Video Object Detection Using Fast And Accurate Change Detection And Thresholding

Chang Su

A Thesis

in

The Department

of

Electrical and Computer Engineering

Presented in Partial Fulfillment of the Requirements

for the Degree of Master of Applied Science (Electrical and Computer Engineering) at

Concordia University

Montréal, Québec, Canada

March, 2007

Your file Votre référence
ISBN: 978-0-494-28929-7
Our file Notre référence
ISBN: 978-0-494-28929-7

Abstract

Video Object Detection Using Fast And Accurate Change Detection And Thresholding

Chang Su

Video object detection is an important video processing technique. Change detection and thresholding based video object detection techniques are widely used due to their efficiency. However, change detection and thresholding in real-world video sequences is challenging due to the complexity of video contents and of environmental artifacts. This thesis proposes a color-based change detection and a video-content adaptive thresholding method for accurate and fast video object detection.

The proposed color-based change detection algorithm is based on the YUV color model, which has been proved as the most effective color model for object detection. First, frame-differencing is carried out in each channel of a video frame. Then, the pixel intensities in both gray-level channel Y and the color channels U and V of the difference frames are statistically modeled. Second, based on the statistical model of the gray-levels in Y channel, an entropy-based blocks-of-interest scatter estimation algorithm is proposed for locating the frame blocks potentially containing moving objects; and based on the statistical models of the color intensities in color channels, a statistical model of the maximum-intensity between U and V channels are obtained. Third, significance test is applied to the detected blocks-of-interest in both gray-level channel and color channels based on the gray-level statistical model of Y channel and the maximum-intensity statistical model of U and V channels. The gray-levels of the non-significant pixels in Y channel but significant in the U or the V channels are then compensated according to their significance probabilities in the color channels. Finally, change masks can be obtained by a thresholding algorithm.

The proposed thresholding algorithm for change detection is based on a change region scatter estimation algorithm and a video-content assessment algorithm to detect

the empty frames and estimate the strength of local unimportant changes. According to the proposed video-content assessment, the global threshold of a difference frame is discriminatively computed. For an empty frame, a noise-statistic based thresholding algorithm with a low false alarm is applied to obtain the threshold. Otherwise, the global threshold is obtained by an optimum-thresholding based artifact-robust thresholding algorithm.

Experimental results show that 1) with the support from the scatter estimation of the blocks-of-interest, the proposed change detection algorithm is efficient and robust to multiple video contents; 2) the proposed thresholding algorithm clearly outperforms the widely used intensity-distribution based thresholding methods and more efficient and more stable than the state-of-the-art spatial-property based thresholding methods for change detection; and 3) the video object detection technique consisting of the proposed change detection and the proposed thresholding algorithms is robust to artifacts and multiple video contents, and is especially suitable for real-world on-line video applications such as video surveillance.

Acknowledgments

I would like to present my deepest appreciation to my supervisor, Dr. Aishy Amer. She not only opened a splendid door to video processing technology for me, but also guide me step by step from a video-technology enthusiast to be a researcher with her intensive wisdom, rigorous work style, and great patience. Her continuous and effective help covers each field of my works, even my technical writing. By her great helping, I have seen my great improvement on my research. I would also like to thank all professors who taught me theory knowledge, and all staffs provided me timely help during my research in Concordia University.

I would like thank my colleagues Firas Achkar and Mohammed Ghazal. They provided me timely and effective supports when I met difficulties in my works. I also thank the VidPro group member Hanif Azhar, Julius Popoola and the former member Bin Qi, they share their excellent ideas with me, and provide me their wise suggestions.

I would like specially thank my wife, Yuan Li. During the whole period of my research, I focus all myself on my works. She understood me, took most of houseworks, and tried her best to guarantee my health. I could not put all my heart into my research without her great support. My deepest gratitude to my parents, my sister, and all my relatives in China. Their loves always support me to pass through each difficulty of life. I would also thank professor Liqiang Han and Dr. Chunning Jin. They always support me and encourage me to fight with all difficulties and realize my dreams.

Contents

List of Figures

List of Tables

List of Notations

Abbreviations

BOI	blocks of interested
CD	change detection
cdf	cumulative distribution function
FD	frame differencing
GEM	generalized exponential model
GM	global motion
GUC	global unimportant changes
HVS	human vision system
JC	Jaccard similarity coefficient
LBP	local binary pattern
LUC	local unimportant changes
MI	maximum intensity
MOG	mixture of Gaussian
OS	operation system
PCC	percentage correctly classified
pdf	probability density function
RCG	regions of change
RV	random variable
VOD	video object detection
YC	Yule coefficient

General Symbols

F_n	the current video frame taken at time constant n
R_n	the reference frame of F_n
$\grave{D}_n$	the signed difference frame between F_n and R_n
D_n	the unsigned difference frame between F_n and R_n
B_n	the binary frame of F_n
$\mathbf{i}$	the spatial location of a pixel
$T(\cdot)$	a thresholding algorithm
$\mathcal{H}_0$	no-change hypothesis
$\mathcal{H}_1$	changed hypothesis
p_h	a high probability
σ_0^2	noise variance in $\grave{D}_n$
W_k	the k-th frame block of D_n
W_k^c	a block of interested containing significant changes
W_k^b	a background frame block
N_k^b	the number of W_k^b in D_n
BK_n	the background of F_n

Effective fast change detection for video object detection

D_n^Y	difference frame in Y channel
D_n^U	difference frame in U channel
D_n^V	difference frame in V channel
$D_n^{Y_c}$	gray-level compensated difference frame of D_n
$\grave{X}$	the random variable for modeling the difference values in $\grave{D}_n$ under $\mathcal{H}_0$
X	the random variable for modeling the gray-levels in D_n under $\mathcal{H}_0$
$\grave{X}_C$	the random variable for modeling the difference values in $\grave{D}_n$ under $\mathcal{H}_1$
X_C	the random variable for modeling the gray-levels in D_n under $\mathcal{H}_1$
σ_ν^2	the noise variance of F_n
σ_b^2	the noise variance of $\grave{X}$
σ_c^2	the noise variance of $\grave{X}_C$
μ_k	the block mean of W_k
H_X	the entropy of X
H_X^k	the entropy of the W_k
T_e	entropy threshold for BOI scatter estimation
$h_d(\cdot)$	the histogram of D_n
$h_w(\cdot)$	the histogram of a block W_k
Y, U and V	random variables for modeling pixel intensities in D_n^Y, D_n^U, and D_n^V, respectively
Z	the random variable for modeling the maximum between U and V
λ_u and λ_v	the mean of U and V, respectively
g	the value of a gray-level
$g_{\mathbf{i}}$	the gray-level of the pixel located in $\mathbf{i}$
$g_{\mathbf{i}}^c$	the gray-level compensated pixel in $\mathbf{i}$
$s_{\mathbf{i}}$	the maximum color intensity in color channels at $\mathbf{i}$
p_s	the significance probability of $s_{\mathbf{i}}$
a_c	the multiplication factor for gray-level compensation
G_s	the maximum intensity that a pixel may have in D_n
α_s	the false alarm for significance test
α_G	the false alarm for determining G_s
t_1, t_2	the block-mean thresholds, and $t_1 > t_2$
α_{t_1}, α_{t_2}	the false alarm for determining t_1 and t_2

Thresholding for change detection

D_n^{ϕ}	an empty difference frame
T_n	gray-level threshold of D_n
T_n^{ϕ}	the threshold of D_n^{ϕ}
$T_n^{'}$	the initial optimum threshold of D_n
$\dot{X}_l$	the random variable for modeling the LUC in $\dot{D}_n$
X_l	the random variable for modeling the LUC in D_n
σ_l^2	the variance of $\dot{X}_l$
H_{X_l}	the entropy of X_l
κ_l	noise enlarge factor for measuring LUC
$H_{X_l}^k$	the block entropy of a W_k^b under $\mathcal{H}_0$ for LUC measurement
κ_l^k	the noise enlarge factor of a W_k^b
μ_n^{ϕ}	the intensity mean of a D_n^{ϕ}
γ_l	the adjusting factor of T_n for LUC adaptation
P_b	the probabilities of the unimportant changes in W_k^c
P_c	the probabilities of the important changes in W_k^c
κ_c	the ratio between σ_c and σ_b
C_l	constant adjust factor for LUC adaptation
a_l	multiplication factor for LUC adaptation
g	a gray-level in the histogram
$h_w^c(\cdot)$	the histogram of $\{W_k^c\}$
p_r	a probability for estimating GUC distribution
α_e	the false alarm for thresholding D_n^{ϕ}
α_l	the false alarm for estimating a_l

Chapter 1

Introduction

1.1 Motivation

In recent years, content-based video systems are driven by the trends of both technology and markets, and are widely used in many applications, including video coding [1], surveillance [2,3], machine vision [4], and medical diagnosis [5]. Compared to the traditional pixel-based and block-based systems, the content-based video systems are, in general, more efficient and more accurate [6].

Video object detection (VOD) is the core of any content-based video system. A VOD technique detects the regions in a video frame that have different semantic meaning based on some criteria (e.g., motion). VOD is usually the first step toward the high-level goal of video-content understanding, e.g., object tracking, event detection, and machine vision. Any failure of VOD may serious affect the performance of the whole video system, and lead to unreliable final outputs. An ideal VOD method should be 1) precise, 2) efficient, and 3) automatic.

Many VOD algorithms have been proposed in literatures. However, classification of the VOD algorithms varies significantly in literatures [7, 8]. After investigating many algorithms, we classify VOD algorithms into four categories: 1) model-matching

based, 2) spatio-temporal based, 3) motion-parameter (or motion estimation) based, and 4) motion-adaptive based.

Model-matching based VOD [9, 10] models video objects as a set of features, or predefined templates. Then VOD is performed by finding the best match between video frames and the object models (or templates). In real-world applications, feature abstraction is often degraded by object occlusion or distortion. While a robust model-matching VOD algorithm is, in general, very time consumming. It may fail when a video sequence contains multiple occluded moving objects. Therefore, model-matching VOD methods are seldom used in on-line applications with multiple objects.

Spatio-temporal based VOD [11–14] generally computes an initial object mask according to temporal measures (e.g., the motion energy) first, then refine the initial object mask by spatial operations. Spatial morphological operations and watershed segmentation techniques are often involved. However, the computation of temporal measures are complex and time consumming. The spatial (such as morphological) operators used in those algorithms may degrade the quality of object details, especially the edges of objects, which may play a pivotal rule in the later processing such as contour tracing. Although some detail-protected morphological operators have been proposed, e.g., [15], their computational complexity is high.

Motion-parameter based VOD algorithms [16–19] are very popular in video applications. They compute the object masks of a video sequence by analyzing the motion vectors. First, the motion parameters of a video sequence are estimated by a motion estimation algorithm. Then, each frame of the video is classified into multiple regions with coherent motion features according to the estimated motion parameters thus object masks are obtained. Although motion-parameter based VOD methods are widely used, they are seriously coupled with motion estimation algorithms which often suffer from non-rigid, occlusion, and aperture problems [8] where estimated

motion fields are inaccurate. In addition, motion-parameter based VOD methods are, in general, computationally expensive due to the time consumming motion estimation procedures. Although some fast motion estimation algorithms [20, 21] exist, their computational complexities and system requirements (e.g., the size of memory buffers) are still relatively high. Thus the motion-parameter based VOD methods have difficulties to be applied in on-line applications.

In motion-adaptive (or change-adaptive) VOD, instead of explicitly estimating motion parameters or modeling video objects, they detect regions of change between the current frame and a reference frame of the same scene taken at different time instants. Change (or binary) masks are initially computed based on either frame differencing or background modeling, and the object masks are then obtained by post-processing procedures such as morphological operations.

Compared to other video VOD categories, motion-adaptive VOD methods have many advantages. Motion-adaptive VOD is, in general, the most efficient compared to other methods. Motion-parameter based VOD has difficulties to detect non-rigid, slow moving, moving-then-stopping, or new appearing objects, yet the motion-adaptive VOD can perform well for such video objects by applying background estimation techniques.

A change (binary) mask obtained by a motion-adaptive method may contain not only the regions caused by the important changes, e.g., moving objects, but also the regions caused by the unimportant changes, including noise, shadows, partial background movement, or local light changes due to door opening. The unimportant regions in change masks lead to artifacts and eventually degrade the quality of object masks. Any artifacts in change masks will increase the workloads of the subsequence high-level postprocessing and decreases the quality of the final object masks. Accurate change detection or background modeling methods are thus essential in

motion-adaptive VOD. If a change detection or a background modeling algorithm uses thresholding to binarize difference frames, then high-performance thresholding is also essential.

An ideal motion-adaptive VOD technique should be 1) robust to the unimportant changes, 2) spatially stable to obtain accurate change masks, and 3) computationally efficient. In this thesis, we propose a motion-adaptive VOD technique, that is based on change detection and thresholding, which meet these three criteria.

1.2 Definition

- Important changes and unimportant changes: we regard the changes caused by moving objects between the current frame and a reference frame as important changes. Changes due to other factors such as noise, illumination changes, shadows, local light changes, partial background movement, etc., are unimportant changes.

- Global unimportant changes and local unimportant changes: the unimportant changes which globally affect the intensities of a video frame, e.g., the changes caused by noise, are regarded as global unimportant changes. If the unimportant changes only affect local areas of a video frames, e.g., the changes caused by shadows, they are regarded as local unimportant changes.

- Noise: there are many types of noise exist in video frames. For example, photon shot noise is caused by the random-arrival photons at the camera sensor which is governed by Poisson statistics. Other types noise include output amplifier noise, camera noise, etc. Due to the high counting effect of photon arrivals and according to the central limit theorem, the aggregate noise effect can be approximated by Gaussian statistics. In addition, the additive white Gaussian noise

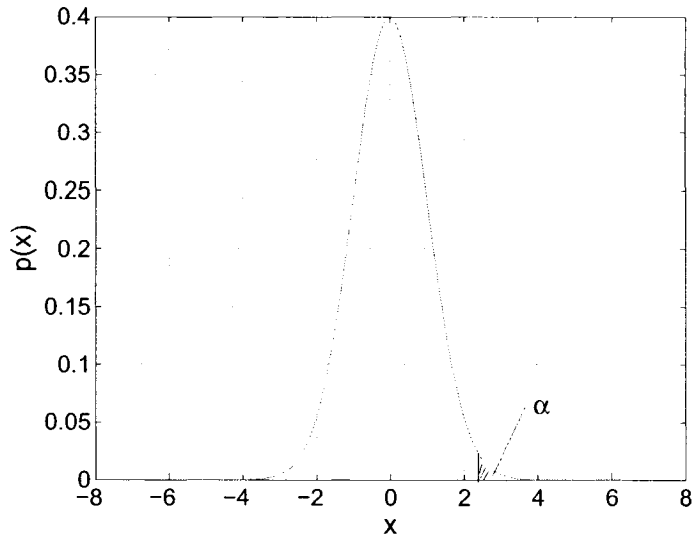


Figure 1.1: The false alarm of a *pdf*.

is the most common noise model for terrestrial TV broadcasting. Therefore, in this thesis, we assume the noise in video frames as AWGN [22, 23].

- Artifact: any factors that cause unimportant changes are artifacts, such as noise, illumination changes, shadows, local light changes due to door opening, partial background movement due to tree leavies waving.

- Change (binary) mask: a change mask identifies the locations of the pixels that are have significantly different intensities or color between the current frame and a reference frame. The changed pixels in a change mask may caused by either important changes or unimportant changes.

- Object mask: an improved change binary mask, where only the pixels caused by important changes are contained, are then called an object mask.

- False alarm: in statistics, false alarm α, $0 \leq \alpha \leq 1$, is the area under the probability density corresponding the right side of a threshold. Fig.1.1 shows an example of a false alarm, where $p(\cdot)$ is the *pdf* of a distribution.

1.3 Related Work

In this section, we review motion-adaptive VOD techniques that are related to the proposed VOD method. Detailed related work review to change detection and thresholding is given in chapters 2 and 3.

Due to the simplicity and low computational cost, motion-adaptive VOD techniques are widely used in real-time video applications [24]. They follow a similar strategy to obtain object masks shown in Fig.1.2: preprocessing, change (or motion) detection, reference-frame updating, and postprocessing. The goal of the preprocessing is to improve the quality of an input frame or perform necessary transformations for later processing. Examples are noise reduction, noise estimation, and global motion estimation and compensation. Change (or motion) detection detects the changes between an input frame (i.e., the current frame) and a reference frame (either a background frame, or the previous frame). If a background frame is not available, or the video conditions (e.g., illumination) vary, background update algorithms are often applied to update the background. Since the change/binary *masks* obtained by the change detection stage may include unimportant changes, (which lead to spurious blobs, gaps and holes), spatial (e.g., morphological) and temporal operations (e.g., change consistency testing) are often employed in a postprocessing stage to improve the reliability of the change masks. The improved change masks are then called object masks since the better resample the physical objects.

1.3.1 Change-detection based VOD

In the past decade, many change detection (CD) algorithms have been proposed in literatures. Radke *et. al* [25] give a good survey of the CD methods. In this thesis, we classify the CD based VOD algorithms into two categories, 1) differencing-thresholding based, and 2) statistical based.

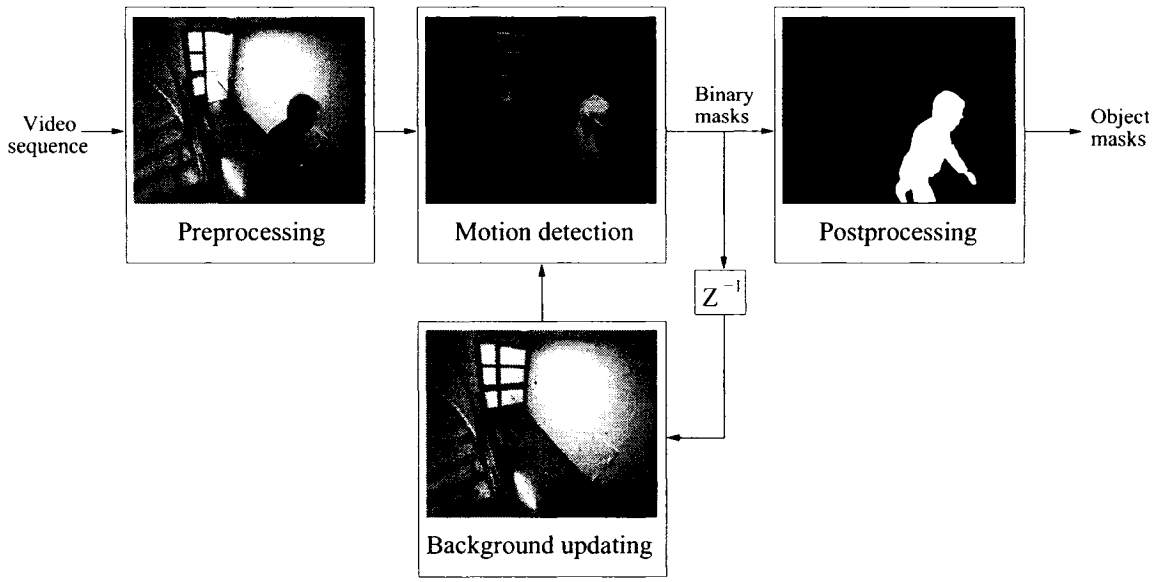


Figure 1.2: The flow diagram of motion-adaptive video object detection

Differencing-thresholding based CD

Frame-differencing (FD) is still popular in CD algorithms due to its simplicity and efficiency. Its basic idea is to 1) obtain the difference frame between the current frame and a *reference* frame by FD followed by some spatial filters such as maximum filters, and 2) detect regions of change in the difference frame by thresholding [24]. Spatial (e.g., morphological) filters are often involved to improve the quality of region (binary) masks obtained. The reference frame can be either a background frame (the CD is then called background subtraction) or the previous frame (the CD is then called temporal differencing) of the input video sequence. Background subtraction is more effective than temporal differencing in detecting changes under complex video object motion, e.g., occlusion, disappearance, new appearance, and non-rigid motions.

C. Kim *et. al* [26,27] detect regions of change in difference frames as follows. First, a robust double-edge map is computed based on temporal differencing CD, then the edges belonging to the previous frame are removed. Object masks are obtained from the remaining edge map, i.e., the moving edges.

The method in [24] obtains first the difference frame between the current frame and a background frame by frame-differencing followed by a spatial low-pass filter and a spatial maximum filter. Then the change mask is computed by thresholding the difference frames. To improve the reliability of the change masks, spatio-temporal adaption is applied to adaptively compute the thresholds. Second, an edge frame is computed by applying a morphological edge detection algorithm to the change masks. A complex-contour tracing algorithm is then applied to the edge frame, and result in the contours map of the moving objects. The object labels are finally obtained by contour filling.

T. H. Chen *et. al* [28] used both background subtraction and temporal differencing. For each current frame, the two change masks of background subtraction and temporal differencing are obtained, respectively. An initial object mask is then obtained by applying four object-region detection rules to the two change masks. The final object mask is then computed by refining the object boundaries in the initial object masks. The background frame can be updated according to the object mask.

Although frame-differencing methods are efficient, they are sensitive to unimportant changes caused by noise, shadows, and or local light changes. Statistical CD aims at making CD more robust to unimportant changes.

Statistical CD

The statistical CD computes change masks by statistical classification algorithms. First, a statistical CD statistically models the properties of video frames (e.g., noise) under some hypothesises. Based on the statistical models, decision rules (e.g., the Bayes rule) are then applied to each pixel of a video frame to classify it into either changed or unchanged class. Due to the robustness to the unimportant changes, statistical based CD algorithms significantly improve the quality of change masks.

Significance test CD is a widely used statistical CD [29–31]. Generally, the significance test of a given pixel can be performed based on two competing hypotheses, the no-change hypothesis $\mathcal{H}_0$, where we assume that the difference between the current frame F_n and the reference frame R_n is caused by noise only, and the changed hypothesis $\mathcal{H}_1$. Significant test CD is based on $\mathcal{H}_0$, and can be carried out as

$$p(\grave{D}_n(\mathbf{i})|\mathcal{H}_0) \gtrless_{\mathcal{H}_1}^{\mathcal{H}_0} p_h, \tag{1.1}$$

where $p(\cdot)$ is a conditional probability density function (pdf), $\mathbf{i}$ is the spatial location of a pixel, $\grave{D}_n$ is the frame difference (signed) between the current frame F_n and the reference frame R_n, p_h is a relatively high probability and can be computed from a desired false alarm. To improve the robustness of the classification to noise, significance test is usually performed on a block centered at $\mathbf{i}$.

Significance test can be performed on pixel-wise or block-wise. Pixel-wise significance test is efficient, but it is sensitive to unimportant changes. Block-wise significance test is more robust to the unimportant changes than the pixel-wise significance test, it is time consuming.

A. Ach *et. al* [32, 33] model the intensity distribution of $\grave{D}_n(\mathbf{i})$ under $\mathcal{H}_0$ as a Gaussian distribution with zero mean and variable σ_0^2. To robust significance test, the block-wise significance test is used. The statistic of the block-based significance test follows a χ^2 distribution. The decision threshold is then determined by the χ^2 table with a specific false alarm. Examples of the significance test CD based VOD are in [34–37]. In addition to CD under $\mathcal{H}_0$, statistical based CD can also be performed under $\mathcal{H}_1$, e.g., [33, 38].

Since significance test CD only takes noise into account, it is sensitive to local unimportant changes, e.g., shadows, local light changes, and partial background movement such as leaves waving. Therefore, the VOD methods based on the signifi-

cance test CD have difficulties to be applied to the outdoor systems.

Significance test CD has difficulties when being applied to real-world video systems. The test algorithms based on hypothesis $\mathcal{H}_0$ assumes that the artifacts in unchanged areas of a difference frame are only caused by noise. Since the artifacts in real-world video sequences are caused by not only noise but also other global unimportant changes (e.g., illumination variation) and local unimportant changes (e.g., shadows). The $\mathcal{H}_0$ based significance-test CD is very sensitive to such non-noise artifacts. While the testing algorithms based on $\mathcal{H}_1$ have difficulties on estimating the paremeters of the statistical models when being applied to real-world applications.

1.3.2 Background-modeling based VOD

In background-modeling based VOD, the background of a video sequence is first adaptively modeled. Then change (binary) masks are computed based on the background model by a CD algorithm, e.g., frame differencing, or statistical classification. Finally, object masks are obtained by postprocessing where high-level analysis for object refinement and artifact reduction are involved.

Binary-mask based background modeling

The simplest way to dynamically model the background of a video sequence is to estimate the background based on the estimated change binary masks in the difference frames.

E. P. Ong *et. al* [39] compute object masks based on background subtraction and updating. First, change masks of a video sequence are obtained by background subtraction followed by thresholding. Then the reference (background) is updated by temporal averaging algorithm based on the change masks. Third, the moving edges are obtained by edge differencing between the current frame and the updated

reference frame. The object masks are finally obtained by abstracting individual connected components from the edge map and refining them using a graph-based edge linking algorithm.

Similarly, J. Zhang *et. al* [40] also detect moving objects using background updating. After obtaining the change mask of the current frame by background subtraction, they update the reference frame by thresholding two memory matrices, which are computed from change masks. Then, moving edges are obtained by an edge differencing method between the reference frame and the current frame. To improve the quality of moving edges, morphological filters and a connected component analysis algorithm are employed.

Temporal-median based background modeling

Temporal median filter technique is a common used background modeling method in VOD algorithms. R. Cucchiara *et. al* [41] obtain a background model by computing the temporal median of each pixel of L frames in memory buffer. Change masks are then obtained by thresholding the frame difference between the adaptive background model and the current frame. Blob analysis for detecting moving objects, shadows and ghosts is carried out by applying a set of decision rules to the region-based labeled change masks, and eventually obtain the object masks. D. Farin *et. al* [42] also propose an VOD algorithm based on temporal median filter based background modeling. Although temporal median filter based background-modeling is efficient and effective, they are sensitive to the video contents, e.g., the moving directions objects.

Statistical background modeling

An effective background modeling approach uses statistical properties of changes between frames.

A. Cavallaro *et. al* [43] and H. Han *et. al* [44] model the background of a video sequence as a single Gaussian distribution. Foreground pixels are determined by testing if their intensities are matched with the Gaussian model. However, the video conditions are complex in real-world applications. Modeling background with a single Gaussian distribution may fail for the video sequences containing serious object occlusion, shadows, local light changes, etc.

C. Stauffer *et. al* [45] employ a mixture Gaussian distribution to model the background of a video sequence to overcome the problems mentioned above. They consider the value of a given pixel as a pixel process and model the recent history of each pixel $\{F_1(\mathbf{i}), F_2(\mathbf{i}), \cdots, F_n(\mathbf{i})\}$ as a mixture of K Gaussian (MOG) distribution

$$P(F_n(\mathbf{i})) = \sum_{k=1}^{K} \omega_{k,n} \cdot \eta(F_n(\mathbf{i}), \mu_{k,n}, \Sigma_{k,n}), \tag{1.2}$$

where K is the number of Gaussian distributions, $\omega_{k,n}$ is the portion of the data accounted for the k-th Gaussian distribution, and $\eta(F_n(\mathbf{i}), \mu_{k,n}, \Sigma_{k,n})$ is the *pdf* of the k-th Gaussian distribution with the mean $\mu_{k,n}$ and variance $\Sigma_{k,n}$. For each given new pixel $F_n(\mathbf{i})$, a matching detection is applied to find a match between $F_n(\mathbf{i})$ and the K exist Gaussian distributions. If no match are found, the weight $\omega_{k,n}$ of the least probable Gaussian distribution is updated by a pre-defined learning rate. Otherwise, $\mu_{k,n}$ and $\Sigma_{k,n}$ are updated based on $F_n(\mathbf{i})$ and a learning factor. The first few Gaussian distributions are then chosen as the background model.

The background-modeling algorithm proposed in [45] works well for VOD methods in real-world applications such as video surveillance [46], and becomes the standard

formulation for the mixture approach of VOD algorithms in recent years. However, it is sensitive to the local unimportant changes such as shadows. Another shortcoming of [45] is that it has slow convergence when one of the Gaussian distributions adapts to a new cluster. To overcome these problems, many revised MOG CD based VOD algorithms are proposed. For example, P. Kaewtrakulpong *et. al* [47] propose a new algorithm for updating the component parameters as well as a moving shadow detection algorithm, Z. Zivkovic [48] introduce the dynamic learning rate to cope with the problems caused by transient components and an adaptive algorithm algorithm to determine the number of components. D. S. Lee *et. al* [49] proposes an effective learning algorithm to overcome the slow convergence problem of [45] and improve the estimation accuracy. A component classification algorithm based on a posterior probability analysis is also proposed in [49].

