VIDEO PROCESSING METHODS ROBUST TO ILLUMINATION VARIATIONS

A THESIS
SUBMITTED TO THE DEPARTMENT OF ELECTRICAL AND
ELECTRONICS ENGINEERING
AND THE INSTITUTE OF ENGINEERING AND SCIENCES
OF BILKENT UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF SCIENCE

By
Fuat Çoğun
09.08.2010

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

Prof. Dr. A. Enis Çetin (Supervisor)

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

Asst. Prof. Dr. Sinan Gezici

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

Asst. Prof. Dr. İbrahim Körpeoğlu

Approved for the Institute of Engineering and Sciences:

Prof. Dr. Levent Onural
Director of Institute of Engineering and Sciences

ABSTRACT

VIDEO PROCESSING METHODS ROBUST TO ILLUMINATION VARIATIONS

Fuat Çoğun
M.S. in Electrical and Electronics Engineering
Supervisor: Prof. Dr. A. Enis Çetin
09.08.2010

Moving shadows constitute problems in various applications such as image segmentation, smoke detection and object tracking. Main cause of these problems is the misclassification of the shadow pixels as target pixels. Therefore, the use of an accurate and reliable shadow detection method is essential to realize intelligent video processing applications. In the first part of the thesis, a cepstrum based method for moving shadow detection is presented. The proposed method is tested on outdoor and indoor video sequences using well-known benchmark test sets. To show the improvements over previous approaches, quantitative metrics are introduced and comparisons based on these metrics are made.

Most video processing applications require object tracking as it is the base operation for real-time implementations such as surveillance, monitoring and video compression. Therefore, accurate tracking of an object under varying scene and illumination conditions is crucial for robustness. It is well known that illumination variations on the observed scene and target are an obstacle against robust object tracking causing the tracker lose the target. In the second part of the thesis, a two dimensional (2D) cepstrum based approach is proposed to overcome this problem. Cepstral domain features extracted from the target region are introduced into the covariance tracking algorithm and it is experimentally observed that 2D-cepstrum analysis of the target region provides robustness to varying illumination conditions. Another contribution is the development of the co-difference matrix based object tracking instead of the recently introduced covariance matrix based method.

One of the problems with most target tracking methods is that they do not have a well-established control mechanism for target loss which usually occur

when illumination conditions suddenly change. In the final part of the thesis, a confidence interval based statistical method is developed for target loss detection. Upper and lower bound functions on the cumulative density function (cdf) of the target feature vector are estimated for a given confidence level. Whenever the estimated cdf of the detected region exceeds the bounds it means that the target is no longer tracked by the tracking algorithm. The method is applicable to most tracking algorithms using features of the target image region.

Keywords: Moving Shadow Detection, 2D-Cepstrum Analysis, Object Tracking under Illumination Conditions, Target Loss Detection.

ÖZET

IŞIK DEĞİŞİMLERİNE DAYANIKLI VİDEO İŞLEME YÖNTEMLERİ

Fuat Çoğun
Elektrik ve Elektronik Mühendisliği Yüksek Lisans
Tez Yöneticisi: Prof. Dr. A. Enis Çetin
09.08.2010

Hareketli gölgeler imge ayrıştırma ve nesne takibi gibi çesitli uygulamalarda sorunlar teşkil etmektedir. Bu sorunların ana sebebi gölge piksellerinin hedef pikselleri olarak yanlış sınıflandırılmasıdır. Bu nedenle, doğru ve güvenilir bir gölge tesbit yönteminin kullanımı, ileri video işleme uygulamalarının gerçekleştirilebilmesi için şarttır. Tezin ilk bölümünde, hareketli gölge tesbiti için cepstrum tabanlı bir yöntem sunulmaktadır. Önerilen yöntem, iyi bilinen açik hava ve kapali ortam video örneklerinde kıyaslama amacıyla test edilmiştir. Önceki yaklaşımların üzerine sağlanan iyileştirmelerin gösterilmesi amacıyla, nicel ölçütler tanıtılmış ve bu ölçütlere dayalı karşılaştırmalar yapılmıştır.

Çoğu video işleme uygulaması; gözetleme, izleme ve video sıkıştırma gibi gerçek zamanlı uygulamalarda nesne takibinin temel işlem olması sebebiyle nesne takibini gerektirir. Bu nedenle, bir nesnenin değişen ışık koşulları altında doğru takibi güvenilirlik açısından çok önemlidir. Bilindiği üzere gözlenen ortam ve hedef üzerinde oluşan ışık değişiklikleri, takipçinin hedefi kaybetmesine sebep olması yönüyle güvenilir nesne takibinin önünde bir engel teşkil etmektedir. Tezin ikinci bölümünde, bu sorunun üstesinden gelmek amacıyla iki boyutlu (2B) cepstrum kullamına dayalı bir yaklaşım önerilmiştir. Hedef bölgesinden elde edilen cepstral alan özellikleri kovaryans takip algoritmasına tanıtılmış ve deneysel olarak hedef bölgesinin 2B-cepstrum analizinin değişen ışık koşullarına karşı getirdiği dayanıklılık gözlemlenmiştir. Diğer bir katkı ise kovaryans matris tabanlı nesne takip yönteminin yerine kullanılabilecek ko-fark matris tabanlı nesne takip yönteminin geliştirilmesidır.

Çoğu hedef takip yönteminin sorunlarından biri, genelde ışık değişikliklerinden kaynaklanan hedef kaybı durumu için oluşturulmuş, iyi kurulmuş bir tesbit

mekanizmasının bulunmayışıdır. Tezin son bölümünde, hedef kaybı tesbiti için güven aralığı tabanlı istatistiksel bir yöntem gelistirilmiştir. Hedefin özellik vektörünün yığmalı dağılım fonksiyonunun üst ve alt sınır fonksiyonları, verilen bir güven düzeyi için tahmin edilir. Hedef bölgesi olarak tahmin edilen bölgenin tahmini yığmalı dağılım fonksiyonunun sınırları aşması durumunda hedefin takip algoritması tarafından artık takip edilmediği sonucuna varılır. Bu yöntem hedef imge bölgesinin özelliklerini kullanan çoğu hedef takip algoritmasına uygulanabilir.

Anahtar Kelimeler: Hareketli Gölge Tesbiti, 2B-Cepstrum Analizi, Değişken Işık Koşullarında Nesne Takibi, Hedef Kaybı Tesbiti.

ACKNOWLEDGMENTS

I would like to thank my supervisor Prof. Dr. A. Enis Çetin for his guidance and support throughout my graduate education and my thesis research, Asst. Prof. Dr. Sinan Gezici and Asst. Prof. Dr. İbrahim Körpeoğlu for being members of my thesis defense committee and Asst. Prof. Dr. Soosan Behesti for her contributions in Chapter 4.

I would also like to thank my wife and my family for their encouragements and endless support during my graduate studies.

