of the most popular texture features used to build a robust descriptor in order to classify occluded pixels in image sequences. E. Shannon was the first who introduced the notion of entropy in quantifying the information [80]. It is able to extract the same features of the same scene or object repeatedly under variety of viewing conditions. The theory was based on the probabilistic modeling of the information containing messages [79]. Therefore, the entropy algorithm measures the randomness of intensity distribution from the neighborhood of each pixel P(x,y) in the image; in our work, we chose the neighborhood size of 5×5.

4.2.4 k-means clustering Method

To attract and control the evolving contour towards a moving objects [81], we proposed a combination of feature vectors of each pixel in order to organize isolated, sensible, and compact clusters. The four feature vectors considered are magnitude, direction, texture and color.

K-means algorithm is a widely used technique for clustering, understanding and learning. In our method, we measured the similarity by minimizing the squared distances between all points and the center of cluster using the euclidean distance [82].

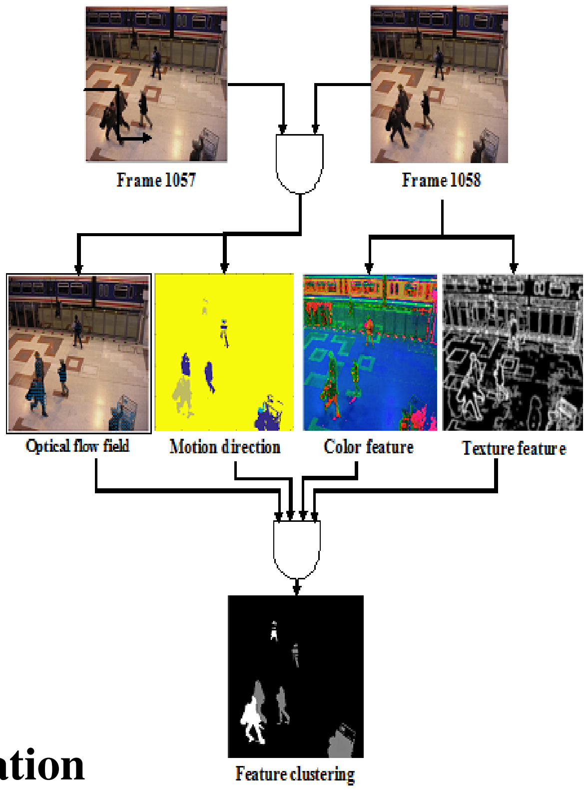


Figure 4.8: Feature detection and clustering steps

4.3 Segmentation

This step is necessary to distinguish zones where there are homogeneous temporal and spatial features from the sequence. We need an active contour model where there is no need to smooth the initial image, even if it is noisy. Also, it can detect objects with boundaries that are not necessarily defined by gradients. Moreover, the initial curve does not necessarily start around the objects to be detected. Thus, we chose the active contour model without edges. The model combines between mean curvature motion techniques and the Mumford-Shah model [83]. The aim of the algorithm is to minimize the energy function $F(c_1, c_2, C)$

defined by:

$$
\begin{aligned}
F(c_1, c_2, C) = \\
& \mu.length(C) + v.Area(inside(C)) \\
& + \lambda_1 \int_{inside(C)} \mid \mu_0(x,y) - c_1 \mid^2 \\
& + \lambda_2 \int_{outside(C)} \mid \mu_0(x,y) - c_2 \mid^2 dxdy
\end{aligned}
\tag{4.10}
$$

where C is the boundary of a closed set, c_1 and c_2 are averages of the image I. $\lambda_1, \lambda_2, v, and \mu$ are positive constants. In equation (4.10), the first term controls the regularity by penalizing the length. The second term penalizes the enclosed area of C to control its size. The third and fourth terms penalize discrepancy between the piecewise constant model μ and the input image (I) ([81, 84–86]. In our approach, we incorporate the norm of the obtained clusters on Chan-Vese model [85] in order to attract the evolving contour to a moving object contour.

4.4 Results and discussions

To evaluate the quantitative and qualitative performances of our technique, we used Matlab software. Our approach has been tested with several indoor and outdoor databases. The database is relatively long and umbra and penumbra shadows are cast by multiple foreground objects. In order to evaluate the performance of the proposed approach in a quantitative and qualitative way, ground-truth segmentation masks were generated by manual segmentation.

4.4.1 Qualitative Analysis

The first sequence considered in our experiment comes from the context aware vision using Image-based Active Recognition database (CAVIAR) database [87], which shows people walking at different distances from the camera.

This video presents extensive shadows and local illumination problems. Our technique removes the shadow from the foreground regions of frames as illustrated in figure (4.9, 4.10).

Figure 4.9: (a) Input image.(b) Optical flow field.(c) Result of k-mean clustering algorithm.(d) Motion segmentation

However, some parts of people are not accurately segmented due to the camouflage problem in intensity caused by resemblance between people and background.

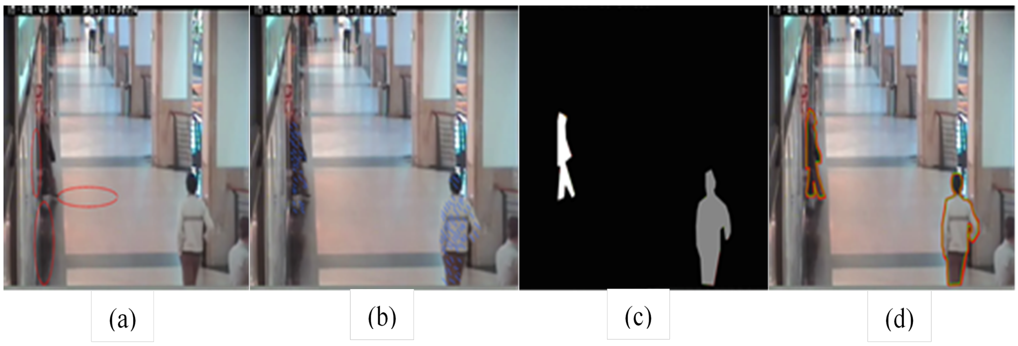
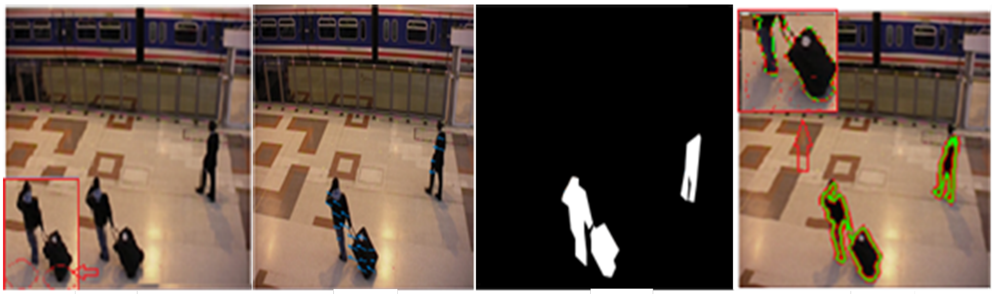


Figure 4.10: (a) Input image.(b) Optical flow field.(c) Result of k-mean clustering algorithm.(d) Motion segmentation

The second sequence is the pets 2006 [88] dataset shows walking pedestrians in opposite and (or) same directions with different color appearance, and at different distances from the camera. This scene presents multiple problems such as different sources of illumination; the shadows which are reflected over the floor , dark, and light camouflages, etc. in figures (4.11, 4.12), the proposed method removes the shadow of the object and successfully segments the multi-direction object regions. In figure (4.12), we can see how the mother and her daughter presented in the scene are correctly segmented in spite of similarity between people and background. Furthermore, the illumination changes.



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

Figure 4.11: (a) Input image.(b) Optical flow field.(c) Result of k-mean clustering algorithm.(d) Motion segmentation

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

Figure 4.12: (a) Input image.(b) Optical flow field.(c) Result of k-mean clustering algorithm.(d) Motion segmentation

Our algorithm is robust to the updating system which has the sleeping foreground object and the foreground aperture problems (figure (4.13)).

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

Figure 4.13: (a) Input image.(b) Optical flow field.(c) Result of k-mean clustering algorithm.(d) Motion segmentation

In figures (4.14) , (4.15) we test the case of occlusions. The proposed method can manage the occluded persons using the direction classification (figure (4.14.c) and figure (4.15.c)).

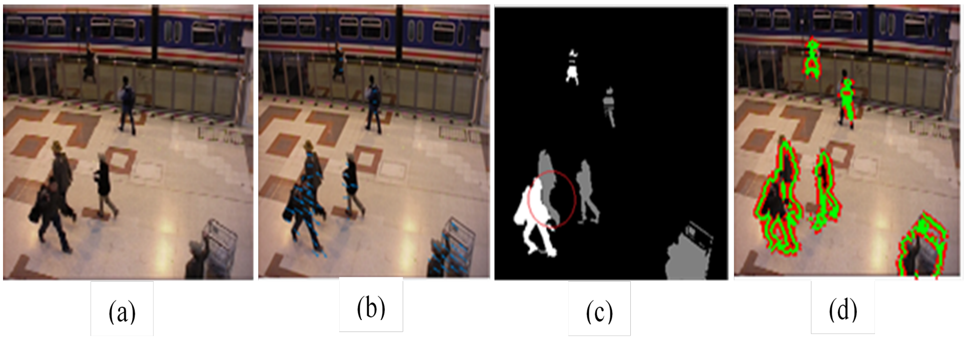


Figure 4.14: (a) Input image.(b) Optical flow field.(c) Result of k-mean clustering algorithm.(d) Motion segmentation

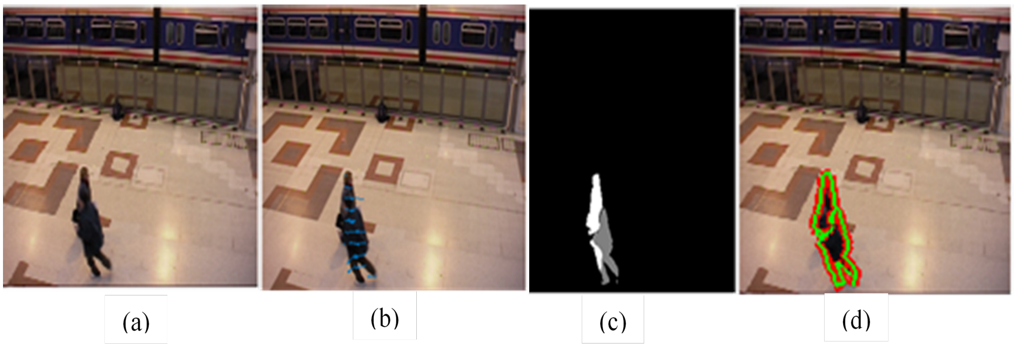


Figure 4.15: (a) Input image.(b) Optical flow field.(c) Result of k-mean clustering algorithm.(d) Motion segmentation

The third sequence dataset called the Caviar Fight one Man Down [89] is very noisy. It illustrates pedestrians meeting others.

In figures (4.16) and (4.17), we observed the strong illumination changes caused by the sunshine through the glass which gives a wrong background. The proposed method segments successfully correct object regions even with strong illumination changes.

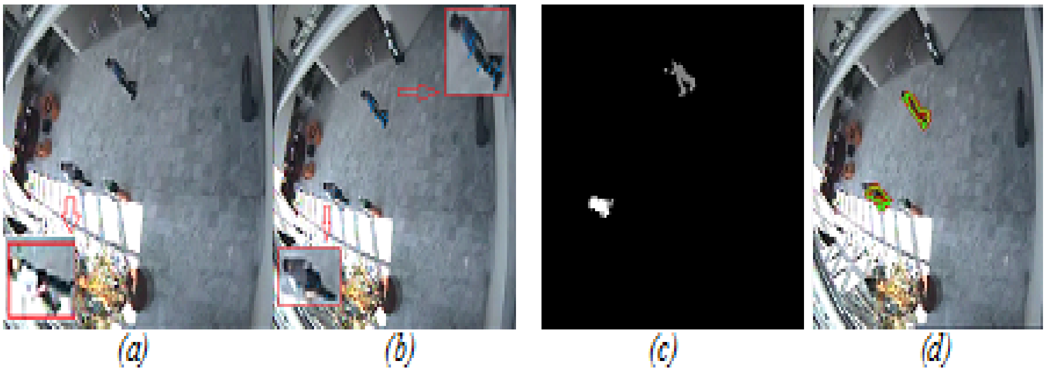


Figure 4.16: (a) Input image.(b) Optical flow field.(c) Result of k-mean clustering algorithm.(d) Motion segmentation

In the fourth sequence, we used the ATON dataset [90]. Figure (4.18) shows the result of segmentation of the intelligent room. In figure (4.18.d), we see how our technique is able to successfully segment both parts of men despite of low image quality and chair occluding.
Figure (4.19) illustrates foreground segmentation results from ATON Campus [90] sequence. The people are successfully segmented despite of blurred image problems, strong shadows and sign occlusions.

In the fifth sequence, we used change detection.net dataset (CDNET) [?]. Figures (4.20), (4.21) and (4.22) show the segmentation results of an office video sequence which comes from baseline category. The video sequence shows challenging aspects due to illumination changes, and slow motion. The moved parts of the person are correctly detected and segmented.