In addition to modeling the pixel intensities as shown in [45, 47–49], the background model can also be obtained by modeling textures. M. Keikkila *et. al* [50] propose an VOD method based on texture-based background modeling algorithm. First, a pixel in a video frame is modeled by a set of local binary pattern (LBP) histograms with different weights. Second, the LBP histogram of the pixel located in the same position in a new frame is compared to the LBP model histograms. If matches are found, the best matching model histogram as well as its weight is updated adaptively with a user-settable learning rate. Third, according to the persistence of the background LBP, the histograms with relatively high weights are selected as background histogram. The detection of moving objects is done before background updating by testing if their LBP histogram matches at least one background histogram.

The background-modeling based VOD methods are usually sensitive to the video contents. They may suffer when the moving directions of objects are parallel to the axes of the camera. They are also sensitive to the local unimportant changes.

This is because that local unimportant changes usually statistically different from the background thus the VOD methods may mistakenly classify a local unimportant change as an important change.

1.3.3 Hybrid methods

Y. Feng *et. al* [51] compute the moving edges in video frames by thresholding the edge-distance between the difference edge maps generated based on frame-differencing CD and the edges of the original frames. The edges in the original frames with the edge-distance less than the threshold are regarded as the moving edges. Then Kim's [26] object detection algorithm is applied to the moving edge maps thus obtain the object masks. G. Zhang *et. al* [52] detect video objects based on both motion-parameter estimation and change detection. First, a coarse video mask is obtained by thresholding the fuzzy matrix, which is obtained based on the change mask and the estimated motion field of a video frame. Then the initial object mask is then obtained by filling its coarse version. Postprocessing is employed to refine the initial object mask. Although the method is robust to multiple video contents, it is computational expensive for including both CD and motion parameter estimation.

1.4 Overview Of The Proposed VOD Method

We propose a VOD method based on an effective color-based CD algorithm and an artifact-robust thresholding algorithm for CD.

The proposed CD algorithm aims at 1) presenting a fast and accurate solution to the problem when the foreground and the background of a video sequence are similar in gray-level, and 2) improving the robustness of change masks to strong local-unimportant such as shadows and local light changes due to door opening. The

YUV color model is employed in the proposed CD algorithm for effective CD [53].

First, frame-differencing followed by absolute-value operation is performed in each of Y, U, and V channels thus obtain the difference frame of each channel. To improve the robustness of the CD to unimportant changes, the gray-levels in Y channel and the color intensities in U and V channels are statistically modeled as a Gaussian distribution (Y) and two exponential distributions (U and V). The statistical model of the maximum color intensities between U and V channels is then obtained based on the two exponential distributions. Second, an entropy based scatter estimation of the blocks-of-interest (BOI) (in short, BOI estimation) is applied to the difference frame of Y channel to indicate the frame blocks which potentially contain moving objects. Third, a significance test algorithm is applied to the pixels in BOI based on the obtained statistical models. The gray-levels of the pixels which are non-significant in the Y channel but significant in U or V channel are compensated according to their significance probabilities in chrominance channels. Finally, the change mask is obtained by applying a thresholding algorithm to the gray-level compensated frame difference.

Note that the proposed BOI estimation is independent of the proposed CD method. It can be also employed in other content-dependent video applications, e.g., video-content assessment [1].

The proposed thresholding algorithm is for CD aims at improving the robustness of change masks to both global unimportant changes (e.g., noise and illumination changes) and the local unimportant changes (e.g., shadows and local light changes). First, a video-content assessment algorithm based on the BOI estimation mentioned above is proposed to 1) detect the empty frames where no moving objects or only extremely small moving objects exist, and 2) adaptively estimate the strength of the local unimportant changes. The strength of the local unimportant changes is

defined as the intensity-variance in this thesis. Then a content-adaptive thresholding algorithm is proposed to discriminatively compute the global threshold binarize the difference frame. If the current frame is an empty frame, we compute the threshold by a noise-statistic based thresholding method with a low false alarm. Otherwise, we compute an initial threshold based on the analysis of optimum thresholding, and then refine the initial threshold by adapting it to the strength of local unimportant changes.

1.5 Summary Of Contributions

Effective fast color-based change detection

- An entropy-based scatter estimation of the blocks-of-interest of video frames.
- An exponential-distribution based statistical model for the pixel intensities in chrominance channels of difference frames under no-change hypothesis.
- A maximum-intensity distribution based significance test algorithm for pixel-classification in chrominance channels under no-change hypothesis.
- A color-based gray-level compensation algorithm based on the significance probability of maximum-intensity of difference frames.

Fast and artifact-robust thresholding for CD

- A fast video contents assessment algorithm based on the blocks-of-interest scatter estimation and noise estimation. This algorithm consists of
  - a simple but effective empty frame detection algorithm
  - a fast local unimportant changes measurement

- A fast optimum threshold estimation for CD without statistical parameter estimation.
- A video-content assessment based discriminative thresholding algorithm for CD.

Also several widely used thresholding algorithms [54–57] for CD are studied and implemented, and their performance is compared with the proposed thresholding method.

1.6 Thesis Organization

This thesis is organized as follows. Chapter 2 describes the proposed color-based CD algorithm and presents related work. Chapter 3 presents the proposed video-content adaptive thresholding algorithm and its related work. Experimental results to CD and thresholding are present in Section 4.3 and 4.4 of Chapter 4. Comparion between the proposed VOD approach (combine the proposed CD and the proposed thresholding) and a state-of-the-art VOD algorithm [49] are given in Section 4.5 of Chapter 4. Conclusion and further works are given in Chapter 5.

Chapter 2

Effective Fast Color-based Change Detection For Object Detection

This chapter is organized as follows. Section 2.1 describes the commonly used color models. Section 2.2 presents the related work. Section 2.3 presents an overview of the proposed CD algorithm. Section 2.5 presents a fast blocks-of-interest scatter estimation algorithm for detecting the frame blocks potentially containing moving objects. The details of the proposed algorithm are given in Section 2.6. Section 2.7 summaries this chapter.

2.1 Color Models For Video Object Detection

Color information is becoming widely used today in video processing due to accuracy. Different color models are used. The RGB model is the most commonly used color model in practice, and used in all three TV systems (NTSC, PAL, and SECAM) as primary color [6]. It describes a color by three components, i.e., red, green, and blue. Although the RGB model is suitable for capturing or displaying video frames, it does not separate the luminance component and chrominance components. This

is disadvantageous for video object detection due to the dependence between the illumination and the chrominance components.

To separate the illumination component and the chrominance components in video frames, several color models are proposed. The HSV model describes a color by hue, saturation, and value. Hue component is related to the gradation of color within the visible spectra of light. Saturation component describes the purity of a hue. In general, high saturated hue leads to vivid color, and low saturated leads to muted and gray. Value component is the brightness of the color. The HSV model is commonly used in computer graphics applications.

The CIE XYZ model is obtained from the RGB model as shown in (2.1) and it separates the illumination and chrominance components by using the luminance component Y and chrominance components X and Z. It can specify almost all visible colors, however it is not realizable by actual color stimulation [6]. Therefore, the XYZ model is not directly used in practice but used to define other color models, e.g., the YUV color model.

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 2.365 & -0.515 & 0.005 \\ -0.897 & 1.426 & -0.014 \\ -0.468 & 0.089 & 1.009 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}. \tag{2.1}$$

The LUV model, where L is the luminance component and U and V are chrominance components, is proposed to describe all the colors visible to the human vision system. First, the RGB model is converted to the XYZ model, and then the LUV model is obtained based on the XYZ model. The LUV color model is efficient for video object detection, however, A. Chikando *et. al* [53] show that it does not outperform the YUV color model.

The YUV color model represents the luminance by component Y, and two color-

difference based chrominance components by U and V. Although YUV model is used in PAL, the color models used in NTSC and SECAM systems are also derived from YUV model [6]. Based on the relation between RGB model and the XYZ model as shown in (2.1), we can get the Y component. The U and V components are computed based on color differences $B - Y$ and $R - Y$, respectively. The convertion between the RGB model and the YUV model is then

$$\begin{bmatrix} Y \\ U \\ V \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ -0.147 & -0.289 & 0.436 \\ 0.615 & -0.515 & 0.100 \end{bmatrix} \begin{bmatrix} \tilde{R} \\ \tilde{G} \\ \tilde{B} \end{bmatrix}, \tag{2.2}$$

where $\tilde{R}, \tilde{G}$, and $\tilde{B}$ are normalized R, G, and B intensities after gamma-correction.

After investigating the most frequently used color models such as RGB, HSV, YUV, etc., A. Chikando *et. al* [53] show that the YUV model is more efficient for object detection compared to other color models (e.g., RGB model). This is because the distributions of the different object regions in a video frame described the YUV format have less overlap than the distributions of the same object regions described by other color models. We therefore use the YUV color model in this thesis.

2.2 Related Work

Change detection (CD) can be carried out in both gray-level and chrominance channels. Due to their low complexity (both storage and computation), most CD methods are gray-level based, e.g., CD in [32, 33]. However, gray-level only contains illumination information of a video frame thus the gray-level based CD suffers when the foreground and the background of a video sequence are similar to each other in gray-level. This is very often happened in real-world video applications. A color frame can

provide much more information than a gray-level one. Since a recognizable object is, in general, different from the background in at least one chrominance channel, color information is a good aid for accurate CD in real-world video sequences. As the computational abilities of computer systems increase in recent years, color-based CD methods become more and more popular.

Color-based CD methods can be classified into three categories: 1) logic-operator based, 2) color-distance based, and 3) color-vector modeling based color CD. The logic-operator based color CD is the simplest solution to employ color information in CD. First, change detection is carried out in all chrominance channels of a video frame thus obtain the change mask of each chrominance channel. Then, the final change masks are obtained by combining all the change masks of different chrominance channels by logic operators, e.g., "OR", "AND".

Logic-operator based color CD can directly use gray-level based CD methods [24, 32, 33]. A. Cavallaro *et. al* [58] propose a color-edge based CD. The difference frame of each chrominance channel is obtained by simple-differencing CD first. Then the Sobel edge detector is applied to each difference frame, and the moving edge mask is obtained by combining all edge maps obtained in different chrominance channel by logic "OR" operator. The change mask is finally obtained by a morphologic filling filter. Stefano *et al.* [59] propose a content-adaptive CD using image structure and color, and they also compute a frame-level change mask by combining all change masks of chrominance channels with logic "OR" operator.

Logic-operator based color CD is simple. However, it is computationally and storage expensive since it performs CD in all chrominance channels. Another disadvantage of the method is that it is very sensitive to artifacts. For example, when "OR" operator is used, the false positives in the change masks of any chrominance channels will be included into the final change mask, and when "AND" operator is

used, the false negatives in the change masks of any chrominance channel will be included into the final change mask.

To improve the efficiency and the robustness of logic-operator based color CD, T. Alexandropoulos *et al.* [60] propose a block-wise cluster-distance based CD method for real-time video applications. First, they statistically model the noise in video frames by a block clustering algorithm, then the change masks of each chrominance channel is obtained by thresholding the cluster-distance between a given frame block and the largest cluster which carries noise model information. The final change mask is then obtained by applying logic "OR" operator to all change masks of different chrominance channels. The algorithm is proposed for real-time video surveillance applications in public.

The color-distance based color CD performs CD based on the color distance between the current frame and the reference frame in color space, e.g., the RGB space, or YUV space. Instead of computing the intensity difference in each chrominance channel, the color distances of the pixel pair between the current frame and the reference frame are computed first. Then the change mask is obtained by statistically classifying the color-distance into different categories. The Euclidean distance and the Mahalanobis distance are often used in the CD method. Y. Hwang *et. al* [61,62] propose a color-based CD based on Euclidean color distance. First, they model the pixel intensities in each chrominance channel as a generalized exponential model (GEM). According to the GEM, they deduce the statistical model of the Euclidean color distance between the current frame and the reference frame under hypothesis $\mathcal{H}_0$. Based on the statistical model of color-distance, change masks are obtained by a pattern classification algorithm.

The color-distance based CD comprehensively takes color information from different chrominance channels into account instead of independently performing CD

in each chrominance channel like logic-operator CD does, thus it improves the robustness of the change masks to artifacts. However, the distance computation is also time consumming, and the CD method is not an efficient solution for big-size video sequence. Also, the color-distance measure is sensitive to local unimportant changes.

The color-vector modeling based color CD is very popular in video surveillance applications due to its temporal stability and the robustness to global unimportant changes. First, the intensities of a pixel in a video frame in all chrominance channels are statistically modeled as a vector random variable and described by the *pdf*. Then the change mask is obtained by statistical CD algorithms. The recently wide interested background-modeling VOD algorithms [45, 49, 63] are good examples for the applications of color-vector based CD. E. Durucan *et al.*'s color Gramian matrix based color CD [64] is an another example of the color-vector based algorithm but without statistically modeling color vectors, however, the Gramian-matrix based method is extremely computationally expensive and sensitive to artifacts.

The disadvantages of the color-vector modeling based color CD are 1) they are sensitive to local changes since even slight local changes may lead to significant difference between the current frame and the reference frame in one or more chrominance channel, e.g., shadows. Statistical classification may mistakenly classifies such unimportant changes into foreground. This will increase the workload of the postprocessing; 2) the parameter on-line estimation for statistically modeling the color-vector is time consumming, and may need long time to reach convergence.

2.3 Overview of Proposed Algorithm

In this chapter, we propose to introduce color information into gray-level CD to overcome the problems of 1) inaccuracy caused by similarity between foreground and background in gray-level, 2) unimportant-change sensitivities in different channels, 3)

computational and storage load using the three channels. Different from the previous color-based CD work, we content-adaptively include color information into gray-level CD by a gray-level compensation algorithm without performing CD in all channels or compute color-distances.

The proposed algorithm is based on the observation that a recognizable object in a color video sequence is different from the background in at least one of the channels Y, U, and V.

As shown in [25], the changes existed in a difference frame D_n can be classified into two categories, one is the important changes which are caused by object motion, the other is the unimportant changes which may be both global (e.g., noise, and illumination changes) and local (e.g., shadows, and local illumination changes due to door opening).

Fig.2.1 shows the flow chart of the proposed algorithm. First, three difference frames D_n^Y, D_n^U and D_n^V are obtained by frame-differencing followed by the absolute-value operation, and a spatial average filter with size $w \times w$ in the Y, U, and V channels between the current frame F_n at time instant n and its reference frame R_n. The spatial average filter is important to reduce noise, and its size w can be 3 or 5. The gray-levels in the Y channel under hypothesis $\mathcal{H}_0$ is statistically modeled as a Gaussian random variable (RV), and the color intensities in U and V channels under $\mathcal{H}_0$ are modeled as two exponential RVs by taking not only noise but also other unimportant changes into account (Sec.2.6.1). A maximum-intensity (MI) statistical model between the U and the V channels is then obtained to statistically detect the significant changes in U and V channels. Second, based on the statistical model of the Y channel, a blocks-of-interest (BOI) scatter estimation algorithm (Sec.2.5) is proposed to indicate the frame blocks that potentially contain moving objects in D_n^Y. Third, the pixel-based significance test algorithms based on the Gaussian model of the

Y channel and the MI model of the U and the V channels are applied (Sec.2.6.2) to the BOI, and the proposed gray-level compensation algorithm (Sec.2.6.3) is applied to the pixels which are non-significant in the BOI blocks of D_n^Y but significant in D_n^U or D_n^V according to the significance probabilities they have in the chrominance channels. Thus, the gray-level compensated difference frame $D_n^{Y_c}$ is obtained. Finally, the change masks are obtained by a classification algorithm such as thresholding [55] or statistical classification [32, 33].

2.4 Computation of Difference Frames

Difference frames are the base of the proposed as well as most change detection methods. A difference frame may be signed or unsigned dependent on its applications, and an unsigned difference frame can be obtained from its signed version. A signed difference frame $\grave{D}_n$ is obtained by frame differencing between the current frame F_n and a reference frame R_n, as

$$\grave{D}_n(\mathbf{i}) = F_n(\mathbf{i}) - R_n(\mathbf{i}), \tag{2.3}$$

where $\mathbf{i}$ is the spatial location of a pixel.

Although a signed $\grave{D}_n$ can be used in statistical CD methods, where statistical classification algorithms are used to obtain change masks, it can not be used in frame-differencing and thresholding based CD methods (e.g., simple-differencing CD). This is because the important changes in a signed $\grave{D}_n$ between different frames may be either positive or negative, and a threshold obtained by a thresholding method may mistakenly classify the negative important changes into the background. By applying the absolute-value operation to a signed $\grave{D}_n$, we can obtain an unsigned difference frame D_n where only positive values exist. Thus the change mask of an unsigned D_n

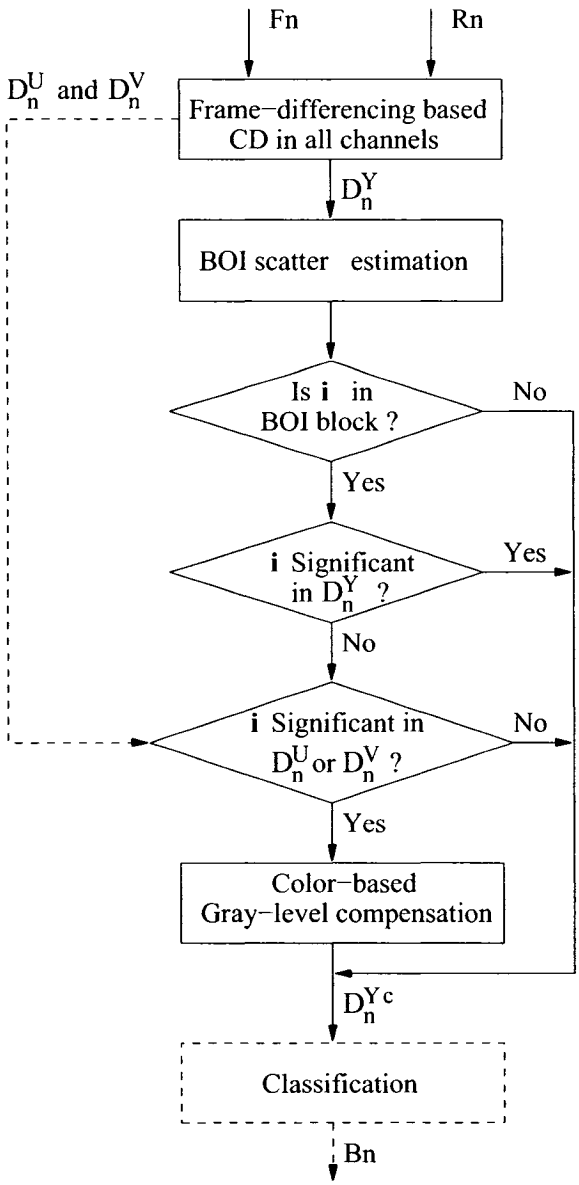


Figure 2.1: Flow chart of the proposed algorithm.

can be obtained by both statistical-classification based and thresholding-based CD methods. In this thesis, we use unsigned difference frame D_n as

$$D_n(\mathbf{i}) = |\grave{D}_n(\mathbf{i})|. \tag{2.4}$$

2.5 Blocks-of-interest scatter estimation in Y

In this section, we first propose an entropy-based algorithm to estimate the scatter of the BOI in a difference frame. This is important for further content-dependent processing of the proposed approaches. The goal of BOI estimation is to indicate the frame blocks which contain strong changes thus potentially contain moving objects. The BOI estimation is a pivotal step in the proposed CD algorithms.

Objects in video sequences have either smooth surfaces or textures. Compared to pixel intensities belonging to background regions in a difference frame, the pixel intensities belonging to object regions have 1) relatively high difference values if the objects have smooth surfaces, and/or 2) relatively large range of variation if the objects have textures. The two properties of the object regions can be statistically represented by the mean and the variance of the pixel intensities. Therefore, the BOI estimation can be done by analyzing the statistical descriptions of frame blocks.

2.5.1 Statistical modeling of gray-level in difference frames

The values in a signed difference frame $\grave{D}_n$ (given in Eq.(2.3)) can be between -255 to 255 where 1) low absolute values (e.g., -5 to 5) are high probable caused by noise and classified as no-change areas; and 2) high absolute values (e.g., -6 to -255 or 6 to 255) are high probable caused by moving objects or by strong artifacts such as sudden illumination charges and classified as changed areas. Thus the intensities modeling

in difference frames is usually performed in the two competitive hypothesises, the no-change hypothesis $\mathcal{H}_0$ and the changed hypothesis $\mathcal{H}_1$. The no-change areas in signed $\grave{D}_n$ can be modeled as a Gaussian RV with zero mean and variance σ_b^2, and the changed areas in signed $\grave{D}_n$ can be modeled as a Gaussian RV with zero mean and variance σ_c^2 [32, 33].

Under no-change hypothesis $\mathcal{H}_0$, to obtain the statistical model of the gray-levels in an unsigned difference frame D_n, we first obtain the statistical model of a signed difference frame $\grave{D}_n$. Based on the assumptions that the noise in a video sequence is AWGN, we model the noise in F_n and R_n as two independent identical Gaussian RVs X_f and X_r with zero mean and variance σ_ν^2. Under $\mathcal{H}_0$, the noise in the signed difference frame $\grave{D}_n$ can be modeled as a RV $\grave{X}$ as

$$\grave{X} = X_f - X_r. \tag{2.5}$$

Since the output of a linear function of two Gaussian RVs is still a Gaussian RV [65], $\grave{X}$ is a Gaussian RV. The mean of $\grave{X}$ is

$$E[\grave{X}] = E[X_f - X_r] = 0, \tag{2.6}$$

and the variance of $\grave{X}$ is

$$Var[\grave{X}] = E[\grave{X}^2] - E[\grave{X}]^2, \tag{2.7}$$

since $E[\grave{X}] = 0$, we have

$$\begin{aligned} Var[\grave{X}] &= E[(X_f - X_r)^2] \\ &= E[X_f^2] - 2E[X_f X_r] + E^2[X_r]. \end{aligned} \tag{2.8}$$

Because X_f and X_r are RVs with zero mean and variance σ_ν^2, we have $E[X_f^2]$ and

$E[X_r^2]$ are equal to σ_ν^2. Since X_f and X_r are independent, we get

$$E[X_f X_r] = E[X_f]E[X_r] = 0. \tag{2.9}$$

Eq.(2.8) becomes

$$Var[\grave{X}] = 2\sigma_\nu^2. \tag{2.10}$$

Therefore, $\grave{X}$ is a Gaussian RV with zero mean and variance $2\sigma_\nu^2$, and the *pdf* of $\grave{X}$ is then

$$\grave{X} \sim \frac{1}{\sqrt{2\pi}\sqrt{2}\sigma_\nu} e^{-\frac{x^2}{4\sigma_\nu^2}}. \tag{2.11}$$

After applying the absolute-value operation to $\grave{D}_n$, the unsigned difference frame D_n is obtained (given in Eq.2.4). Compared to $\grave{D}_n$, the occurrence of a specific intensity value in D_n is statistically doubled since a negative difference value in $\grave{D}_n$ is converted to a positive difference value with the same absolute value. Assume that the gray-levels in D_n is modeled as a RV X, $X \geq 0$, from (2.11) we get the *pdf* of X as [57]

$$X \sim \frac{2}{\sqrt{2\pi}\sigma_b} e^{-\frac{x^2}{2\sigma_b^2}}, \quad X \geq 0, \tag{2.12}$$

where $\sigma_b^2 = 2\sigma_\nu^2$, and σ_ν^2 can be obtained by a noise estimation algorithm such as [66].

Gray-level modeling in changed areas is challenging due to the complexity of video contents. Simulations in [32, 33] show that the difference values in a signed $\grave{D}_n$ under both $\mathcal{H}_0$ and $\mathcal{H}_1$ can based modeled as two zero-mean Gaussian RVs with different variances. We obtain the gray-level statistical model in unsigned D_n under $\mathcal{H}_1$ based on the Aach *et. al*'s gray-level model in signed $\grave{D}_n$ under $\mathcal{H}_1$ as follows. Under hypothesis $\mathcal{H}_1$, assume that the difference values in signed $\grave{D}_n$ is modeled as a Gaussian RV $\grave{X}_C$ with zero-mean and variance σ_c^2 [32, 33]. Similar to the derivation of (2.12), the gray-level in unsigned D_n under $\mathcal{H}_1$ is then modeled as a RV X_C with

the *pdf*

$$X_C \sim \frac{2}{\sqrt{2\pi}\sigma_c} e^{-\frac{x_c^2}{2\sigma_c^2}}, \quad X_C \geq 0, \tag{2.13}$$

where σ_c^2 is the variance of the changed areas in signed $\grave{D}_n$. As having been proved in [32, 33, 57], $\sigma_c^2 \gg \sigma_b^2$. This means that the intensity variation in the object regions is much greater than the intensity variation in the background regions.

2.5.2 BOI scatter estimation

The BOI scatter estimation is based on the statistical models of the gray-levels in D_n under $\mathcal{H}_0$ and $\mathcal{H}_1$. We first divide D_n into blocks W_k, $k = 1, 2, \cdots, N$, of equal size, N is the number of blocks in a frame.

From (2.12) and (2.13), the means of X and X_C are

$$\begin{aligned} E[X] &= \int_0^\infty x \cdot 2\frac{1}{\sqrt{2\pi}\sigma_b} e^{-\frac{x^2}{2\sigma_b^2}} dx \\ &= -\frac{2\sigma_b}{\sqrt{2\pi}} \left[e^{-\frac{x^2}{2\sigma_b^2}} \right]_0^\infty \\ &= \sqrt{\frac{2}{\pi}}\sigma_b \\ E[X_C] &= \sqrt{\frac{2}{\pi}}\sigma_c. \end{aligned} \tag{2.14}$$

As can be seen, in (2.14), $E[X_C] \gg E[X]$. Therefore, we can detect if a frame block potentially contains moving objects by testing if the mean and the variance of the frame block are both high enough. However, the variance of a block may change greatly for different block size, and the variance computation is computationally expensive.

The entropy of a random variable is the expected value of the uncertainty of the outcomes of the random variable [65]. Therefore, entropy is good measure to test the uncertainty of a random variable. It is suitable for detecting if the intensities in a frame block vary drastically. If the block is an object block, then this block

shows either high entropy in the difference frame if the original object has texture, or high mean (in both cases where the original object has either texture or smooth surface). For reliable estimation, we combine the entropy measure and the intensity mean measure to detect BOI frame blocks.

First, we compute the differential entropy H_X of X under no-change hypothesis $\mathcal{H}_0$ as

$$H_X = -\int_{-\infty}^{\infty} p_X(x)\ln(p_X(x))\,dx = -E[\ln(p_X(x))], \tag{2.15}$$

where $p_X(x)$ is the pdf of RV X, and $\ln(\cdot)$ is the natural logarithm function. From (2.12), we get

$$\begin{aligned} H_X &= -\int_0^{\infty} 2\tfrac{1}{\sqrt{2\pi}\sigma_b} e^{-\frac{x^2}{2\sigma_b^2}} \cdot \ln 2\tfrac{1}{\sqrt{2\pi}\sigma_b} e^{-\frac{x^2}{2\sigma_b^2}}\,dx \\ &= -\ln\tfrac{2}{\sqrt{2\pi}\sigma_b} + \tfrac{1}{\sqrt{2\pi}\sigma_b^3}\int_0^{\infty} x^2 e^{-\frac{x^2}{2\sigma_b^2}}\,dx \\ &= \tfrac{1}{2}\ln(\tfrac{1}{2}\pi e\sigma_b^2). \end{aligned} \tag{2.16}$$

From (2.16), we note that the value of H_X depends on σ_b^2, where $\sigma_b^2 = 2\sigma_\nu^2$. Theoretically, H_X should be constant for each block in the unchanged areas in a video frame with a know noise level. Since the intensity variation in the changed areas is much more greater than the intensity variation in the unchanged ares [32,33], i.e., $\sigma_c^2 \gg \sigma_b^2$, we can detect a block-of-interest by testing if the block entropy H_X^k of W_k is high enough compared to H_X.