Contents

1 Introduction **1**

1.1 Contributions of this Thesis 1

1.2 Thesis Organization 2

2 Moving Shadow Detection **3**

2.1 Related Work 4

2.2 Moving Shadow Detection Algorithm 5

2.2.1 Hybrid Background Subtraction Method 5

2.2.2 Cepstrum Analysis for Shadow Identification 7

2.3 Experimental Results 10

2.4 Summary 12

3 Object Tracking under Illumination Variations **16**

3.1 Related Work 17

3.2 Moving Object Tracking Algorithm 18

3.2.1 Covariance and Co-difference Tracking Methods 19

3.2.2 Two-dimensional (2D) Cepstrum Analysis of the Target . . 21

3.3 Experimental Results 22

3.4 Summary 25

4 Target Loss Detection **28**

4.1 Target Loss Detection Algorithm 28

4.1.1 Object Signature Function 28

4.2 Bounds and the Confidence Region 31

4.3 Experimental Results 31

4.4 Summary 34

5 Conclusions **41**

List of Figures

2.1 "Laboratory" video sequence 12

2.2 "Intelligent room" video sequence 13

2.3 "Campus" video sequence 14

2.4 "Highway I" video sequence 15

3.1 A man walking into a building. (a) Covariance Tracker (b) Cepstrum-based Covariance Tracker (c) Cepstrum-based Co-difference Tracker 25

3.2 A man walking to a shadowed area. (a) Co-difference Tracker (b) Cepstrum-based Covariance Tracker (c) Cepstrum-based Co-difference Tracker 26

3.3 A man walking into a covered area. (a) Covariance Tracker (b) Cepstrum-based Covariance Tracker (c) Cepstrum-based Co-difference Tracker 27

3.4 A woman walking into a darker region. (a) Co-difference Tracker (b) Cepstrum-based Covariance Tracker (c) Cepstrum-based Co-difference Tracker 27

4.1 Cumulative distribution function estimated by sorting observations v_i and the corresponding gray scale object 30

4.2 Upper and lower bounds on $g(z, v_N)$ for the confidence probability p=0.9997 and the corresponding object 32

4.3 Tracker loses the ball and determines the head of the player as the object 35

4.4 Highway video sequence is used as source to the object tracking algorithm 36

4.5 A man walking into a building is lost by the tracker due to varying illumination conditions 37

4.6 A man walking into a building is lost by the tracker due to varying illumination conditions 38

4.7 A man walking into a building is lost by the tracker due to varying illumination conditions 39

4.8 A hand moving over a complex background is lost by the tracker . 40

List of Tables

2.1 The used benchmark test 11

2.2 Shadow detection accuracy (η) values in percentage 14

2.3 Shadow discrimination accuracy (ξ) values in percentage 14

3.1 Performance of the Covariance Tracking Method 23

3.2 Performance of the Cepstrum-based Covariance Tracking Method 23

3.3 Performance of the Cepstrum-based Co-difference Tracking Method 24

4.1 Performance of the Target Loss Detection Method 33

Dedicated to my wife ...

Chapter 1

Introduction

Image and video processing have become popular research areas in recent years. Many interesting applications have been developed for different purposes. Some of these applications can be listed as video compression, face detection, computer vision, image enhancement, video denoising and fire detection. These applications have different complications as they all require different approaches towards the problem. It is crucial to develop necessary methods fulfilling the requirements of the problem at hand and to overcome the complications.

1.1 Contributions of this Thesis

In this thesis, various research areas of video processing are taken into consideration. Methods addressing different problems in the literature are developed.

First, moving shadows problem in video sequences is considered. Moving shadows cause difficulties in accurate moving object detection since all the moving points of both objects and shadows are detected at the same time causing distortions of the moving object shape and model. As a result of this, some applications such as classification and assessment of moving object position give erroneous results. Surveillance tasks such as counting and classifying objects in the scene is adversely affected by moving shadows. Therefore, it is desirable to differentiate moving shadows from moving objects. A method is developed for moving shadow detection and it is observed that it performs better than the methods proposed in the literature.

Another problem in the literature is object tracking under illumination variations [1]. Illumination variations may cause misclassification of background objects as moving objects or lost targets due to changes in the target model [2]. Most of the image processing applications require robustness in moving object tracking. Since illumination variations are unavoidable in real-world environments, it is preferable to have an object tracking algorithm that is not highly affected by environmental changes. For robust object tracking, features that are not affected by illumination variations should be used. An object tracking method robust to illumination variations is proposed and it is seen that the method has better performance under illumination variations compared to methods that do not use illumination invariant features.

Recently, many interesting video target tracking methods have been developed. However, these tracking methods lack control mechanisms for target loss detection. Therefore, a measure controlling the reliability of the target tracking algorithms is absent. With an additional target loss detection mechanism, the object tracking methods would become more robust knowing that the target is lost at a frame and the tracker is not tracking the target anymore. For this purpose, a detection mechanism for target losses in object tracking applications is developed as a final contribution. Experimentally, it is observed that the target loss detection algorithm is successful in finding target losses that occurs in video target tracking applications.

1.2 Thesis Organization

The organization of this thesis is as follows. In Chapter 2, a method based on cepstral domain analysis is proposed for moving shadow detection. The method is applied to benchmark test sets and comparisons with some previous studies are done. A novel method introducing the usage of 2D-cepstral features of the target region in tracking algorithms is developed in Chapter 3. Comparisons with tracking algorithms that do not use the 2D-cepstral features of the target region are done. Chapter 4 presents a confidence interval based statistical approach for target loss detection. Performance of the target loss detection method is tested on various video sequences by using different object tracking algorithms. Final chapter concludes the thesis by summarizing the contributions to the image processing literature.

Chapter 2

Moving Shadow Detection

In many computer vision applications, detection and tracking of moving objects is one of the key steps because all consequent processes are based on its resulting regions. Therefore, accurate detection and tracking of moving objects is required. On the other hand, moving shadows cause problems in accurate moving object detection.

In video sequences, shadow points and object points share two characteristic visual features: motion model and detectability. Since the most common technique for foreground object detection requires inter-frame differentiation or background subtraction, all the moving points of both objects and shadows are detected at the same time. In addition, moving shadow pixels are normally adjacent to moving object pixels. Hence, moving shadow pixels and object pixels merge in a single blob causing distortions of the object shape and model. Thus, object shape is falsified and the geometrical properties of the object are adversely affected by shadows. As a result of this, some applications such as classification and assessment of moving object position (normally given by the shape centroid) give erroneous results. Another problem arises when shadows of two or more close objects create false adjacency between different moving objects resulting in detection of a single combined moving blob instead of multi moving objects. This problem has adverse effects on most of the higher level surveillance tasks such as counting and classifying objects in the scene.

It is desirable to find a way to differentiate shadow points from object points to overcome the challenges that are mentioned above. The detection of the shadow points should be performed successfully in the segmentation process since the correct identification and classification of the shadow points is crucial. The

main cause of the problems in segmentation and extraction of moving objects is the misclassification of shadow points as foreground. Recently, there has been growing interest on this subject and several methods are proposed to overcome this misclassification. In this part of the thesis, a two-dimensional cepstrum analysis based shadow detection method is proposed. The method is composed of two steps. In the first step, hybrid background subtraction based moving object detection is done to determine the candidate regions for further analysis. The second step involves the use of a non-linear method based on cepstrum analysis of the candidate regions for detecting the shadow points inside those regions.

The next section underlines some related work on shadow detection. Section 3 presents the proposed cepstrum based shadow detection method. Results of the proposed method and comparisons with previous approaches are presented in Section 4. The final section presents conclusions.

2.1 Related Work

In recent years, many algorithms have been proposed to deal with shadows of moving objects. In [3] it is pointed out that detection of shadows is necessary to prevent false alarms caused by cloud shadows moving over forests in video based forest fire detection systems. Shadows of slow moving clouds are the major source of false alarms and in order to reduce the false alarm rate, shadow detection and elimination is necessary.

It is also well-known in the computer vision literature that, shadow regions retain underlying texture, surface pattern, color and edges in images [4]. Some approaches like [5] prefer HSV color space analysis as a shadow cast on a background does not change significantly its hue. There have been some further studies on HSV color space analysis for shadow detection such as [6] and [7].

In the study of Jiang and Ward [8], classification is done on the basis of an approach that shadows are composed of two parts: self-shadow and the cast shadow. Self-shadow is defined as the part of the object which is not illuminated by the light source. The cast shadow is the area projected on the scene by the object. Cast shadow is further classified in umbra and penumbra. This detailed classification is also used in another work [9].

Some other approaches in the field are also used. The usage of multiple cameras for shadow detection is proposed by Onoguchi [10]. Shadow points are separated using the fact that shadows are on the ground plane, whereas foreground objects are not. Another method proposed uses geometry to find shadowed regions [11]. It produces height estimates of objects using their shadow position and size by applying geometry.

There has been also some useful comparative evaluations and classifications of existing approaches. The shadow detection approaches are classified as statistical and deterministic type and comparisons of these approaches are made in [12], [13] and [14].