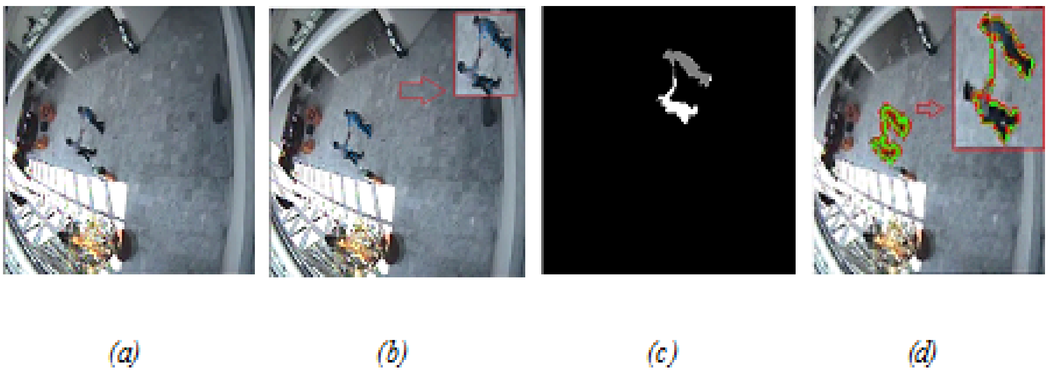
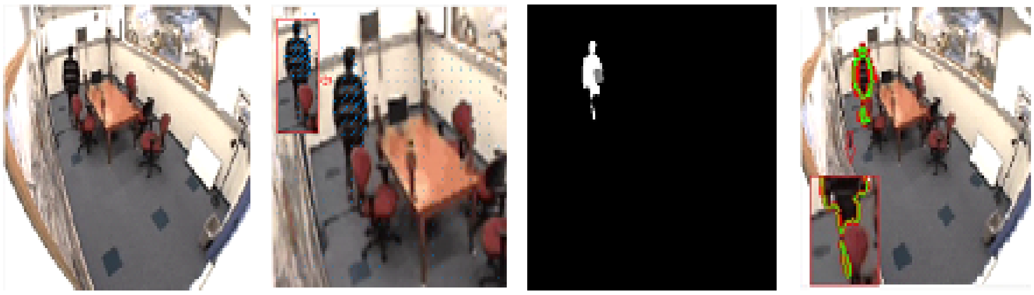
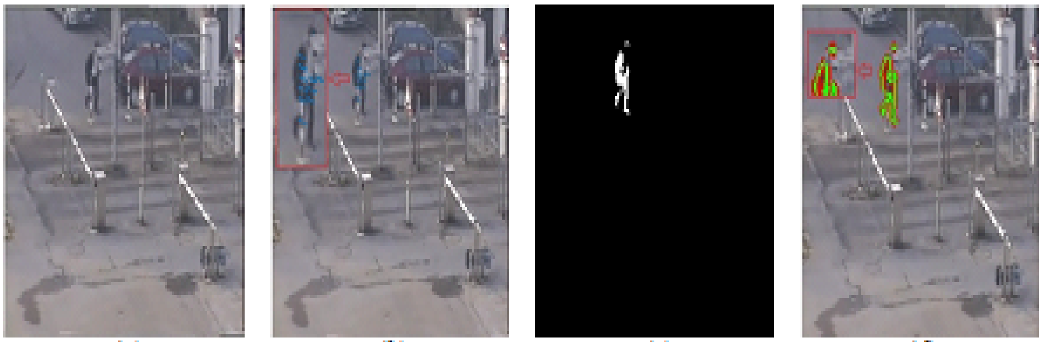


Figure 4.17: (a) Input image.(b) Optical flow field.(c) Result of k-mean clustering algorithm.(d) Motion segmentation



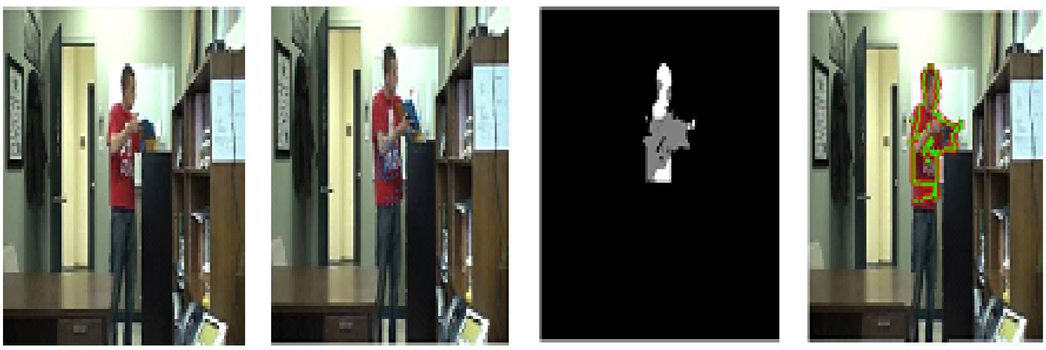
(a) (b) (c) (d)

Figure 4.18: (a) Input image.(b) Optical flow field.(c) Result of k-mean clustering algorithm.(d) Motion segmentation



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

Figure 4.19: (a) Input image.(b) Optical flow field.(c) Result of k-mean clustering algorithm.(d) Motion segmentation



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

Figure 4.20: (a) Input image.(b) Optical flow field.(c) Result of k-mean clustering algorithm.(d) Motion segmentation

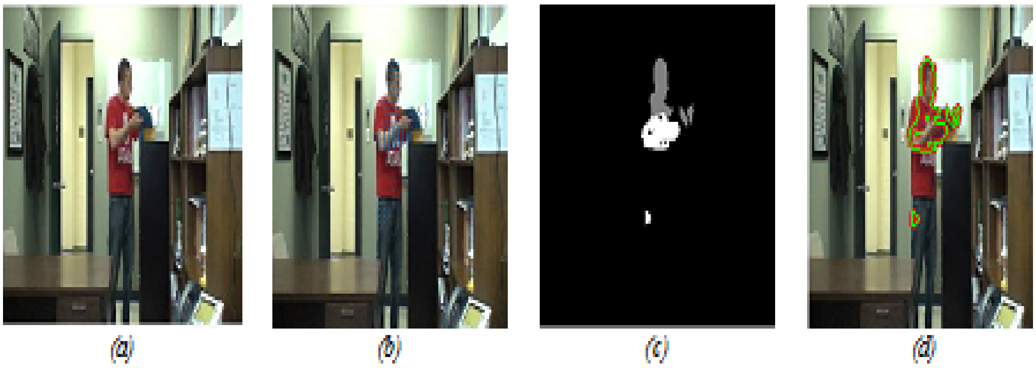


Figure 4.21: (a) Input image.(b) Optical flow field.(c) Result of k-mean clustering algorithm.(d) Motion segmentation

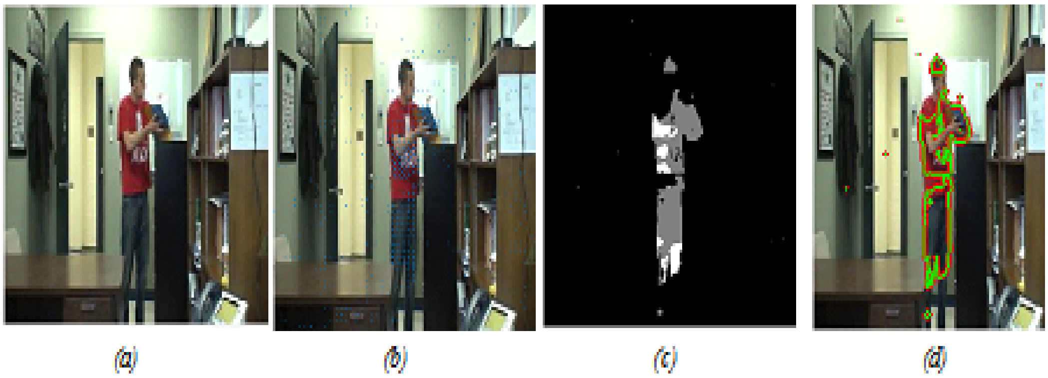


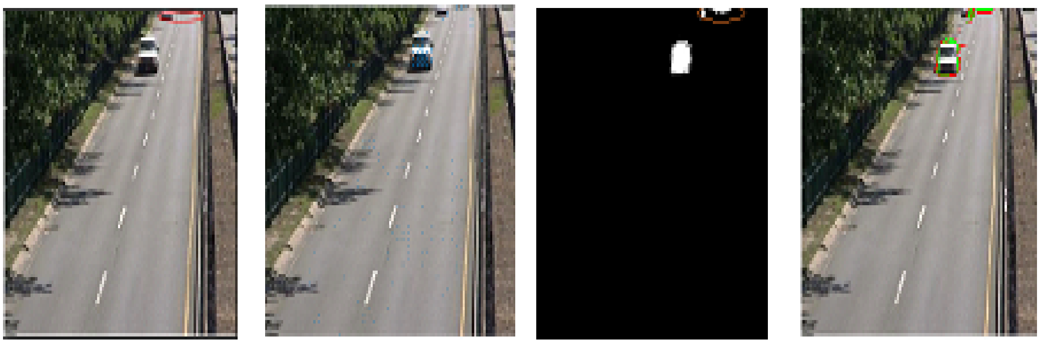
Figure 4.22: (a) Input image.(b) Optical flow field.(c) Result of k-mean clustering algorithm.(d) Motion segmentation

Figures (4.23), (4.24), and (4.25) show some significant detection results from the highway video sequence which belongs to baseline category. The cars in the sequence are accurately segmented despite of the reflected shadows over the road. The inserted foreground is successfully segmented, (figures (4.24) and (4.25)).

Significant video frames are illustrated in figures (4.26), (4.27) and (4.28). In these figures, we can see how the women is successfully segmented despite of image low quality which presents a noise, shadows and blurred image.

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

Figure 4.23: (a) Input image.(b) Optical flow field.(c) Result of k-mean clustering algorithm.(d) Motion segmentation



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

Figure 4.24: (a) Input image.(b) Optical flow field.(c) Result of k-mean clustering algorithm.(d) Motion segmentation

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

Figure 4.25: (a) Input image.(b) Optical flow field.(c) Result of k-mean clustering algorithm.(d) Motion segmentation

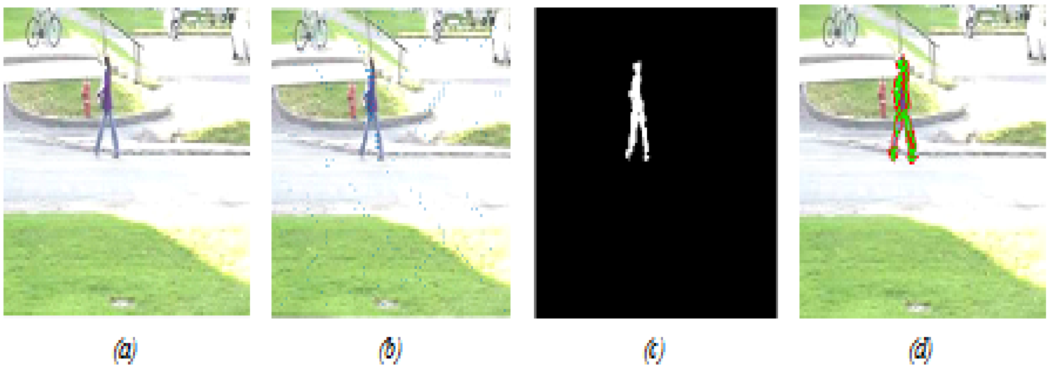


Figure 4.26: (a) Input image.(b) Optical flow field.(c) Result of k-mean clustering algorithm.(d) Motion segmentation

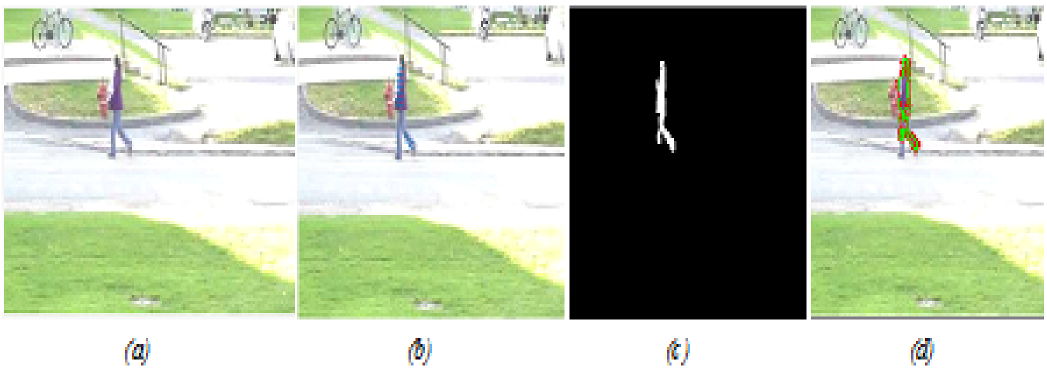


Figure 4.27: (a) Input image.(b) Optical flow field.(c) Result of k-mean clustering algorithm.(d) Motion segmentation

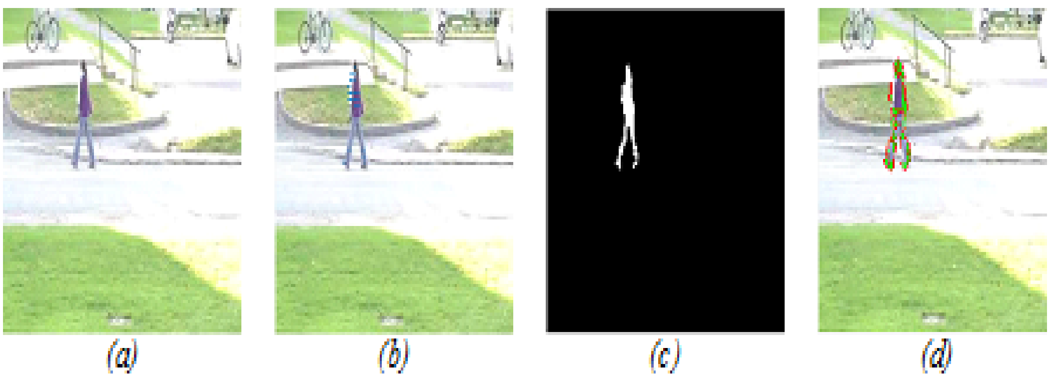


Figure 4.28: (a) Input image.(b) Optical flow field.(c) Result of k-mean clustering algorithm.(d) Motion segmentation

Figures (4.29), (4.30) and (4.31) show some significant detection results from the copy machine indoor video sequence. This video sequence comes from shadow category .It contains videos with prevalent hard and soft shadows and intermittent shades. The person is correctly segmented despite of the dark and light camouflage resulted from the curtain and the shadow.