Block entropy H_X^k can be obtained from (2.15). Since the block histogram is an estimation of the real *pdf* of the intensities in a block, in practice, we compute the entropy H_X^k of the k-block in D_n as

$$H_X^k = -\sum_{g=0}^{g_{max}} h_w(g)\ln(h_w(g)), \tag{2.17}$$

where g is a gray-level, $h_w(g)$ is the probability of gray-level g in W_k, and g_{max} is the maximum non-zero frequency gray-level in $h_w(g)$. From (2.16), we note that the

entropies of the blocks belonging to the background should be similar to each other, i.e., their value are stable. An effective and efficient way to detect BOI is to test if H_X^k is greater than a threshold T_e.

From (2.16), we can simply set $T_e = H_X$, however, this is not reliable in real-world applications because the non-zero intensities in the unchanged area in D_n are caused by not only noise but also other artifacts (e.g., illumination changes, shadows, etc.), the H_X^k of a background block may greater than H_X. For robust BOI scatter estimation, we adaptively estimate T_e as follows. First, we sort $\{H_X^k\}$ in descending order to obtain an unimodal curve of block entropy. Then, we apply the unimodal thresholding algorithm in [67], which is proposed for fixing the corner of an unimodal distribution curve, to fix T_e.

The entropy-measure is suitable for detecting BOI when the original objects have textures. By taking the case that the original objects have either smooth surfaces or textures, we classify a block W_k in a difference frame D_n (given in Eq.2.4) to be either BOI W_k^c or background blocks W_k^b as follows: we regard W_k as a W_k^c if its intensity mean μ_k is higher than a very high threshold t_1; or its μ_k is higher than a relatively high threshold t_2 and its H_X^k is greater than T_e, otherwise, we regard the W_k as a W_k^b, i.e.,

$$W_k = \begin{cases} W_k^c & : \ (\mu_k > t_1) \bigvee \{(\mu_k > t_2) \bigwedge (H_X^k > T_e)\} \\ W_k^b & : \ \text{otherwise}, \end{cases} \tag{2.18}$$

where t_1 and t_2 are two thresholds, and $t_1 > t_2$. The values of t_1 and t_2 can be determined adaptively to σ_b^2 as

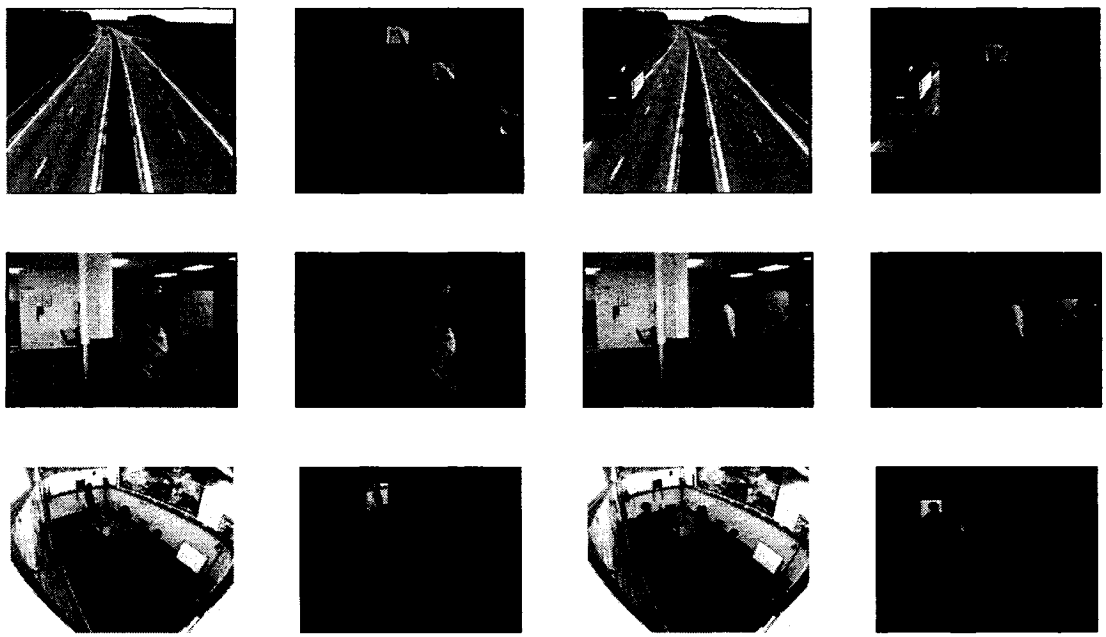
$$\begin{aligned} t_1 &= a_1 \sigma_b \\ t_2 &= a_2 \sigma_b, \end{aligned} \tag{2.19}$$

where a_1 and a_2 are two coefficients and can be determined by the significance

test algorithm with two desired false alarms α_{t_1} and α_{t_2}, respectively (see details in Sec.4.3.1).

In this thesis, an average filter is used in the computation of obtaining D_n. Since $\sigma_b^2 = 2\sigma_\nu^2$ and σ_ν^2 is estimated from the original frame, the value of σ_b is not affected by the average filter. Thus, we improve the robustness of the BOI estimation to noise by using the average filter, which reduces the noise variance in D_n.

Fig.2.2 shows the output of the proposed BOI estimation algorithm applied to the video sequences "Road", "Ekrlb", and "Intelligent room". As can be seen, the proposed algorithm can successfully detect the BOI in difference frames. Note that pixels in BOI may belong to an object or to a background. This is decided in Chapter 3.



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

Figure 2.2: BOI estimation, (a) and (c): the original frames of "Road" (F_{26} and F_{244}), "Ekrlb" (F_{70} and F_{305}), and "Intelligent room" (F_{143} and F_{252}), (b) and (d): output of the proposed BOI estimation algorithm applied to (a) and (c), respectively.

2.6 Content-adaptive color CD

2.6.1 Statistically modeling D_n^Y, D_n^U, and D_n^V

Based on the statistical model of the gray-level distribution of D_n under $\mathcal{H}_0$ in (2.12), we model the gray-level distribution of D_n^Y under $\mathcal{H}_0$ as a Gaussian RV Y with the *pdf*

$$Y \sim \frac{2}{\sqrt{2\pi}\sigma_b} e^{-\frac{y^2}{2\sigma_b^2}}, \quad Y \geq 0, \tag{2.20}$$

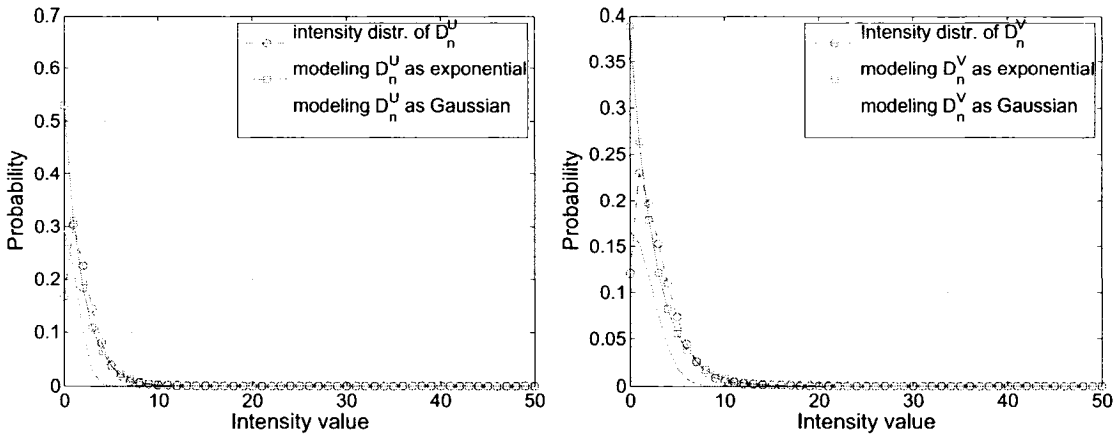
where $\sigma_b^2 = 2\sigma_\nu^2$, and σ_ν^2 is the noise variance in the original frame F_n.

To avoid including artifacts in U and V channels into the final difference frame, we take not only noise but also illumination changes and shadows into account when modeling the maximum intensity (MI) distributions between D_n^U and D_n^V under $\mathcal{H}_0$. Maximum-intensity distribution is defined as the distribution of the maximum value in each intensity pair (u, v) of a pixel in the chrominance channels, u and v are color intensities of the pixel in U and V channels, respectively. First, we model the color intensity distribution in D_n^U and D_n^V under $\mathcal{H}_0$. Note that D_n^U and D_n^V only have positive values due to frame differencing followed by absolute-value operation. For precise significance-testing, the tail sections of the models must be consistent with the intensity distribution of D_n^U and D_n^V under $\mathcal{H}_0$. In this thesis, we model the intensity distributions of D_n^U and D_n^V under $\mathcal{H}_0$ as two exponential RVs U and V, respectively, i.e.,

$$\begin{aligned} U &\sim \lambda_u e^{-\lambda_u u} \\ V &\sim \lambda_v e^{-\lambda_v v}, \end{aligned} \tag{2.21}$$

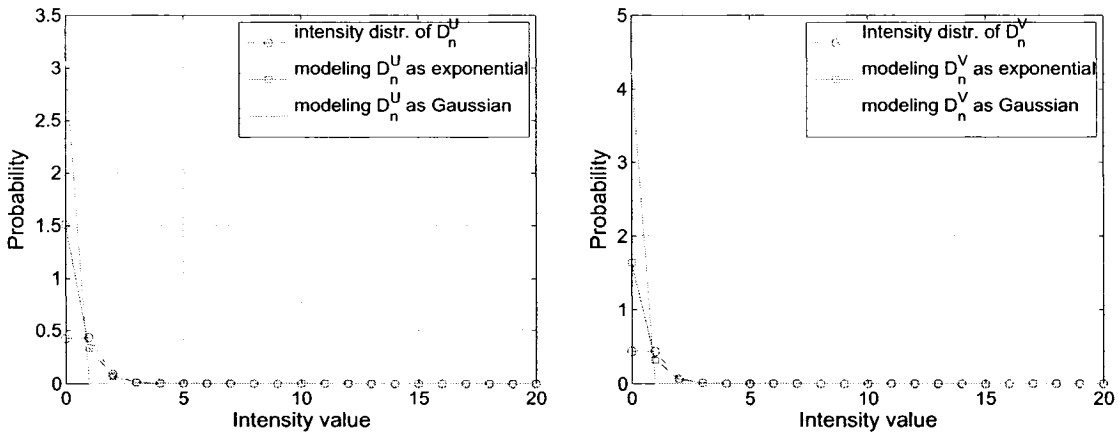
where λ_u and λ_v are the mean of the U and V, respectively. We choose exponential *pdf* for U and V because the tail sections of the *pdf* match the intensity distributions in D_n^U and D_n^V well. Fig.2.3 shows an example of intensity modeling in D_n^U and D_n^V under $\mathcal{H}_0$ using 30 frames of video "Hall". In Fig.2.3, we first compute the histogram

over 30 frames, and then compare it with the theoretical *pdf* in (2.21). As can be seen, the tail sections of the exponential models are more consistent with the real intensity distribution than the Gaussian models does under $\mathcal{H}_0$. Fig.2.4 shows another example which confirms our observations in Fig.2.3.



(a) Modeling in D_n^U (b) Modeling in D_n^V

Figure 2.3: Comparison between Gaussion and exponential for modeling intensity distributions under $\mathcal{H}_0$ in chrominance channels (video "Hall").



(a) Modeling in D_n^U (b) Modeling in D_n^V

Figure 2.4: Comparison between Gaussion and exponential for modeling intensity distributions under $\mathcal{H}_0$ in chrominance channels (video "Ekrlb").

Let $Z = \max(U, V)$, then under $\mathcal{H}_0$, the cumulative distribution function (cdf) of

Z is

$$\begin{aligned} F_Z(z|\mathcal{H}_0) &= P[Z \leq z|\mathcal{H}_0] \\ &= P[\max(U,V) \leq z|\mathcal{H}_0]. \end{aligned} \tag{2.22}$$

Since $U > V$ and $U \leq V$ are mutually exclusive to each other, we have

$$\begin{aligned} P[\max(U,V) \leq z|\mathcal{H}_0] &= P[(U \leq z, U > V) \cup (V \leq z, U \leq V)] \\ &= P[U \leq z|U > V] + P[V \leq z|U \leq V]. \end{aligned} \tag{2.23}$$

Because U and V are independent [61,62], and $Z \geq 0$, we have

$$\begin{aligned} F_Z(z|\mathcal{H}_0) &= \iint_{D(z)} p_U(u)p_V(v)\,dudv \\ &= 1 - e^{-\lambda_u z} - e^{-\lambda_v z} + e^{-(\lambda_u+\lambda_v)z}, \end{aligned} \tag{2.24}$$

where $p_U(u)$ and $p_V(v)$ are pdf of U and V shown in (2.21), and $D(z)$ are integral region for the function $\max(U,V)$.

2.6.2 Significance test

Significance test is based on the no-change hypothesis $\mathcal{H}_0$. For a given pixel, significance test estimates how well its intensity value matches the hypothesis $\mathcal{H}_0$. If it matches $\mathcal{H}_0$ well, it is a non-significant pixel and high probably belongs to background, otherwise it is a significant pixel and high probably belongs to foreground. A false alarm α is used to determine if the pixel matches $\mathcal{H}_0$ well. A false alarm α in significance test is such a value that one can $(1-\alpha) \times 100\%$ sure that the no-change hypothesis $\mathcal{H}_0$ for a given pixel is rejected. Thus the pixel is significant and should be classified as a changed pixel.

Significance test can be carried out based on either pixel-wise or block-wise. Block-wise significance test is more reliable than pixel-wise test, and usually used to directly

obtain change masks from the test results [25,32,33]. However, a block in a block-wise significance test may include pixels from both important changes and unimportant changes. This may result in spurious blobs in change masks because decision is made based on all pixels of a block. Also, the block-wise significance test is computationally expensive due to the block-based square computation. Significance test in this section is to fast fix the significant pixels instead of getting change masks, therefore, we use the pixel-wise significance test.

Significance test is first performed in the BOI of D_n^Y under $\mathcal{H}_0$. We regard a pixel $\mathbf{i}$ in D_n^Y with gray-level $g_\mathbf{i}$ a significant pixel if $g_\mathbf{i}$ is high enough, i.e.,

$$P[Y \leq g_\mathbf{i}|\mathcal{H}_0] > p_h, \tag{2.25}$$

where p_h is a high probability (e.g., 0.9975) that is computed by a false alarm α_s, $p_h = 1 - \alpha_s$. A low false leads to a high p_h thus improves the robustness of the significance test to artifacts, but it may lead to lose some relatively weak important changes. A high false alarm leads to a low p_h thus avoids to lose many important changes, however, it may lead to mistakenly classify some relatively strong unimportant changes as important changes.

From (2.20), we get

$$P[Y \leq g_\mathbf{i}|\mathcal{H}_0] = \int_0^{g_\mathbf{i}} \frac{1}{\sqrt{2\pi}\sqrt{2}\sigma_\nu} e^{-\frac{y^2}{4\sigma_\nu^2}}\, dy < \frac{p_h}{2}. \tag{2.26}$$

(2.26) gives

$$Q\left(\frac{g_\mathbf{i}}{\sqrt{2}\sigma_\nu}\right) < \left(0.5 - \frac{p_h}{2}\right), \tag{2.27}$$

where $Q(\cdot) = 1 - \Phi(\cdot)$, and $\Phi(\cdot)$ is the standard Gaussian *cdf*. We can determine if $\mathbf{i}$ is significant by testing if $g_\mathbf{i}$ satisfies (2.27), e.g., for $p_h = 0.9975$, (2.27) gives that $\mathbf{i}$ is significant if $g_\mathbf{i} > 4.27\sigma_\nu$. (In this thesis, σ_ν^2 is estimated by the noise estimation

method in [66].)

Similarly, significance test in D_n^U and D_n^V is performed by testing if $s_{\mathbf{i}} = \max(D_n^U(\mathbf{i}), D_n^V(\mathbf{i}))$ with $\mathbf{i}$ is the spatial location of a pixel in chrominance channel is high probable greater than Z, i.e.,

$$P[Z \leq s_{\mathbf{i}} | \mathcal{H}_0] > p_h. \tag{2.28}$$

Note that the $\mathbf{i}$ in $D_n^U(\mathbf{i})$ or $D_n^V(\mathbf{i})$ may not be the $\mathbf{i}$ in $D_n^Y(\mathbf{i})$ because the size of D_n^Y may not equal to the size of D_n^U and D_n^V. From (2.24), we can determine if pixel $\mathbf{i}$ is significant in chrominance channels by testing if (2.29) is satisfied.

$$e^{-\lambda_u s_{\mathbf{i}}} + e^{-\lambda_v s_{\mathbf{i}}} - e^{-(\lambda_u + \lambda_v) s_{\mathbf{i}}} < (1 - p_h). \tag{2.29}$$

2.6.3 Color-based gray-level compensation

A pixel in BOI may be significant or non-significant (e.g., the artifact or noise). Pixels classified as non-significant in BOI of D_n^Y but significant in D_n^U and D_n^V belong to objects but have similar gray-levels as the background. We compensate the gray-levels of those pixels based on their significance probabilities p_s in chrominance channels, where p_s is

$$p_s = F_Z(s_{\mathbf{i}} | \mathcal{H}_0), \tag{2.30}$$

and $s_{\mathbf{i}} = \max(D_n^U(\mathbf{i}), D_n^V(\mathbf{i}))$. Thus, we compensate the gray-level $g_{\mathbf{i}}$ of a pixel $\mathbf{i}$ using

$$g_{\mathbf{i}}^c = g_{\mathbf{i}} + a_c \cdot G_s, \tag{2.31}$$

where $g_{\mathbf{i}}$ is the original gray-level of $\mathbf{i}$ in D_n^Y, $g_{\mathbf{i}}^c$ is the compensated gray-level of $\mathbf{i}$, G_s is the gray-level that a significant pixel high-probably have in D_n^Y, and a_c is a

compensation coefficient that is determined by p_s using a quadratic function as

$$a_c = \left(\frac{p_s - p_h}{1 - p_h} \right)^2, \tag{2.32}$$

where p_h is the high probability in (2.25), and p_s is as in (2.30). We can see in (2.32), the more significant a pixel in chrominance channels is, the higher the p_s is, thus the higher the a_c is.

The simplest way to get the value of G_s is to set it to the maximum gray-level in D_n^Y. However, this is not reliable because some artifacts may generate very high gray-levels in D_n^Y. A reasonable way to estimate G_s is based on the gray-level distribution of D_n^Y with a desired false alarm α_G (e.g., 0.0025), or it can be obtained by the statistical model of the gray-level distribution of D_n^Y under change hypothesis $\mathcal{H}_1$. In this thesis, G_s is the minimum gray-level which satisfies (2.25). Since the histogram is the estimation of a real *pdf*, we can estimate the probability given in (2.25) by using the histogram $h_d(g)$ of D_n. From (2.25), G_s is the minimum gray-level that satisfies

$$\sum_{g=0}^{G_s} h_d(g) > (1 - \alpha_G), \tag{2.33}$$

where g is a gray-level.

The proposed gray-level compensation in (2.31) maps the significant intensities in the U and the V channels to the gray-level channel thus avoid the computationally expensive multi-channel CD [58–60] or computation of color-distance [61, 62]. Based on the pixel-wise significance test with the different statistical models in Y and chrominance channels, the proposed CD algorithm is robust to the noise in difference channels. With the aid from BOI estimation, the proposed CD performs gray-level compensation only in BOI thus it is robust to the local unimportant changes in the background.

2.7 Summary

A gray-level compensation based color change detection algorithm is proposed for fast video object detection in this thesis using the YUV color model. Under no-change hypothesis, the gray-level distribution of Y channel and the maximum-intensity distribution in U and V channels are statistically modeled. Based on the scatter estimation of the blocks-of-interest, pixel-based significance tests are performed in the blocks that potentially containing moving objects using the statistical models. The gray-levels of non-significant pixels belonging to the blocks-of-interest are compensated based on their significance probability if they are significant in chrominance channels.

The proposed BOI scatter estimation detects the frame blocks that potentially contain moving objects thus focus the later processing steps on the BOI. It not only significantly decreases the data volume to be processed but also improve the robustness of the proposed VOD method to the artifacts in the background area. The significance test in the Y channel with the Gaussian statistical model and in the chrominance channels with the exponential distribution based maximum-intensity statistical model efficiently detects the pixels that high probably belonging to moving objects but similar to the background in gray-level. By adaptively compensating the gray-levels of those pixels, the quality of the change masks are significantly improved.

As will be shown in chapter 4.3.2, color information significantly improves the change detection in cases where objects have similar gray-levels as the background. It also shows that complex operations are not necessarily need to be performed in all chrominance channels when using color information in video processing.

Chapter 3

Proposed Thresholding For Change Detection

The structure of this chapter is: introduction and related work are presented in Section 3.1 and 3.2, respectively; Section 3.4 presents the proposed video-content assessment algorithm. Section 3.5 describes the proposed thresholding algorithm in detail; Section 3.6 summaries the chapter.

3.1 Introduction

Due to its simplicity and efficiency, frame-differencing based change detection (CD) methods are still popular in many video application systems, including surveillance [24, 68], coding [69], and noise reduction [70]. Thresholding plays a pivotal role in frame-differencing based CD.

Global thresholding is the simplest but the most efficient thresholding algorithm for CD. It classifies the pixels in a difference frame D_n, the output of frame-differencing based CD, into two categories by testing if their intensities are greater or less than a threshold T_n at time instant n. The pixels whose intensities are greater than T_n are

marked as the changed pixels. Thus, the change mask B_n (binary) of D_n is obtained, i.e.,

$$B_n(\mathbf{i}) = \begin{cases} 1 & : \quad D_n(\mathbf{i}) > T_n, \\ 0 & : \quad \text{otherwise}, \end{cases} \tag{3.1}$$

where $\mathbf{i}$ is the spatial location of a pixel in D_n. Object masks can be obtained from B_n by a post-processing procedures such as contour tracing followed object filling [24] or morphological operations [39].

The threshold T_n in (3.1) is a critical parameter in frame-differencing based CD. The non-zero pixels in D_n are caused by, not only the important changes, e.g., motion of objects, but also by unimportant changes, e.g., noise, illumination changes, shadows, or local light changes due to door opening. A relatively low threshold can give relatively complete regions of change (RCG) but may include many spurious blobs in change masks. A relatively high threshold generates few spurious blobs in change masks, however, it usually loses parts of objects. An ideal thresholding algorithm for CD should be 1) artifacts robust, 2) temporally stable, and 3) has low computational cost.

3.2 Related Work

Although many thresholding methods have been proposed in the literature, most of them are proposed for intensity images [54, 55, 71]. Different thresholding algorithms make different assumptions about image contents [72]. However, when being applied to difference frames, the assumptions that these thresholding algorithms make for intensity image contents will no longer hold. A typical example is that Ridler *et. al* [71] assume that the histogram of an image is bimodal thus a global threshold can be obtained by iteratively computing the average value of the mean of each of the two classes of the histogram. However, the histogram of most of difference frames

are unimodal [67]. Rosin *et. al* [57, 72] have shown that most of the thresholding algorithms proposed for intensity images are not suitable for CD.

After investigating many thresholding algorithms subjectively and objectively, Rosin *et al.* [57, 72] recommend three thresholding methods which perform best for CD: the Kapur, the Euler-number, and the Poisson-noise model thresholding.

The Kapur thresholding [55] is a gray-level distribution based threshold algorithm. It obtains a threshold by maximum entropy criteria. It assumes that two classes of events, i.e., the changed regions and the unchanged regions, occur in a difference frame. Each event can be described by a probability density function (pdf). The threshold is then selected such that the sum of the entropy of the two pdfs is maximized. The Kapur thresholding performs well for many difference frames, however, it is sensitive to the variation in contrast between the foreground and the background of a video sequence.

The Euler-number thresholding [73] is based on the assumption that the number of RCG in a difference frame will tend to be stable over a wide range of thresholds. When applying a low threshold to a difference frame, the number of RCG is high since many small RCG generated by noise or illumination changes are included into the change mask. As the threshold increasing, the number of RCG decreases rapidly, and become stable when the threshold is high enough. In practice, the number of RCG is usually replaced by the Euler number of the frame [56]. Thus, a threshold can be obtained by finding the corner of "Euler-number vs. threshold" curve. A searching algorithm for fixing the corner of a curve is proposed in [67].

Poisson-noise model thresholding [57] is based on the assumption that the number of observations in a difference frame (i.e., number of the pixels over a threshold) follows Poisson distribution. A difference frame is divided into N blocks first, where N is an integer. The number of observations of each block is then counted. The relative

variance of the difference frame is then defined as the ratio between the mean and the variance of the numbers of observations. Since a Poisson distribution has its mean equal to its variance, the relative variance of a Poisson distribution is equal to 1. The RCG in the difference frame occur as clusters, this leads the relative variance greater than 1. A reasonable threshold of the difference frame is selected such that the relative variance is maximized.

The Euler-number and the Poisson-noise model thresholding methods are based on spatial properties of difference frames. Simulations in [57, 72] have shown that the two thresholding methods perform well for difference frames. However, compared to the Kapur thresholding, the Euler-number and Poisson-noise model thresholding methods are computionally expensive, and thus not suitable to real-time applications. L. Snidaro *et. al* [74] present a fast algorithm to accelerate the computation of the Euler numbers. This makes the Euler thresholding faster than in [56]. The two spatial-properties based thresholding methods are sensitive to the spatial properties of a difference frame and may be not stable under multiple video conditions, e.g., serious shadows. In addition, Poisson-noise model thresholding is a parametric algorithm with the parameter as the number of frame blocks, N. It is very sensitive to N. A carefully manually selected N is required in practice. This makes the Poisson-noise model thresholding is not suitable for on-line video systems. In this thesis, we implement the Poisson method using $N = 32$ that recommended in [72].

3.3 Overview Of The Proposed Thresholding

In this chapter, we propose a fast artifact-robust thresholding algorithm based on video-content assessment. We assume the noise in F_n is additive white Gaussian noise (AWGN) [22, 23] with zero mean and variance σ_ν^2 estimated, e.g., by the algorithm in [66]. We define the video frames which do not contain any important changes, or

only contain extremely small size important changes as empty frames.

As shown in [25], the changes existed in a difference frame D_n can be classified into two categories, one is the important changes which are caused by object motion, the other is the unimportant changes which may be both global (e.g., noise, and illumination changes) and local (e.g., shadows, and local illumination changes due to door opening). The basic idea of the proposed artifact-robust adaptive thresholding algorithm is to fix such a threshold that suppresses most unimportant-changes while protects most important changes based on the assessment of video contents.

Fig.3.1 shows the block diagram of the proposed algorithm. First, a difference frame D_n resulted from a CD method is divided into K equal-sized blocks $\{W_k\}$, where K is an integer. Then a video-content assessment algorithm (Sec. 3.4) is applied to obtain a content description of D_n including 1) the scatter of the blocks-of-interest (BOI), 2) if D_n is an empty frame D_n^ϕ, and 3) the strength of the local unimportant changes (LUC) compared to noise level in D_n (i.e., local unimportant changes measurement). Then, based on the video content assessment, each frame block W_k, $k = 1, 2, \cdots, K$, is marked as either a region block W_k^c which potentially containing moving objects, or a background W_k^b which are high probable that only contain background regions (see Eq.(2.18)). A discriminative global thresholding algorithm is then applied to D_n. If D_n is an empty frame, a noise-statistic based thresholding (Sec.3.5.1) with a low false alarm α is applied to obtain the global threshold T_n^ϕ of D_n^ϕ. If D_n is non-empty, threshold T_n is obtained by a LUC adaptive thresholding algorithm (Sec. 3.5.2).