2.2 Moving Shadow Detection Algorithm

The proposed method for moving shadow detection consists of two parts. In the first part, the hybrid background subtraction method is used to determine the moving regions. After determining moving regions, cepstrum analysis is carried out on detected moving regions. This analysis yields the regions with shadows. The following subsections present the steps of the proposed method.

2.2.1 Hybrid Background Subtraction Method

Stationary points in video forms the background scene. In other words, background can be defined as temporally stationary part of the video. Background subtraction is commonly used for segmenting out the moving regions for surveillance purposes. The hybrid background subtraction that we use is based on [15] with some modifications. Other background subtraction methods include [16] and [17].

As computers can not know the background of a scene exactly, background has to be estimated. If the scene is observed for some time, background scene can be estimated based on observations. A simple way to estimate the background is averaging the observed video frames of the scene.

The recursive background estimation algorithm used in this study is a simple IIR filter with control mechanism applied to each pixel independently to update the background.

Let $\mathbf{v}_{\mathbf{x},n}$ represent the component vector containing R, G and B values of the pixel at location $\mathbf{x} = (x_1, x_2)$ in the n^{th} video frame. Estimated background vector at $(n+1)^{\text{th}}$ video frame at the same positioned pixel, $\mathbf{b}_{\mathbf{x},n+1}$, is calculated as:

$$\mathbf{b}_{\mathbf{x},n+1} = \begin{cases} \alpha \mathbf{b}_{\mathbf{x},n} + (1-\alpha)\mathbf{v}_{\mathbf{x},n} & \text{if } \mathbf{x} \text{ is a stationary pixel} \\ \mathbf{b}_{\mathbf{x},n} & \text{if } \mathbf{x} \text{ is a moving pixel} \end{cases} \tag{2.1}$$

where $\mathbf{b}_{\mathbf{x},n}$ is the previous estimate of the background vector containing R, G and B values of the background pixel at location $\mathbf{x}$. The parameter α is a positive real number close to 1. Initially, $\mathbf{b}_{\mathbf{x},0}$ is set to the first video frame $\mathbf{v}_{\mathbf{x},0}$ for all $\mathbf{x}$.

It is seen from Eqn. 2.1 that the parameter α determines the weight of the observed frame on background estimation. With higher α values, the contribution of the observed frame is reduced on background estimation for the next frame.

A pixel positioned at $\mathbf{x}$ is assumed to be moving if:

$$\|\mathbf{v}_{\mathbf{x},n} - \mathbf{v}_{\mathbf{x},n-1}\| > \tau_{\mathbf{x},n} \tag{2.2}$$

where $\mathbf{v}_{\mathbf{x},n-1}$ is the component vector containing R, G and B values of the pixel at location $\mathbf{x} = (x_1, x_2)$ in the $(n-1)^{\text{th}}$ video frame and $\tau_{\mathbf{x},n}$ is a recursively updated threshold at each frame, describing a statistically significant difference vector length in the RGB space at pixel position $\mathbf{x}$. The recursive threshold update mechanism is given as:

$$\tau_{\mathbf{x},n+1} = \begin{cases} \alpha \tau_{\mathbf{x},n} + (1-\alpha)(c\|\mathbf{v}_{\mathbf{x},n} - \mathbf{b}_{\mathbf{x},n}\|) & \text{if } \mathbf{x} \text{ is stationary} \\ \tau_{\mathbf{x},n} & \text{if } \mathbf{x} \text{ is moving} \end{cases} \tag{2.3}$$

where c is a real number greater than 1 and update parameter α is a positive real number close to 1. Initial threshold values are set to pre-determined nonzero values.

It is seen from Eqn. 2.3 that, the parameter c determines the sensitivity of the detection scheme. For higher parameter values c, the sensitivity of the detection scheme is lower.

It is assumed that regions that are significantly different from the background are moving regions. So, the estimated background for the current frame is subtracted from the current video frame to detect moving regions that correspond to the set of pixels satisfying:

$$\|\mathbf{v}_{\mathbf{x},n} - \mathbf{b}_{\mathbf{x},n}\| > \tau_{\mathbf{x},n} \tag{2.4}$$

Equation 2.4 is a binary pixel map indicating whether the pixel at location $\mathbf{x}$ in n^{th} frame is moving or not. The pixels satisfying Eqn. 2.4 are mapped to 1 (marked as moving pixels) and others are mapped to 0. It should be noted that 2D median filter is also applied to the binary pixel map output for robustness.

2.2.2 Cepstrum Analysis for Shadow Identification

Cepstrum analysis is applied to the moving regions detected by the first step of the method in a similar way introduced in [3] with some modifications. The proposed cepstrum analysis method for shadow detection is composed of two parts. The first part includes the separation of the moving regions into 8x8 blocks and the application of the 2D cepstrum to the blocks of interest and their corresponding background blocks to decide whether the texture and color properties are preserved for that moving block or not. If it is decided that the properties are preserved for the block, the algorithm proceeds with the second part. If not, the detection algorithm marks the block as moving object block. Note that this decision mechanism is based on [4], in which it is stated that ideally shadows retain the color and texture information in images. In the second part, a more detailed pixel-based approach is taken. 1D cepstrum is applied to each pixel belonging to the block to decide if the pixel is a moving shadow pixel or object pixel. The following subsections present the parts of the proposed cepstrum analysis method.

Cepstrum Analysis of Blocks

Cepstral domain analysis of signals are widely used in speech processing applications [18]. The cepstrum was originally found by Tukey [19]. The cepstrum $\hat{x}[n]$ of a signal x is defined as the inverse Fourier transform of the log-magnitude Fourier spectrum of x.

Let $x[n]$ be a discrete signal, its cepstrum $\hat{x}[n]$ is defined as follows:

$$\hat{x}[n] = F^{-1}\{\ln(|F\{x[n]\}|)\} \tag{2.5}$$

where $F\{.\}$ represents the discrete-time Fourier Transform, $|.|$ is the magnitude, $\ln(.)$ is the natural logarithm and $F^{-1}\{.\}$ represents the inverse discrete-time Fourier Transform operator. In our approach, we use both one-dimensional and two-dimensional (2D) cepstrums for shadow detection.

Moving regions in video are divided into 8x8 moving blocks as a subset of the whole moving region. Let the i-th moving 8x8 block be defined as $\mathbf{R}_{\mathrm{i}}$. Then, 2D cepstrum of $\mathbf{R}_{\mathrm{i}}$, $\hat{\mathbf{R}}_{\mathrm{i}}$ is defined as follows:

$$\hat{\mathbf{R}}_{\mathrm{i}} = F_{\mathrm{2D}}^{-1}\{\ln(|F_{\mathrm{2D}}\{\mathbf{R}_{\mathrm{i}}\}|)\} \tag{2.6}$$

where $F_{\mathrm{2D}}\{.\}$ is the 2D discrete-time Fourier Transform and $F_{\mathrm{2D}}^{-1}\{.\}$ is the inverse discrete-time Fourier Transform operator.

Similarly, let the i-th corresponding background block for the current image frame be defined as B_{i} and its 2D cepstrum as $\hat{\mathbf{B}}_{\mathrm{i}}$. A difference matrix D_{i} for the i-th block can be defined as:

$$\mathbf{D}_{\mathrm{i}} = |\ \hat{\mathbf{R}}_{\mathrm{i}} - \ \hat{\mathbf{B}}_{\mathrm{i}}| \tag{2.7}$$

Based on [4], if the block of interest is a shadowed block, it should have the following property:

$$\mathbf{R}_{\mathrm{i}} = \alpha\ \mathbf{B}_{\mathrm{i}} \tag{2.8}$$

where α is a positive real number less than 1. The shadowed block is an α scaled version of the background block. The effect of this on the difference matrix in the 2D cepstral domain is: $\mathbf{D}_{\mathrm{i}}$ having only the (1,1)-indexed value different than zero because of the scaling by constant α. Theoretically, the other entries of $\mathbf{D}_{\mathrm{i}}$ should be equal to zero. So we define our distance metric as:

$$m_{\mathrm{i}} = \sum_{(a,b)\neq(1,1)} \mathbf{D}_{\mathrm{i}}(\mathrm{a},\mathrm{b}) \tag{2.9}$$

Note that this operation is done for R, G and B values of the block separately. Therefore, the distance metric M_{i} used for the decision of the i-th block is selected as follows:

$$M_{\mathrm{i}} = \sqrt{m_{\mathrm{i,r}}^2 + m_{\mathrm{i,g}}^2 + m_{\mathrm{i,b}}^2} \tag{2.10}$$

where $m_{\mathrm{i,r}}$ is the R component distance metric, $m_{\mathrm{i,g}}$ is the G component distance metric and $m_{\mathrm{i,b}}$ is the B component distance metric.