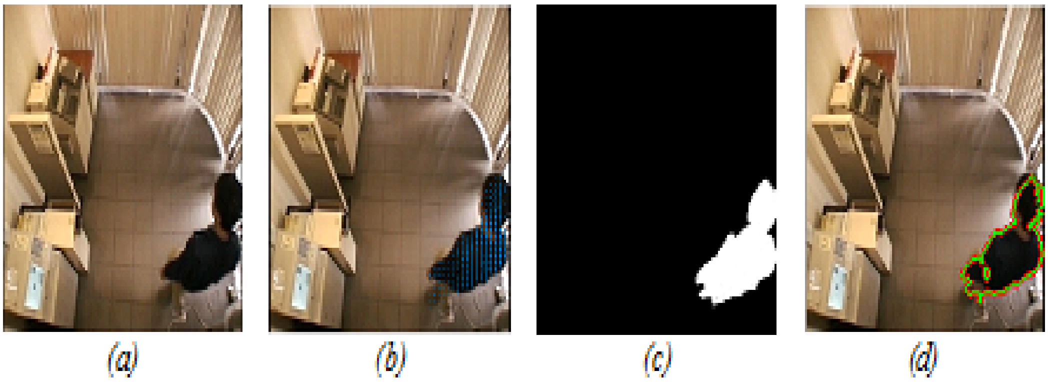


Figure 4.29: (a) Input image.(b) Optical flow field.(c) Result of k-mean clustering algorithm.(d) Motion segmentation

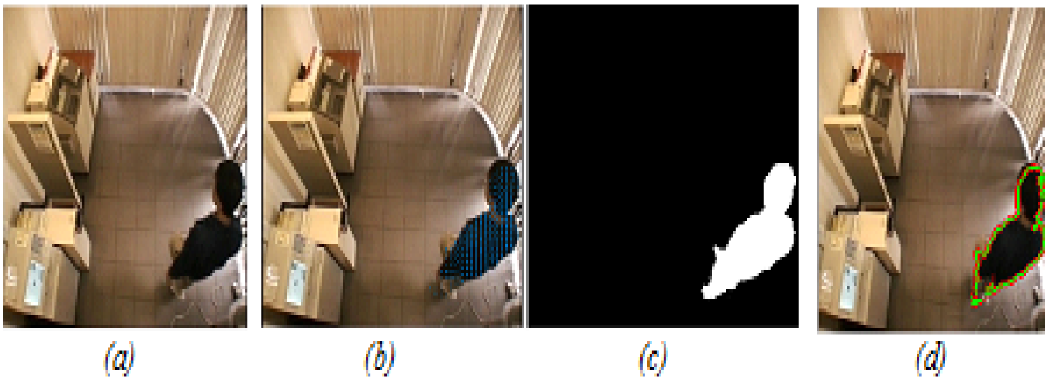


Figure 4.30: (a) Input image.(b) Optical flow field.(c) Result of k-mean clustering algorithm.(d) Motion segmentation

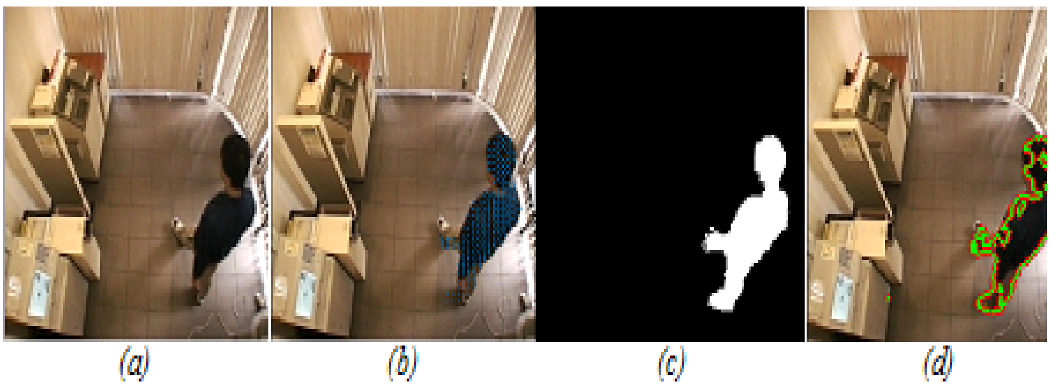


Figure 4.31: (a) Input image.(b) Optical flow field.(c) Result of k-mean clustering algorithm.(d) Motion segmentation

In the figures (4.32), (4.33) and (4.34), we used the outdoor backdoor video sequence which comes from shadow category. Despite of the soft and intermittent shadows, the woman is correctly segmented using our technique.

The figures (4.35) and (4.36), show the segmentation results of the bus station video sequence using our proposed technique. The bus station video sequence comes from the Shadow category. In these figures, the three people are correctly segmented, despite camouflage problems and the shadow. Thus, some parts of people are sometimes not segmented due to the camouflage problem and slow motion.

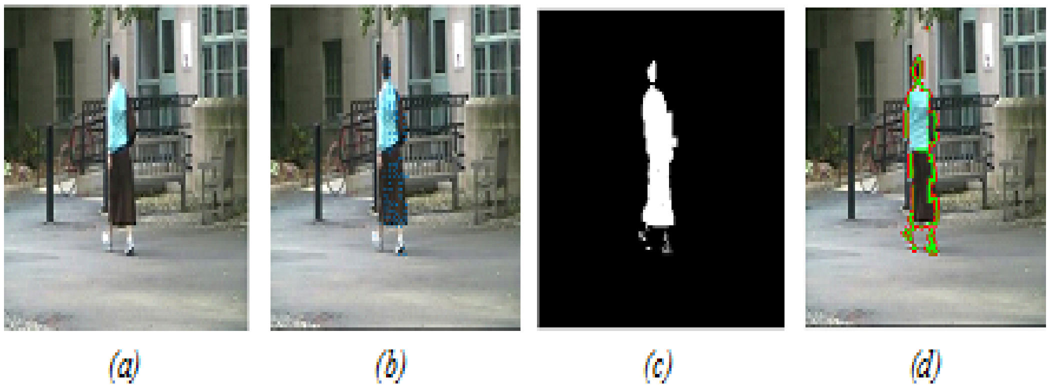


Figure 4.32: (a) Input image.(b) Optical flow field.(c) Result of k-mean clustering algorithm.(d) Motion segmentation

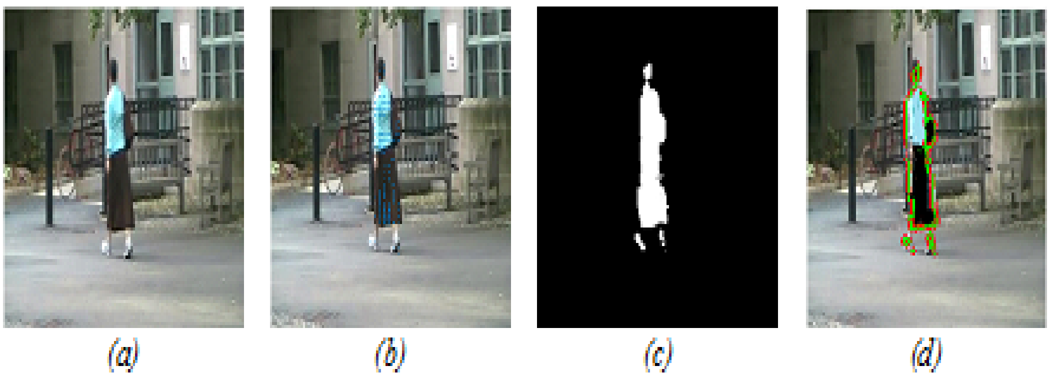


Figure 4.33: (a) Input image.(b) Optical flow field.(c) Result of k-mean clustering algorithm.(d) Motion segmentation

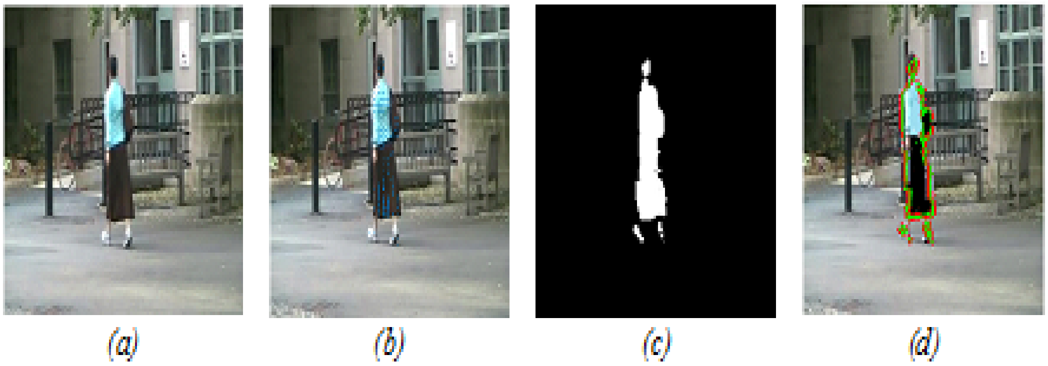


Figure 4.34: (a) Input image.(b) Optical flow field.(c) Result of k-mean clustering algorithm.(d) Motion segmentation

Figure 4.35: (a) Input image.(b) Optical flow field.(c) Result of k-mean clustering algorithm.(d) Motion segmentation

Figure 4.36: (a) Input image.(b) Optical flow field.(c) Result of k-mean clustering algorithm.(d) Motion segmentation

From the CDNET, we used the bridge entry video sequence came form the night videos category. In this scene, cars and trucks go in opposite and / or identical directions and at different distances from the camera. This scene presents multiple problems such as illumination from different sources , the shadows which are reflected over the rood , blurred image. In figure (4.37), the proposed method removes the foreground object shadows and successfully detects the multi-directional object regions.

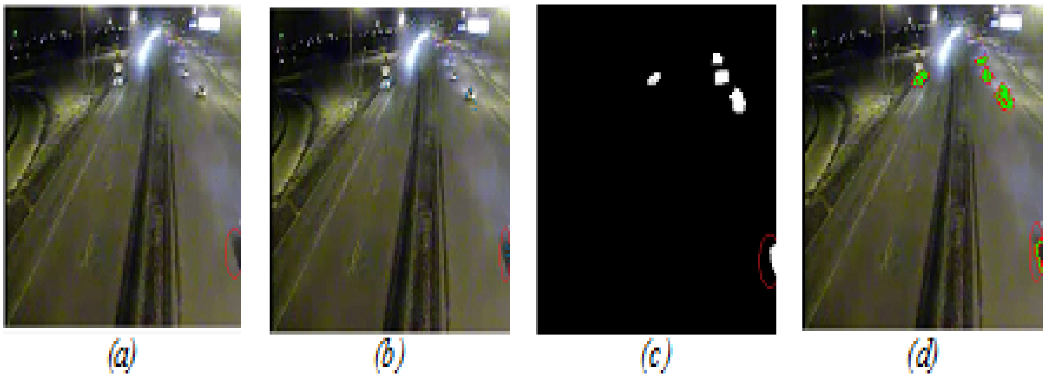


Figure 4.37: (a) Input image.(b) Optical flow field.(c) Result of k-mean clustering algorithm.(d) Motion segmentation

Finally, from the CDNET, we used the sofa video sequence came form the intermittent object motion category. This scene presents multiple problems such as slipping foreground (the box stops for a while and then moves for an other while). Furthermore, several oriented lighting sources with different illuminant are present. The color distribution of the background is very similar to the person shirt. In figures (4.38) and (4.39), the problems presented in the scenes are correctly solved using our approach.

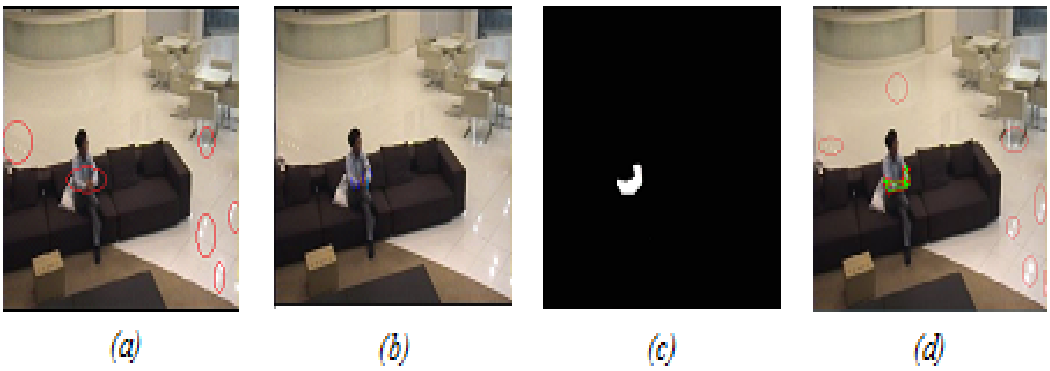


Figure 4.38: (a) Input image.(b) Optical flow field.(c) Result of k-mean clustering algorithm.(d) Motion segmentation

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

Figure 4.39: (a) Input image.(b) Optical flow field.(c) Result of k-mean clustering algorithm.(d) Motion segmentation

4.4.2 Quantitative Analysis

We have evaluated our algorithm quantitatively by comparing our results with the $LBAadaptiveSOM$ [91], $T2FMRF_UM$ [92], $T2FGMM_UV$ [93], and GMG [94]. We have used quantitative measurements such as Precision, Recall, and F-measure defined as follows:

$$Precision = \frac{TP}{TP + FP} \tag{4.11}$$

$$Recall = \frac{TP}{(TP + FN} \tag{4.12}$$

$$F - measure = \frac{2 \times precision \times recall}{precision + recall} \tag{4.13}$$

Where FP and FN refer to pixels misclassified as foreground FP or background FN respectively and TP and TN account for accurately classified pixels respectively as foreground and background. The performance analysis of the proposed algorithms obtained by using various complex dynamic video sequences (ChangeDetection2014 dataset [**?**] and PETS2006 [88]. For all of these measurements, we have initially construct manually some segmented ground-truth images for moving objects.