3.4 Proposed Video-content Assessment Algorithm

Video-content assessment includes BOI scatter estimation, empty-frame detection, and local unimportant changes (LUC) measurement. The BOI scatter estimation

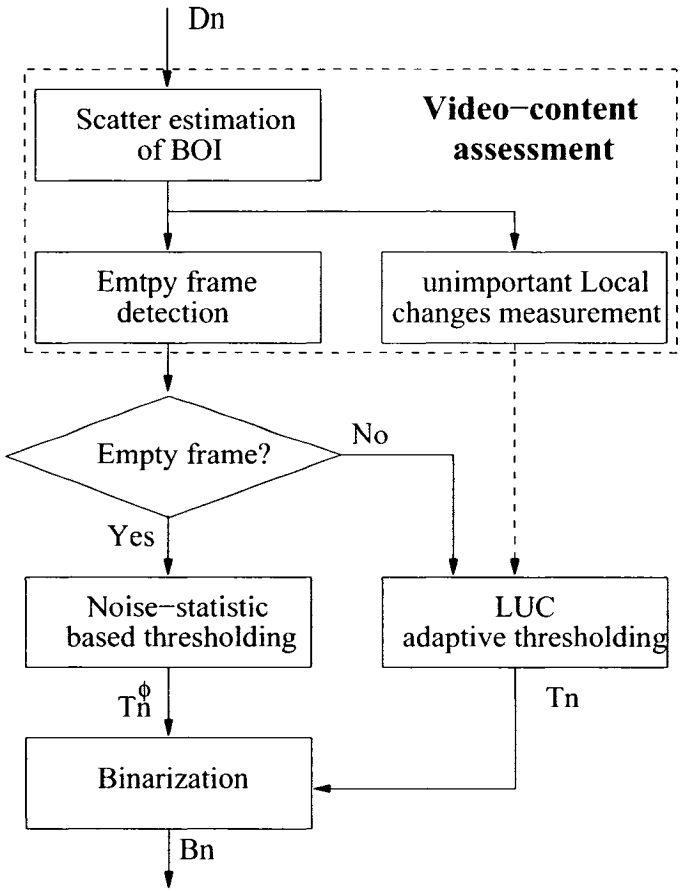


Figure 3.1: Flow chart of the proposed thresholding algorithm. D_n is the output of a frame differencing based CD method.

is described in Sec.2.5. After BOI scatter estimation, the frame blocks in D_n are classified into two categories: background blocks W_k^b and significant changed blocks W_k^c which potentially contain moving objects, as shown in (2.18). Empty-frame detection and LUC measurement are performed based on the BOI scatter estimation as follows.

3.4.1 Empty-frame Detection

For real-world video sequences, two video conditions may lead to empty frames, one is no object exists in a frame, the other is that the size of the objects are too small to be detected. An important application of the proposed BOI scatter estimation

algorithm (Section 2.5) is empty-frame detection, where the video frames without any changes, or only contain extremely small changes are detected.

Based on the BOI scatter estimator given in (2.18), we can easily determine if D_n is an empty frame D_n^{ϕ} by counting the number of W_k^c. We regard a D_n as a D_n^{ϕ} if its all blocks are marked as W_k^b by (2.18).

3.4.2 Local unimportant changes estimation

The LUC is one of the primary factors that seriously degrade the quality of change masks. The intensity distributions of LUC are, in general, different from both the intensity distribution of the background and the intensity distribution of objects. Thus a global thresholding algorithm may seriously underthreshold a D_n by classifying the LUC into the foreground, or serious overthreshold a D_n by using a relatively high threshold for removing all LUC. To obtain reliable change masks, the estimation of the strength of LUC is important for content-adaptive thresholding algorithms.

In this section, we propose a fast algorithm to estimate the strength of LUC in terms of the noise variance σ_b^2 of the difference frame D_n. First, we assume that the non-zero pixels in a signed difference frame $\grave{D}_n$ under $\mathcal{H}_0$ are caused by only LUC. Since Aach *et. al* [32, 33] have shown that the changed areas (including both important and unimportant changes) can be modeled as a Gaussian distribution in signed $\grave{D}_n$ (see page 30), we can model the pixel intensities in signed $\grave{D}_n$ under $\mathcal{H}_0$ as a Gaussian RV $\grave{X}_l$ with zero mean and variance σ_l^2. Thus, the intensities in unsigned D_n, which is obtained by applying absolute-value operation to $\grave{D}_n$, can be modeled as a RV X_l with the *pdf* (see details in Sec.2.5.1)

$$X_l \sim \frac{2}{\sqrt{2\pi}\sigma_l} e^{-\frac{x_l^2}{2\sigma_l^2}}, \quad X_l \geq 0. \tag{3.2}$$

Thus from (2.16) the entropy of X_l is

$$H_{X_l} = \frac{1}{2}\ln(\frac{1}{2}\pi e\sigma_l^2). \tag{3.3}$$

Let $\sigma_l^2 = \kappa_l \cdot \sigma_b^2$, then (3.3) becomes

$$\begin{aligned} H_{X_l} &= \tfrac{1}{2}\ln(\tfrac{1}{2}\pi e\sigma_b^2) + \tfrac{1}{2}\ln\kappa_l \\ &= H_X + \tfrac{1}{2}\ln\kappa_l. \end{aligned} \tag{3.4}$$

H_X is the entropy of noise in D_n and is defined in (2.16). The factor κ_l is less or equal 1 if LUC is non-dominant compared to noise, or greater than 1 otherwise, as shown in(3.5)

$$\kappa_l = e^{2(H_{X_l} - H_X)}, \tag{3.5}$$

where H_X is the entropy of X under $\mathcal{H}_0$ that assume the non-zero pixels are only caused by noise, and H_{X_l} is the entropy of X_l under $\mathcal{H}_0$ that assume the non-zero pixels are only caused by LUC. As can be seen in (3.5), for a known noise level, the higher H_{X_l} is, the higher κ_l is, thus the stronger the strength of LUC is. Therefore, factor κ_l is a good measure of the strength of LUC.

In practice, κ_l can be estimated from the background blocks $\{W_k^b\}$ (given in Eq.(2.18)) as follows. First, based on the assumption that non-zero pixels in D_n under $\mathcal{H}_0$ are caused by LUC only, we approximate H_{X_l} of each W_k^b by computing the block entropy $H_{X_l}^k$, which can be computed based on the block histogram by (2.17) (see page 31). Then, we get the factor κ_l^k of each W_k^b from (3.5). If D_n contains very weak LUC, i.e., the statistical distribution of X_l is the statistical distribution of noise X, H_{X_l} will be similar to H_X, and κ_l is close to 1. Here, H_X is computed by (2.16), where $\sigma_b^2 = 2\sigma_\nu$, and the noise variance σ_ν in the original frame can be estimated by a noise estimation algorithm, e.g., [66]. If D_n contains serious LUC, H_{X_l} will be

much more greater than H_X, and $\kappa_l \gg 1$. The value of H_X can be computed by (2.16), where the noise variance of the original frame is estimated via [66]. Finally, κ_l is obtained by averaging all factors of $\{W_k^b\}$, i.e.,

$$\begin{aligned} \kappa_l^k &= e^{2(H_{X_l}^k - H_X)}, \\ \kappa_l &= \frac{1}{N_k^b} \sum_{W_k \in \{W_k^b\}} \kappa_l^k, \end{aligned} \tag{3.6}$$

where N_k^b is the number of W_k^b.

3.5 Discriminatively Thresholding Region And Background Blocks

The proposed thresholding algorithm discriminatively takes different thresholding strategies for different video contents. For the empty video frames, a noise-statistic based thresholding strategy is employed to compute a global threshold which suppresses most unimportant-changes but protects the small size important changes. For non-empty video frames, a LUC adaptive thresholding strategy is applied to obtain an artifact-robust threshold.

3.5.1 Thresholding empty or almost empty frames

The simplest way to threshold an empty frame D_n^ϕ is to set all the pixels of the frame to zero. However, this is not a robust solution for the video sequences containing extremely small objects. Therefore, we compute the threshold T_n^ϕ for an empty frame by a noise-statistic based algorithm, where T_n^ϕ is highly probable higher than the

pixels of an empty frame, i.e.,

$$P[X \leq T_n^{\phi}] > p_h, \tag{3.7}$$

where $X = D_n^{\phi}(\mathbf{i})$ is the intensity value of a pixel located at $\mathbf{i}$, and p_h is a high probability value.

The RV X is given in 2.12, we get

$$P[X \leq T_n^{\phi}] = \int_0^{T_n^{\phi}} 2\frac{1}{\sqrt{2\pi}\sigma_b} e^{-\frac{x^2}{2\sigma_b^2}}\, dx. \tag{3.8}$$

Let $z = \frac{x}{\sigma_b}$, from (3.8), we get

$$\begin{aligned} P[X \leq T_n^{\phi}] &= 2[\textstyle\int_0^{\infty} \frac{1}{\sqrt{2\pi}\sigma_b} e^{-\frac{x^2}{2\sigma_b^2}}\, dx \\ &\quad - \textstyle\int_{T_n^{\phi}}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_b} e^{-\frac{x^2}{2\sigma_b^2}}\, dx] \\ &= 1 - 2\textstyle\int_{\frac{T_n^{\phi}}{\sigma_b}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}\, dx \\ &= 1 - 2Q(\tfrac{T_n^{\phi}}{\sigma_b}) \end{aligned} \tag{3.9}$$

where $Q(\cdot) = 1 - \Phi(\cdot)$, and $\Phi(\cdot)$ is the *cdf* of standard normal distribution. From (3.7), we get $Q(\frac{T_n^{\phi}}{\sigma_b}) < \frac{1-p_h}{2}$. In statistics, false alarm α, $0 \leq \alpha \leq 1$, is defined as the area under the probability density corresponding no-change hypothese to the right side of a threshold. False alarm is often used to yield an acceptable detection probability in significance test [32]. Therefore, p_h can be determined by a false alarm α_e, i.e., $p_h = 1 - \alpha_e$. Thus, we get

$$Q(\frac{T_n^{\phi}}{\sigma_b}) < \frac{\alpha_e}{2}. \tag{3.10}$$

Using the Q-function table [65], we can find a value q which satisfies $Q(q) \approx \frac{\alpha_e}{2}$. Since the Q-function is a non-increasing function, $P[X \leq T_n^{\phi}] > p_h$ is equivalent

to $\frac{T_n^\phi}{\sigma_b} > q$, where the standard deviation σ_b can be computed from the standard deviation of the noise in the original frame σ_ν as $\sigma_b = \sqrt{2}\sigma_\nu$ [57]. Thus, the threshold of D_n^ϕ should satisfy

$$T_n^\phi > q\sigma_b. \tag{3.11}$$

The T_n^ϕ obtained from (3.11) only takes noise into account. In real-world video sequences, the non-zero pixels in D_n^ϕ are not only caused by noise, but also by illumination changes and unimportant local changes. Taking them into account, we refine T_n^ϕ as

$$T_n^\phi = \mu_n^\phi + q\sigma_b, \tag{3.12}$$

where μ_n^ϕ is the intensity mean of D_n^ϕ, and q can be determined by a low false alarm α_e (see Sec.4.4.1).

3.5.2 Thresholding non-empty frames

For a non-empty difference frame, we compute the threshold T_n by two steps as shown in (3.13): first we estimate the optimum threshold T_n' of a D_n based on the assumption that the non-zero pixels in background regions are only caused by noise, then we refine T_n' by an adjusting factor γ_l according to the measurement of the strength of LUC (estimated using Eq.(3.6)) to obtain the artifact-robust threshold of D_n,

$$T_n = T_n' + \gamma_l, \tag{3.13}$$

where T_n' is for suppressing the global unimportant changes which are mainly caused by noise, and γ_l is for suppressing LUC, which may be caused by artifacts such as shadows, local light changes, or partial background movement.

Only by taking noise into account, we can fix an initial threshold T_n' which is estimated based on the intensity distribution in the blocks-of-interest $\{W_k^c\}$ as follows.

The intensity distribution in $\{W_k^c\}$ detected by BOI scatter estimator can be modeled as the mixture of the distribution of the noise and the distribution of the important changes in D_n. Since the no-change hypothesis $\mathcal{H}_0$ and the changed hypothesis $\mathcal{H}_1$ are mutually exclusive, the distribution of the intensities in $\{W_k^c\}$ can be obtained by using the total probability theorem, the *pdf* of the mixture distribution is then

$$p(x_d) = P_b \times p_b(x_d|\mathcal{H}_0) + P_c \times p_c(x_d|\mathcal{H}_1), \tag{3.14}$$

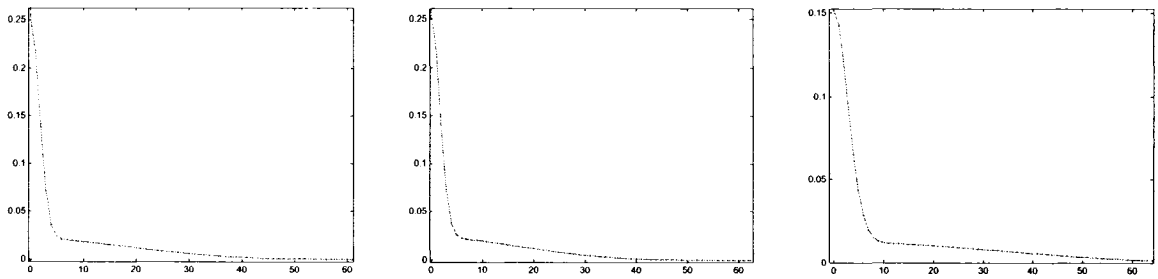
where x_d is the intensity of a pixel belongs to $\{W_k^c\}$, note that the $\{W_k^c\}$ are the blocks detected by the BOI scatter estimate instead of the moving objects areas, as shown in Fig.2.2 (see page 33), they may contain both changed and unchanged areas; P_b and P_c are probabilities of the changed and unchanged areas in $\{W_k^c\}$, respectively; $p_b(\cdot)$ and $p_c(\cdot)$ are the *pdf* of the intensity distributions under the condition $\mathcal{H}_0$ and $\mathcal{H}_1$, respectively.

We have shown that the *pdf* of the intensity distribution under hypothesis $\mathcal{H}_0$ is given in (2.12), and the *pdf* of the intensity distribution under hypothesis $\mathcal{H}_1$ is given in (2.13). Based on (2.12) and (2.13), we get

$$p(x_d) = \frac{2P_b}{\sqrt{2\pi}\sigma_b} e^{-\frac{x_d^2}{2\sigma_b^2}} + \frac{2P_c}{\sqrt{2\pi}\sigma_c} e^{-\frac{x_d^2}{2\sigma_c^2}}. \tag{3.15}$$

As can be seen in (3.15), $p(x_d)$ is determined by four parameters, the probabilities P_b and P_c, the noise variance σ_b^2 in signed $\grave{D}_n$, and the intensity variance of the changed areas σ_c^2 in signed $\grave{D}_n$. The value of P_b and P_c are related to video contents. For example, to the video sequences with small size objects, P_b may be greater than P_c, while to the video sequences with big size objects, P_b may be smaller than P_c. Assume that the occurrence of the changed and the unchanged areas have the same probability, we set $P_b = Pc = 0.5$ in this thesis. The value of σ_b^2 can be estimated

by a noise estimation algorithm, e.g., [66]. The value of σ_c^2 is difficult to estimate, however, we know $\sigma_c^2 \gg \sigma_b^2$. The simulations in [32, 33] show that $\sigma_c^2 > 100\sigma_b^2$. As will be shown later in this section (page 56), we need not estimate σ_c^2. Fig.3.2 shows the examples of the theoretical intensity distributions given in (3.15) with different parameters but $P_b = P_c = 0.5$.



(a) $\sigma_b^2 = 4.11$, and $\sigma_c^2 = 150\sigma_b^2$ (b) $\sigma_b^2 = 4.11$, and $\sigma_c^2 = 100\sigma_b^2$ (c) $\sigma_b^2 = 8.206$, and $\sigma_c^2 = 120\sigma_b^2$

Figure 3.2: Examples of the theoretical intensity distribution in W_k^c with different parameters and $P_b = P_c = 0.5$.

For a threshold t (e.g., T_n' in (3.13)), the probability of error e_t in classification of important and unimportant changes can be computed based on (3.15) as

$$e_t = 2P_b \int_t^{\infty} \frac{1}{\sqrt{2\pi}\sigma_b} e^{-\frac{x_d^2}{2\sigma_b^2}} dx_d + 2P_c \int_0^t \frac{1}{\sqrt{2\pi}\sigma_c} e^{-\frac{x_d^2}{2\sigma_c^2}} dx_d. \tag{3.16}$$

Let $x_b = \frac{x_d}{\sigma_b}$ and $x_c = \frac{x_d}{\sigma_c}$, we have

$$e_t = 2P_b Q(\frac{t}{\sigma_b}) + \frac{P_c}{\sqrt{2}} \mathbf{erf}(\frac{t}{\sigma_c}), \tag{3.17}$$

where $\mathbf{erf}(\cdot)$ is the error function of a Gaussian RV. As can be seen in (3.17), $Q(\cdot)$ is a non-increasing function, and $\mathbf{erf}(\cdot)$ is a non-decreasing function. Therefore, both a too low t and a too high t may greatly increase the classification error since it may greatly increase the value of $Q(\cdot)$ or $\mathbf{erf}(\cdot)$, respectively.

An initial threshold T_n' in (3.13) can be fixed by minimizing e_t. We use Fermat's

theorem to fix an optimum t that minimizes e_t, i.e., $T_n^{'}$. First, we take the first-order derivative of (3.17), i.e.,

$$\frac{de_t}{dt} = -\frac{2P_b}{\sqrt{2\pi}\sigma_b} e^{-\frac{t^2}{2\sigma_b^2}} + \frac{2P_c}{\sqrt{2\pi}\sigma_c} e^{-\frac{t^2}{2\sigma_c^2}}, \tag{3.18}$$

second, we set (3.18) equal to 0, we have

$$\frac{P_c}{\sigma_c} \cdot e^{-\frac{t^2}{2\sigma_c^2}} - \frac{P_b}{\sigma_b} \cdot e^{-\frac{t^2}{2\sigma_b^2}} = 0. \tag{3.19}$$

Since $\sigma_c^2 \gg \sigma_b^2$ [33], we assume $\sigma_c = \kappa_c \sigma_b$, where κ_c is a real number which represents the variance ratio between the important changes and the noise in D_n, and $\kappa_c \gg 1$. Solving (3.19) for t, we get

$$\frac{t^2}{\sigma_b^2} = 2\ln(\frac{P_b}{P_c}\kappa_c) \cdot \frac{\kappa_c^2}{(\kappa_c^2 - 1)}. \tag{3.20}$$

Since $\frac{\kappa_c^2}{(\kappa_c^2-1)} \approx 1$ when $\kappa_c \gg 1$, we get the relationship between t and σ_b^2 as

$$\frac{t}{\sigma_b} = \sqrt{2\ln(\frac{P_b}{P_c}\kappa_c)}. \tag{3.21}$$

Fig.3.3 shows the $\frac{t}{\sigma_b}$ vs. κ_c curve, where κ_c varies from 20 to 1000, and $P_b = P_c$. As can be seen, the value of $\frac{t}{\sigma_b}$ varies in a relatively small range when κ_c varying in a large range, i.e., $\frac{t}{\sigma_b}$ is relatively stable compared to κ_c. Then the initial threshold $T_n^{'}$ in 3.13 can be obtained from (3.21) as

$$T_n^{'} = \sigma_b \cdot \sqrt{2\ln(\frac{P_b}{P_c}\kappa_c)}. \tag{3.22}$$

It is difficult to estimate the variance of important changes σ_c^2 in D_n due to the complexity of the video contents, thus it is difficult to estimate κ_c. We can estimate

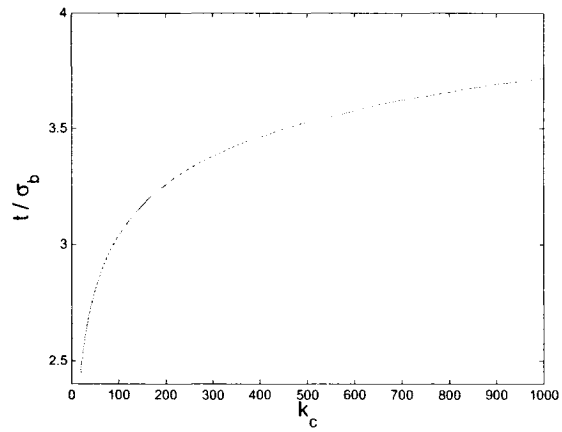


Figure 3.3: The $\frac{t}{\sigma_b}$ vs. κ_c curve (κ_c varies from 20 to 1000, and $P_b = P_c$).

T_n' by using the stabilization fact of $\frac{t}{\sigma_b}$ shown in Fig.3.3 as follows.

First, we estimate the range that an optimum threshold t may vary in a non-empty D_n. From (3.21), we get that when κ_c varies from 20 to 1000, $\frac{t}{\sigma_b}$ varies from 2.45 to 3.72, i.e., the threshold t varies between $2.45\sigma_b$ and $3.72\sigma_b$, where $\sigma_b^2 = 2\sigma_\nu$, and σ_ν is the noise variance in the original frame. As can be seen, t depend on the noise level in a video frame. Second, we fix the location that the range of t in the intensity distribution curve of D_n. We note that under no-change hypothesis $\mathcal{H}_0$, the *cdf* of the intensity distribution in D_n can be obtained from the *pdf* given in (2.12) as

$$P(X \leq x) = \int_0^x \frac{2}{\sqrt{2\pi}\sigma_b} e^{-\frac{x'}{2\sigma_b^2}} dx', \quad X \geq 0. \tag{3.23}$$

Let $x'' = \frac{x'}{\sigma_b}$, (3.23) becomes

$$P(X \leq \frac{x}{\sigma_b}) = \int_0^{\frac{x}{\sigma_b}} \frac{2}{\sqrt{2\pi}} e^{-\frac{x''^2}{2}} dx'', \quad X \geq 0. \tag{3.24}$$

From (3.24), we convert the *pdf* given in (2.12) to the *pdf* as

$$p(x'') = \frac{2}{\sqrt{2\pi}} e^{-\frac{x''^2}{2}}, \quad x'' \geq 0, \tag{3.25}$$

as shown in Fig.3.4. As can be seen, the range that t varies between $2.45\sigma_b$ and $3.72\sigma_b$ is located at the corner of the distribution curve. Thus, without statistically estimating κ_c, T_n' can be efficiently estimated by fixing the corner of the distribution curve of the noise in D_n, as shown in Fig.3.4.

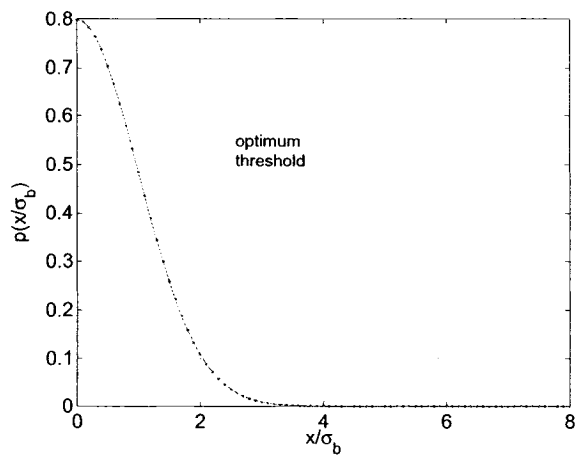
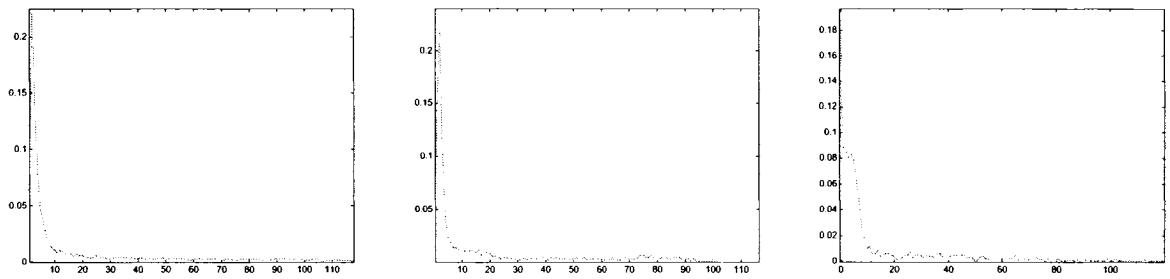


Figure 3.4: The *pdf* of $\frac{x}{\sigma_b}$ and the range that t varies.

In real-world video applications, the real *pdf* of the intensity-distribution in $\{W_k^c\}$ may be different from the *pdf* given in (3.15) because (3.15) only takes noise into account. The histogram of $\{W_k^c\}$ is an estimation of the real *pdf* of $\{W_k^c\}$. Fig.3.5 shows the histograms of the $\{W_k^c\}$ in three frames of three real-world video sequences with different video contents. Comparing Fig.3.2 and Fig.3.5, the histograms shown in Fig.3.5 are similar to the *pdf* given in (3.15) with $\sigma_c^2 \gg \sigma_b^2$. Thus, we estimate the initial optimum threshold T_n' from the histogram of $\{W_k^c\}$.

The histogram of the BOI $\{W_k^c\}$ is obtained based on the BOI scatter estimation, and denoted $h_w^c(g)$. Then we assume that the distribution located in the low gray-level partition of the histogram is caused by the global unimportant changes (GUC) including noise and slight illumination changes, since the important changes are usually stronger than the unimportant changes in difference frames. The distribution can be determined by fixing the dominant distribution in the histogram partition



(a) "Hall" (F_{140}) (b) "Take object" (F_{243}) (c) "3Meet" (F_{71})

Figure 3.5: The histograms of $\{W_k^c\}$ in different video sequences ((a) video "Hall" with $\sigma_b^2 = 4.583$, (b) video "Take object" with $\sigma_b^2 = 2.872$), and (c) video "3Meet" with $\sigma_b^2 = 2.061$.

$[0, g]$ where g is the minimum gray-level that satisfies $\sum h_w^c(g) > p_r$, where p_r is a pre-defined probability between $0.3 \sim 0.5$. We set $p_r = 0.5$ in this thesis. Finally, the thresholding algorithm proposed in [67] is applied to the distribution of the GUC to fix the corner of the GUC distribution curve.

Recall that the final threshold in (3.13) is $T_n = T_n' + \gamma_l$. We determine the adjusting factor γ_l in (3.13) to suppress the the LUC in D_n without damaging the important changes. To improve the robustness of γ_l to multiple video contents, we compute γ_l adaptively to the strength of the LUC. The higher the LUC are, the higher the γ_l is. As can be seen in (3.5), κ_l can be used as a measure of LUC. The higher κ_l is, the more significant the LUC are compared to noise. If $\kappa_l \leq 1$, i.e., the LUC are not significant in D_n compared to noise, γ_l should be relatively low. Otherwise, γ_l should be relatively high and set adaptively to $\sqrt{\kappa_l \sigma_b^2}$, i.e., the standard deviation of X_l in (3.2). Thus, γ_l is

$$\gamma_l = \begin{cases} C_l & : \ \kappa_l \leq 1 \\ a_l \sqrt{\kappa_l \sigma_b^2} & : \ \kappa_l > 1, \end{cases} \tag{3.26}$$

where C_l is a constant between $0 \sim 5$, in this thesis, we set $C_l = 3$, and a_l is a multiplication factor can be adaptively estimated by a desired false alarm α_l (see details in Sec.4.4.1).

3.6 Summary

Frame differencing followed by thresholding is an efficient change detection method, however, it does not outperform advance statistical change detection due to sensitivity to noise, illumination changes, and local changes. A fast artifact-robust thresholding algorithm is proposed in this chapter based on video content assessment. First, the scatter of the blocks-of-interest are estimated. Second, an initial threshold is estimated based on the analysis of optimum thresholding to suppress global unimportant changes. Finally, the global threshold is obtained by adjusting the initial threshold according to the strength of local unimportant changes. To improve the reliability of object masks, an empty frame detection is employed to make the global threshold adaptive to different video conditions.

The proposed video-content assessment method detects the empty frames which are often happened in surveillance applications, and measures the local unimportant changes in terms of the noise variance. Thus it greatly improve the robustness of the proposed thresholding method to video contents by aiding the proposed discriminative thresholding algorithm.

As shown in chapter 4.4, the proposed thresholding method is more precise than the intensity-distribution based thresholding, and more stable and much more efficient than the state-of-the-art spatial-property based thresholding methods.

Chapter 4

Experimental Results

In this chapter, we evaluate the proposed methods by applying them as well as the reference approaches to real-world video sequences in both subjective and objective ways. Section 4.2 describes the objective measures we used in this chapter. Section 4.3 evaluates the proposed change detection (CD) algorithm, Section 4.4 evaluates the proposed thresholding algorithm. The video object detection consisting of the proposed CD and the proposed thresholding algorithm is evaluated in Section 4.5. Section 4.6 summarizes the chapter.