Therefore, the decision algorithm for the first part is:

$$R_{\mathrm{i}} : \begin{cases} \text{moving shadow block,} & \text{if } M_{\mathrm{i}} < \kappa \\ \text{moving object block,} & \text{otherwise} \end{cases} \tag{2.11}$$

where κ is a determined threshold. After detecting possible candidate 8 by 8 shadow regions, we examine each pixel one by one to determine the exact boundary of shadow pixels.

Cepstrum Analysis of Pixels

The signal used in this part is the R, G and B values of each pixel belonging to the passed 8x8 block. R, G and B values of the pixel positioned at $\mathbf{x} = (x_1, x_2)$ in the n^{th} frame is defined as:

$$\mathbf{v}_{\mathbf{x},n} = (r_{\mathbf{x},n} \;\; g_{\mathbf{x},n} \;\; b_{\mathbf{x},n}) \tag{2.12}$$

note that the length of the signal $\mathbf{v}_{\mathbf{x},n}$ is 3 where $\mathbf{v}_{\mathbf{x},n}[1] = r_{\mathbf{x},n}$, $\mathbf{v}_{\mathbf{x},n}[2] = g_{\mathbf{x},n}$ and $\mathbf{v}_{\mathbf{x},n}[3] = b_{\mathbf{x},n}$ For the same positioned pixel in the same frame, background value is defined as:

$$\mathbf{b}_{\mathbf{x},n} = (br_{\mathbf{x},n} \;\; bg_{\mathbf{x},n} \;\; bb_{\mathbf{x},n}) \tag{2.13}$$

Based on [4], a shadow pixel positioned at $\mathbf{x}$ in n^{th} frame should have the following property:

$$\mathbf{v}_{\mathbf{x},n} = \alpha \mathbf{b}_{\mathbf{x},n} \tag{2.14}$$

where α is a positive real number less than 1. Thus, the shadow pixel frame value is an α scaled version of the same positioned background pixel value in the RGB-space. Using Eqn. 2.14, we state:

$$\hat{\mathbf{v}}_{\mathbf{x},n}[1] \neq \hat{\mathbf{b}}_{\mathbf{x},n}[1] \tag{2.15}$$

$$\hat{\mathbf{v}}_{\mathbf{x},n}[2] = \hat{\mathbf{b}}_{\mathbf{x},n}[2] \tag{2.16}$$

$$\hat{\mathbf{v}}_{\mathbf{x},n}[3] = \hat{\mathbf{b}}_{\mathbf{x},n}[3] \tag{2.17}$$

$$\hat{\mathbf{v}}_{\mathbf{x},n}[4] = \hat{\mathbf{b}}_{\mathbf{x},n}[4] \tag{2.18}$$

Note that the length of the cepstral coefficients is 4. This is because of the fact that the size of the discrete Fourier Transform is selected 4 for faster processing, enabling the usage of the Fast Fourier Transform algorithm in cepstrum calculation.

As it is seen from Eqns. 2.15-18, theoretically second, third and fourth cepstral coefficients of pixel frame value, $\hat{\mathbf{v}}_{\mathbf{x},n}[2]$, $\hat{\mathbf{v}}_{\mathbf{x},n}[3]$, $\hat{\mathbf{v}}_{\mathbf{x},n}[4]$ and their counterpart cepstral coefficients of background pixel value, $\hat{\mathbf{b}}_{\mathbf{x},n}[2]$, $\hat{\mathbf{b}}_{\mathbf{x},n}[3]$, $\hat{\mathbf{b}}_{\mathbf{x},n}[4]$ should be equal if the pixel of interest is a shadow pixel. First cepstral coefficients, $\hat{\mathbf{v}}_{\mathbf{x},n}[1]$

and $\hat{\mathbf{b}}_{\mathbf{x},n}[1]$ are different due to the effect of the natural logarithm of coefficient α. Using this fact we define a difference vector:

$$\mathbf{d}_{\mathbf{x},n} = |\hat{\mathbf{v}}_{\hat{\mathbf{x}},n} - \hat{\mathbf{b}}_{\hat{\mathbf{x}},n}| \tag{2.19}$$

Decision mechanism for detecting shadow pixels uses Eqn. 2.19. Shadow detection method for moving pixels inside the block is given as:

$$\mathbf{x} : \begin{cases} \text{moving shadow pixel,} & \text{if } \mathbf{d}_{\mathbf{x},n}[2] \,\&\, \mathbf{d}_{\mathbf{x},n}[3] \,\&\, \mathbf{d}_{\mathbf{x},n}[4] < \tau \\ \text{moving object pixel,} & \text{otherwise} \end{cases} \tag{2.20}$$

where τ is an adaptive threshold changing its value as a function of the background pixel value for the current image frame. It is formed by using the calculated threshold values for each color component as:

$$\tau_{\mathbf{x},n} = \kappa \mathbf{b}_{\mathbf{x},n} \tag{2.21}$$

$$\tau = \sqrt{\sum_{r,g,b} \tau_{\mathbf{x},n}^2} \tag{2.22}$$

where κ is a predetermined constant.

2.3 Experimental Results

In this section, the outcomes of the proposed algorithm are presented and comparisons with some of the previous approaches are made.

The benchmark test set used is taken from [20] since its widely referenced by most of the researchers working in the field. Each video sequence in the benchmark test set has different sequence type, shadow strength, shadow size, object class, object size, object speed and noise level. The video sequences and their properties are given in Table 2.1 in detail.

It is essential to test algorithm performances at different conditions to decide for the environments the algorithm works better. So, the proposed method is applied separately to all of the video sequences.

In Fig. 2.1a, a man walking in laboratory is the moving object. It is seen that moving object and shadow pixels (Fig. 2.1b) are identified successfully. The result is very promising since the shadow strength is very low for the video sequence in concern.

Table 2.1: The used benchmark test

Properties	Campus	Highway I	Highway II	Intel. Room	Laboratory
Seq. Type	outdoor	outdoor	outdoor	indoor	indoor
Shadow Str.	low	medium	high	low	very low
Shadow Size	very large	large	small	large	medium
Object Class	veh./people	vehicles	vehicles	people	people/oth.
Object Size	medium	large	small	medium	medium
Object Speed	low	medium	high	low	low
Noise Level	high	medium	medium	medium	low

In Fig. 2.2a, a man far away from the camera is the shadow source. Despite of the distance and low shadow strength, the segmentation is done successfully. The proposed shadow detection method in this study gives much better results than [21] for the video sequence.

The video sequences of campus raw (Fig. 2.3a) have very low shadow strength as well as high noise level. In Fig. 2.3b, it is clearly seen that two moving objects are detected perfectly and most of the moving shadow points on the ground are marked with success.

The shadow pixels of the outdoor video sequence presented in Fig. 2.4a is determined accurately (Fig. 2.4b). There are only few object points falsely determined as shadow points in this sequence.