Table. 4.1 illustrates the quantitative evaluation of the proposed algorithm $NMES$. The values of Precision, Recall, and F-measure should be maximum for better segmentation. These results show that the segmentation rate for the GMG [94] is neither sufficient nor satisfactory. In comparison to the methods referred in [92, 93],and [91], our method gives better results for the F-measure.

Processing time is also evaluated during the experiments. As demonstrated in table.4.2, the processing time of other methods outperforms to our method. From these comparisons, we conclude that our proposed algorithm is better than the other methods considering the quality of detection results. However, our proposed algorithm shows relatively poor performance considering the processing time

.

Table 4.1: Quantitative evaluation of our proposed approach compared to other methods

Video sequences	Metrics	LBAdaptive _SOM [41]	T2FMRF _UM [42]	T2FGMM _UV [43]	GMG [44]	Our NMES method
Highway	**Recall**	0.581	0. 456	0.709	0.346	**0.841**
	Precision	0.902	**0.967**	0.783	0.794	0.908
	F-measure	0.706	0.619	0.744	0.482	**0.873**
Pedestrians	**Recall**	0.672	**0.905**	0.897	0.861	0.851
	Precision	0.891	0.912	0.935	0.857	**0.996**
	F-measure	0.766	0.908	0.915	0.859	**0.917**
Intelligent room	**Recall**	0.865	0.668	0.795	0.574	**0.901**
	Precision	0.843	0.986	**0.990**	0.848	0.937
	F-measure	0.853	0.796	0.881	0.684	**0.918**
PETS2006	**Recall**	0.854	0.605	0.872	0.681	**0.958**
	Precision	0.781	0.977	0.924	0.807	**0.996**
	F-measure	0.815	0.747	0.897	0.738	**0.976**

Table 4.2: Execution time comparison of our proposed approach to other methods

Algorithms	LBAdaptive _SOM [41]	T2FMRF _UM [42]	T2FGMM _UV [43]	GMG [44]	Our NMES method
Execution time	186 ms/frame	108 ms/frame	116 ms/frame	**49 ms/frame**	170 ms/frame

4.5 Conclusion

In this chapter, we gave a detailed description of our proposed approaches. The quantitative and qualitative results revealed that our approach is able to work in all kind of scenes under challenging environments. Our proposed algorithm is able to segment the moving objects without taking by a consideration the scene type (indoor, outdoor), the camera resolution (high, low) or localization, the surface geometry or textures, the quality of the images (blurred images), the size, the shape, the type or even the appearance of the objects.

However, our method is relatively disadvantages in terms of computations time.

5

ChapterVI. Moving object segmentation from complex videos based on feature extraction

Contents

5.1 Introduction

In this chapter, we will focus on developing two new techniques to discriminate moving objects from a static or dynamic background. Our first technique is the (MSFs). It is based on two essential steps. In the first step, we will use the k-means clustering, and Horn and Schunck algorithms to segment the images based on color space and extract the optical flow respectively. In the second step, we will make a comparison between the results of optical flow, and the color space segmentation of each pixel in order to detect moving objects.

The second proposed technique called (BMFS-LBP) consists of combining local binary pattern and block matching full Search algorithm. Firstly, the local binary pattern operator labels the pixels of reference and current macro-blocks region. Next, we will use Pearson product-moment correlation to compute linear relationship between the labeling pixels of current and previous macro blocks. In the end of this task, we will obtain a motion vector matrix. Based on the motion vector matrix, we will cluster the current frame to get a binary mask.

To evaluate the performance of our proposed techniques, we will experiment it on challenging sequences.

5.2 Combination between motion and shape features (MSFs)

In this section, an new method to detect foreground objects from both static and dynamic background is proposed. The proposed algorithm is divided into two essential steps. The block diagram of the proposed method is shown in figure (5.1).

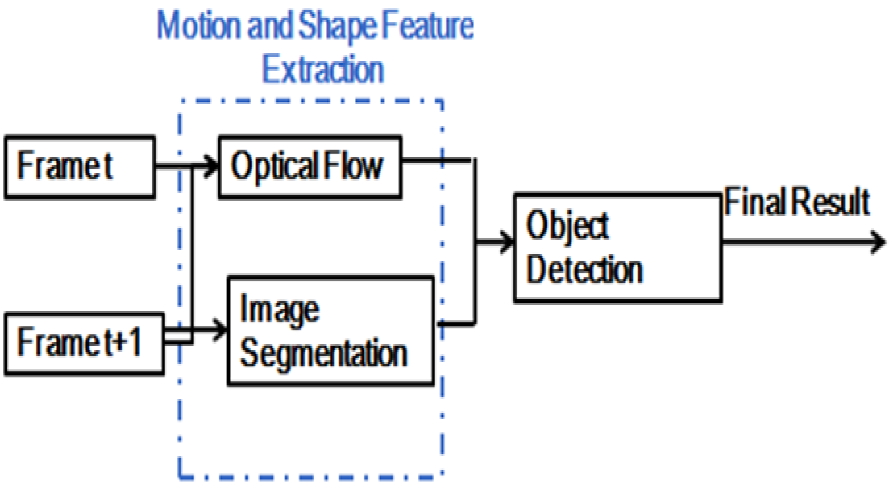


Figure 5.1: Block-diagram of the proposed moving object detection method

5.2.1 Optical flow computing

The first step of our MSFs approach is to compute the optical flow between frames. The optical flow is used to detect moving regions between two frames. It is a dense field of displacement vectors which describes the distribution of the apparent velocities of brightness patterns in image sequence. The optical flow is apparently similar to the dense motion field derived from the techniques of motion estimation.

Horn & Schunck [95] are the initiators of global methods for motion estimation, they minimize a global energy function. This latter is composed from a quadratic norm of optical flow constraint (OFC) and a smooth constraint which assumes that optical flow field varies smoothly almost everywhere in the image.

$$HS(u,v) = \int ((I_x.u + I_y.v + I_t)^2 + \alpha(|\nabla_u|^2 + |\nabla_v|^2))d_x d_y \tag{5.1}$$

Where u, v are the two components of the optical flow vector; they referr to the horizontal and vertical displacement of a pixel $p(x, y)$ from t to $t + 1$ [96], and $\alpha > 0$ is a weighting parameter which regularizes the balance between the OFC and the smoothness constraint. We should note that a larger value of α can produce Blurred contour. The solution of the energy function is achieved by solving the corresponding Euler-Lagrange equation with an iterative implementation Gauss-Seiled. The optical flow can also be defined by its magnitude $V(x, y)$ as:

$$V(x, y) = \sqrt{u^2 + v^2} \tag{5.2}$$

However, the optical flow alone is not sufficient since it cannot help to solve occlusions, shadows, illumination changes and dynamic background problems. Therefore, we have segmented the frame $t + 1$ by using the k-means clustering algorithm.

5.2.2 Color image segmentation using k-means clustering algorithm

Segmentation is a fundamental process in image analysis techniques which is defined as the search for homogenous regions in an image [97]. It is a valuable tool in many fields including medical image processing, compression, remote sensing traffic image, microscopy, and pattern recognition. The two basic properties in image segmentation are discontinuity and similarity [98]. The aim of segmentation is to change the representation of an image into mode to analyse [99]. One of the most efficient segmentation methods is the K-means clustering method. K-means clustering algorithm is a supervised algorithm based on features, used to classify the object into K- groups. It is one of the simplest and computationally clustering algorithm. K-means clustering algorithm can work for large number of variables [100].

In order to segment images using k-mean clustering algorithm, we relied on CIE-LAB color model. Initially, we converted the original image into CIE-LAB color space. Then, we employed the K-means clustering in order to measure the Euclidean distance between each pixel

$P(x, y)$ to another as:

$$d = \mid P(x, y) - C_k \mid \tag{5.3}$$

Then, the pixels close to each other as possible are defined as clusters. Then, a new position of the central cluster C_k will be recalculated using the relation given below.

$$C_k = \frac{1}{K} \sum_{(y \in C_k)} \sum_{(x \in C_k)} P(x, y) \tag{5.4}$$

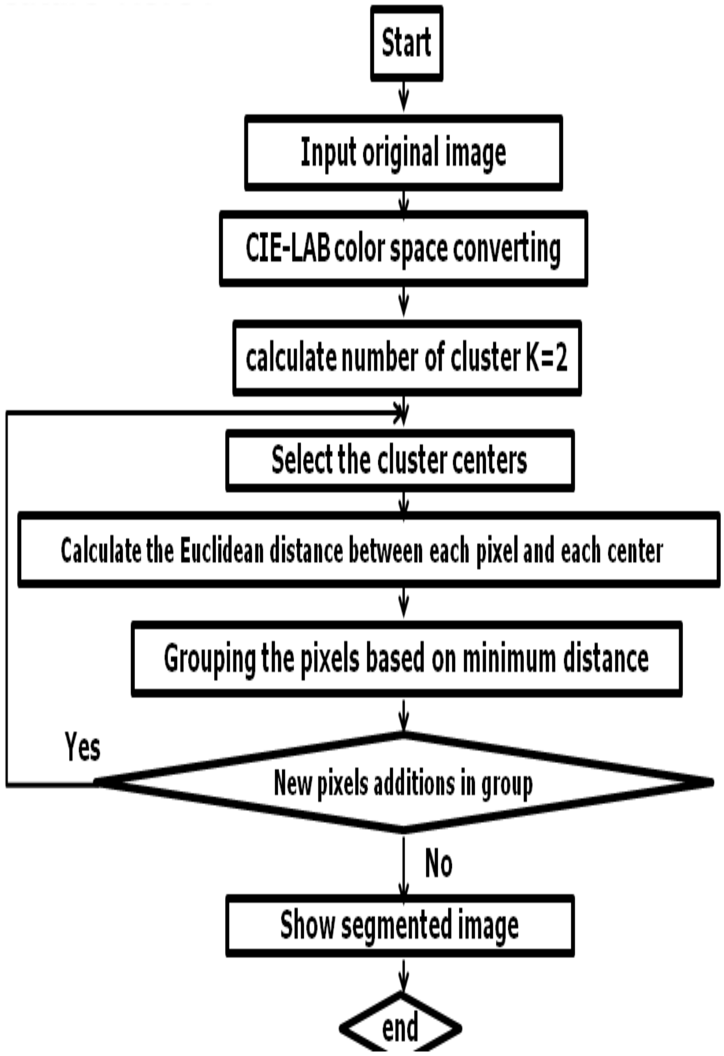


Figure 5.2: Flow-chart of K-means clustering algorithm

Finally, the cluster index value for every pixel in the image will return to segment the image into clusters $I_s(x, y)$. Figure (5.2) explains how the K-means clustering algorithm works.

5.2.3 Object detection

We proposed an adaptive concept that combines the magnitude of the optical flow and color image segmentation using k-means clustering algorithm in order to detect the moving objects as illustrated in figure (5.3). We created a binary mask $M(x, y)$ by comparing between the magnitude $V(x, y)$ of each pixel and the color image segmentation $I_s(x, y)$ as:

$$M(x,y) = \begin{cases} 1, ifV(x,y) \leq 0.5andI_s(x,y) = 1 \\ 0, ifV(x,y) > 0.5andI_s(x,y) = 0) \end{cases} \tag{5.5}$$

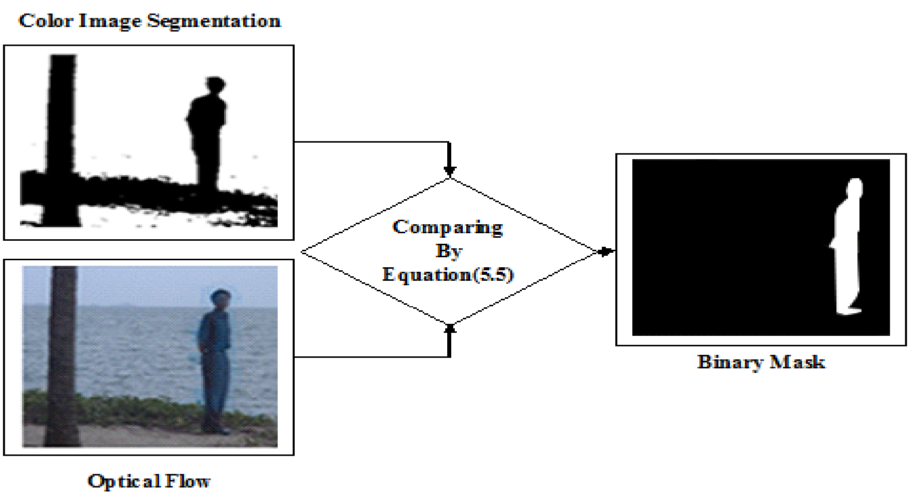


Figure 5.3: Illustration of the object detection concept

5.3 Hybridization between motion and texture features (BMFS-LBP)

We proposed the hybridization between two methods LBP operator and block matching full search (MSFs) algorithm. For a good description to our proposed technique, the details are given in section (3.2.1.A).

The main idea of BMFS-LBP algorithm is to divide the current frame into matrixes of macro-blocks. Afterwards, the LBP operator labels the pixels of each macro-block region by thresholding the neighborhood of each pixel with its center value, and considering the result as a binary number. Then, each binary number value is multiplied by powers of two and then summed.