4.1 Video Sequences Used

Fifteen real-world indoor and outdoor *test* video sequences containing different video contents are used in our simulations. Most video sequences used are publicly available and widely used in video object detection literatures. The indoor video sequences are

1. "Hall": 300 frames of size 352×288, stable illumination, low noise level, and some local light changes.

2. "Intelligent room": 300 frames of size 320×240, serious shadows, and consid-

erable variation in the contrast between the foreground and the background.

3. "2Meet": 690 frames of size 320 × 240, considerable shadows, low contrast between the foreground and the background.

4. "Ekrlb": 678 frames of size 360 × 244, multiple objects, serious local changes, very low contrast between the foreground and the background.

5. "Stair": 1475 frames of size 352 × 288, serious local changes, relatively high noise, and considerable variation in the contrast between the foreground and the background.

6. "Put object": 655 frames of size 320 × 240, well illumination, considerable variation in the contrast between the foreground and the background.

7. "ScriptLab2": 2283 frames of size 360 × 240, many empty frames, and very low contrast between the foreground and the background.

8. "Tennis": 53 frames of size 720 × 576 with zoom global motion.

The outdoor video sequences are

1. "Survey": 1000 frames of size 320 × 240, multiple objects, serious local changes and noise, high variation in contrast between the foreground and the background.

2. "Road": 300 frames of size 352 × 288, multiple fast moving objects, partial background movement, parts of the foreground are similar to background in gray-level.

3. "Vand_paint": 299 frames of size 320 × 240, fast moving objects, and illumination changes.

4. "Vnj": 293 frames of size 360 × 244, multiple moving objects as well as serious illumination changes, shadows, and partial background movement.

5. "Snow": a shot of 427 frames with size 320 × 240, captured on-line at the Ste-Catherine street, Montreal, by VidPro member A. Firas, multiple moving objects, serious background turbulence.

6. "Pavement": a shot of 3000 frames with size 320×240, captured on-line near the cross of the Ste-Catherine street and the Rue Ste-Mark, Montreal, by the author, well illuminated, multiple moving objects, serious shadows, and considerable variation in contrast between the foreground and the background.

7. "Car": 60 frames of size 720 × 576 with rotational and translational global motion.

Five real-world video sequences are used as training video sequences in our simulations to experimentally determine the parameters that need to be pre-defined in the proposed methods, e.g., the false alarms. The training video sequences are

1. Indoor "3Meet": 929 frames with size 384 × 288, small size objects, serious local changes and illumination changes.

2. Part of indoor "Intelligent room": The first 81 frames are empty frames, and used in training set. The non-empty frames are not in the training set.

3. Outdoor "Road1": 300 frames with size 352 × 288, multiple fast moving objects, and parts of the foreground are similar to background in gray-level.

4. On-line "Cross-St-Catherine": a shot of 2000 frames with size 320 × 240, captured on-line at the cross of the street Ste-Catherine and the Rue Ste-Mathieu, Montreal, well illuminated, multiple moving objects, considerable variation in contrast between the foreground and the background.

5. On-line "Cloudy": a shot of 3017 frames with size 320 × 240, captured on-line on the Ste-Catherine street, Montreal, under illuminated, multiple moving objects, and considerable variation in contrast between the foreground and the background.

4.2 Objective Evaluation Measures

Visually evaluating the output of a video processing system is widely used in literatures. The performance of the video processing system is subjectively evaluated based on the observations of the viewers. However, subjective evaluation is not always reliable due to the limits of the human vision system (HVS). A typical example is that the sensitivities of a HVS to video frame quality may decrease after a long time observation due to tiredness. It is important that objectively evaluate the performance of a CD method. Ground truth plays a pivotal rule in objective evaluation of CD methods.

A ground truth is such a video frame where the pixels are correctly classified into objects and the background. Fig.4.1 shows an example of a ground truth. Ground truth is necessary for many objective measures. Although many methods have been proposed to generate ground truth, e.g., [75], it is difficult to generate ground truth for real-world video sequences. Manual annotation is the most common way to generate ground truth for real-world video sequences. Based on the ground truth of a video, many objective measures can be computed to evaluate the performance of a CD method.

An important application of the ground truth is to compute the classification-based objective measures. The basic of a classification based objective measure is C_{AB}, the number of pixels belonging to class A that have been classified as class B. For binary images, $(A, B) = (0, 1)$, and c_{11} is true positives (TP), c_{00} is the

(a) Original frame 45 (b) the ground truth of (a)

Figure 4.1: An example: the ground truth for frame 45 of video "Hall".

true negatives (TN), c_{01} is the false positives (FP), and c_{10} is the false negatives (FN). Based on C_{AB}, several advance objective measures are computed and widely used [25,57,72]. They are

1. Percentage correctly classified (PCC)

$$PCC = \frac{TP + TN}{TP + FP + TN + FN}. \tag{4.1}$$

2. Yule coefficient (YC)

$$YC = |\frac{TP}{TP + FP} + \frac{TN}{TN + FN} - 1|. \tag{4.2}$$

3. Jaccard similarity coefficient (JC)

$$JC = \frac{TP}{TP + FP + FN}. \tag{4.3}$$

The higher the three advance measures are, the better the performance of a CD method is.

Ground truth is not necessary for some objective measures. Erdem *et al.* [76] propose three objective measures which need not ground truth. However, the per-

formance of the measures are tightly coupled with the motion estimation and the contour tracing algorithms they used. Therefore, they may not always reliable in real-world applications due to the complexity of video contents. In this chapter, the classification based measures are used for evaluating the proposed approaches.

4.3 Evaluation of Change Detection Under Background Subtraction

Here, we show the effectivity of the proposed CD algorithm for including the color information into CD. We visually compare the change masks computed by the proposed color CD method with the change masks obtained by gray-level based simple-differencing CD in Sec. 4.3.2. Second, we evaluate the performance of the proposed CD method by applying it as well as the recently proposed Alexandropoulos (Alex.) CD method [60] to seven real-world video sequences containing different video contents in Sec. 4.3.3. We select the Alexandropoulous CD as the reference method because it is 1) an artifact-robust color-based CD (based on cluster-distance classification in each chrominance channel), and 2) real-time.

The seven real-world video sequences used in this section are "Hall", "Intelligent room", "2Meet", "Ekrlb", "Put object", "Road", and "Vnj" (see Section 4.1).

4.3.1 Algorithm parameters

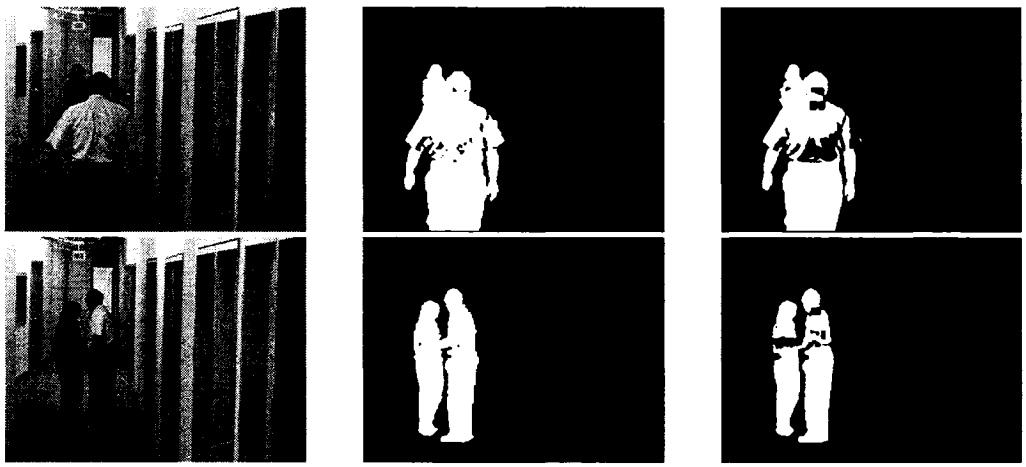
The proposed CD method include following parameters: 1) the size of the spatial average filter for reducing noise used in frame differencing (page 24), 2) the number of frame blocks in BOI scatter estimation (page 30), 3) the block mean thresholds t_1 and t_2 for BOI estimation (see Eq.(2.18)), 4) the low and and the high false alarms for computing t_1 and t_2 (see Eq.(2.19)), and 5) the false alarm α_s to compute the

high probability value p_h for significance test (see Eq.(2.25) and Eq.(2.28)). Training video sequences "3Meet", "Road1", "Cross-St-Catherine", and "Cloudy" are used to estimate the parameters.

The size of the spatial average filter can be set to 3×3, 5×5, or 7×7. However, a big filter size tends to degrade the object boundaries thus degrade the object details. In our simulations, we set the size of the average filter to 3×3. The number of blocks in D_n can be set from 4×4 to 20×20. However, a low block number leads to the big block which may contain many background areas, and a high block number leads to a small block containing few pixels, thus the BOI estimate becomes sensitive to relatively strong artifacts. In our simulations, we divide a frame into 10×10 blocks. In significance test given in (2.25) and (2.28), a high false alarm α_s may mistakenly classify the relatively strong artifacts as important changes, and a very low α_s may lose some relatively weak important changes. Based on experimental results obtained from the four training video sequences, we set α_s to 0.0025. The block mean thresholds t_1 and t_2 in (2.18) are automatically computed by (2.19), and the factors a_1 and a_2 are determined automatically by two different false alarms α_{t_1} and α_{t_2}. Simulations in training set suggest that $\alpha_{t_1} \in [0.0001, 0.0004]$ and $\alpha_{t_2} \in [0.1, 0.3]$. Similar to the analysis of α_s mentioned above, we experimentally set $\alpha_{t_1} = 0.0002$ and $\alpha_{t_2} = 0.2$ thus $a_1 = 3.7$ and $a_2 = 1.3$ using the Q-function table [65]. Note that very low noise in a video leads to a low σ_b^2 thus t_1 and t_2 become low. This may mistakenly classify some blocks with serious local unimportant changes as BOI. To improve the robustness of the BOI estimator, based on observing the means of BOI in the training video sequences, we experimentally set the minimum possible values of $t_1 \in [12, 18]$ and $t_2 \in [5, 10]$. In this thesis, we use $t_1 = 15$ and $t_2 = 7.5$. The above values were used in all simulations we carried out. The major parameters that the proposed CD method depends on are t_1, t_2, and α_s.

4.3.2 Usefulness of adding color to gray-level CD

To show the effectivity of the proposed CD compared to the gray-level based CD, video "2Meet" and "Ekrlb" where the foreground and the background are similar to each other in gray-level are employed. The thresholding algorithm proposed in [77] is used to obtained the binary frames. Fig.4.2 shows the comparison results of "2Meet". As can be seen, there are considerable holes and gaps existed in the change masks obtained by the gray-level based simple differencing CD due to the similarity between the foreground and the background. The proposed method successfully overcomes the problem and obtains clear and complete change masks.

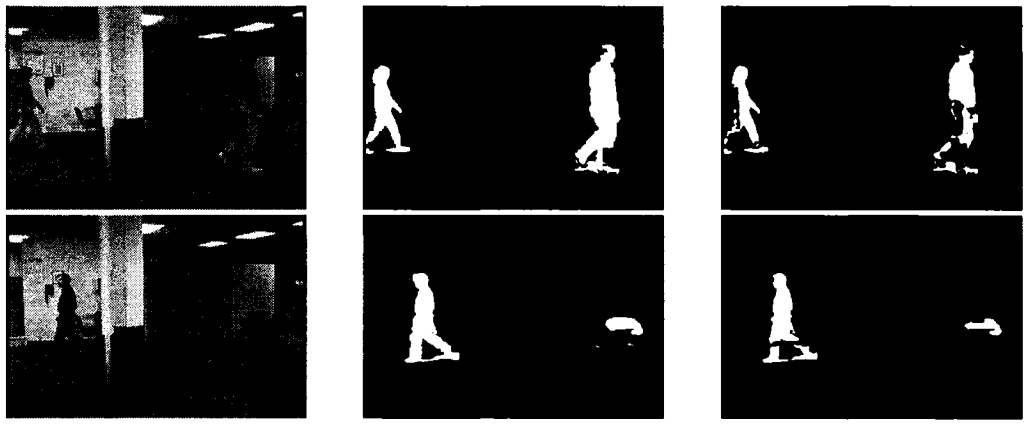


(a) Org. (b) proposed CD (c) gray-level CD

Figure 4.2: Comparison between the prop. CD and gray-level CD applied to "2Meet" with the original frame F_{225}, and F_{359}.

Fig.4.3 shows comparison results of "Ekrlb". As can be seen, without the aid of color, simple-differencing CD mistakenly divides an object into several parts by the gaps caused by the similarities between the objects and background in gray-level. The proposed method significantly improves the quality of the change masks by including color information into CD.

As having shown in Fig.4.2 and Fig.4.3, the proposed algorithm effectively introduces the color informant to CD, and obtains the content-robust change masks.



(a) Org. (b) proposed CD (c) gray-level CD

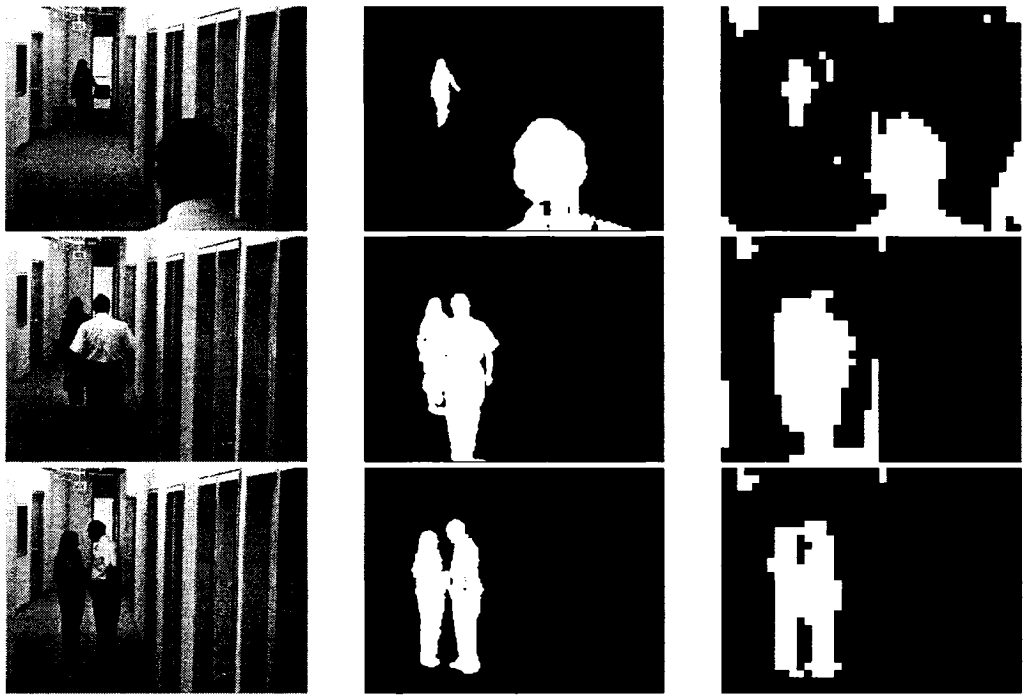
Figure 4.3: Comparison between the prop. CD and gray-level CD applied to "Ekrlb" with the original frame F_{85}, and F_{415}.

4.3.3 Comparison between color-based CD methods

In this section, we compare the proposed CD algorithm with the recently proposed real-time cluster-distance based Alexandropoulos CD method [60]. A fast thresholding algorithm proposed in [77] is applied to the proposed CD method to obtain binary frames. Five real-world video sequences, the indoor "2Meet", "Ekrlb", and "Put object", and the outdoor "Road", and "Vnj" are used in the simulations.

Fig.4.4 shows the comparison results of video "2Meet". As can be seen, the proposed CD performs best. It obtains complete and stable change masks for video "2Meet" where the unimportant changes such as shadows are serious and the contrast between parts of the foreground and the background is low. The [60] CD also works well under the case that the foreground is similar to the background in gray-level, however, it is sensitive to the shadows and includes considerable spurious blobs into the change masks.

Comparison results of video "Ekrlb" are shown in Fig.4.5. The proposed CD clearly outperforms the [60] CD method. Since the [60] CD performs CD in each chrominance channel of a video frame, it is sensitive to the artifacts in any color



(a) Org. (b) prop. CD (c) [60] CD

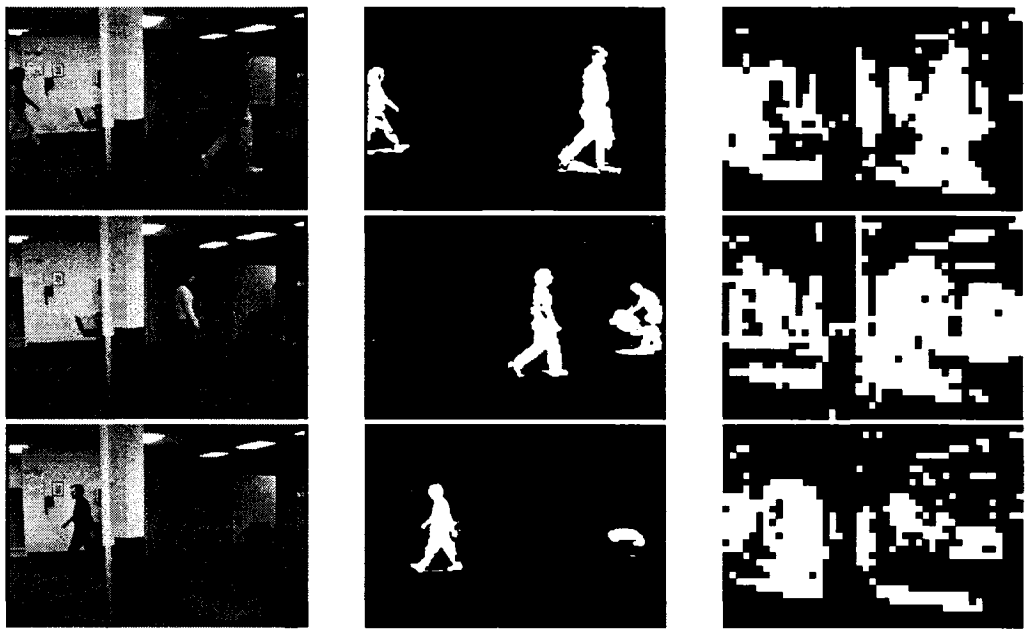
Figure 4.4: CD Comparison applied to "2Meet" with the original frame F_{102}, F_{303}, and F_{375}.

channels. Although it can obtain the complete moving objects, it includes serious spurious blobs caused by the local unimportant changes into the change masks.

Fig.4.6 shows the comparison results of video "Put object". Although both of the CD methods obtain stable and complete change masks for the video, the proposed CD method is more robust to the artifacts than the [60] CD method. The [60] CD is sensitive to the shadows and includes many spurious blobs into the change masks.

Fig.4.7 shows the comparison results of video "Road". As can be seen, both of the CD methods performs well. However, the [60] CD is sensitive to the shadows and includes considerable spurious blobs. The proposed CD performs the best.

The comparison results of "Vnj" are shown in Fig.4.8. Serious local unimportant changes caused by shadows and partial background movement are existed in the video. Although the proposed CD method includes some spurious blobs caused by the partial



(a) Org. (b) prop. CD (c) Alex. CD

Figure 4.5: CD Comparison applied to "Ekrlb" with the original frame F_{82}, F_{301}, and F_{410}.

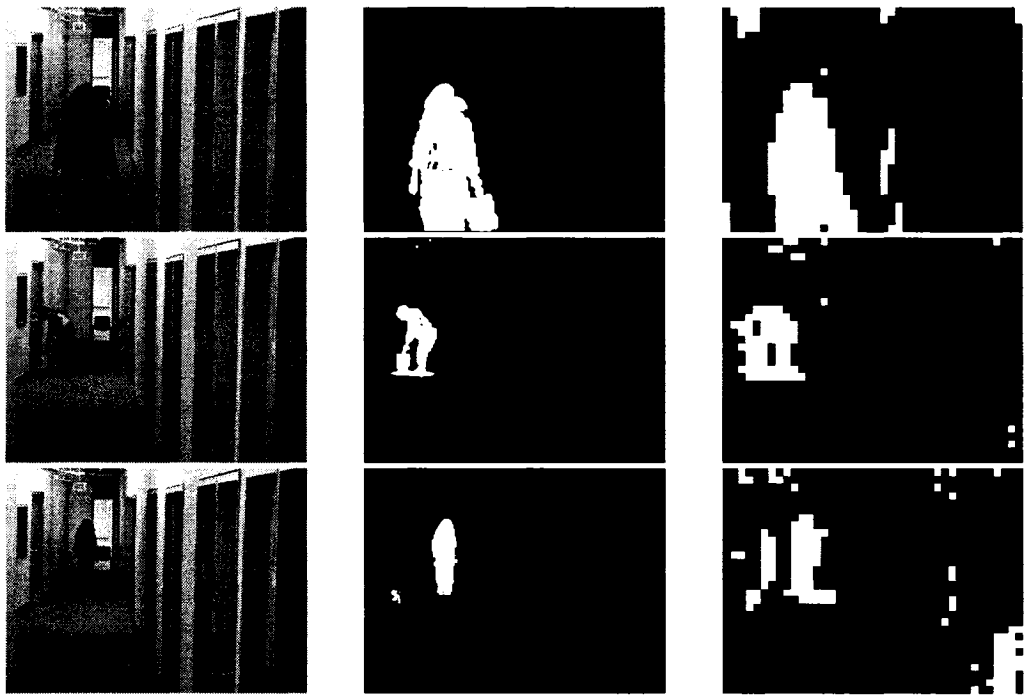
movement of background, it outperforms the [60] CD for the video. The [60] CD is sensitive to shadows and include many spurious blobs into the change masks.

4.3.4 Objective evaluation

In addition to the visually evaluation, we objectively evaluate the proposed CD algorithm as well as the two reference CD methods by the objective measures PCC, YC, and JC introduced in Section 4.2. The objective evaluation is performed based on video "Hall" and "Intelligent room" for which the ground truth sequences are available.

Fig.4.9 shows the objective comparsion results of the two CD algorithms. All the three objective measures clearly show that the proposed method outperforms the [60] CD.

Fig.4.10 shows the objective measures of the two CD algorithms for video "Intel-



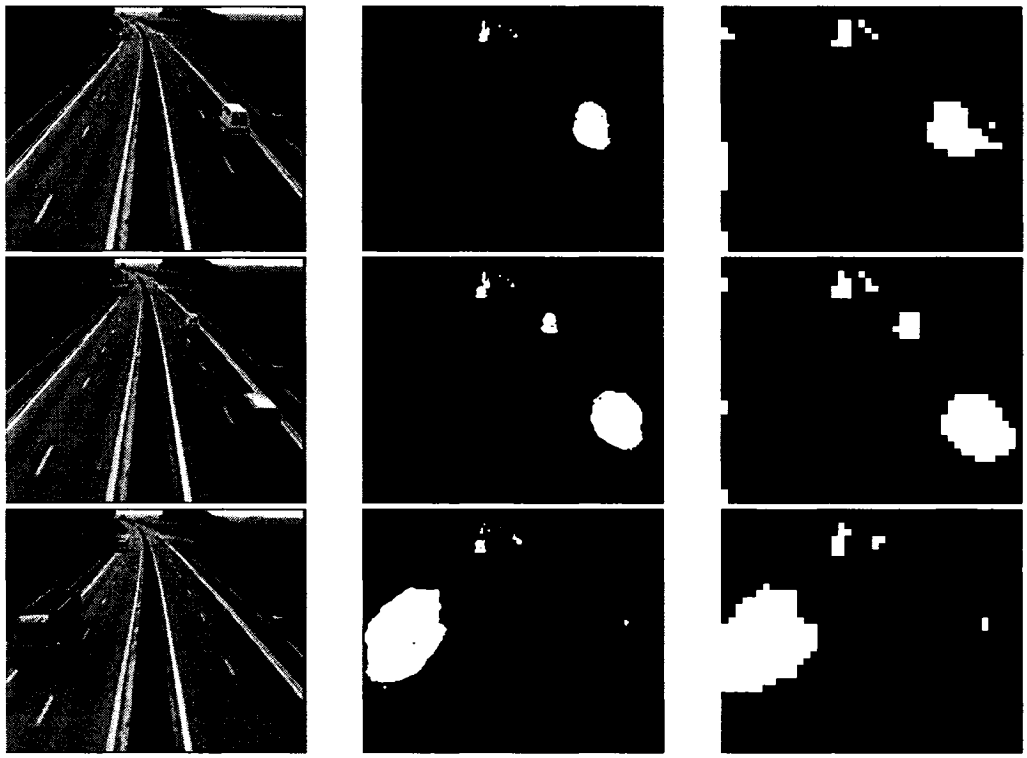
(a) Org. (b) prop. CD (c) [60] CD

Figure 4.6: CD Comparison applied to "Put object" with the original frame F_{143}, F_{410}, and F_{512}.

ligent room". The *PCC* measure shows that the performance of the proposed and the [60] CD are similar. Both *YC* and *JC* measures show that the proposed CD outperforms the [60] CD at most frames, and they have similar performance when the size of moving objects are small (see the objective measures between frame 150 and 200). The problem can be solved by applying more video-content robust thresholding algorithm, see Sec.4.4.2 and Sec.4.4.4 for details.

Fig.4.11 shows the objective evaluation of video "Ekrlb", where shadows and partial background movement are serious, and the foreground is similar to the background in both gray-level and chrominance channels. As can be seen, the proposed algorithm works stable and clearly outperforms the [60] CD.

Without performing complex CD in all chrominance channels of a video, the proposed method improves the quality of gray-level CD while it slightly increase the



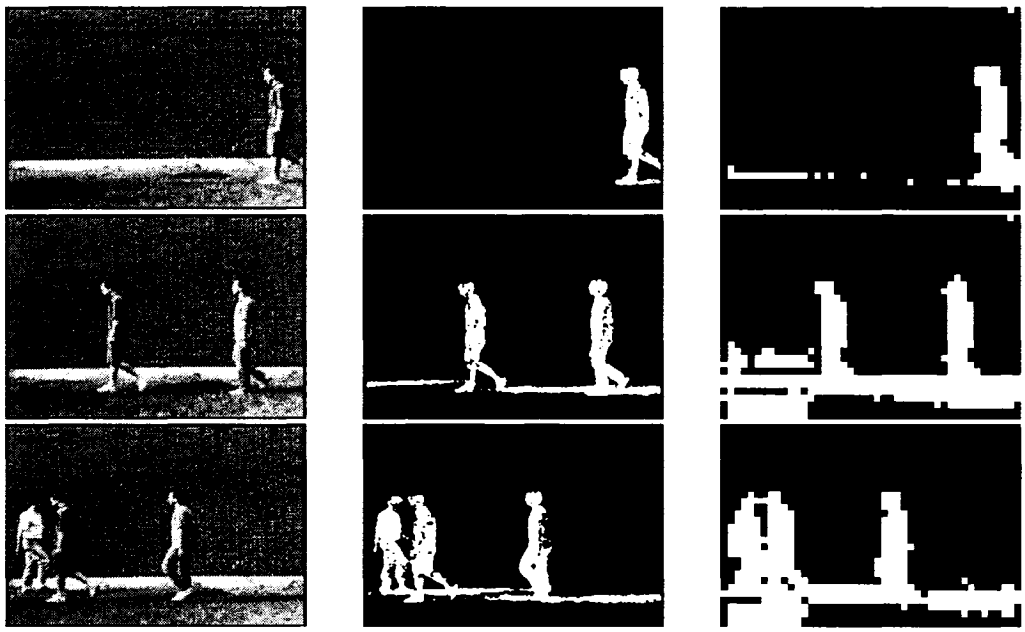
(a) Org. (b) prop. CD (c) [60] CD

Figure 4.7: CD Comparison applied to "Road" with the original frame F_{115}, F_{140}, and F_{251}.

computation time. Under Linux OS using C++, the average computation time of the proposed algorithm for CIF video sequences $0.0456s$ per frame (including thresholding). The [60] CD method requires $0.0518s$ per frame.