In order to compare the performance of the proposed method with the others, quantitative measures are used. In this study, shadow detection accuracy η and shadow discrimination accuracy ξ metrics introduced in [13] are used as the quantitative measures for comparison purposes. The reason for selecting [13] for comparison is due to the existence of detailed classification schemes and utilization of different approaches available in the literature for shadow detection in its content. Table 2.2 and Table 2.3 summarizes the performance of the proposed method and the other methods using the same benchmark test set. In the tables, the abbreviations SNP, SP, DNM1, DNM2 and CB stands for the statistical non-parametric approach, statistical parametric approach, deterministic non-model based approach using color exploitation, deterministic non-model based approach using spatial redundancy exploitation and the proposed cepstrum based approach, respectively. The ξ and η values in percentage for the proposed approach are commonly better than the SNP, SP, DNM1 and DNM2 approaches used by the other researchers in the literature.

(a) Original video frame

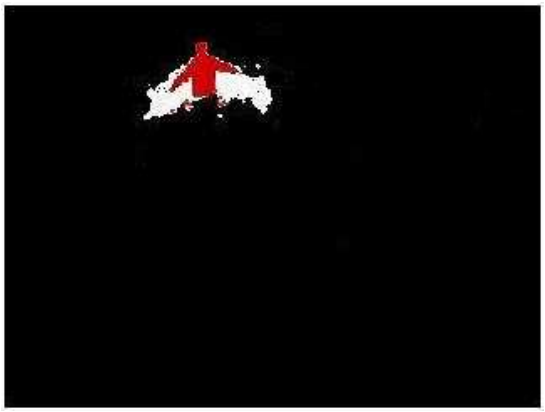
(b) Detected moving object (shaded in red) and shadow regions (dark shaded)

Figure 2.1: "Laboratory" video sequence

2.4 Summary

In this part of the thesis, a shadow detection method based on cepstral domain analysis is proposed. The method is a two step approach with moving object detection followed by cepstral analysis for moving shadow detection. The cepstral analysis steps are based on the fact that shadow regions retain the underlying color and texture of the background region. After determining the possible shadow blocks in the first step, a pixel based decision mechanism is used to determine the exact shadow boundaries. The method is applied to benchmark test sets and it is seen that proposed method gives successful results. The shadow pixels and object pixels are segmented accurately in all video sequences. Finally, quantitative measures are defined for comparison with previous approaches. The detection and discrimination rate comparisons show that the proposed method

(a) Original video frame



(b) Detected moving object and shadow regions

Figure 2.2: "Intelligent room" video sequence

commonly gives better results than other approaches used by other researchers in the literature.

(a) Original video frame

(b) Detected moving object and shadow regions

Figure 2.3: "Campus" video sequence

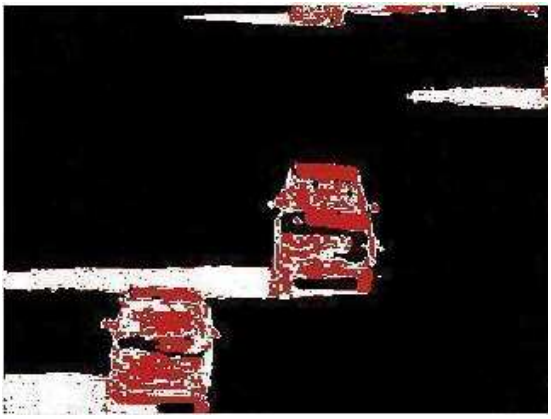
Table 2.2: Shadow detection accuracy (η) values in percentage

	Campus	Highway I	Highway II	Intelligent Room	Laboratory
SNP	80.58	81.59	51.20	78.63	84.03
SP	72.43	59.59	46.93	78.50	64.85
DNM1	82.87	69.72	54.07	76.52	76.26
DNM2	69.10	75.49	60.24	71.68	60.34
CB	**84.21**	**77.38**	**62.73**	**80.67**	**83.26**

Table 2.3: Shadow discrimination accuracy (ξ) values in percentage

	Campus	Highway I	Highway II	Intelligent Room	Laboratory
SNP	69.37	63.76	78.92	89.92	92.35
SP	74.08	84.70	91.49	91.99	95.39
DNM1	86.65	76.93	78.93	92.32	89.87
DNM2	62.96	62.38	72.50	86.02	81.57
CB	**81.35**	**85.34**	**86.88**	**93.56**	**94.90**

(a) Original video frame



(b) Detected moving object and shadow regions

Figure 2.4: "Highway I" video sequence

Chapter 3

Object Tracking under Illumination Variations

Visual object tracking is defined as the process of estimating the location of the moving object in the current image frame given all previous frames of a video sequence. Tracking of moving objects is one of the most important tasks in computer vision as object tracking algorithms are used in many applications such as security surveillance, traffic flow analysis, and content-based video compression.

Robustness to varying illumination conditions is crucial for visual tracking algorithms because illumination variations are unavoidable in real-world environments. Globally or locally changing illumination conditions of the observed scene constitutes an important challenge for many video processing applications including object tracking algorithms. Illumination variations may cause misclassification of some background objects as moving objects or lost targets due to changes in the target model parameters.

There has been many algorithms proposed for object tracking in the past. Covariance matrix based object tracking method proposed by Porikli et al. [22] is used as the base line tracker in this study. Covariance tracking uses a covariance matrix based object description measure. Since the target model is represented as the covariance matrix of features covering the target region, the method is applicable to non-stationary camera sequences and less susceptible to noise than other methods mainly based on background subtraction. In addition, computationally efficient co-difference matrix based object tracking method [23] introducing a new operator to replace the multiplication operator of the covariance matrix is also examined. Co-difference tracking method differs from the

covariance tracking method in the formation of the matrix describing the object region. Co-difference method can be implemented without performing any multiplication leading to a computationally efficient tracking method.

Ideally, the color and texture information of objects are retained in images under changing light intensities. Therefore, for robustness to varying illumination conditions, measures independent of the light intensity that uses color and texture information to represent the target are required. In this study, two-dimensional (2D) cepstrum analysis of the target is used because the cepstrum is an amplitude invariant feature extraction method widely used in speech processing.

In this part of the thesis, a novel object tracking algorithm increasing the robustness of both covariance and co-difference tracking methods under varying illumination conditions is proposed. The proposed object tracking algorithm introduces the 2D-cepstrum analysis of the target region to the covariance and co-difference tracking methods. The light intensity-independent 2D-cepstrum coefficients of the target region are used to increase the robustness of the object tracking algorithms to varying illumination conditions.

The next section underlines some related work on object tracking under varying illumination conditions. Section 3.2 gives an overview of the covariance and co-difference based object tracking methods and presents the proposed 2D-cepstrum based object tracking method. Results of the proposed method implementation and comparisons with covariance and co-difference tracking methods are presented in section 3.3. The final section presents conclusions.

3.1 Related Work

In recent years, many algorithms have been proposed to deal with robust object tracking under varying illumination conditions.

Background subtraction approaches are widely used to detect moving objects [24][25]. However, these approaches are susceptible to illumination variations of the background.

Some work in the literature discards the illumination-sensitive color information by using other features which are less-sensitive to illumination variations such as edges and textures. In [26], a model-based method where edge information is used to capture hand articulation by learning hand natural constraints

is implemented. In [27], the head is modeled as a texture mapped cylinder and tracking is formulated as an image registration problem for the cylinder's texture map image.

Other methods using only the color information to represent targets and their background are existing. In [1], a color modeling approach including intensity information in HSI color space using B-spline curves is used. Yang and Waibel [28] detected human faces by using a normalized RG plane and proposed a color-model adaptation algorithm based on the observation that shape of the histograms remain similar under illumination change. In [2], Gaussian mixture models are used to estimate probability densities of color for target and background objects. In their work, a technique for dynamically updating the models to accommodate changes in color due to varying illumination conditions is also introduced.

Methods combining both the color and texture information for robust object tracking are also available in the literature. The fusion of appearance and structural information in [29] is done using the condensation algorithm. In [30], illumination insensitive features are extracted from both target and background objects using the Bayesian framework for robust tracking. A method for moving object detection based on the background modeling and subtraction which uses both color and edge information is proposed in [31]. In this study, confidence maps are introduced to fuse intermediate results and represent the results of the background subtraction. Li and Leung [32] proposed a method which uses texture information by calculating the quotient between the cross-correlation and auto-correlation of a gradient vector to eliminate brightness variations.