The LBP operator repeats the same operation with the shifted regions (previous macro-block) of the same size and its adjacent neighbors from the previous frame. To create a vector that stipulates the movement of a macro block from one location to another in the previous frame, we made a linear relationship between the labeling pixels of the current macro-block and the previous one. This task is estimated by computing the Pearson product-moment correlation coefficient. The value of coefficient near zero indicates that the two macro blocks are uncorrelated and the value near one indicates that the two macro blocks are correlated (matched macro-blocks) [101]. The Pearson product-moment correlation coefficient is computed by equation (5.6):

$$r_{C,R} = \frac{COV(C_{ij}, R_{ij})}{\sigma_C \times \sigma_R} \tag{5.6}$$

Where cov (C_{ij}, R_{ij}) is the covariance of the macro-blocks,C_{ij} and R_{ij} are the labeled pixels which are compared in current macro-block and reference macro-block, respectively. σ_C define the standard deviations of the labeled pixels of the current macro-block, σ_R define the standard deviations of the labeled pixels of the reference macro-block.

A good macro-block match is constraint to the search parameter. A larger motion require a larger search parameter. The matching of one macro-block with another is based on the output of the Pearson product-moment correlation coefficient. Let (MV_x, MV_y) is the motion vector of each pixel $P(x, y)$ in the x and y axes. The motion vector can also be defined by its magnitude

$V(x, y)$ (displacement) as:

$$V(x, y) = \sqrt{(MVx)^2 + (MVy)^2} \tag{5.7}$$

The backgrounds can represent disconnected sets of pixel values. This disconnection of pixels is caused by the movement of the background (a fountain, cloud movements, swaying of tree branches, water waves). Such movements can be periodical or irregular like traffic lights and waving trees. These sets of pixels should not be considered part of the foreground. In order to solve this challenge, we assumed that a moving object have a large displacements then the dynamic backgrounds. In our experiments, we have observed that the number of displacement in the same region belonging to a moving background can be lower than the number of displacements in the same region belonging to the foreground. For that, we eliminated the displacement of set of pixels low than 4*4. Consequently, we have normalized the magnitude of motion vector value [19] by:

$$V_{max}(x, y) = max(\sqrt{(MVx)^2 + (MVy)^2}) \tag{5.8}$$

$$V_{min}(x, y) = min((\sqrt{(MVx)^2 + (MVy)^2}) \tag{5.9}$$

Figure (5.4) is a flow-chart of the proposed object segmentation method.

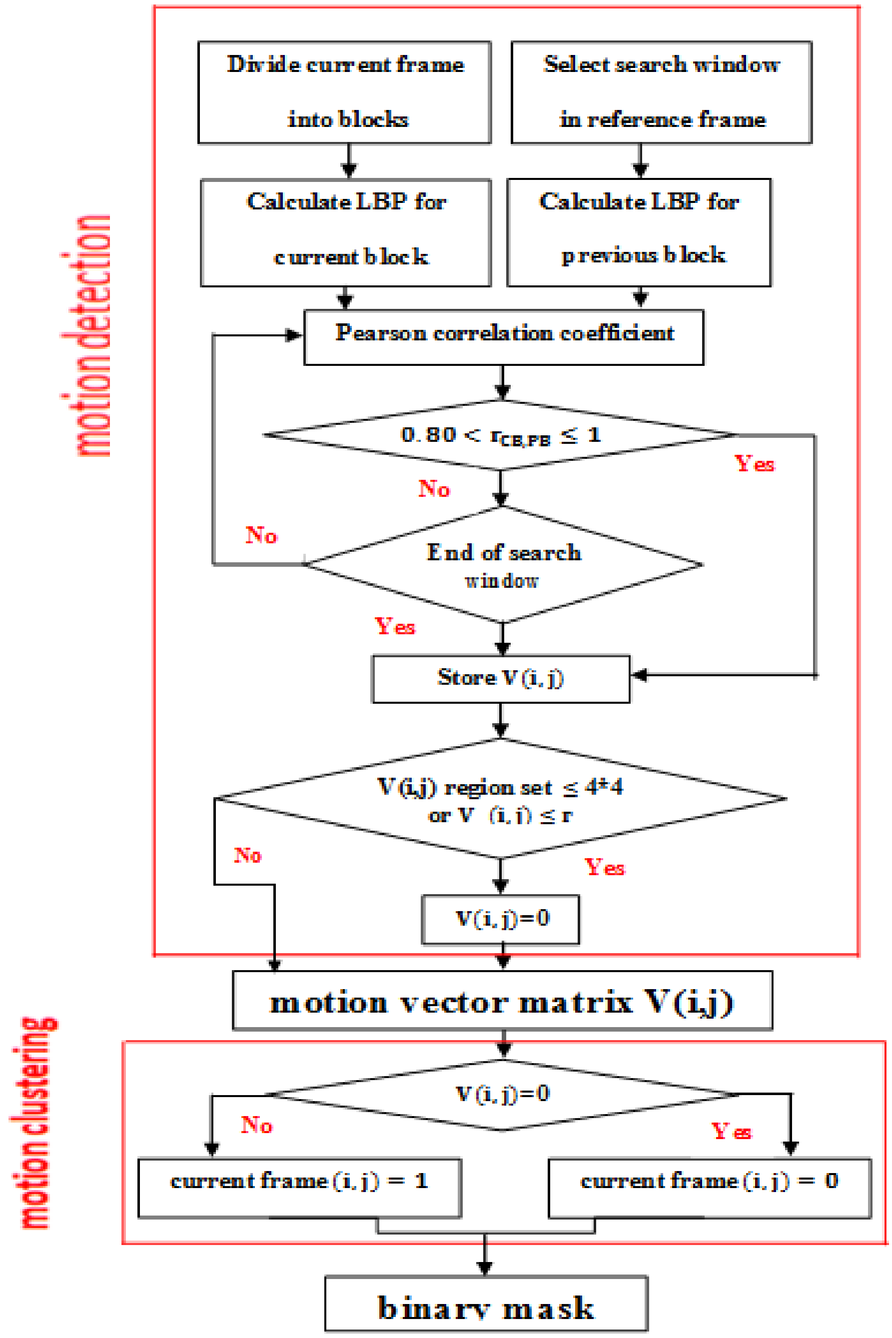


Figure 5.4: Flow-chart of the proposed object segmentation method

5.4 Experimental results and discussions

To evaluate the performance of the proposed approachs, we have implemented our algorithms by MATLAB software. The databases used to evaluate quantitatively and qualitatively our techniques are selected from well-known databases. Our techniques are tested on sequences of outdoor and indoor scenarios.Our proposed algorithms was compared with GMM, BGS, $T2FGMM_UM$ and $T2FMRF_UM$ algorithms. In order to compare our proposed algorithms with the above mentioned algorithems we used Background Subtraction Library [19]. The **BGS library!** (**BGS library!**) is available free of charge to all users, academic and commercial; the library contains 43 algorithms.

In BMFS-LBP technique, the parameters used for LBP computation are set to $R = 2$ and $P = 6$; we tested the algorithms under different datasets including the $CVPR$ 2014 Change Detection dataset [20].

5.4.1 Qualitative Analysis

The segmentation results are shown in figures (5.5) to (5.10).

The first sequence considered in our experiment is from the Campus environments; the Campus shows cars, people, trees, and waving yellow flag. This video illustrates dynamic background occurred by motion of tree branches and changes of tree shadows. The goal is to detect the foreground and classify the moving tree brushes in wind as background. Our technique removes the movement of the background as illustrated in figure (5.5).

The second sequence is from the fountain environments. This scene presents multiple problems such as non-stationary background and blurred image problems. We observed in figure (5.6) that the proposed method is effective in solving this problem.

In the third sequence, we used curtain environments. This indoor scene presents multiple problems such as camouflage (the colors of the peaple clothes and the dog are the same), light switch, shadows, and dynamic background (waving curtains). The results of the proposed method are qualitatively better than those obtained by the other methods. Examples from test sequences are shown in figure (5.7).

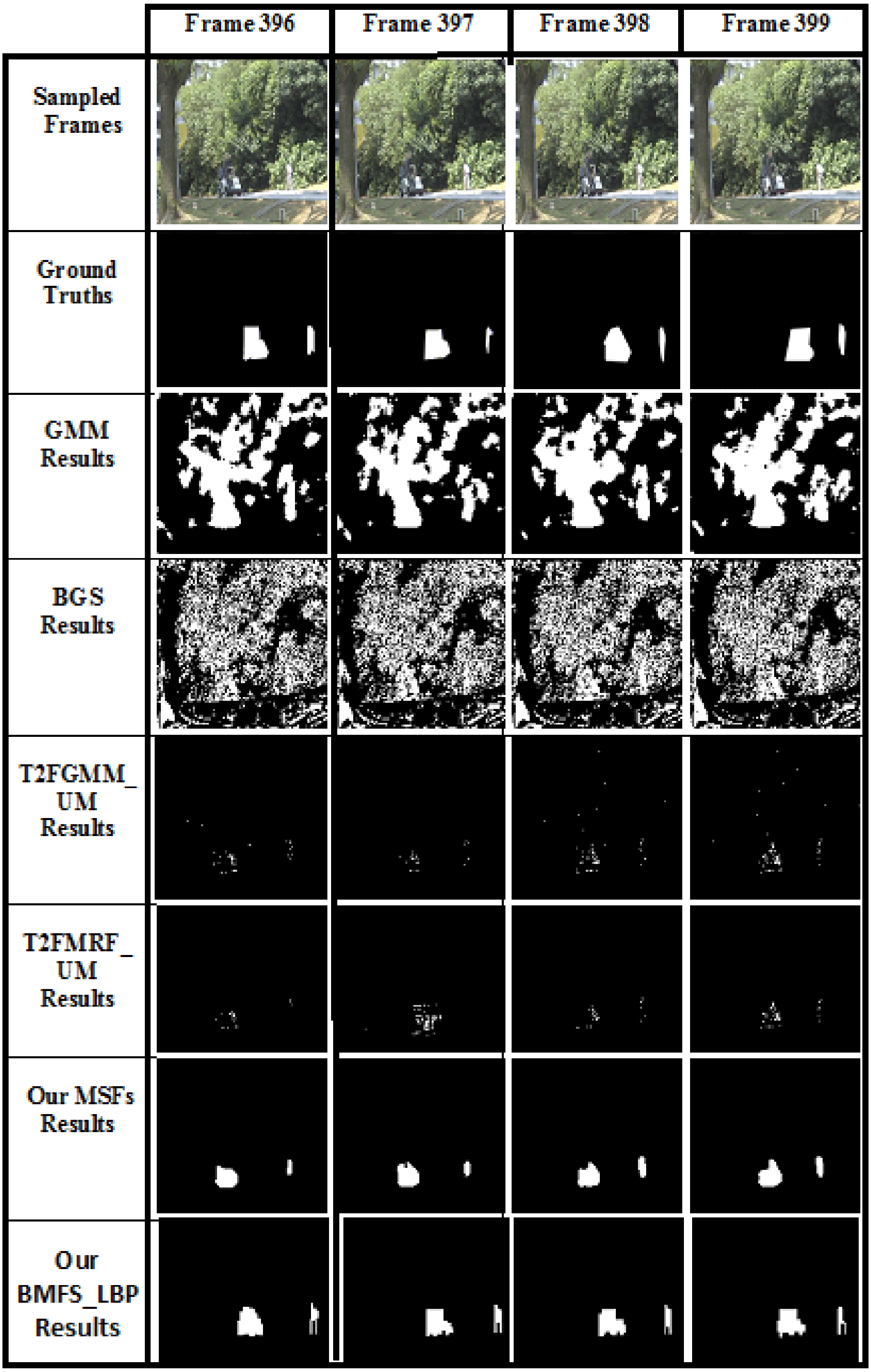


Figure 5.5: Comparison of experimental results (campuses sequence)

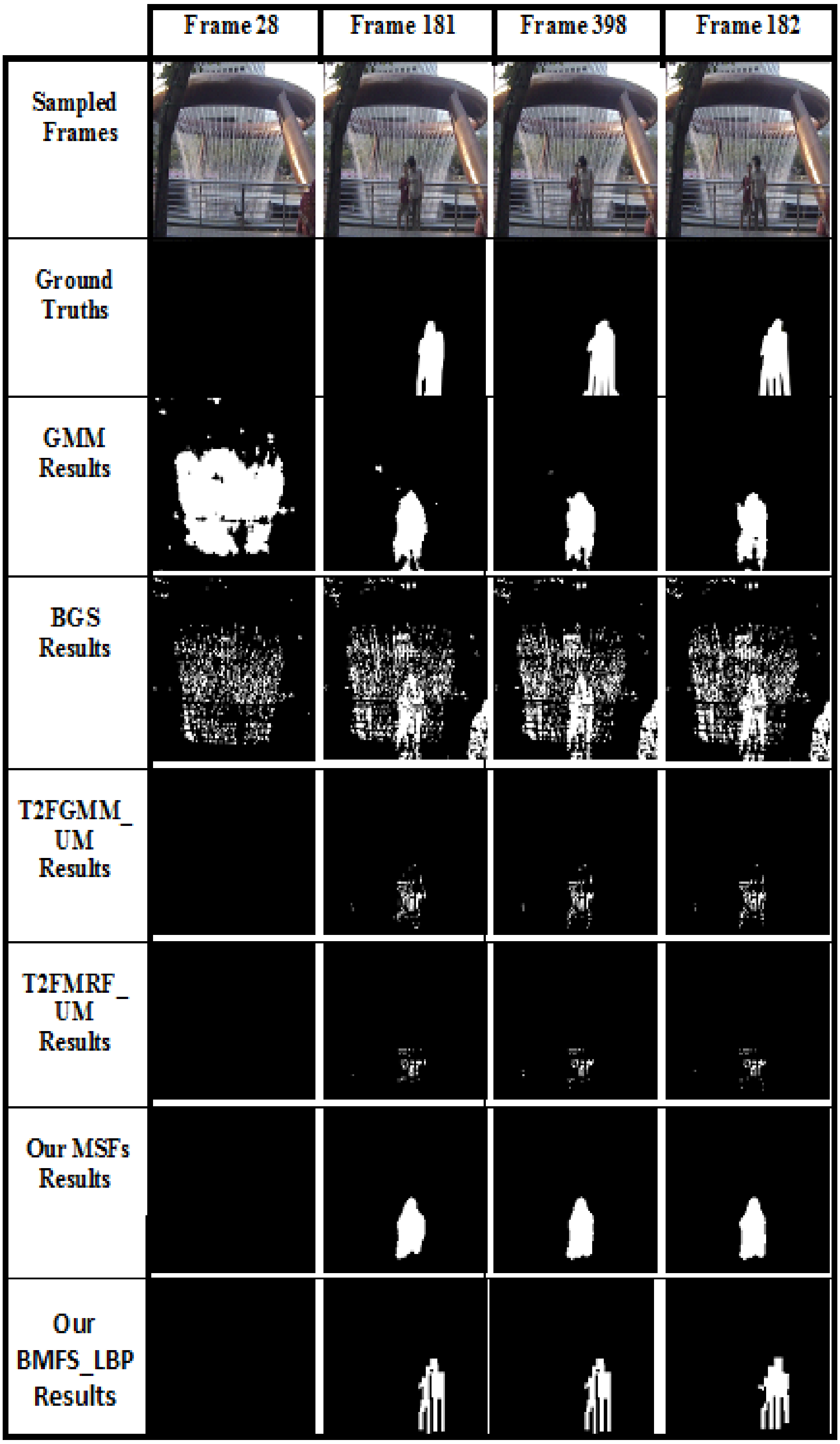


Figure 5.6: Comparison of experimental results (fountain video sequence)