4.3.5 Analysis of the algorithm performance and its limitation

Our simulations show that color information significantly improves the change detection in cases where objects have similar gray-levels as the background. The proposed change detection method employs the color information by compensating the pixel intensities which are non-significant in the blocks-of-interest in gray-level channel but significant in chrominance channels. Experimental results show that the proposed

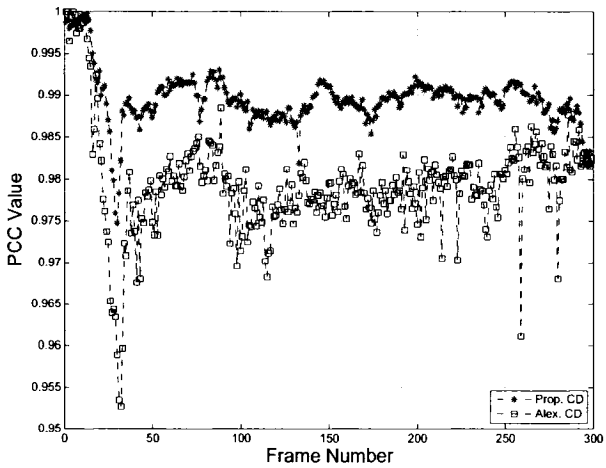


(a) Org. (b) prop. CD (d) [60] CD

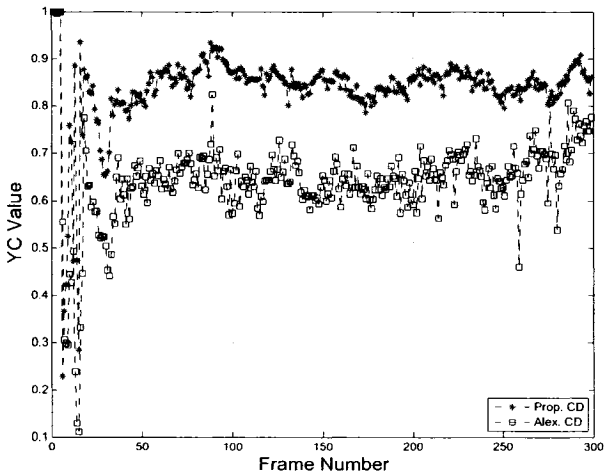
Figure 4.8: CD Comparison applied to "Vnj" with the original frame F_{107}, F_{166}, and F_{186}.

change detection method effectively include the color information into CD and performs well for real-world video sequences. Due to the proposed scatter estimation algorithm of the blocks-of-interest, the proposed change detection method is robust to multiple video contents. The reference change detection [60] is proposed for real-time application. It is sensitive the unimportant changes in video sequences.

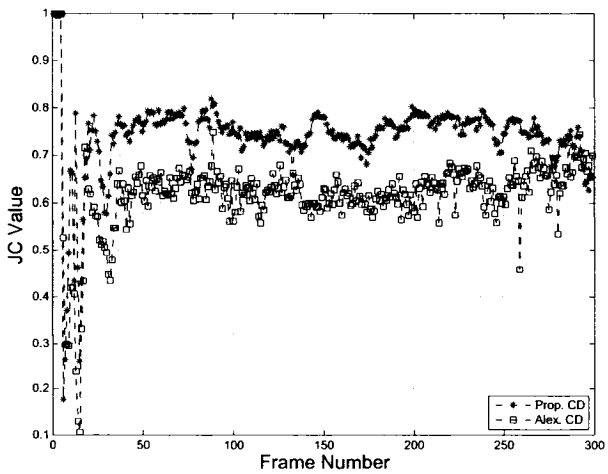
The proposed CD method has two limitations. First, for fast CD, the BOI scatter estimation is only performed in Y channel. It may miss the BOI if two conditions apply 1) the foreground and the background are very similar in gray-level and 2) the foreground has uniform gray-level (i.e., no texture). Second, if strong unimportant changes around object boundaries in Y are detected inside BOI and they are detected as important changes in chrominance channels then the proposed CD will classify these strong unimportant changes as important changes. A possible solution is discussed in Sec.5.2.



(a) PCC measures

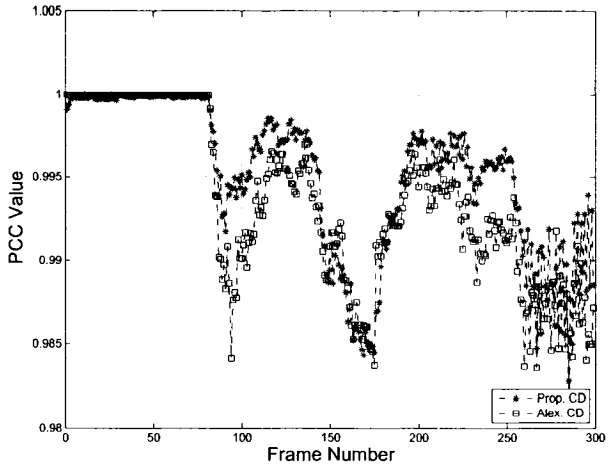


(b) YC measures

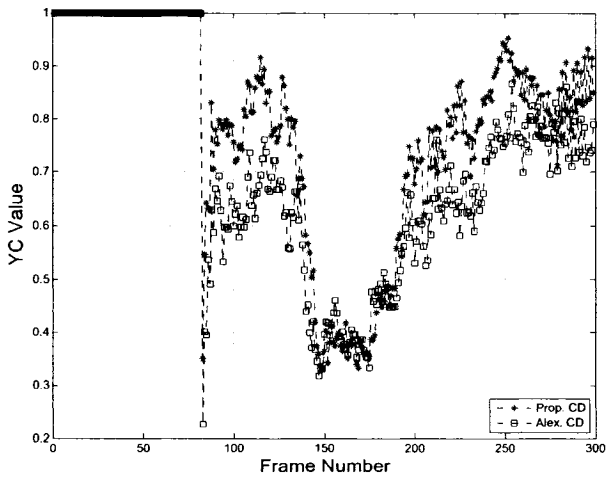


(c) JC measures

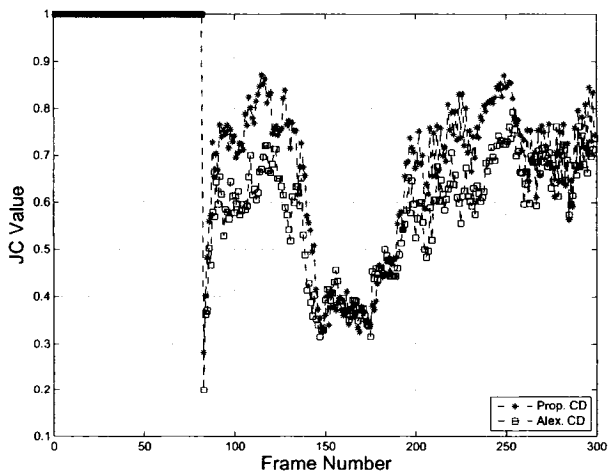
Figure 4.9: CD Objective comparison applied to "Hall".



(a) PCC measures

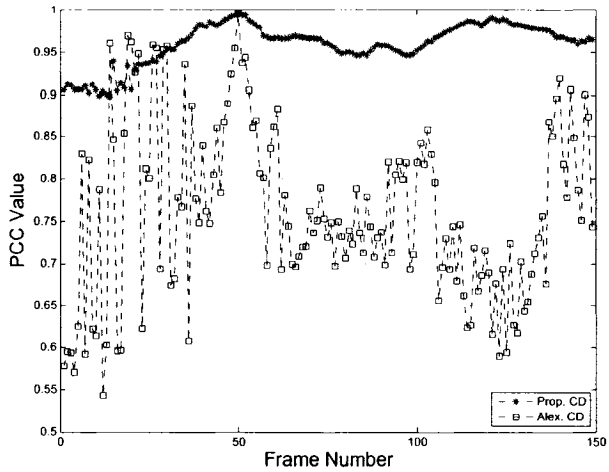


(b) YC measures

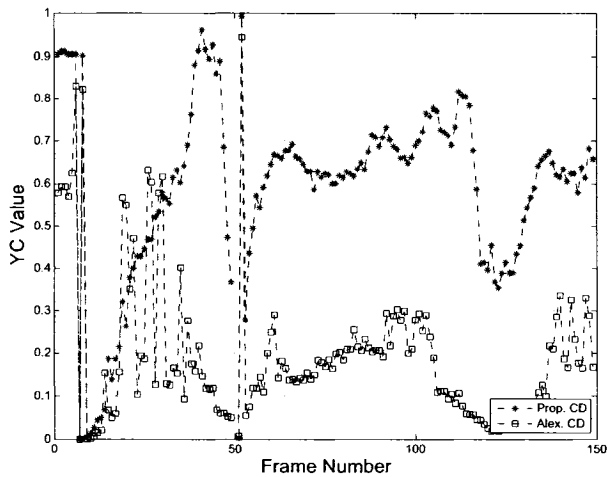


(c) JC measures

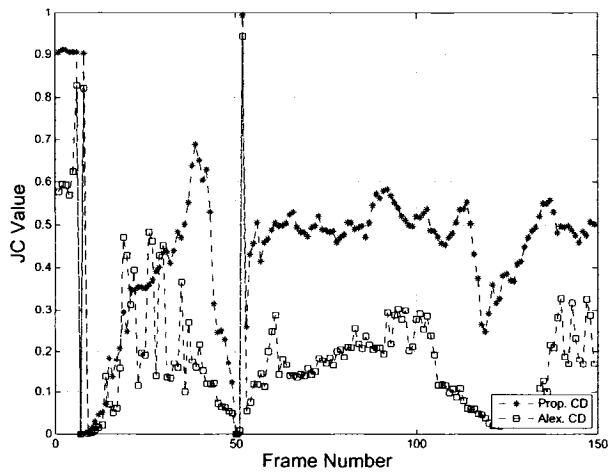
Figure 4.10: CD Objective comparison applied to "Intelligent room" (first 81 frames are in the training set).



(a) PCC measures



(b) YC measures



(c) JC measures

Figure 4.11: CD Objective comparison applied to "Ekrlb".

4.4 Evaluation Of Thresholding

The evaluation of the proposed thresholding algorithm is performed by applying it as well as three reference thresholding methods to real-world video sequences containing different video conditions. The three reference methods, i.e., Poisson-noise-model (Poisson), stable Euler-number (Euler), and Kapur methods, are shown in [57, 72] to perform best for change detection. (Our simulations confirm as in [72], that the classical Otsu [54] thresholding performs poorly for change detection.) In this section, a fast CD method proposed in [24] is applied on nine video sequences ("Hall", "Intelligent room", "ScriptLab2", "Stair", "Tennis", "Survey", "Vand_paint", "Vnj", and "Car") to obtain the difference frames.

In this section, A background frame BK_n is used to obtain the difference frames of each video. Although BK_n may be not available in some real-world applications, many background abstraction (or updating) algorithms [78, 79] had been proposed in literatures. A real-time background abstraction algorithm proposed in [80] is used in this thesis to obtain the background frames for on-line testing video sequences.

The parameters of the proposed thresholding do not changes for different video sequences.

4.4.1 Algorithm parameters

The proposed thresholding method needs four parameters: 1) the probability p_r for fixing the dominant distribution in low gray-level partation of a histogram (page 57), 2) the false alarm α_e for significance test in empty frames in (3.10), 3) the constant C_l for getting the adjust factor γ_l in Eq.(3.26) (for the video frames with *non-significant* local unimportant changes), and 4) the multiplication factor a_l for getting γ_l in Eq.(3.26) (for the video frames with *significant* local unimportant changes). In the training video set, the empty frames in video "3Meet" and "Intelligent room"are

used to estimate α_e, video "3Meet", "Road1", "Cross-St-Catherine", and "Cloudy" are used to estimate other parameters.

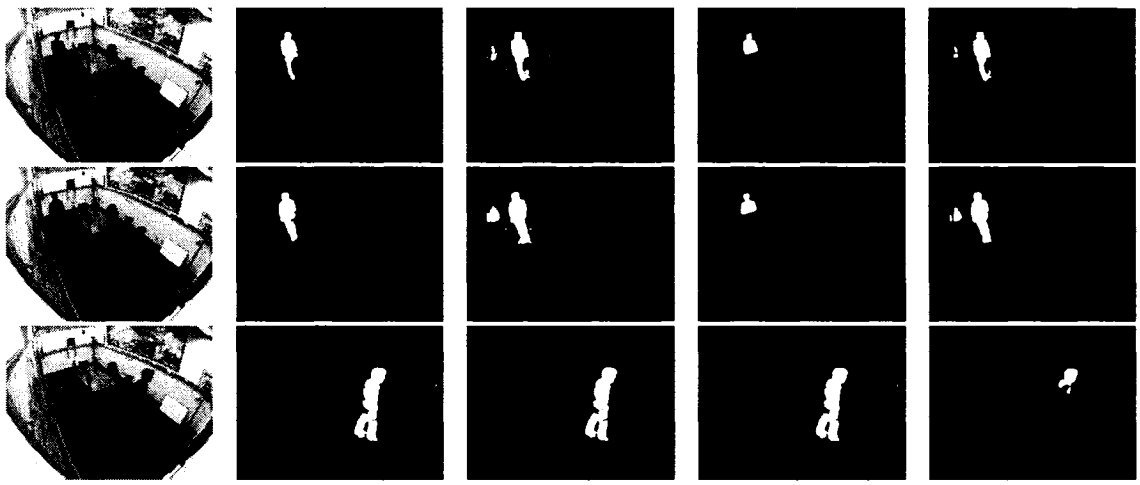
Based on observing the intensity distribution of the non-empty frames in video "3Meet", "Road1", "Cross-St-Catherine", and "Cloudy", we experimentally estimate $p_r \in [0.3, 0.5]$, we use $p_r = 0.5$ in this thesis. To robustly thresholding empty frames, the false alarm α_e should be very low. We tested considerable values of α_e between 5.0×10^{-6} to 5.0×10^{-12} using the empty frames of video "3Meet" and "Intelligent room", and experimentally set $\alpha_e = 2.5 \times 10^{-9}$. From (3.11), we get that $q > 5.96$ in Eq.(3.12) and in our simulations, q is set to 6.25. For the D_n with significant LUC ($\kappa_l > 1$), the adjust factor γ_l should be adaptive to the intensity distribution of D_n. This is because if the intensities in D_n vary in a relatively small range, i.e., the intensity distributions of the important and the unimportant changes are close to each other, γ_l should be relatively low to protect the important changes; otherwise, γ_l should be high to suppress the local unimportant changes. The value of γ_l is depended on the factor a_l, and from (3.26), we note that a_l can be determined by a desired false alarm α_l. By using the training video sequences "3Meet", "Road1", "Cross-St-Catherine", and "Cloudy", we tested a set of values of α_l, and experimentally estimated that α_l was between 0.2 and 0.0004. Using the Q-function table [65], we can get that a_l is between 1.25 and 3.54. To automatically compute a_l and avoid looking up the false alarm in the Q-function table frequently, we approximate a_l by a linear function shown in (4.4) according to the maximum intensity that a pixel may have in D_n, i.e., the G_s obtained by (2.33) (see page 39). Thus we have

$$\begin{aligned} a_l' &= 0.083 \cdot G_s - 5, \\ a_l &= \begin{cases} 1.25 & : \ a_l' < 1.25 \\ a_l' & : \ 1.25 \leq a_l' < 3.54 \\ 3.54 & : \ a_l' > 3.54. \end{cases} \end{aligned} \tag{4.4}$$

We use the same parameter values in all simulations.

4.4.2 Subjective evaluation under background subtraction

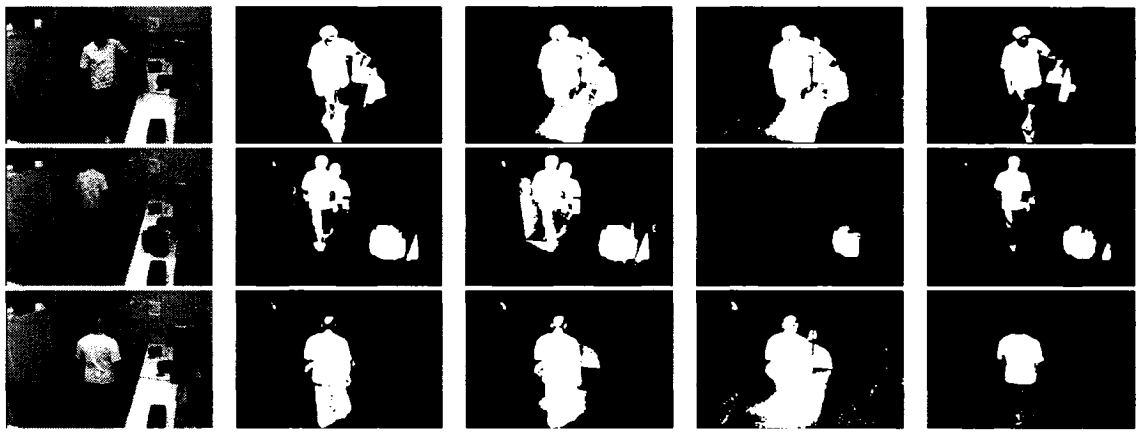
Fig.4.12 shows the comparison results of "Intelligent room". Since only the empty frames in the video are in the training set, it is fair that we compare the binary results of the non-empty frames obtained by the proposed as well as the reference methods. As can be seen, the proposed algorithm performs the best. The Euler method is the next best, however it is sensitive to LUC such as shadows. The Poisson and the Kapur methods are not stable and lead to seriously overthresholding.



(a) Org. (b) proposed (c) Euler (d) Poisson (e) Kapur

Figure 4.12: Comparison applied to "Intelligent room" with the original frame F_{105}, F_{233}, and F_{291}.

Fig.4.13 shows the comparison results of "Script2, which is challenging due to the similarities between the background and the foreground. Although slightly overthresholding the video, the proposed algorithm performs the best. The Euler method suffers due to its high sensitivity to shadows and local light changes. The Kapur method overthresholds the video thus loses parts of objects. The Poisson method is not stable. It underthresholds some frames while overthresholds some others.



(a) Org. (b) proposed (c) Euler (d) Poisson (e) Kapur

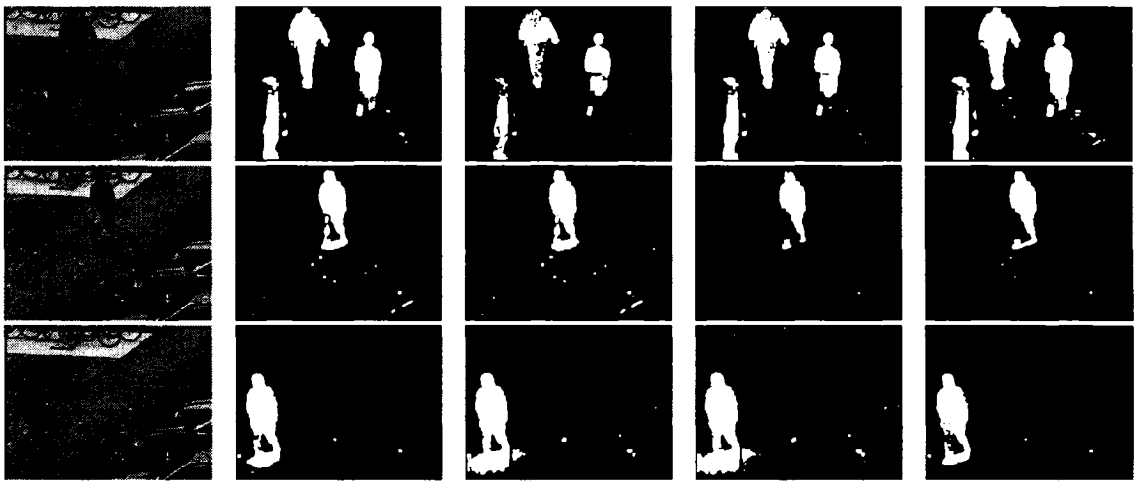
Figure 4.13: Comparison applied to "Script2" with the original frame F_{95}, F_{200}, and F_{2193}.

Fig.4.14 shows the superiority of the proposed algorithm with video "Stair". As can be seen, the proposed algorithm performs best. The reference methods breakdown for the empty frames. The Euler method includes many spurious blobs caused by shadows and local light changes. The Poisson method is not stable for failing in some frames. The Kapur method seriously overthresholds the video.

(a) Org. (b) proposed (c) Euler (d) Poisson (e) Kapur

Figure 4.14: Comparison applied to "Stair" with the original frame F_3, F_{202}, and F_{682}.

The comparison results of "Survey" is shown in Fig.4.15. The proposed algorithm performs best and obtains the clear and stable change masks. The Euler and the Poisson methods are not stable that they underthreshold some frames yet overthreshold some others. The Kapur method tends to overthreshold the video.

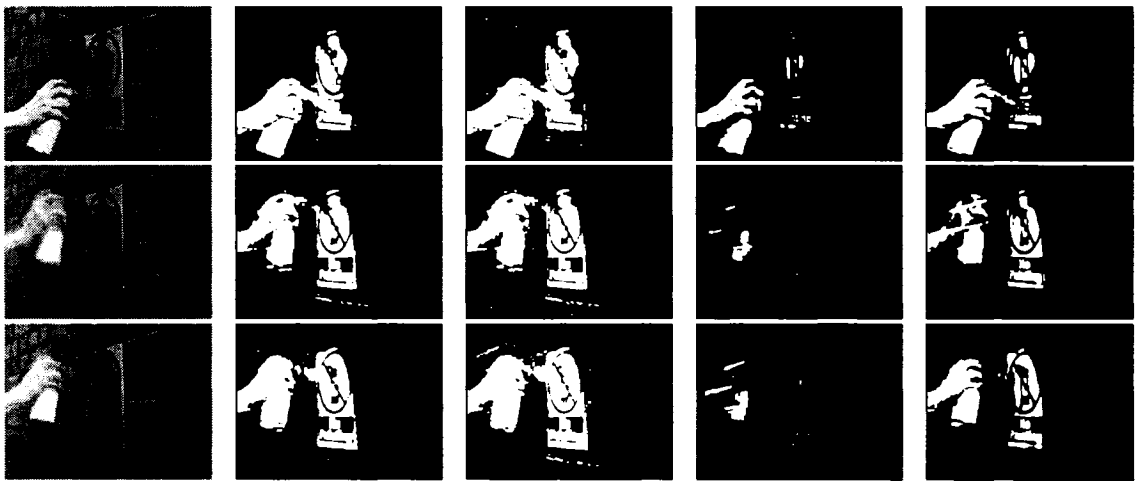


(a) Org. (b) proposed (c) Euler (d) Poisson (e) Kapur

Figure 4.15: Comparison applied to "Survey" with the original frame F_{111}, F_{626}, F_{655}, and F_{716}

Fig.4.16 shows the comparison results of "Vand_paint". We can see the proposed and the Euler methods performs best, however the Euler method is sensitive to the shadows. The Kapur method overthresholds some frames. The Poisson method performs poorly due to seriously overthresholding.

The comparison results of "Vnj" shown in Fig.4.17 also show the superiority of the proposed algorithm. As can be seen, although including some spurious blobs due to the serious LUC, the proposed algorithm performs best. The Kapur method seriously overthresholds the video. The Euler and the Poisson methods are sensitive to the LUC and breaks down form quite a few frames. In addition, all reference methods break down for the empty frames.



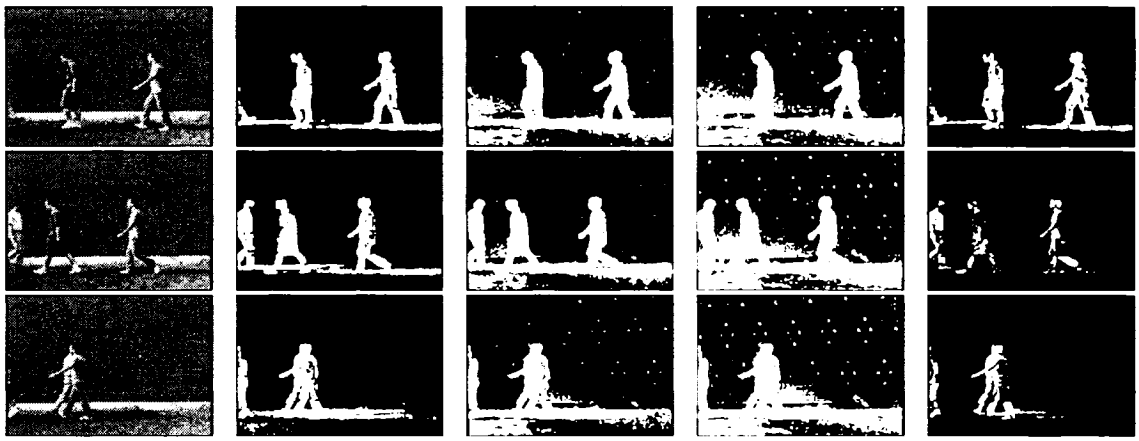
(a) Org. (b) proposed (c) Euler (d) Poisson (e) Kapur

Figure 4.16: Comparison applied to "Vand_Paint" with the original frame F_{105}, F_{129}, and F_{146}.

4.4.3 Subjective evaluation under global motion compensation

In this section, we evaluate the proposed thresholding method with the video sequences "Tennis" and "Car" with global motion (GM). The global motion estimation algorithm proposed in [81] is employed to compensate the global motion. Then the CD method in [24] is applied between the current frame and the GM-compensated frame.

Fig.4.18 and Fig.4.19 show the output of the proposed thresholding for video "Tennis" and "Car". As can be seen, the proposed thresholding is able to detect important changes and disregard unimportant changes, e.g., the spurious blobs caused by the inaccurate GM estimation and compensation. Since the Euler thresholding clearly outperforms the Poisson and the Kapur thresholding methods as shown in Sec.4.4.2, we only applied it to the two video sequences. Experimental results show that the performance of the Euler method is similar to the proposed method, but the proposed method is over 70 times faster than the Euler method (Note: the two video



(a) Org. (b) proposed (c) Euler (d) Poisson (e) Kapur

Figure 4.17: Comparison applied to "Vnj" with the original frame F_{172}, F_{181}, and F_{211}.

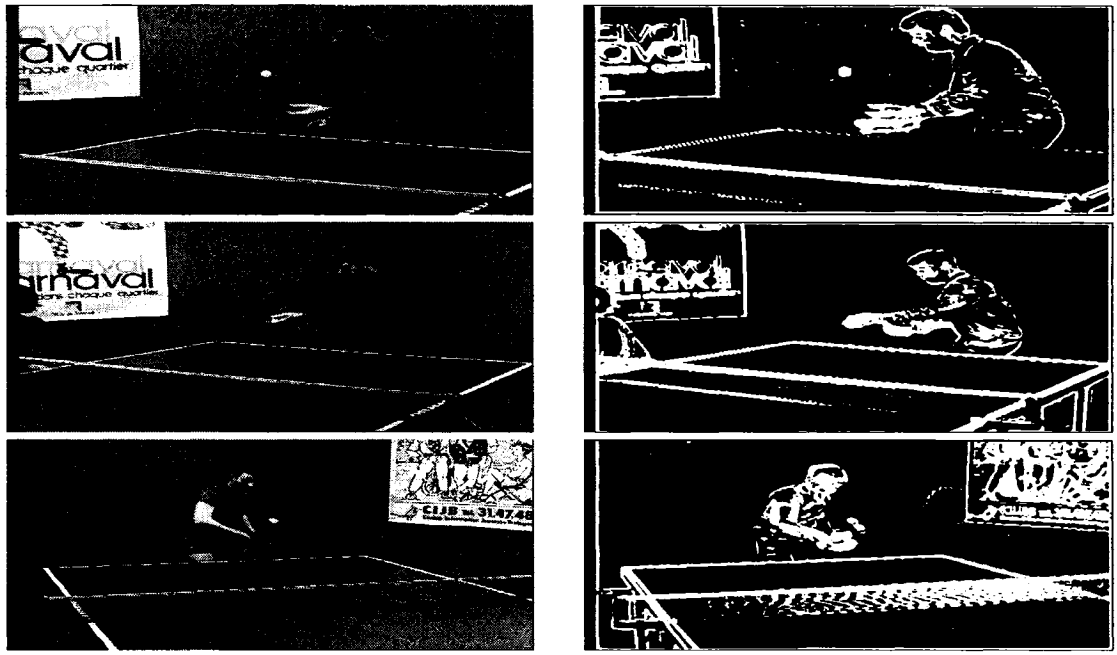
sequences are of frame size 720×576).