3.2 Moving Object Tracking Algorithm

The proposed tracking method introduces 2D-cepstral domain coefficients of the target region into the covariance and co-difference matrices that are used to describe the target characteristics. The tracking algorithms become more reliable under changing illumination conditions because cepstrum is an amplitude invariant feature extraction method.

In this section, first the covariance and co-difference tracking algorithms are introduced. In the second subsection, the 2D-cepstrum analysis of the target region and its incorporation to the tracking algorithms are presented.

3.2.1 Covariance and Co-difference Tracking Methods

Both covariance and co-difference tracking algorithms require the computation of a matrix representing the given target region by using the feature images formed for each frame to construct the target model. The covariance of feature vectors describing the target is called covariance matrix in the covariance tracking method. Similarly, the co-difference matrix is computed from feature vectors in the co-difference tracking method to model the moving target. In both covariance and co-difference methods, the aim is to find the region in a given image frame having the minimum distance from the target model matrix and assign this region as the estimated location of the moving target at that frame. The first step of tracking algorithms is feature vector construction from a given image or image region.

Feature Images and Vectors

Let the observed m-dimensional image denoted as $\mathbf{I}$. Then, the corresponding m-dimensional feature image, $\mathbf{F}$, can be written as:

$$\mathbf{F}(x, y) = \gamma(\mathbf{I}, x, y) \tag{3.1}$$

where $\gamma(.)$ can be any mapping feature such as color, filter responses, image gradients $\mathbf{I}_x$, $\mathbf{I}_y$, $\mathbf{I}_{xx}$, ..., temporal frame differences, edge magnitudes, etc.

For a given window region $\mathbf{R} \subset \mathbf{F}$, let $\{\mathbf{f}_k\}_{k=1,...,n}$ be the d-dimensional feature vectors inside $\mathbf{R}$. Feature vector $\{\mathbf{f}_k\}$ is constructed using two types of mappings; spatial mappings using the pixel coordinates and appearance mappings using color and gradient values. The feature vector used in this work is:

$$\mathbf{f}_k = [\, x \;\; y \;\; \mathbf{I}(x, y) \;\; \mathbf{I}_x(x, y) \;\; \mathbf{I}_y(x, y) \;\; \mathbf{I}_{xx}(x, y) \;\; \mathbf{I}_{yy}(x, y) \,]^T \tag{3.2}$$

where x and y are the pixel coordinates, $\mathbf{I}(x, y)$ is the color value and $\mathbf{I}_x(x, y)$, $\mathbf{I}_y(x, y)$, $\mathbf{I}_{xx}(x, y)$, $\mathbf{I}_{yy}(x, y)$ are the gradients along the x and y directions.

The Covariance Matrix

The second step of the tracking algorithms is the computation of the covariance matrix which is formed by using the feature vectors constructed in the previous

step. To represent an $M \times N$ rectangular region $\mathbf{R}$, the covariance matrix of feature points of $\mathbf{R}$ is defined as:

$$\mathbf{C_R} = \frac{1}{MN} \sum_{k=1}^{MN} (\mathbf{f}_k - \mu_\mathbf{R})(\mathbf{f}_k - \mu_\mathbf{R})^T \tag{3.3}$$

where $\mu_\mathbf{R}$ is the mean of the feature vectors belonging to region $\mathbf{R}$.

The Co-difference Matrix

In the co-difference tracking method the co-difference matrix is computed from the feature vectors described in 3.2.1. To represent an $M \times N$ rectangular region $\mathbf{R}$, the co-difference matrix of feature points of $\mathbf{R}$ is defined as:

$$\mathbf{D_R} = \frac{1}{MN} \sum_{k=1}^{MN} (\mathbf{f}_k - \mu_\mathbf{R}) \odot (\mathbf{f}_k - \mu_\mathbf{R})^T \tag{3.4}$$

The operator $\odot$ acts like a matrix multiplication operator, however, the scalar multiplication is replaced by an additive operator $\oplus$ which is defined as follows

$$a \oplus b = \text{sign}(a \times b)(|a| + |b|) \tag{3.5}$$

Since $a \oplus b = b \oplus a$, the co-difference matrix is also symmetric. The new operator decreases the computational cost of tracking by replacing the multiplication operator by the addition operator in some processors.

For d-dimensional feature vector sets, the corresponding covariance and co-difference matrices of features has size $d \times d$. These matrices of the feature points inside $\mathbf{R}$ are used to represent region $\mathbf{R}$. Initially, the covariance and co-difference matrices are used to model the target and in the following image frames of the video they are used to find the estimated target locations.

Distance Metric and Target Location Estimation

In both covariance and co-difference tracking methods, to obtain the most similar region to the given target window, distances between matrices corresponding to the target window and the candidate regions are calculated. Since the space of covariance and co-difference matrices are not vector spaces, subtraction of two matrices would not be a valid measure. Therefore, a distance metric needs to be introduced for finding the best candidate region match and estimating

that region as the target location. The distance metric used in both tracking algorithms computes the dissimilarity between matrices as

$$\rho(\mathbf{C}_i, \mathbf{C}_j) = \sqrt{\sum_{k=1}^{d} \ln^2 \lambda_k(\mathbf{C}_i, \mathbf{C}_j)} \tag{3.6}$$

where $\{\lambda_k(\mathbf{C}_i, \mathbf{C}_j)\}$ is the set of generalized eigenvalues of matrices $\mathbf{C}_i$ and $\mathbf{C}_j$.

At each frame, neighboring regions of the previously estimated location of the target are defined as the candidate regions. The descriptive matrix of these candidate regions are computed and the region with the smallest distance to the matrix representing the target is assigned as the estimated target location in that image frame. This operation is repeated for each frame.

3.2.2 Two-dimensional (2D) Cepstrum Analysis of the Target

The proposed cepstrum analysis method includes the computation of the 2D-cepstrum of the initial target window and the candidate regions at each frame. The 2D-cepstrum analysis is used as cepstrum is an amplitude invariant feature extraction method, therefore, cepstral domain coefficients of a region remains unchanged under light intensity variations. This property of cepstrum provides robustness to illumination variations at the target region.

In our approach, 2D-cepstrum is used. For a region $\mathbf{R}$, 2D-cepstrum of $\mathbf{R}$, $\hat{\mathbf{R}}$ is defined as follows:

$$\hat{\mathbf{R}} = F_{2\mathrm{D}}^{-1}\{\ln(|F_{2\mathrm{D}}\{\mathbf{R}\}|)\} \tag{3.7}$$

where $F_{2\mathrm{D}}\{.\}$ is the 2D discrete-time Fourier Transform and $F_{2\mathrm{D}}^{-1}\{.\}$ is the inverse 2D discrete-time Fourier Transform operator.

Let the initial $M \times N$ target window denoted as $\mathbf{W}$ and the shadowed version of the target window be represented as $\mathbf{W_s}$. According to [33], when the region of interest is shadowed, its intensity is scaled by a constant factor throughout that region. In our case, this statement corresponds to:

$$\mathbf{W_s} = \alpha \mathbf{W} \tag{3.8}$$

where α is a positive real number less than 1 for the target window $\mathbf{W}$. When we compute the 2D-cepstrum of both sides in Eq. 3.8 we obtain:

$$\hat{\mathbf{W}}_\mathbf{s} = K_\alpha \delta(x, y) + \hat{\mathbf{W}} \tag{3.9}$$

where δ is the dirac delta function, K_α is a constant, $\hat{\mathbf{W}}_\mathbf{s}$ and $\hat{\mathbf{W}}$ are the 2D-cepstrums of $\mathbf{W_s}$ and $\mathbf{Ws}$, respectively. In the proposed 2D-cepstrum analysis, Eqn. 3.9 reveals the fact that the output 2D-cepstrum coefficients except the (0,0)-indexed (magnitude) coefficient remains unchanged under intensity variations for the analyzed region. That is:

$$\hat{\mathbf{W}}_\mathbf{s}(i,j) = \hat{\mathbf{W}}(i,j), \;\; \forall (i,j) \neq (0,0) \tag{3.10}$$

Therefore, to obtain additional target characteristics robust to illumination variations, the output 2D-cepstrum coefficients except the magnitude coefficient should be used as cepstral domain feature parameters.