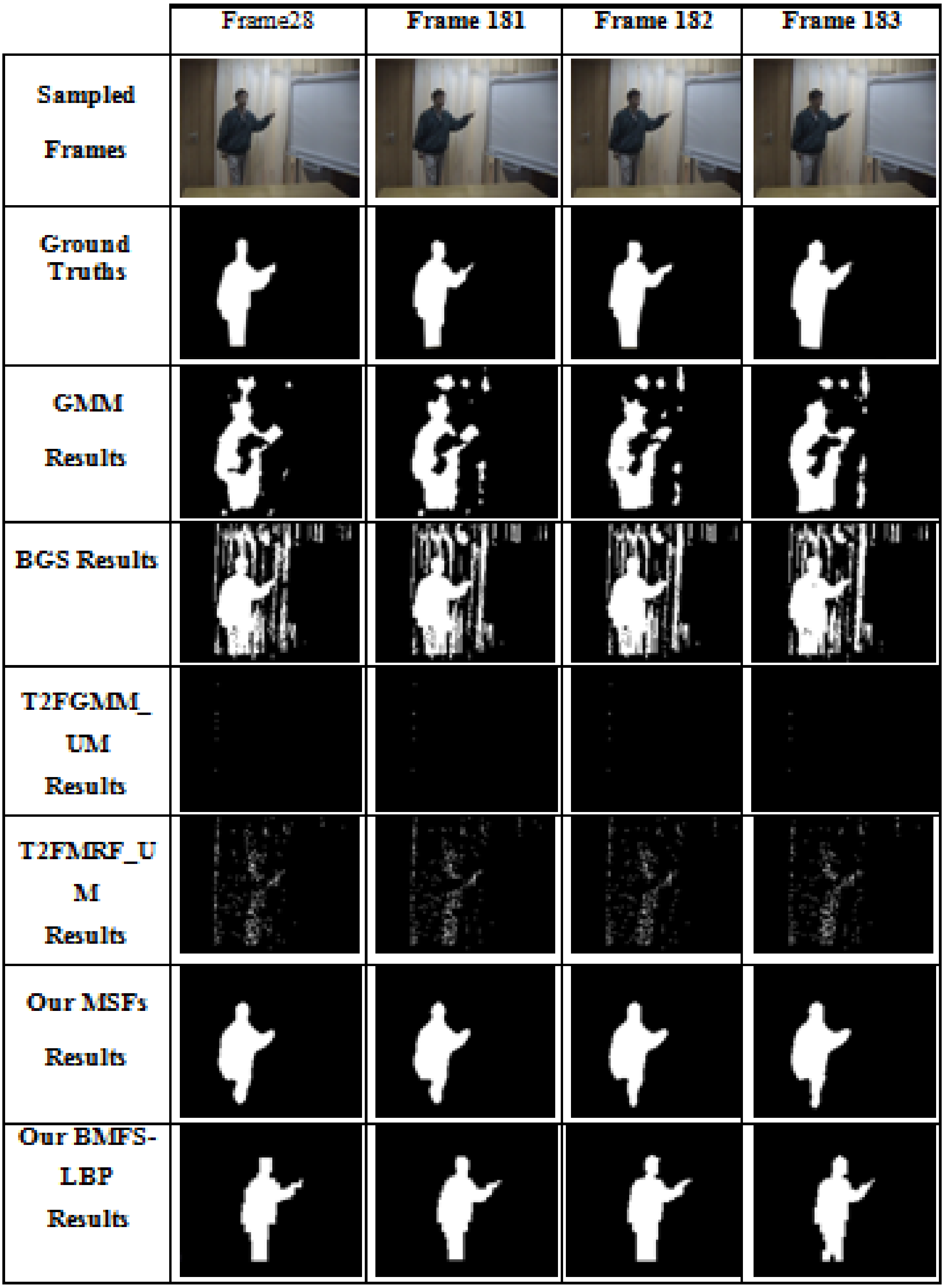


Figure 5.7: Comparison of experimental results (curtains video sequence)

The forth example displayed in figure (5.8) comes from water surface video sequence. This scene presents multiple problems such as camouflage (the colors of the person's clothes and the background are the same), shadows, and non-stationary background (flow of water in the river). The goal is to detect the person and classify the flowing water as background.

The fifth example shown in figure (5.9) is from fountain2 video sequence. This video sequence show cars moving in front of the fountain.

The sixth example in Figure (5.10) come from fall1 video sequence. It contains persons walking by a swaying tree. This scene presents multiple problems such as different illumination, shadows, and non-stationary background (moving tree brushes in strong wind). Our proposed algorithm is able to segment successfully the cars and peaple. The motion of tree branches can be seen from the results of GMM and $MLBSBCT$. As to the results of $SOBS$ are acceptable though many parts of the foreground object were misclassified as the background.

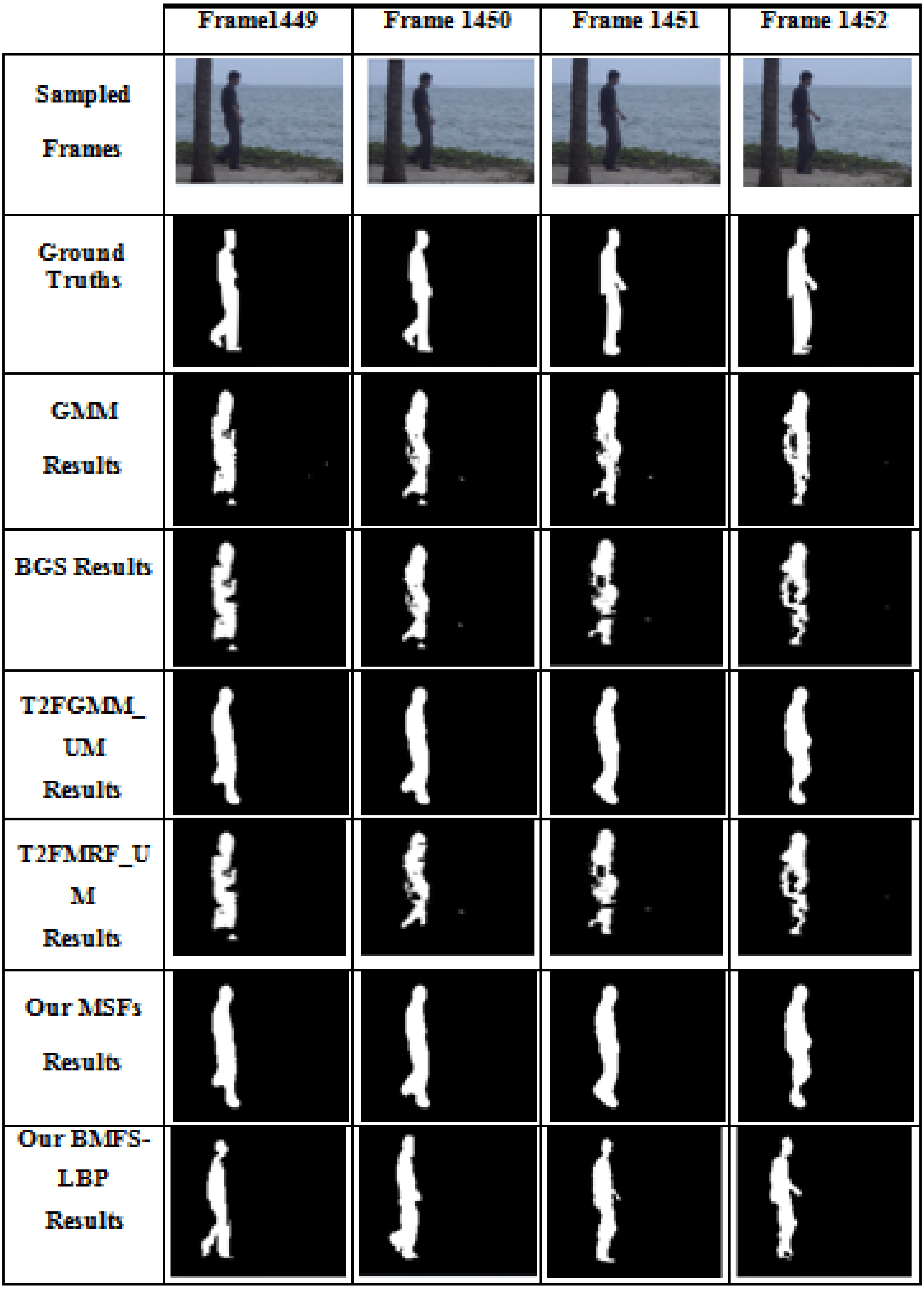


Figure 5.8: Comparison of experimental results (Water Surface video sequence)

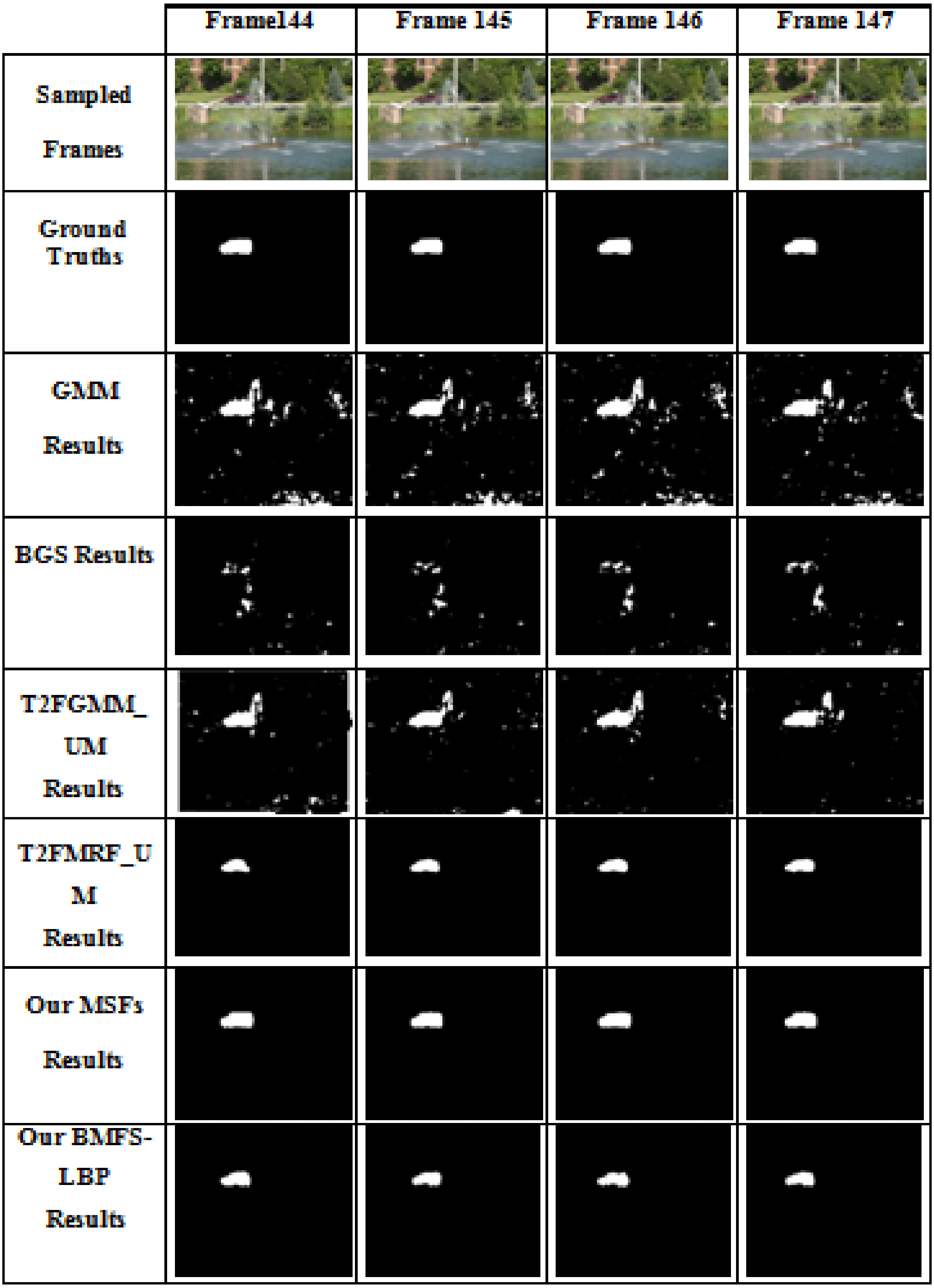


Figure 5.9: Comparison of experimental results (fountain2 video sequence)