4.4.4 Objective evaluation under background subtraction

In addition to the subjective evaluations, the proposed algorithm and the reference methods are evaluated objectively by applying objective measures to the change masks they generated. In Section 4.2, three true/false positives and negatives comparison objective measures PCC, YC, and JC are introduced. In this section, we will use YC and JC measures since PCC measure may suffer when video sequences contain relatively small changes [72].

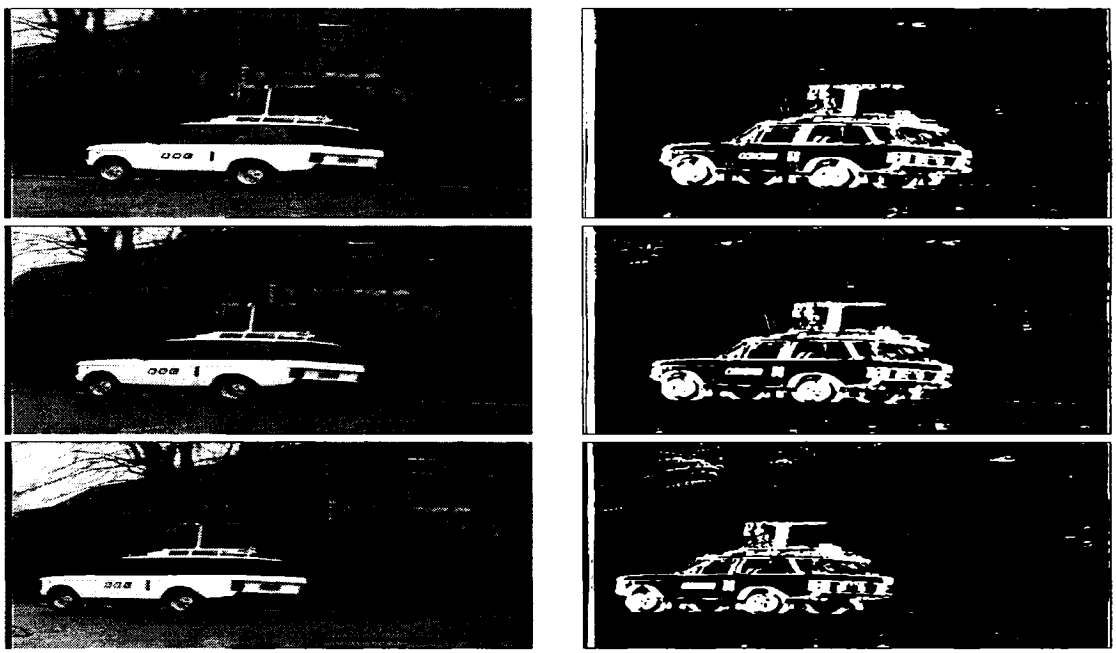
We have shown in the subjective evaluation in Sec.4.4.2 that the proposed and the Euler algorithms clearly outperform the Poisson and the Kapur methods. In this section, we will show the sample of objective comparison results between the proposed and the Euler methods. Fig.4.20 shows the PCC, YC, and JC measures of "Hall". As can be seen, the performance of the proposed and the Euler methods are similar, but the proposed method is slightly better than the Euler method.

The superiority of the proposed algorithm is shown in the objective comparison



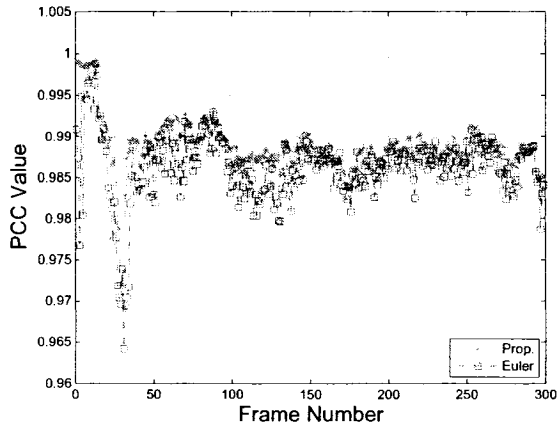
(a) Org. (b) binary mask

Figure 4.18: Comparison applied to "Tennis" with GM, the original frame F_{11}, F_{32}, and F_{44}.

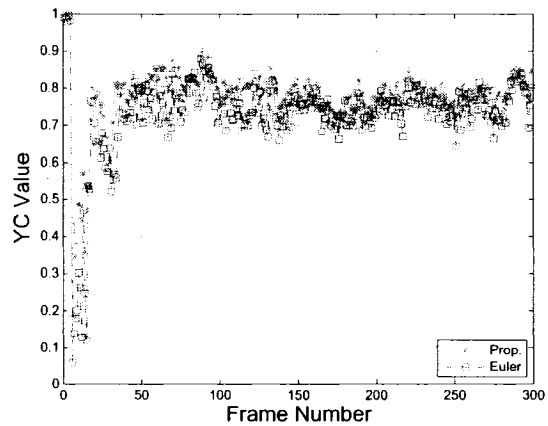


(a) Org. (b) binary results

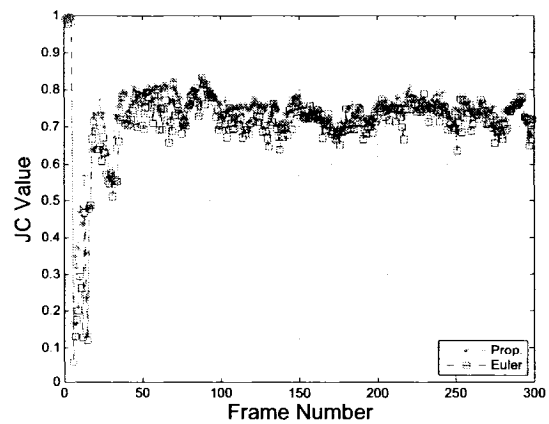
Figure 4.19: Comparison applied to "Car" with GM, the original frame F_{19}, F_{25}, and F_{41}.



(a) PCC measure



(b) YC measure



(c) JC measure

Figure 4.20: Objective comparison applied to "Hall".

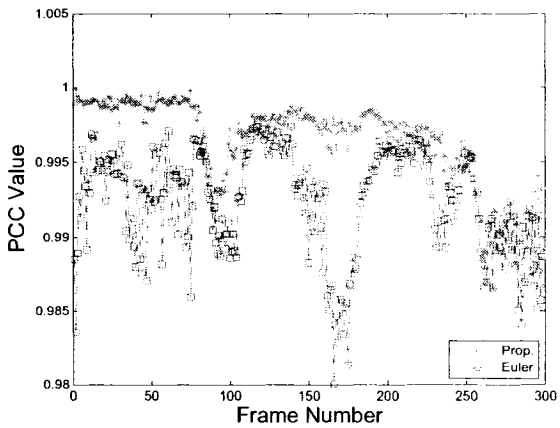
Table 4.1: Relative average computation time.

Video	Prop.	Poisson	Clas. Euler	Kapur
CIF	1	40.02	179.07	1.98
SIF	1	37.07	103.84	1.83

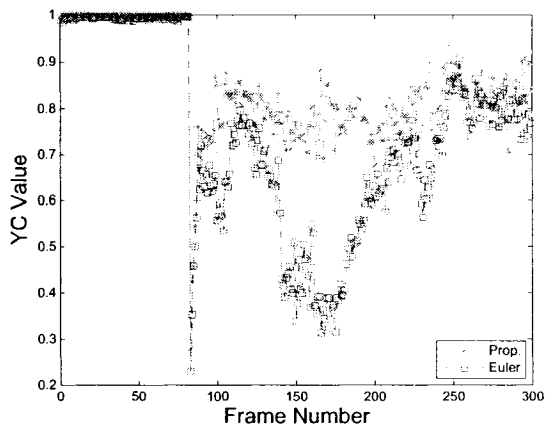
using the non-empty frames of "Intelligent room". Fig.4.21 shows the objective comparison results of the video. As can be seen, All objective measures show that the proposed algorithm is clearly better than the Euler method. In addition, the *PCC* measure shows that the proposed method clearly outperforms the Euler method for the empty video frames due to the proposed empty-frame detection algorithm. Note that the first 81 frames of "Intelligent room" are empty frames and used as training data.

Table 4.1 shows the average relative computation time of the video sequences used in our simulations. As can be seen, the efficiency of the proposed method is about 37 times higher than the Poisson method, over 103 times higher than the classical Euler method, and slightly higher than the Kapur method. As we have shown, the performance of the proposed method is much better than the Kapur method.

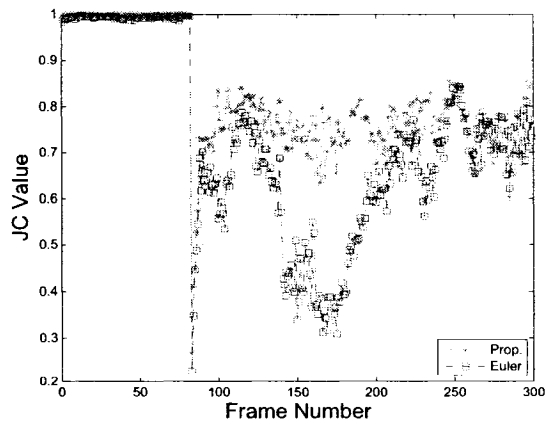
As shown in [74], the fast Euler thresholding is 12.5 times faster than the classical Euler thresholding and gives the same results. But 1) the system requirement (e.g., the memory buffer size) of [74] is much higher than the proposed algorithm, 2) as shown in Fig.4.14, we have seen that the Euler method is sensitive to local unimportant changes, and 3) the proposed method is still faster than the fast Euler method. The fast Euler method runs a 25 frames per second per 256×256 frame while the classical Euler method only runs at 2 frames per second on a 1G Hz CPU. The proposed method runs at 1162 frames per second per 352×288 frame while the classical Euler method runs at 6.49 frames per second on a 3G Hz CPU.



(a) PCC measure



(b) YC measure



(c) JC measure

Figure 4.21: Objective comparison applied to "Intelligent room" (the first 81 empty frames are used in training set).

4.4.5 Analysis of the algorithm performance and its limitation

Eight real-world video sequences are used to evaluate the proposed artfact-robust thresholding as well as the three state-of-the-art thresholding methods. Both visual assessment and objective assessment have shown that the proposed thresholding method is more robust and precise than the intensity-distribution based Kapur thresholding, and more stable than the spatial-property based Poisson and the Euler thresholding. The Kapur method is sensitive to gray-level distribution, which may be seriously affected by noise and local unimportant changes (LUC). The Poisson and the Euler methods are not temporally stable, and sensitive to LUC. The computational time consumming of the Poisson and the classical Euler methods are much higher than the proposed algorithm, and the fast Euler thresholding does not improve the robustness of the Euler method to LUC. Further more, the Poisson method is sensitive its parameter, and it breaks down at empty frames.

The limitation of the proposed thresholding method is that it is not sensitive to small size important changes. The small size important changes in a video frames are usually appeared similar to local unimportant changes, thus the proposed thresholding may overthreshold the frames by suppressing the local changes. This may be more worse for some extreme video conditions, e.g., a video sequence contains very small size important changes but with serious shadows.

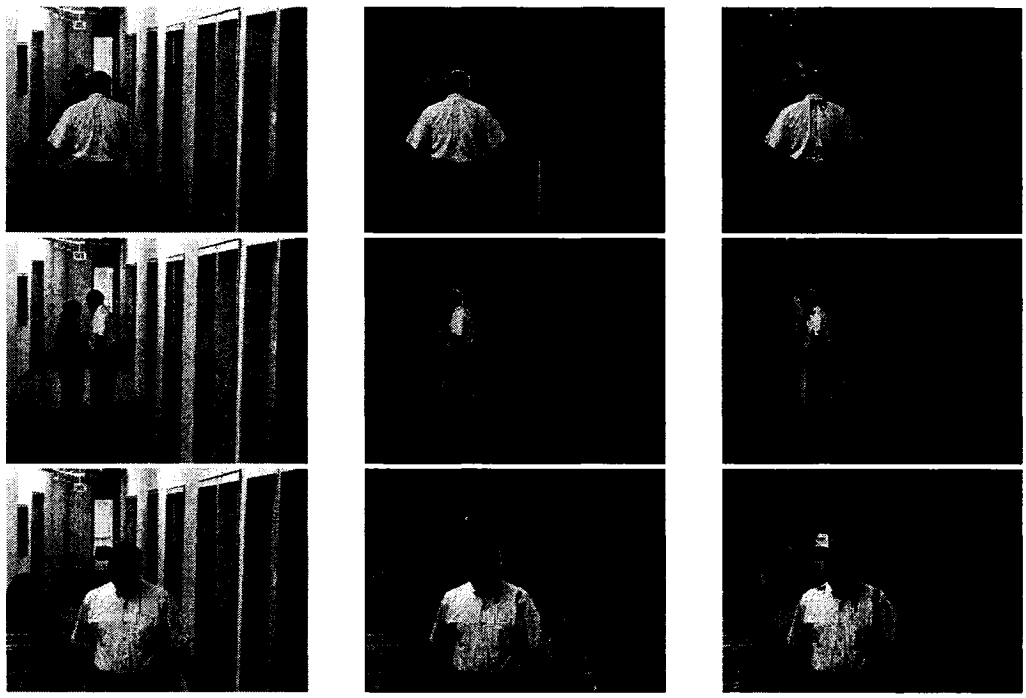
4.5 Combined CD and Thresholding (VOD)

Based on the proposed CD algorithm and the proposed thresholding method, an "one-pass" video object detection (VOD) is built for fast video applications. A background frame BK_n is used in this section. For the video sequences whose background frames are not available, a fast background modeling algorithm proposed in [80] is employed to obtain the background frames.

Four video sequences ("2Meet", "Ekrlb", "Pavement", and "Snow") containing different contents are used in subjective evaluation, and three video sequences ("Hall", "Intelligent room", and "Ekrlb") with ground truth sequences are used in objective evaluation. An improved background modeling based Lee VOD algorithm [49] based on the widely used Stauffer background-modeling VOD method [45], is used as the reference object detection method in this section. Note that [49] uses color-vector modeling based CD technique. It automatically updates the background frame while the proposed method use a fixed background frame. Note also that [49] includes postprocessing steps (e.g., data validate and components connection) that we do not implement to keep the comparison fair because the proposed method uses no postprocessing.

Fig.4.22 shows the comparison results of "2Meet". As can be seen in Fig.4.22, the proposed VOD method clearly outperforms the Lee VOD method at the integrality of the object masks and the robustness to video contents. The Lee VOD method loses parts of objects due to 1) the similarity between the foreground and the background in both gray-level and chrominance channels, and 2) the sensitivity to video contents. The video-content sensitivity of the Lee method is shown by the dependence to the moving directions of the video objects. In "2Meet", the moving directions of the two people is almost parallel to the axes of the camera, thus the pixels of parts of moving objects may vary their intensities very slow. This leads to the mistakenly

classification that the Lee method mistakenly classify some foreground regions into background regions. Also, the Lee VOD is sensitive to the local unimportant changes. We can see considerable spurious blobs are included into the object masks caused by shadows in Fig.4.22.

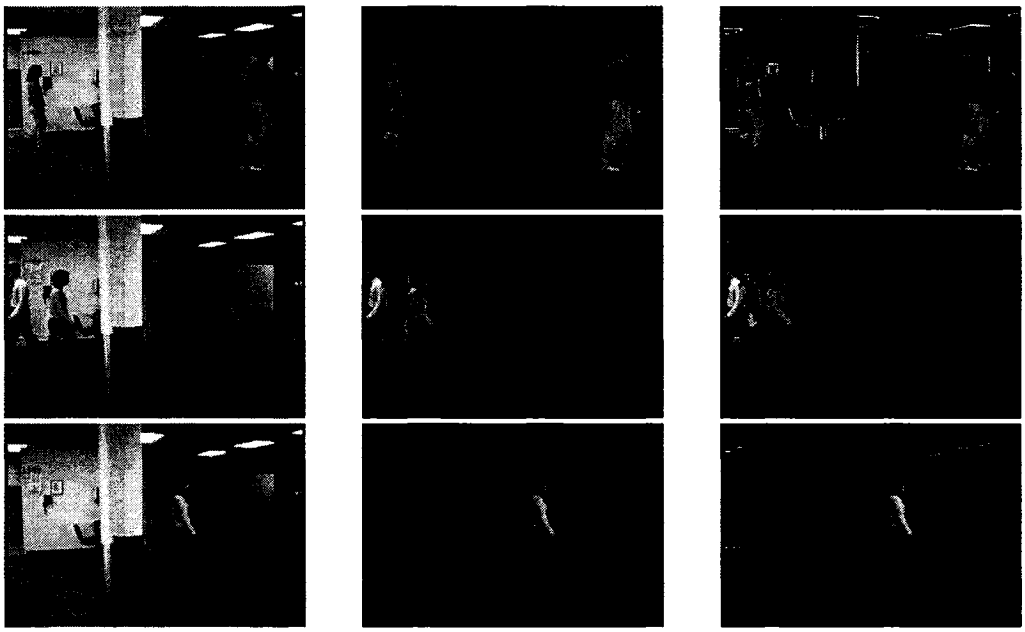


(a) Org. (b) proposed VOD (c) Lee. VOD

Figure 4.22: Comparison between the proposed VOD and the Lee VOD applied to "2Meet" with the original frame F_{226}, F_{360}, and F_{559}.

Fig.4.23 shows the comparison results of "Ekrlb". As can be seen, the proposed VOD method shows its superiority in the challenging video containing serious local changes, and considerable variation in contrast between the foreground and background. The object masks obtained by the proposed VOD method are stable and complete. The Lee method is sensitive to the local changes and suffers due to the similarity between the foreground and the background in both gray-level and chrominance channels.

Video "Pavement" and "Snow" are on-line surveillance shot at the Ste-Catherine



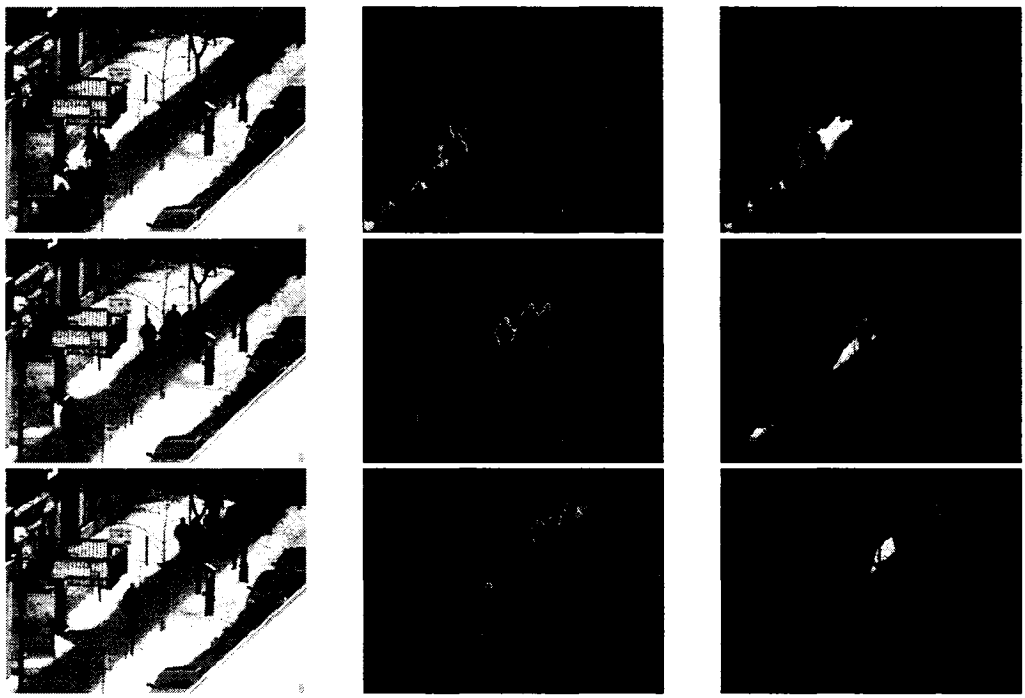
(a) Org. (b) proposed VOD (c) Lee. VOD

Figure 4.23: Comparison between the proposed VOD and the Lee VOD applied to "Ekrlb" with the original frame F_{90}, F_{262}, and F_{300}.

Avenue, Montreal, Quebec under different video conditions. Fig.4.24 shows the comparison results of "Pavement". Although both of the VOD methods can obtain the complete object masks, the Lee method is sensitive to the shadows thus includes considerable spurious blobs into the object masks. The proposed VOD method performs well for the video.

Video "Snow" is very challenging since it contains serious background turbulence caused by snow fall. As can be seen in Fig.4.25, the performance of the two VOD methods are similar, however, the Lee method includes quite a few spurious blobs caused by shadows. The proposed method is more robust to the video contents than the Lee method.

We also objectively evaluate the two VOD methods by applying them to three real-world video sequences whose ground truth sequences are available. Fig.4.26 shows the objective comparison results of "Hall". As can be seen, the PCC measure and



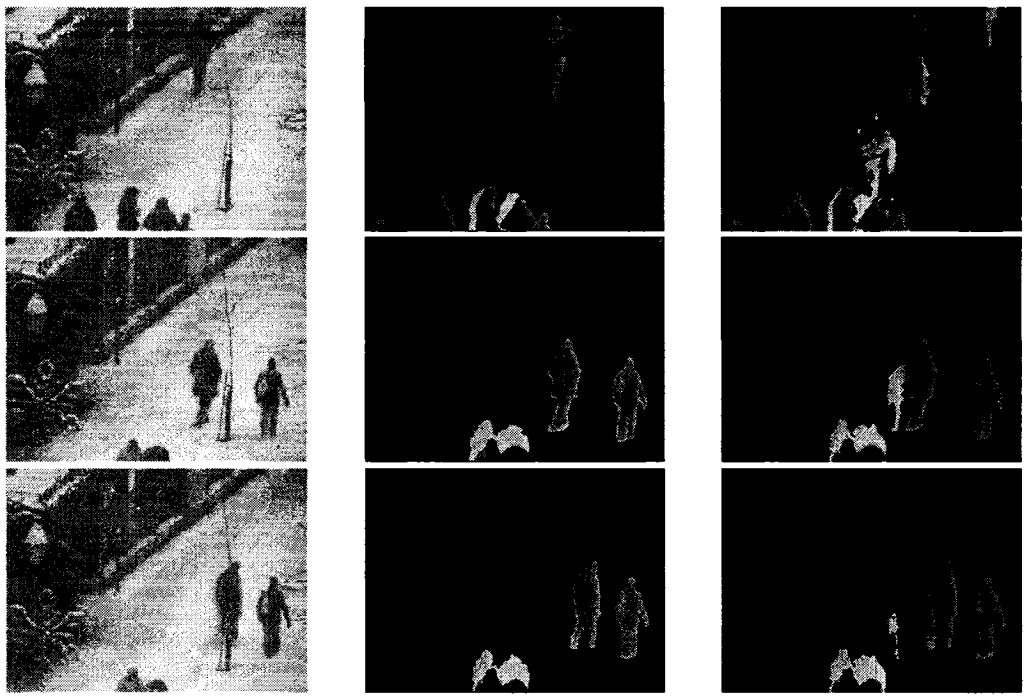
(a) Org. (b) proposed VOD (c) Lee. VOD

Figure 4.24: Comparison between the proposed VOD and the Lee VOD applied to "Pavement" with the original frame F_{1858}, F_{2436}, and F_{2641}.

the YC measure show that the performance of the proposed method is similar but better than the Lee method. The JC measure clearly shows that the proposed method outperforms the Lee method.

The objective comparison results of "Intelligent room" shown in Fig.4.27 show the superiority of the proposed VOD method. As can be seen in Fig.4.27, all the three objective measures show that the proposed method significantly outperforms the Lee method. Note that the first 81 frames of the video are used in training set.

Fig.4.28 shows the objective comparison results of challenging video "Ekrlb". As can be seen, the proposed method significantly performs better than the Lee method.



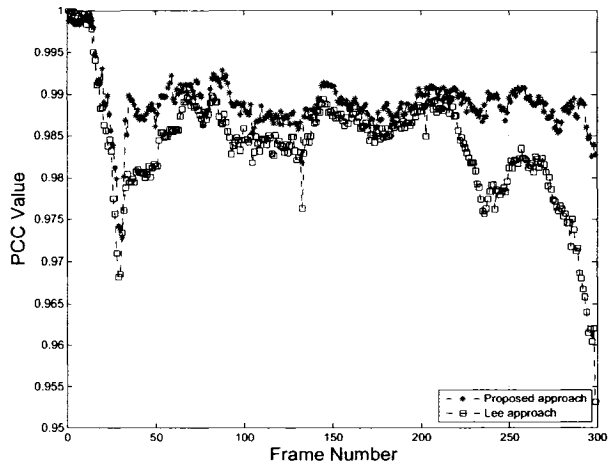
(a) Org. (b) proposed VOD (c) Lee. VOD

Figure 4.25: Comparison between the proposed VOD and the Lee VOD applied to "Snow" with the original frame F_{231}, F_{377} and F_{396}.

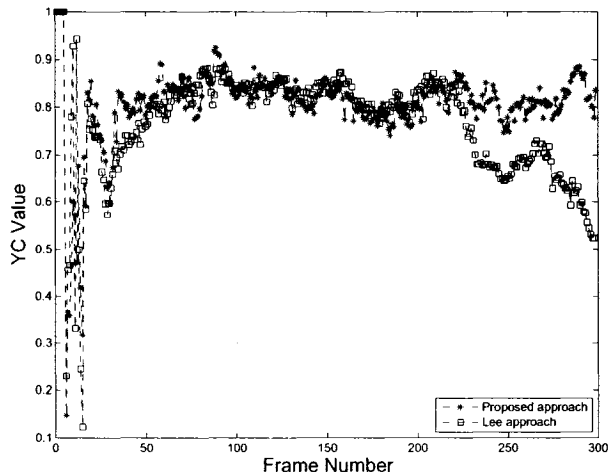
4.6 Summary

The proposed color-based gray-level compensation change detection algorithm, the proposed video-content adaptive thresholding algorithm, and the combined video object detection are evaluated both subjectively and objectively in this chapter. Fourteen real-world video sequences containing different video contents are used to testing the proposed approaches.

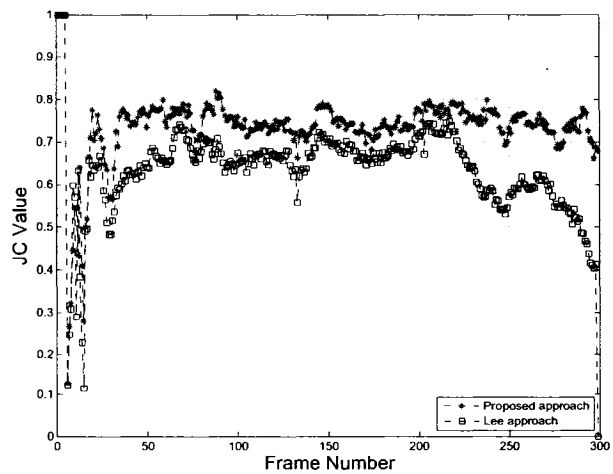
Simulations of change detection show that the proposed change detection algorithm effectively introduces the color information into change detection without performing complex computation in all chrominance channels. Based on the support from the proposed blocks-of-interest scatter estimation algorithm, the proposed change detection is robust to both global and local unimportant changes, and significantly improves the quality of change masks.