Cepstral domain feature parameters of the target region are incorporated to the covariance and co-difference matrix as additional features. The approach introduced in this paper is to increase the size of the matrices and add the 2D-cepstrum coefficients except the magnitude coefficient found by analyzing the target region to the matrix.

For a d-dimensional feature vector set, the corresponding covariance or co-difference matrix of features has size $d \times d$. Let the matrix be denoted as $\mathbf{C}$. The modified covariance or co-difference matrix $\mathbf{C_m}$ including the output 2D-cepstrum coefficients is derived as

$$\mathbf{C_m} = \begin{bmatrix} \mathbf{C} & \mathbf{V} \\ \mathbf{V}^T & \mathbf{0} \end{bmatrix} \tag{3.11}$$

where $\mathbf{V}$ is $d \times z$ matrix containing output 2D-cepstrum coefficients and $\mathbf{V}^T$ is the transpose of $\mathbf{V}$. Therefore, additional $d.z$ values are included in the matrix for robustness. Notice that the modification is done in such a way that the symmetrical property of covariance and co-difference matrices is preserved.

3.3 Experimental Results

In this section, outcomes of the proposed algorithms are presented and comparisons with the covariance tracking approach are presented.

The experimental results are obtained for $\mathbf{V}$ selected as

$$\mathbf{V} = \begin{bmatrix} \hat{\mathbf{W}}(1,2) & \hat{\mathbf{W}}(2,1) & \hat{\mathbf{W}}(2,2) & \hat{\mathbf{W}}(2,3) & ... & \hat{\mathbf{W}}(3,4) \end{bmatrix}^T \tag{3.12}$$

where $\hat{\mathbf{W}}$ is the output 2D-cepstrum matrix of the target window $\mathbf{W}$. In this case, $d = 7$ and $z = 1$ and the corresponding modified covariance matrix has a size of 8×8.

The performance of the covariance, cepstrum-based covariance and cepstrum-based co-difference tracking methods are measured using 10 video sequences adding up to more than 2000 frames. The video sequences are composed of moving and stationary camera recordings. In order to compare the performance of the tracking algorithms quantitatively, the approach taken in [34] is used. The detection rate is defined as the ratio of the number of frames the object location is accurately estimated to the total number of frames in the sequence. The estimated location is considered accurate if the estimate window center is within the 10×10 neighborhood of the tracked object center. Some of the resultant performances of the tracking methods are given in Table 3.1, 3.2 and 3.3. It is observed that the proposed algorithms increase the detection rates of the covariance tracking method and the cepstrum-based covariance tracking method performs slightly better than the cepstrum-based co-difference tracking method in most of the video sequences.

Table 3.1: Performance of the Covariance Tracking Method

	Total Frames	Missed	Detection Rate
Shadowed Street	225	27	88.0
Woman walking	341	25	92.7
Dog running	156	11	92.9
Two friends	571	54	90.5
People group	125	7	94.4
Talking man	401	62	84.5
Pink shirt	364	22	93.9

Table 3.2: Performance of the Cepstrum-based Covariance Tracking Method

	Total Frames	Missed	Detection Rate
Shadowed Street	225	12	94.6
Woman walking	341	7	97.9
Dog running	156	3	98.1
Two friends	571	21	96.3
People group	125	2	98.4
Talking man	401	17	95.8
Pink shirt	364	10	97.3

Some tracking algorithm outcomes are presented in figures. In Fig. 3.1, the man talking on the phone is walking into the building. When the man approaches to entrance of the building, the light intensity on him decreases. The covariance

Table 3.3: Performance of the Cepstrum-based Co-difference Tracking Method

	Total Frames	Missed	Detection Rate
Shadowed Street	225	16	92.9
Woman walking	341	12	96.5
Dog running	156	8	94.9
Two friends	571	16	97.2
People group	125	5	96.0
Talking man	401	22	94.5
Pink shirt	364	16	95.6

tracking method loses the target before he enters the building. However, the modified covariance and co-difference tracking methods track the man accurately until he enters the building.

In Fig. 3.2, two men is walking into a shadowed area. Initially, the man wearing a pink shirt is introduced as target to all tracking methods. It is observed in Fig. 3.2a that as illumination changes, the co-difference tracking method starts to lose the target and eventually the tracker completely loses the track of the target whereas in Fig. 3.2b and Fig. 3.2c, the modified covariance and co-difference tracking methods track the target successfully even there are abrupt light intensity changes in the scene.

Two men walking into a darker region is tracked in Fig. 3.3 using the covariance tracking, the proposed modified covariance and co-difference tracking methods. The video sequence introduces a continuously decreasing target light intensity as the frames advances. The proposed object tracking algorithm manages to track the targets at each frame. However, the covariance tracking algorithm loses one of the targets at some point (Fig. 3.3a).

In Fig. 3.4, a woman is walking into a darker region with her friends. In this case, Fig. 3.4a shows that although the co-difference tracker does not lose the target completely, the tracking is not robust due to the changes in the target model caused by illumination variations. It is observed from figures Fig. 3.4b and Fig. 3.4c that the proposed object tracking methods perform well by using the introduced additional cepstral domain target features.

In conclusion, the proposed tracking methods are successful throughout the video sequence and even though the target gets into the shadowed region, they manage to track the target until it gets out of the video frame.

(a)

(b)

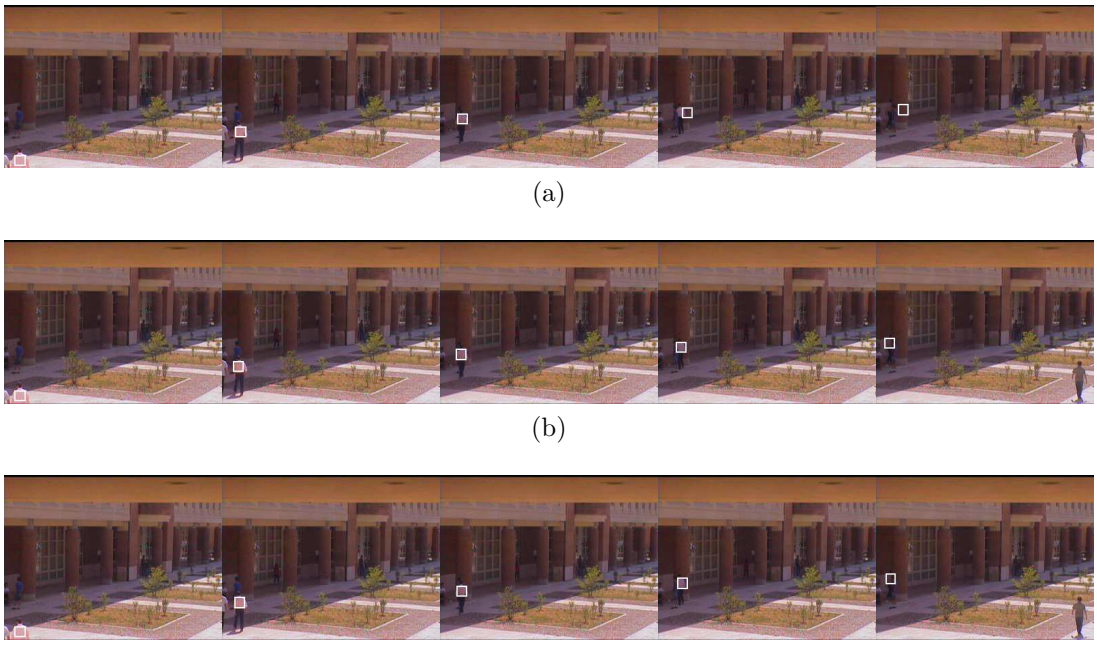
(c)

Figure 3.1: A man walking into a building. (a) Covariance Tracker (b) Cepstrum-based Covariance Tracker (c) Cepstrum-based Co-difference Tracker

Fig. 3.1, 3.2, 3.3 and 3.4 shows that the proposed object tracking methods perform better than the ordinary covariance and co-difference tracking methods under varying illumination conditions. The proposed methods are tested under abrupt illumination changes, continuously varying light intensity conditions and in the presence of a clutter. It is clear from our comparisons that there is need to adapt the changes in the target model under varying illumination conditions for robust object detection and accurate object recognition. It is observed that the introduction of the output 2D-cepstrum values of the target region to the covariance and co-difference matrices increases the robustness of the tracking algorithms to light intensity changes.