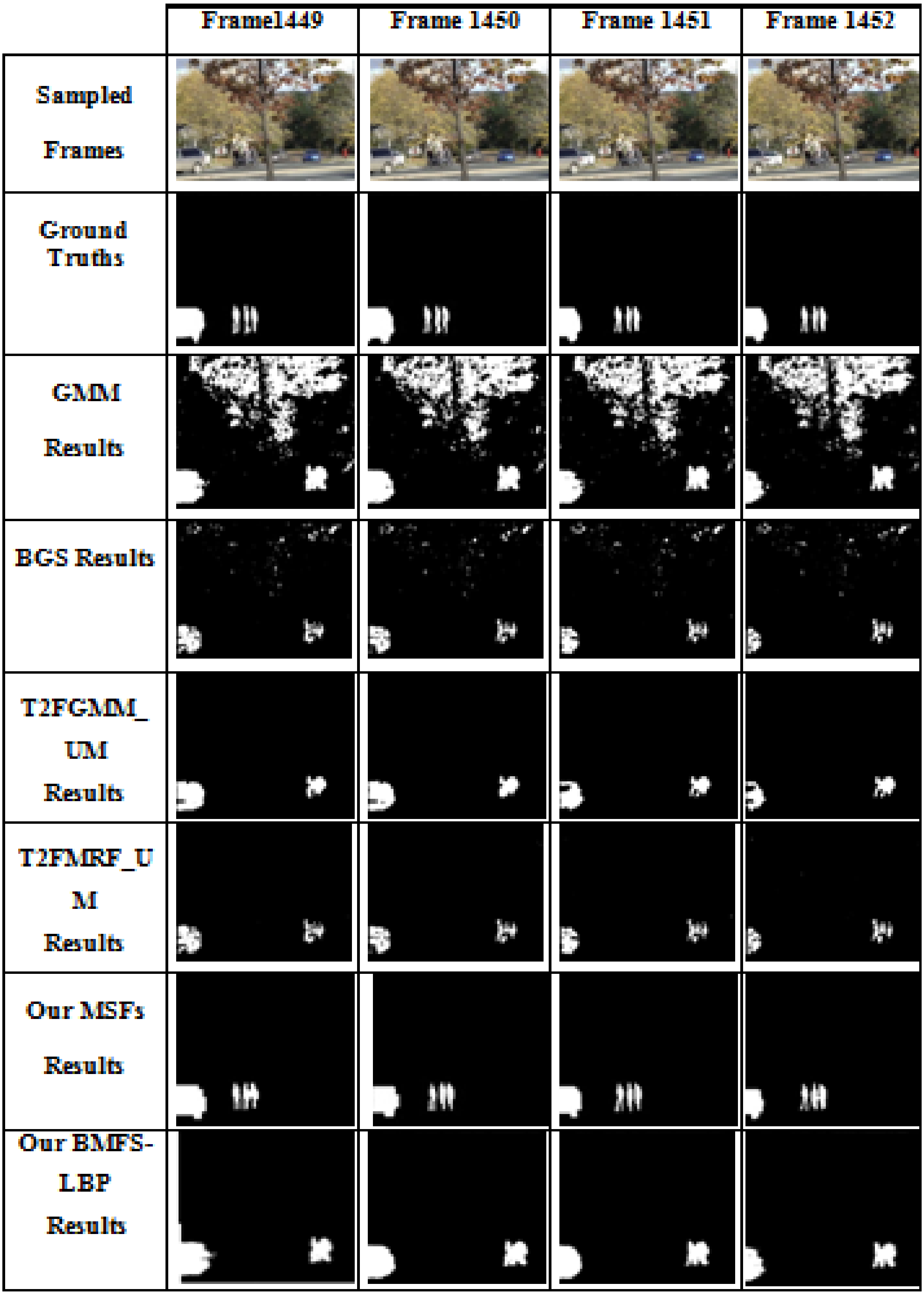


Figure 5.10: Comparison of experimental results (fountain2 video sequence)

Table 5.2: Quantitative evaluation for fountain video sequence

	GMM	SFDBGS	T2FGMM_UM	T2FMRF_UM	MSFs	BMFS-LBP
Recall	36.13	50.84	40.21	42.41	77.57	**89.68**
Precision	56.01	59.67	50.79	30. 75	96.34	**99.95**
F-measure	43.92	54.90	44.88	35.65	85.94	**94.53**

Table 5.3: Quantitative evaluation for curtain video sequence

	GMM	SFDBGS	T2FGMM_UM	T2FMRF_UM	MSFs	BMFS-LBP
Recall	70.89	81.18	48.32	48.57	98.17	**98.58**
Precision	87.48	88.98	54.98	39.99	97.67	**99.99**
F-measure	78.31	84.90	51.43	43.86	97.91	**99.28**

5.4.2 Quantitative evaluation

Quantitative evaluation of our proposed method and comparison with three existing methods were also considerd. We used quantitative measurements such as Precision, Recall and F-measure. The quantitative evaluation results are shown in Tables (5.1), (5.2), (5.3) and (5.4). The quantitative evaluation agrees with the results of the qualitative observation.

Table 5.1: Quantitative Evaluation of Camuses Video Sequence

	GMM	SFDBGS	T2FGMM_UM	T2FMRF_UM	MSFs	BMFS-LBP
Recall	38.92	52.33	39.92	45.24	74.01	**85.07**
Precision	54.61	60.75	63.87	57.07	98.12	**99.92**
F-measure	45.44	56.22	49.13	50.47	84.37	**91.89**

Table 5.4: Quantitative evaluation for Water Surface video sequence

	GMM	SFDBGS	T2FGMM_UM	T2FMRF_UM	MSFs	BMFS-LBP
Recall	84.78	88.99	89.91	93.54	98.74	**98.78**
Precision	89.93	87.87	90.32	89.87	98.95	**99.97**
F-measure	87.28	88.42	90.11	91.66	98.84	**99.37**

Table 5.5: Quantitative evaluation for fountain2 video sequence

	GMM	SFDBGS	T2FGMM_UM	T2FMRF_UM	MSFs	BMFS-LBP
Recall	75.99	60.13	89.97	69.99	99.77	**99.87**
Precision	67.38	50.71	96.87	70.84	98.95	**99.27**
F-measure	71.42	55.01	93.29	70.41	99.54	**99.56**

Table 5.6: Quantitative evaluation for fall1 video sequence

	GMM	SFDBGS	T2FGMM_UM	T2FMRF_UM	MSFs	BMFS-LBP
Recall	89.57	89.65	89.98	87.98	99.81	**99.98**
Precision	69.78	80.78	87.81	90.84	98.05	**99.95**
F-measure	78.44	84.93	88.88	89.39	98.92	**99.96**

5.5 CONCLUSION

In this chapter, we proposed two new approaches to detect foreground objects from complex videos .They contain both stationary (or moving) background objects. The proposed approaches tested on several real scenes contain challenging cases.

The qualitative and quantitative results in comparison to the existing methods show the efficiency of the BMFS-LBP algorithm.

6

Chapter V: CONCLUSIONS AND FUTURE WORKS

Contents

6.1 Conclusions

In this thesis, we focused on motion based segmentation in complex scenes (dynamic background, changes of illumination, camouflage, shadow, occlusions....).

This thesis is divided in four parts.

In the first part, we presented the importance and problems of motion based segmentation methods; then we described the well known traditional motion segmentation techniques. Finally, we presented the data sets usually used. In the second part, we concentrated on the feature extraction techniques and clustering algorithms in order to review, explain and describe the algorithms that we used.

In the third part, we proposed a new technique for Spatio-temporal segmentation of moving objects in image sequences. Our technique involves motion, color and texture features to extract the important information (features) from the image sequences. Then, we used k-means algorithm in order to classify pixels with similar extracted features.Finally, we used chan-vese model to segment moving objects. Our technique was tested on sequences of outdoor and indoor scenarios. The experimental results illustrated the robustness of our method in case of partial occlusions, progressive illumination changes, blurred image, strong shadows and camouflage problems.

However, our technique can not solve the dynamic background problem. Thus, to solve this problem, we proposed two new techniques; MSFs and BMFS-LBP.

In the fourth part of the our thesis, we presented two new techniques: the first one is combination between motion and shape features (MSFs); the second one is the hybridization between motion and texture features (BMFS-LBP) to detect and discriminate the foreground object.

The MSFs technique is based on two essential steps. In the first step, we used the k-means clustering and Horn & Schunck algorithms in order to segment the image based on color space and optical flow extraction respectively to extract the motion and shape features. In the second step, we analyzed the logical comparison between the results of optical flow and the color space segmentation for all pixels in order to create binary mask. The resulting binary mask shows the robustness of our method in case of dynamic background, shadow and color similarity between moving objects and background.

However, in some cases the distinction between foreground objects and background regions can fail. Sometimes, part of foregrounds objects or background are not accurately segmented due to occlusions. The optical flow is erratic and unreliable at the object boundaries. Therefore, the set of occluded pixels may still remain unclassified to be part of foreground objects. In order to solve this problem, the BMFS-LBP technique has been developed.

Qualitative and quantitative results for both outdoor and indoor video sequences validate our approaches. Generally, our BMFS-LBP technique gives better resultes in case of dynamic background, progressive shadow and camouflage in comparison to our MSFs and foned in literature. However, our methods is poor in of compitation time.

6.2 Future work

There are still several issues that need to be developed in future.

Our proposed methods are defined and tested just with fixed cameras, moving cameras.

In the analysis of the third part, the focus was set on algorithm that can be extended to real time implementations. The idea is to use global-local techniques (Bruhn algorithm) to extract motion features instead of block matching algorithm because of the execution speed and the precision of Bruhn algorithm.

In the MSFs technique, we may use graph cut clustering instead to k-means clustering algorithm to extract the shape from image sequences to solve the supervised clustering problem.

Finally, we propose to mix all the features shape, texture, motion and color to give more efficient techniques. Moreover, we propose to use color constancy techniques to improve the color features.

Bibliography

[1] W. Crum, "Feature extraction, shape fitting and image segmentation," in *Proc Int Soc Magn Reson Med*, vol. 14, 2006, p. 6.

[2] I. Koprinska and S. Carrato, "Temporal video segmentation: A survey," *Signal processing: Image communication*, vol. 16, no. 5, pp. 477–500, 2001.

[3] A. Merin Antony and J. Anitha, "A survey of moving object segmentation methods," *International Journalof Advanced Research in Electronics and Communication Engineering (IJARECE) Volume*, vol. 1.

[4] Z. Zhu and Y. Wang, "A hybrid algorithm for automatic segmentation of slowly moving objects," *AEU-International Journal of Electronics and Communications*, vol. 66, no. 3, pp. 249–254, 2012.

[5] 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.

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

[7] J. Wang and E. H. Adelson, "Spatio-temporal segmentation of video data," in *Image and Video Processing II*, vol. 2182. International Society for Optics and Photonics, 1994, pp. 120–132.

[8] D. Zhang and G. Lu, "Segmentation of moving objects in image sequence: A review," *Circuits, Systems and Signal Processing*, vol. 20, no. 2, pp. 143–183, 2001.

[9] S. A. Inigo and P. Suresh, "General study on moving object segmentation methods for video," *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)*, vol. 1, no. 8, pp. pp–265, 2012.

[10] S. C. Sen-Ching and C. Kamath, “Robust techniques for background subtraction in urban traffic video,” in *Visual Communications and Image Processing 2004*, vol. 5308. International Society for Optics and Photonics, 2004, pp. 881–893.

[11] 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.

[12] D. K. Panda, “Motion detection, object classification and tracking for visual surveillance application,” Ph.D. dissertation, 2012.

[13] A. Mittal and D. Huttenlocher, “Scene modeling for wide area surveillance and image synthesis,” in *Computer Vision and Pattern Recognition, 2000. Proceedings. IEEE Conference on*, vol. 2. IEEE, 2000, pp. 160–167.

[14] 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.

[15] M. Piccardi, “Background subtraction techniques: a review,” in *Systems, man and cybernetics, 2004 IEEE international conference on*, vol. 4. IEEE, 2004, pp. 3099–3104.

[16] A. Elgammal, D. Harwood, and L. Davis, “Non-parametric model for background subtraction,” in *European conference on computer vision*. Springer, 2000, pp. 751–767.

[17] 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.

[18] M. Kass, A. Witkin, and D. Terzopoulos, “Snakes: Active contour models,” in *Proc. 1st Int. Conf. on Computer Vision*, vol. 259, 1987, p. 268.

[19] H. Zhang, Z. Duan, Z. Zhu, and Y. Wang, “Fast moving object segmentation based on active contours.” *JCP*, vol. 7, no. 4, pp. 863–869, 2012.

[20] Z. Wu and C. Chen, “User-assisted video segmentation system for visual communication,” in *Visual Communications and Image Processing 2002*, vol. 4671. International Society for Optics and Photonics, 2002, pp. 585–593.

[21] Y. Zinbi, Y. Chahir, and A. Elmoataz, “Moving object segmentation; using optical flow with active contour model,” in *Information and Communication Technologies: From Theory to Applications, 2008. ICTTA 2008. 3rd International Conference on*. IEEE, 2008, pp. 1–5.

[22] B. K. Horn and B. G. Schunck, “Determining optical flow,” *Artificial intelligence*, vol. 17, no. 1-3, pp. 185–203, 1981.