(a) PCC measure

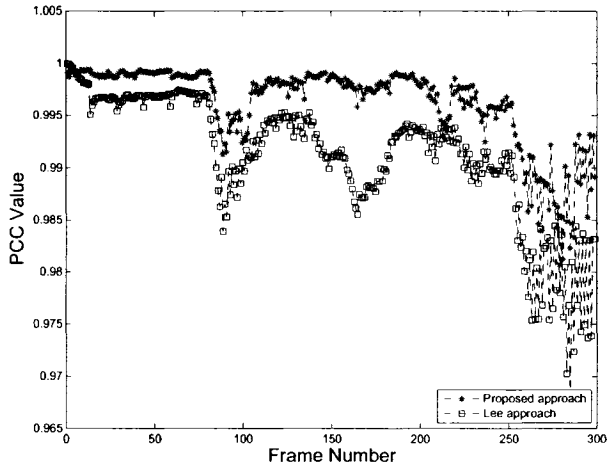


(b) YC measure

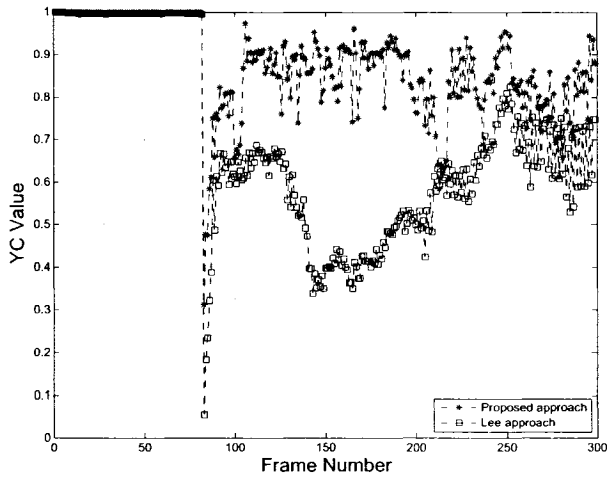


(c) JC measure

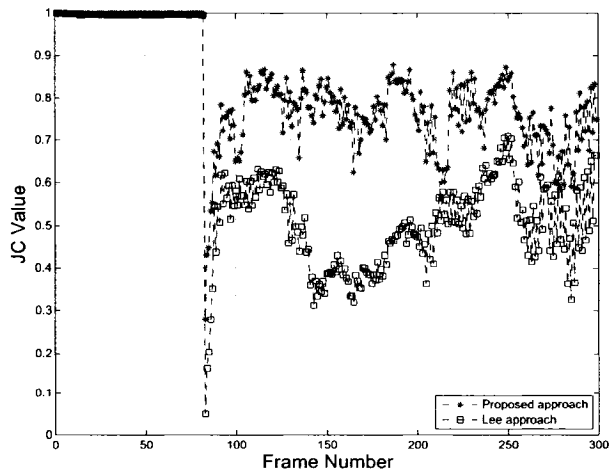
Figure 4.26: VOD Objective comparison applied to "Hall".



(a) PCC measure

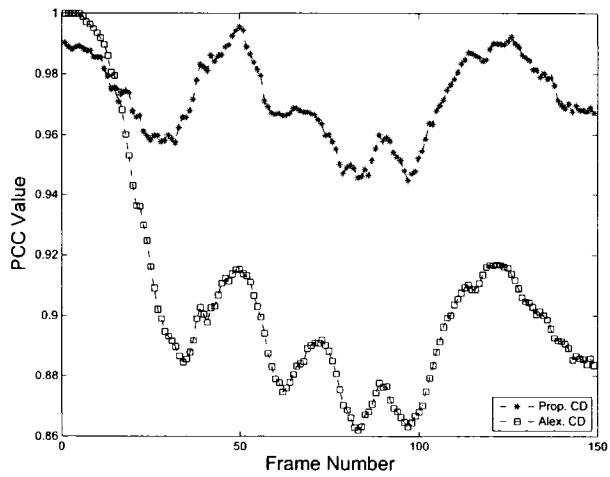


(b) YC measure

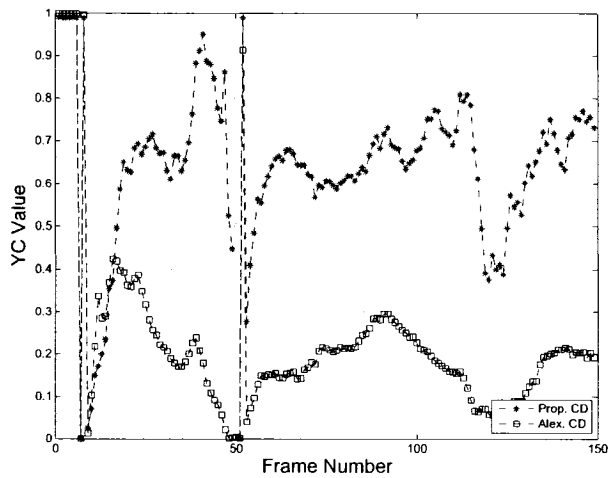


(c) JC measure

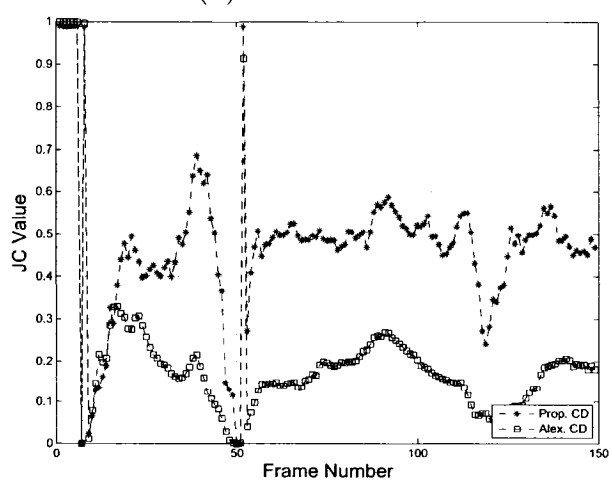
Figure 4.27: VOD Objective comparison applied to "Intelligent room" (the first 81 frames are used in training set).



(a) PCC measure



(b) YC measure



(c) JC measure

Figure 4.28: VOD Objective comparison applied to "Ekrlb". Note that frame 50 is an empty frame.

Simulations of thresholding show that the proposed thresholding algorithm is robust to multiple video contents. The proposed thresholding algorithm is more precise than the intensity-distribution based thresholding methods without increasing computation time, and more stable and much more efficient than the spatial-property based thresholding methods.

The video object detection consisting of the proposed change detection and the proposed thresholding algorithms clearly outperforms the background-modeling based reference method. On one hand, the proposed system is very sensitive to color information thus it successfully detects the objects which are similar to the background in all chrominance channels without including more spurious blobs into the object masks. On the other hand, the proposed system is robust to multiple unimportant changes. It works well under non-serious shadows without any supports from high-level analysis, while the reference method suffers in such a case.

Due to its efficiency and its robustness to artifacts, the proposed methods are suitable for on-line real-time video applications.

Chapter 5

Conclusion And Future Work

5.1 Conclusion

Motion-adaptive video object detection using change detection is efficient. However, change detection in real-world applications is challenging due to the complexity of video contents. In this thesis, we propose a content-adaptive video object detection algorithm consisting of change detection and thresholding.

The proposed fast color-based change detection algorithm uses the YUV color model, which has been proved as the most effective color model for video object detection. First, frame-differencing followed by absolute-value operation is performed in Y, U, and V channels of a video frame to obtain the difference frames of each channel. Under no-change hypothesis, the gray-levels in the Y channel are modeled as a Gaussian random variable, and the unimportant changes in the U and the V channels are modeled as two exponential random variables; then the maximum-intensity statistical model between the U and the V channels is then obtained. Second, an entropy based block-of-interest scatter estimation algorithm is proposed to locate the blocks in a frame potentially containing moving objects in the Y channel. Significance test computation is then applied to the blocks-of-interest in Y. The gray-levels of the pixels

which are non-significant in the Y blocks-of-interest but significant in the U or the V channel are statistically compensated based on their significance probabilities in color channels. Experimental results show that the proposed change detection algorithm is robust to complex video conditions, and without performing change detection in each color channel, the proposed change detection is efficient for real-world applications.

The proposed artifact-robust thresholding algorithm for change detection is proposed based on video content assessment. Using the proposed blocks-of-interest scatter estimation algorithm, a video-content assessment algorithm is proposed to 1) detect if a video frame is an empty frame, and 2) estimate the strength of local unimportant changes. According to the video-content assessment, the global threshold of a difference frame is computed discriminatively by a noise-statistic based thresholding method for the empty frames, or a local-artifact robust thresholding method for the non-empty frames. Experimental results show that the proposed thresholding algorithm is more precise and more robust to video-contents than the intensity-distribution based thresholding, and is more stable than the spatial properties based thresholding methods for change detection.

The combined proposed change detection and the thresholding algorithms is compared to a state-of-the-art motion-adaptive video object detection using background modeling. The proposed algorithm is 1) more sensitive to the important changes in both gray-level and chrominance channels thus it obtains more stable and complete change masks, and 2) more robust to the local unimportant changes such as shadows thus it outputs less noisy change masks.

This work has shown that 1) using color information for change detection can significantly improve the accuracy of change masks, 2) it is not necessarily to perform complex change detection in all channels, and 3) it is possible that implement fast and accurate video object detection for real-world video applications without explicitly

estimating the statistical model of local unimportant changes. Also we conclude that although thresholding is a very simple operation but if carefully designed it can be both efficient and accurate for classification.

5.2 Future Work

To further improve the performance of the proposed methods, the following steps to enhance can be implemented:

1. Improve the scatter estimation of blocks-of-interest: To efficiently detect the frame blocks potentially containing moving objects, the proposed scatter estimation of block-of-interest (BOI) is carried out in gray-level channel only. It may lose some BOI where the foreground is similar to the background in gray-level. We propose to include color information into the BOI estimation to improve the reliability of the algorithm.

2. Improve the gray-level compensation: The proposed color-based gray-level compensation algorithm is robust to the unimportant changes in non-ROC blocks since it only compensates the gray-level in BOI. However, it may be sensitive to strong unimportant changes such as serious shadows around the boundaries of objects. We propose to include the statistical models of shadows into the gray-level compensation to improve the robustness of the proposed algorithm to serious shadows.

Bibliography

[1] W. K. Ho, W. K. Cheuk, and D. P. K. Lun, "Content-based scalable H.263 video coding for road traffic monitoring," *IEEE Trans. On Multimedia*, vol. 7, no. 4, pp. 615–623, Aug. 2005.

[2] S. Cvetkovic, P. Bakker, J. Schirris, and P. H. N. de With, "Background estimation and adaptation model with light-change removal for heavily down-sampled video surveillance signals," in *IEEE Int. Conf. On Image Processing*, Atlanta, USA, Oct. 2006, pp. 1829 - 1832.

[3] J. Tao, M. Turjo, and Y. P. Tan, "Quickest change detection for health-care video surveillance," in *IEEE Int. Symposium On Circuits And Systems*, Kos, Greece, May 2006, pp. 505 - 508.

[4] G. Metta and P. Fitzpatrick, "Early integration of vision and manipulation," in *Int. Joint Conf. On Neural Networks*, Jul. 2003, vol. 4, p. 2703, Invited presentation.

[5] H. Narasimha-lyer, A. Can, B. Roysam, V. Stewart, H. L. Tanenbaum, A. Majerovics, and H. Singh, "Robust detection and classification of longitudinal changes in color retinal fundus images for monitoring diabetic retinopathy," *IEEE Trans. On Biomedical Engineering*, vol. 53, no. 6, pp. 1084–1098, Jun. 2006.

[6] Y. Wang, J. Ostermann, and Y. Q. Zhang, *Video Processing And Communications*, Prentice-Hall Inc., New Jersey, USA, 2002.

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

[8] H. F. Xu, A. A. Younis, and M. R. Kabuka, "Automatic moving object extraction for content-based applications," *IEEE Trans. On Circuits And Systems For Video Technology*, vol. 14, no. 6, pp. 796 - 812, Jun. 2004.

[9] L. J. Latecki and R. Miezianko, "Object tracking with dynamic template update and occlusion detection," in *IEEE Int. Conf. On Pattern Recognition*, Hong Kong, China, Aug. 2006, vol. 1, pp. 556 - 560.

[10] H. T. Nguyen and A. W. M. Smeulders, "Fast occluded object tracking by a robust appearance filter," *IEEE Trans. On Pattern Analysis And Machine Intelligence*, vol. 26, no. 8, pp. 1099 - 1104, Aug. 2004.

[11] S. Zhang, H. F. Chen, Z. Chi, and P. F. Shi, "An algorithm for segmenting moving vehicles," in *IEEE Int. Conf. On Acoustics, Speech, And Signal Processing*, Hong Kong, China, Apr. 2003, vol. 3, pp. 369 - 372.

[12] J. Xia and Y. Wang, "A spatio-temporal video analysis system for object segmentation," in *Int. Symposium On Image And Signal Processing And Analysis*, Aizu, Japan, Sept. 2003, vol. 2, pp. 812 - 815.

[13] W. Wei, K. N. Ngan, and N. Habili, "Multiple feature clustering algorithm for automatic video object segmentation," in *IEEE Int. Conf. On Acoustics, Speech, And Signal Processing*, Montreal, Canada, May 2004, vol. 3, pp. 625 - 628.

[14] I. Karliga and J. N. Hwang, "A framework for fully automatic moving video-object segmentation based on graph partitioning," in *IEEE Int. Symposium On Circuits And Systems*, Vancouver, Canada, May 2004, vol. 3, pp. 23 - 26.

[15] S. Mukhopadhyay and B. Chanda, "Multiscale morphological segmentation of gray-scale image," *IEEE Trans. On Image Processing*, vol. 12, no. 5, pp. 533 - 549, May 2003.

[16] E. Tuncel and L. Onural, "Utilization of the recursive shortest spanning tree algorithm for video-object segmentation by 2-D affine motion modeling," *IEEE Trans. On Circuits And Systems For Video Technology*, vol. 10, no. 5, pp. 776 - 781, Aug. 2000.

[17] Z. He, "Dynamic programming framework for automatic video object segmentation and vision-assisted video pre-processing," *IEE Proceedings On Vision, Image, And Signal Processing*, vol. 152, no. 5, pp. 597 - 603, Oct. 2005.

[18] S. C. Cheng, W. K. Huang, and T. L. Wu, "A fast global motion estimation for moving object segmentation using moment-preserving technique," in *IEEE Int. Symposium On Circuits And Systems*, Kobe, Japan, May 2005, vol. 4, pp. 2455 - 3458.

[19] N. Thakoor and J. Gao, "Automatic video object shape extraction and its classification with camera in motion," in *IEEE Int. Conf. On Image Processing*, Genoa, Italy, Sept. 2005, vol. 3, pp. 437 - 440.

[20] H. Okuda, M. Hashimoto, K. Sumi, and S. Kaneko, "Optimum motion estimation algorithm for fast and robust digital image stabilization," *IEEE Trans. On Consumer Electronics*, vol. 52, no. 1, pp. 127 - 131, Feb. 2006.

[21] I. Ahmad, W. Zheng, J. Luo, and M. Liou, "A fast adaptive motion estimation algorithm," *IEEE Trans. On Circuits And Systems For Video Technology*, vol. 16, no. 3, pp. 420-438, Mar. 2006.

[22] A. Bovik, *Handbook of Image And Video Processing*, vol. 2, Elsevier, USA, 2nd edition, 2005.

[23] G. de Haan, T. G. Kwaaitaal-Spassova, M. Larragy, and O. A. Ojo, "Memory integrated noise reduction IC for television," *IEEE trans. On Consumer Electronics*, vol. 42, no. 2, pp. 175–181, May 1996.

[24] A. Amer, "Memory-based spatio-temporal real-time object segmentation," in *Proc. SPIE Int. Symposium on Electronic Imaging, Conf. on Real-Time Imaging*, Santa Clara, USA, Jan. 2003, vol. 5012, pp. 10–21.

[25] R. J. Radke, S. Andra, O. Al-Kofahi, and B. Roysam, "Image change detection algorithms: a systematic survey," *IEEE Trans. On Image Processing*, vol. 14, no. 3, pp. 294–307, Mar. 2005.

[26] C. Kim and J. N. Hwang, "A fast and robust moving object segmentation in video sequences," in *IEEE Int. Conf. On Image Processing*, Kobe, Japan, Oct. 1999, vol. 2, pp. 131 – 134.

[27] C. Kim and J. N. Hwang, "Fast and automatic video object segmentation and tracking for content-based applications," *IEEE Trans. On Circuits And Systems For Video Technology*, vol. 12, no. 2, pp. 122 – 129, Feb. 2002.

[28] T. H. Chen, T. Y Chen, and Y. C. Chiou, "An efficient real-time video object segmentation algorithm based on change detection and background updating," in *IEEE Int. Conf. On Image Processing*, Atlanta, GA, USA, Oct. 2006, pp. 1837 – 1840.

[29] Y. Tsaig and A. Averbuch, "A region-based mrf model for unsupervised segmentation of moving objects in image sequences," in *IEEE Int. Conf. On Computer Vision And Pattern Recognition*, Kauai Marriott, Hawaii, Dec. 2001, vol. 1, pp. 1889–1896.

[30] D. Farin and P. H. N. de With, "Misregistration errors in change detection algorithms and how to avoid them," in *IEEE Int. Conf. on Image Processing*, Genoa, Italy, Sept. 2005, vol. 2, pp. 438–441.

[31] C. De Roover, M. Gabbouj, and B. Macq, "An accurate semi-automatic segmentation scheme based on watershed and change detection mask," in *SPIE Int. Conf. On Image and Video Communications and Processing*, San Jose, CA, Jan. 2005, vol. 5685, pp. 479–488.

[32] T. Aach and A. Kaup, "Statistical model-based change detection in moving video," *Signal Processing*, vol. 31, no. 2, pp. 165–180, Mar. 1993.

[33] T. Aach and A. Kaup, "Bayesian algorithms for adaptive change detection in sequences using markov random fields," *Signal Processing: Image Communication*, vol. 7, no. 2, pp. 147–160, Aug. 1995.

[34] R. Mech and M. Wollborn, "A noise robust method for 2D shape estimation of moving objects in video sequences considering a moving camera," *Signal Processing*, vol. 66, no. 2, pp. 203 – 217, Apr. 1998.

[35] A. Cavallaro and T. Ebrahimi, "Accurate video object segmentation through change detection," in *IEEE Int. Conf. On Multimedia And Expo*, Lusanne, Switzerland, Aug. 2002, vol. 1, pp. 445 – 448.

[36] X. Zhang, R. Zhao, and Z. Ma, "A new method for moving object extraction automatically," in *IEEE Int. Symposium On Communication And Information Technology*, Beijing, China, Oct. 2005, vol. 2, pp. 1460 – 1463.

[37] A. Cavallaro, O. Steiger, and T. Ebrahimi, "Semantic video analysis for adaptive content delivery and automatic description," *IEEE Trans. On Circuits And Systems For Video Technology*, vol. 15, no. 10, pp. 1200 - 1209, Oct. 2005.

[38] T. Aach and A. P. Condurache, "Transformation of adaptive thresholds by significance invariance for change detection," in *IEEE Workshop On Statistical Signal Processing*, Bordeaux, France, Jul. 2005, pp. 637 - 642.

[39] E. P. Ong, B. J. Tye, W. S. Lin, and M. Etoh, "An efficient video object segmentation scheme," in *IEEE Int. Conf. On Acoustics, Speech, And Signal Processing*, Orlando, Florida, USA, May 2002, vol. 4, pp. 3361 - 3364.

[40] J. Zhang, L. K. Zhang, and H. M. Tai, "Efficient object segmentation using adaptive background registration and edge-based change detection techniques," in *IEEE Int. Conf. On Multimedia and Expo*, Taipei, Taiwan, China, Jun. 2004, vol. 2, pp. 1467 - 1470.

[41] R. Cucchiara, C. Grana, M. Piccardi, and A. Prati, "Detecting moving objects, ghosts, and shadows in video streams," *IEEE Trans. On Pattern Analysis And Machine Intelligence*, vol. 25, no. 10, pp. 1337 - 1342, Oct. 2003.

[42] D. Farin, P. H. N. de With, and W. Effelsberg, "Robust background estimation for complex video sequences," in *IEEE Int. Conf. On Image Processing*, Barcelona, Spain, Sept. 2003, vol. 1, pp. 45 - 48.

[43] A. Cavallaro and T. Ebrahimi, "Video object extraction based on adaptive background and statistical change detection," in *SPIE Visual Communications And Image Processing*, San Jose, CA. USA, Jan. 2001, pp. 465 - 475.

[44] H. Han, Z. Wang, J Liu, Z. Li, B. Li, and Z. Han, "Adaptive background modeling with shadow suppression," in *IEEE Int. Conf. On Intelligent Transportation*, Orlando, FA, USA, Oct. 2003, vol. 1, pp. 720 - 724.

[45] C. Stauffer and W. E. L. Grimson, "Learning pattern of activity using real-time tracking," *IEEE Trans. On Pattern Analysis And Machine Intelligence*, vol. 22, no. 8, pp. 747 - 757, Aug. 2000.

[46] B. Lei and L. Q. Xu, "From pixels to objects and trajectories: a generic real-time outdoor video surveillance system," in *IEE Int. Symposium On Imaging For Crime Detection And Prevention*, London, UK, Jun. 2005, pp. 117 - 122.

[47] P. Kaewtrakulpong and R. Bowden, "An improved adaptive background mixture model for realtime tracking with shadow detection," in *European Workshop On Advanced Video-based Surveillance Systems*, Kingston, UK, Sept. 2001, pp. 90 - 95.

[48] Z. Zivkovic, "Improved adaptive Gaussian mixture model for background subtraction," in *Int. Conf. On Pattern Recognition*, Cambridge, UK, Aug. 2004, pp. 28 - 31.

[49] D. S. Lee, "Effective Gaussian mixture learning for video background subtraction," *IEEE Trans. On Pattern Analysis And Machine Intelligence*, vol. 27, no. 5, pp. 827 - 832, May 2005.

[50] M. Heikkila and M. Pietikainen, "A texture-based method for modeling the background and detecting moving objects," *IEEE Trans. On Pattern Analysis And Machine Intelligence*, vol. 28, no. 4, pp. 657 - 662, Apr. 2006.

[51] Y. Feng, J. Jiang, and S. S. Ipsm, "Towards automatic segmentation of semantic video objects," in *IEEE Int. Conf. On Computer As A Tool*, Serbia and Montenegro, Belgrade, Nov. 2005, vol. 2, pp. 987 - 990.

[52] G. Zhang and W. Zhu, "Automatic video object segmentation by integrating object registration and background constructing technology," in *IEEE Int. Conf. On Communications, Circuits and Systems*, Guilin, China, Jun. 2006, vol. 1, pp. 437 - 441.

[53] A. Chikando and J. Kinser, "Optimizing image segmentation using color model mixtures," in *Applied Imagery And Pattern Recognition Workshop*, Washington DC, US, Oct. 2005, pp. 230–235.

[54] N. Otsu, "A threshold selection method from gray-level histogram," *IEEE Trans. On Systems, Man and Cybernetics*, vol. 19, pp. 62–66, 1979.

[55] J. Kapur, P. Sahoo, and A. Wong, "A new method for gray-level picture thresholding using the entropy of histogram," *Computer Vision, Graphics Image Process*, vol. 29, no. 3, pp. 273–285, 1985.

[56] P. L. Rosin and T. Ellis, "Image difference threshold strategies and shadow detection," in *British Conf. On Machine Vision*, British, 1995, pp. 347–356, BMVA.

[57] P. L. Rosin, "Thresholding for change detection," *Computer Vision And Image Understanding*, vol. 86, no. 2, pp. 79–95, May 2002.

[58] A. Cavallaro and T. Ebrahimi, "Change detection based on color edges," in *IEEE Int. Symposium On Circuits And Systems*, Sydney, Australia, May 2001, vol. 2, pp. 141 - 144.

[59] L. D. Stefano, S. Mattoccia, and M. Mola, "A change-detection algorithm based on structure and colour," in *IEEE Int. Conf. On Advanced Video And Signal Based Surveillance*, Miami, FL, USA, Jul. 2003, pp. 252–259.

[60] T. Alexandropoulos, S. Boutas, V. Loumos, and E. Kayafas, "Real-time change detection for surveillance in public transportation," in *IEEE Conf. On Advanced Video And Signal Based Surveillance*, Teatro Social, Como, Italy, Sept. 2005, pp. 58–63.

[61] Y. Hwang, J. S. Kim, and I. Kweon, "Change detection using a statistical model of the noise in color images," in *IEEE Int. Conf. On Intelligent Robots And Systems*, Sendai, Japan, Oct. 2004, vol. 3, pp. 2713–2718.

[62] Y. Hwang, J. S. Kim, and I. Kweon, "Determination of color space for accurate change detection," in *IEEE Int. Conf. On Image Processing*, Atlanta, USA, Oct. 2006, pp. 3021 - 3024.

[63] N. Li, J. Bu, and C. Chen, "Real-time video object segmentation using HSV space," in *IEEE Int. Conf. On Image Processing*, New York, USA, Sept. 2002, vol. 2, pp. 445 - 448.

[64] E. Durucan and T. Ebrahimi, "Moving object detection between multiple and color images," in *IEEE Conf. On Advanced Video and Signal Based Surveillance*, Miami, FL, USA, Jul. 2003, pp. 243–251.

[65] A. Leon-Garcia, *Probability And Random Processes For Electrical Engineering*, Addison-Wesley Inc., Boston, MA, USA, 2nd edition, 1994.

[66] A. Amer and E. Dubois, "Fast and reliable structure-oriented video noise estimation," *IEEE trans. on Circuits And Systems Video Technology*, vol. 15, no. 1, pp. 113–118, Jan. 2005.

[67] P. L. Rosin, "Unimodal thresholding," *Pattern Recognition*, vol. 34, no. 11, pp. 2083–2096, Nov. 2001.

[68] I. E. Yairi H. Fujiyoshi and K. Kayama, "Road observation and information providing system for supporting mobility of pedestrian," in *IEEE Int. Conf. On Computer Vision Systems*, Manhattan, NY, USA, Jan. 2006, pp. 137–145.

[69] D. Li Y. Sun, I. Ahmad and Y. Q. Zhang, "Region-based rate control and bit allocation for wireless video transmission," *IEEE Trans. On Multimedia*, vol. 8, no. 1, pp. 1– 10, Feb. 2006.

[70] M. Ghazal, C. Su, and A. Amer, "Motion and region detection for effective recursive temporal noise reduction," in *IEEE. Int. Conf. On Acoustics, Speech, and Signal Processing*, Honolulu, Hawaii, USA, Apr. 2007, accepted.

[71] T. W. Ridler and S. Calvard, "Picture thresholding using an iterative selection method," *IEEE Trans. On Systems, Man and Cybernetics*, vol. 8, no. 8, pp. 630–632, 1978.

[72] P. L. Rosin and E. Ioannidis, "Evaluation of global image thresholding for change detection," *Pattern Recognition Letters*, vol. 24, no. 14, pp. 2345–2356, Oct. 2003.

[73] A. Pikaz and A. Averbuch, "Digital image thresholding based on topological stable state," *Pattern Recognition*, vol. 29, no. 5, pp. 829–843, May 1996.

[74] L. Snidaro and G. L. Foresti, "Real-time thresholding with Euler numbers," *Pattern Recognition Letters*, vol. 24, no. 9-10, pp. 1533 – 1544, Jun. 2003.

[75] A. Fitzgibbon, M. Pilu, and R. B. Fisher, "Direct least square fitting of ellipses," *IEEE Trans. On Pattern Analysis And Machine Intelligence*, vol. 21, no. 5, pp. 476–480, May 1999.

[76] C. E. Erdem, B. Sankur, and A. M. Tekalp, "Performance measures for video object segmentation and tracking," *IEEE Trans. On Image Processing*, vol. 13, no. 7, pp. 937–951, Jul. 2004.

[77] C. Su and A. Amer, "A real-time adaptive thresholding for video change detection," in *IEEE Int. Conf. On Image Processing*, Atlanta, GA, USA, Oct. 2006, pp. 157–160.

[78] L. Li, W. Huang, I. Y. H. Gu, and Q. Tian, "Statistical modeling of complex backgrounds for foreground object detection," *IEEE Trans. On Image Processing*, vol. 13, no. 11, pp. 1459–1472, Nov. 2004.

[79] S. Indupalli, M. A. Ali, and B. Boufama, "A novel clustering-based method for adaptive background segmentation," in *IEEE Int. Conf. On Computer and Robot Vision*, Quebec City, Canada, 2006, pp. 37–43.

[80] F. Achkar and A. Amer, "Hysteresis-based selective gaussian mixture models for real-time background maintenance," in *IS&T/SPIE Symposium on Electronic Imaging, Conf. on Visual Communications and Image Processing*, San Jose, CA, US, Jan. 2007, accepted.

[81] B. Qi and A. Amer, "Robust and fast global motion estimation oriented to video object segmentation," in *IEEE Int. Conf. On Image Processing*, Genoa, Italy, Sept. 2005, vol. 1, pp. 153 – 156.