[23] ——, “” determining optical flow”: A retrospective,” 1993.

[24] S. Vishwakarma and A. Agrawal, “A survey on activity recognition and behavior understanding in video surveillance,” *The Visual Computer*, vol. 29, no. 10, pp. 983–1009, 2013.

[25] L. Zappella, X. Lladó, and J. Salvi, “New trends in motion segmentation,” in *Pattern Recognition*. InTech, 2009.

[26] “Test Images for Wallflower Paper,” http://research.microsoft.com/en-s/um/people/jckrumm/WallFlower/TestImages.htm, accessed: September 1, 1999.

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

[28] Y. Wang, P.-M. Jodoin, F. Porikli, J. Konrad, Y. Benezeth, and P. Ishwar, “Cdnet 2014: An expanded change detection benchmark dataset,” in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2014 IEEE Conference on*. IEEE, 2014, pp. 393–400.

[29] L. Li, W. Huang, I. Y.-H. Gu, and Q. Tian, “Statistical modeling of complex backgrounds for foreground object detection,” *IEEE Transactions on Image Processing*, vol. 13, no. 11, pp. 1459–1472, 2004.

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

[31] D. P. Young and J. M. Ferryman, “Pets metrics: On-line performance evaluation service,” in *Visual Surveillance and Performance Evaluation of Tracking and Surveillance, 2005. 2nd Joint IEEE International Workshop on*. IEEE, 2005, pp. 317–324.

[32] "Caviar test case scenarios," http://homepages.inf.ed.ac.uk/rbf/CAVIARDATA1/, accessed: July 11, 2003.

[33] "Change detection.net (cdnet)," http://www.changedetection.net/, accessed: September 1, 1999.

[34] 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.

[35] D. ping Tian *et al.*, "A review on image feature extraction and representation techniques," *International Journal of Multimedia and Ubiquitous Engineering*, vol. 8, no. 4, pp. 385–396, 2013.

[36] P. Kamavisdar, S. Saluja, and S. Agrawal, "A survey on image classification approaches and techniques," *International Journal of Advanced Research in Computer and Communication Engineering*, vol. 2, no. 1, pp. 1005–1009, 2013.

[37] K. K. Pachouri and A. Barve, "A comparative analysis & survey of various feature extraction techniques."

[38] D. ping Tian *et al.*, "A review on image feature extraction and representation techniques," *International Journal of Multimedia and Ubiquitous Engineering*, vol. 8, no. 4, pp. 385–396, 2013.

[39] J. Brickmann, "B. mandelbrot: The fractal geometry of nature, freeman and co., san francisco 1982. 460 seiten, preis:£ 22, 75." *Berichte der Bunsengesellschaft für physikalische Chemie*, vol. 89, no. 2, pp. 209–209, 1985.

[40] A. P. Pentland, "Fractal-based description of natural scenes," *IEEE transactions on pattern analysis and machine intelligence*, no. 6, pp. 661–674, 1984.

[41] Z. Hai-Ying, X. Zheng-guang, and P. Hong, "A texture feature extraction based on two fractal dimensions for content based image retrieval," in *Computer Science and Information Engineering, 2009 WRI World Congress on*, vol. 3. IEEE, 2009, pp. 117–121.

[42] H. Kaneko, "Fractal feature and texture analysis," *Systems and computers in Japan*, vol. 19, no. 8, pp. 28–37, 1988.

[43] M. Tuceryan and A. Jain, "Texture analysis handbook of pattern recognition and computer vision: 207-248," *World Scientific Publishing Co*, 1998.

[44] P. Mohanaiah, P. Sathyanarayana, and L. GuruKumar, "Image texture feature extraction using glcm approach," *International Journal of Scientific and Research Publications*, vol. 3, no. 5, p. 1, 2013.

[45] "Texture," http://www.massey.ac.nz/~mjjohnso/notes/59731/.../texture.pdf, accessed: September 1, 2017.

[46] A. Suresh and K. Shunmuganathan, "Image texture classification using gray level co-occurrence matrix based statistical features," *European Journal of Scientific Research*, vol. 75, no. 4, pp. 591–597, 2012.

[47] S. Selvarajah and S. Kodituwakku, "Analysis and comparison of texture features for content based image retrieval," *International Journal of Latest Trends in Computing*, vol. 2, no. 1, 2011.

[48] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 24, no. 7, pp. 971–987, 2002.

[49] L. Wang and D.-C. He, "Texture classification using texture spectrum," *Pattern Recognition*, vol. 23, no. 8, pp. 905–910, 1990.

[50] M. Hassaballah, A. A. Abdelmgeid, and H. A. Alshazly, "Image features detection, description and matching," in *Image Feature Detectors and Descriptors*. Springer, 2016, pp. 11–45.

[51] L. W. Leung, B. King, and V. Vohora, "Comparison of image data fusion techniques using entropy and ini," in *Paper presented at the 22nd Asian Conference on Remote Sensing*, vol. 5, 2001, p. 9.

[52] P. Kakumanu, S. Makrogiannis, and N. Bourbakis, "A survey of skin-color modeling and detection methods," *Pattern recognition*, vol. 40, no. 3, pp. 1106–1122, 2007.

[53] P. Stanchev, D. Green Jr, and B. Dimitrov, "High level color similarity retrieval," 2003.

[54] H.-D. Cheng, X. H. Jiang, Y. Sun, and J. Wang, "Color image segmentation: advances and prospects," *Pattern recognition*, vol. 34, no. 12, pp. 2259–2281, 2001.

[55] M. S. Banu and K. Nallaperumal, "Analysis of color feature extraction techniques for pathology image retrieval system," in *Computational Intelligence and Computing Research (ICCIC), 2010 IEEE International Conference on*. IEEE, 2010, pp. 1–7.

[56] Y. Shi, M. Xiao, and J. Yang, "Pixel-based skin color detection considering overlap region," in *Intelligent Signal Processing and Communication Systems, 2007. ISPACS 2007. International Symposium on*. IEEE, 2007, pp. 256–259.

[57] P. Kakumanu, S. Makrogiannis, and N. Bourbakis, "A survey of skin-color modeling and detection methods," *Pattern recognition*, vol. 40, no. 3, pp. 1106–1122, 2007.

[58] A. Ford and A. Roberts, "Colour space conversions," *Westminster University, London*, vol. 1998, pp. 1–31, 1998.

[59] . !! and . !!, "C!!!!!" *!!!*, vol. 1998, pp. 1–31, 1998.

[60] D. J. Bora, A. K. Gupta, and F. A. Khan, "Comparing the performance of l* a* b* and hsv color spaces with respect to color image segmentation," *arXiv preprint arXiv:1506.01472*, 2015.

[61] D. Zhang, M. M. Islam, and G. Lu, "A review on automatic image annotation techniques," *Pattern Recognition*, vol. 45, no. 1, pp. 346–362, 2012.

[62] G. Pass and R. Zabih, "Histogram refinement for content-based image retrieval," in *Applications of Computer Vision, 1996. WACV'96., Proceedings 3rd IEEE Workshop on*. IEEE, 1996, pp. 96–102.

[63] M. Yang, K. Kpalma, and J. Ronsin, "A survey of shape feature extraction techniques," 2008.

[64] M. Larsson and L. Söderström, "Analysis of optical flow algorithms for denoising," *Master's Theses in Mathematical Sciences*, 2015.

[65] S. T. Khawase, S. D. Kamble, N. V. Thakur, and A. S. Patharkar, "An overview of block matching algorithms for motion vector estimation."

[66] ?????, "...?????" *IEEE Transactions Evolution Computation*, vol. 8, no. 3, pp. 225–239, 2004.

[67] ????, "?????" *IEEE Transactions Evolution Computation*, vol. 8, no. 3, pp. 225–239, 2004.

[68] C. Neustaedter, "An evaluation of optical flow using lucas and kanade's algorithm," *University of Calgary Department of Computer Science, Calgary, AB, T2N 1N4, Canada*, 2002.

[69] J. Jain and A. Jain, “Displacement measurement and its application in interframe image coding,” *IEEE Transactions on communications*, vol. 29, no. 12, pp. 1799–1808, 1981.

[70] R. Li, B. Zeng, and M. L. Liou, “A new three-step search algorithm for block motion estimation,” *IEEE transactions on circuits and systems for video technology*, vol. 4, no. 4, pp. 438–442, 1994.

[71] J. Lu and M. L. Liou, “A simple and efficient search algorithm for block-matching motion estimation,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 7, no. 2, pp. 429–433, 1997.

[72] L.-M. Po and W.-C. Ma, “A novel four-step search algorithm for fast block motion estimation,” *IEEE transactions on circuits and systems for video technology*, vol. 6, no. 3, pp. 313–317, 1996.

[73] S. Zhu and K.-K. Ma, “A new diamond search algorithm for fast block-matching motion estimation,” *IEEE transactions on Image Processing*, vol. 9, no. 2, pp. 287–290, 2000.

[74] Y. Nie and K.-K. Ma, “Adaptive rood pattern search for fast block-matching motion estimation,” *IEEE Transactions on image processing*, vol. 11, no. 12, pp. 1442–1449, 2002.

[75] C.-H. Hsieh and T.-P. Lin, “Vlsi architecture for block-matching motion estimation algorithm,” *IEEE Transactions on circuits and systems for video technology*, vol. 2, no. 2, pp. 169–175, 1992.

[76] S. Khan and M. Shah, “Object based segmentation of video using color, motion and spatial information,” in *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, vol. 2. IEEE, 2001, pp. II–II.

[77] S. Lee, N. Kim, K. Jeong, I. Paek, H. Hong, and J. Paik, “Multiple moving object segmentation using motion orientation histogram in adaptively partitioned blocks for high-resolution video surveillance systems,” *Optik-International Journal for Light and Electron Optics*, vol. 126, no. 19, pp. 2063–2069, 2015.

[78] S.-g. Wei, L. Yang, Z. Chen, and Z.-f. Liu, “Motion detection based on optical flow and self-adaptive threshold segmentation,” *Procedia Engineering*, vol. 15, pp. 3471–3476, 2011.

[79] P. Kakumanu, S. Makrogiannis, and N. Bourbakis, “A survey of skin-color modeling and detection methods,” *Pattern recognition*, vol. 40, no. 3, pp. 1106–1122, 2007.

[80] L. W. Leung, B. King, and V. Vohora, “Comparison of image data fusion techniques using entropy and ini,” in *Paper presented at the 22nd Asian Conference on Remote Sensing*, vol. 5, 2001, p. 9.

[81] P. Scholor and S. S. Subramanian, “Extraction and classification of blebs in human embryonic stem cell.”

[82] S. Ray and R. H. Turi, “Determination of number of clusters in k-means clustering and application in colour image segmentation,” in *Proceedings of the 4th international conference on advances in pattern recognition and digital techniques*. Calcutta, India, 1999, pp. 137–143.

[83] T. F. Chan and L. A. Vese, “Active contours without edges,” *IEEE Transactions on image processing*, vol. 10, no. 2, pp. 266–277, 2001.

[84] H. Singh and S. Parveen, “Chan-vese deformable model for different domain images.”

[85] R. Crandall, “Image segmentation using the chan-vese algorithm,” *Project report from ECE*, vol. 532, 2009.

[86] R. Cohen, “The chan-vese algorithm,” *arXiv preprint arXiv:1107.2782*, 2011.

[87] “Caviar test case scenarios,” http://homepages.inf.ed.ac.uk/rbf/CAVIARDATA1/, accessed: July 11, 2003.

[88] “Pets 2006 benchmark data,” http://www.cvg.reading.ac.uk/PETS2006/data.html, accessed: September 1, 2006.

[89] “Shadow detection,” http://cvrr.ucsd.edu/aton/shadow/, accessed: 201.

[90] “..//???” http://perception.i2r.a-star.edu.sg/bk_model/bk_index.html, accessed: ???

[91] 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.

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

[93] T. Bouwmans and F. El Baf, "Modeling of dynamic backgrounds by type-2 fuzzy gaussians mixture models," *MASAUM Journal of of Basic and Applied Sciences*, vol. 1, no. 2, pp. 265–276, 2010.

[94] A. B. Godbehere, A. Matsukawa, and K. Goldberg, "Visual tracking of human visitors under variable-lighting conditions for a responsive audio art installation," in *American Control Conference (ACC), 2012*. IEEE, 2012, pp. 4305–4312.

[95] B. K. Horn and B. G. Schunck, "Determining optical flow," *Artificial intelligence*, vol. 17, no. 1-3, pp. 185–203, 1981.

[96] L. Alvarez, J. Weickert, and J. Sánchez, "Reliable estimation of dense optical flow fields with large displacements," *International Journal of Computer Vision*, vol. 39, no. 1, pp. 41–56, 2000.

[97] N. Dhanachandra, K. Manglem, and Y. J. Chanu, "Image segmentation using k-means clustering algorithm and subtractive clustering algorithm," *Procedia Computer Science*, vol. 54, no. 2015, pp. 764–771, 2015.

[98] G. Mathur and H. Purohit, "Performance analysis of color image segmentation using k-means clustering algorithm in different color spaces," *IOSR Journal of VLSI and Signal Processing (IOSR-JVSP)*, vol. 4, no. 6, pp. 01–04, 2014.

[99] F. A. Shmmala and W. Ashour, "Color based image segmentation using different versions of k-means in two spaces," *Global Advanced Research Journal of Engineering, Technology and Innovation (GARJETI*, vol. 1, no. 1, pp. 030–041, 2013.

[100] A. Barjatya, "Block matching algorithms for motion estimation," *IEEE Transactions Evolution Computation*, vol. 8, no. 3, pp. 225–239, 2004.

[101] R. Rakotomalala, "Analyse de corrélation: Étude des dépendances-variables quantitatives," *Document de Cours. Version*, vol. 1, 2012.