3.4 Summary

An object tracking method based on cepstrum analysis is proposed for robust object tracking under varying illumination conditions in this chapter. The proposed object tracking method combines the covariance tracking method and the 2D-cepstral features of the target region. The 2D-cepstrum is used because the cepstrum retains the underlying color and texture information under light-intensity



(a)

(b)

(c)

Figure 3.2: A man walking to a shadowed area. (a) Co-difference Tracker (b) Cepstrum-based Covariance Tracker (c) Cepstrum-based Co-difference Tracker

variations. The method is applied to video sequences in which the intensity of the target region varies and it is experimentally observed that the proposed method produces better results than the ordinary covariance tracking method. Co-difference method provides a computationally efficient trade-off compared to the covariance method because it does not require any multiplications during tracking.

(a)

(b)

(c)

Figure 3.3: A man walking into a covered area. (a) Covariance Tracker (b) Cepstrum-based Covariance Tracker (c) Cepstrum-based Co-difference Tracker

(a)

(b)

(c)

Figure 3.4: A woman walking into a darker region. (a) Co-difference Tracker (b) Cepstrum-based Covariance Tracker (c) Cepstrum-based Co-difference Tracker

Chapter 4

Target Loss Detection

Recently, many interesting video target tracking methods have been developed [34]- [35]. However, these tracking methods lack control mechanisms for target loss detection which usually occur when illumination conditions suddenly change [36]. In this part of the thesis, a confidence interval based statistical approach is developed for target loss detection in video. Upper and lower bounds for the cumulative distribution function (cdf) of the target feature vector are computed for a given confidence level. Target is classified as a lost target whenever the estimated cdf of the target region exceeds the computed bounds.

This part is organized as follows. In the next section, object signature function is introduced. The estimation of the upper and lower bounds and the corresponding confidence region are given in Section 4.2. Experimental results of the target loss detection algorithm are presented in Section 4.3. The final section concludes the chapter.

4.1 Target Loss Detection Algorithm

4.1.1 Object Signature Function

We assume that object pixels or object feature parameters are observations of a random variable V with finite variance. We define the signature function g of

sample of this random variable for any possible value z as follows:

$$g(z,v) = \begin{cases} 1, & \text{if } v \leq z \\ 0, & \text{otherwise} \end{cases} \tag{4.1}$$

where v is an observation (sample) of the random variable V. Expected value of the signature function g, for any value of z, is $F(z)$, the cumulative distribution function(cdf) of the random variable V, i.e.:

$$E[g(z,V)] = F(z)$$

And the variance of the signature function g is given by

$$Var\{g(z,V)\} = F(z)(1-F(z))$$

The cumulative distribution function $F(z)$ can be estimated by simply sorting the observations v_i, $i = 0, 1, ..., N-1$, as shown in Fig. 4.1a for a given object gray scale pixel values shown in Fig. 4.1b. If feature parameters of the object is a vector then V is a random vector and we define the signature function

$$g(z,v) = \begin{cases} 1, & \text{if } ||v|| \leq z \\ 0, & \text{otherwise} \end{cases} \tag{4.2}$$

In this case, we still have $E[g(z,v)] = F(z)$ where F(z) is the cdf of $||V||$ and, similarly, $Var\{g(z,v)\} = F(z)(1-F(z))$.

When there are multiple observations, the object has N pixels, we denote the pixel values of the object by v_i, $i = 1, ..., N$. The vector $v^N = [v_1, v_2, \cdots, v_N]^T$ represents the object pixels that are samples of the object with CDF F. The signature function of the object for any value $z, 0 \leq z \leq 255$, is defined as

$$g(z,v^N) = \frac{1}{N}\sum_{i=1}^{N} g(z,v_i) \tag{4.3}$$

It is known that this function is a CDF estimator with mean

$$E[g(z,V^N)] = F(z) \tag{4.4}$$

and if the samples (object pixel values) are independent, then the variance of such CDF estimator is

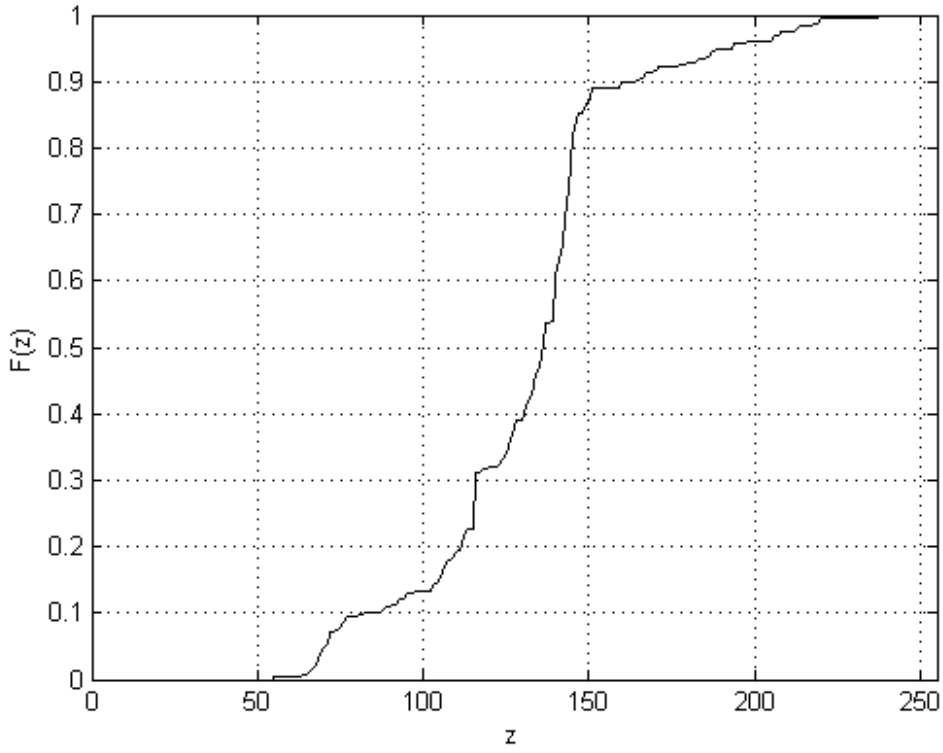
$$Var\{g(z,V^N)\} = \frac{1}{N}F(z)(1-F(z)) \tag{4.5}$$

Independence of pixel values is not a valid assumption for an object in video because neighboring samples are correlated with each other. Regardless of the

correlation Eqn. 4.4 is still valid but the variance becomes [37]

$$Var\{g(z,V^N)\} = \frac{1}{N}F(z)(1-F(z)) + \frac{1}{N^2}\sum_{i=1,j=1,i\neq j}^{N} Cov(g(z,V_i)g(z,V_j)) \tag{4.6}$$

For each z, the value of $g(z,v^N) = m/N$ where m is the number of samples of v^N smaller than the parameter z. Therefore, the observed samples are simply sorted to estimate the CDF [37]. Fig. 4.1a shows an example of a signature function for a given object gray scale pixel values shown in Fig. 4.1b.



(a) Estimate cdf for given object

(b) Given 16x16 gray scale object

Figure 4.1: Cumulative distribution function estimated by sorting observations v_i and the corresponding gray scale object

4.2 Bounds and the Confidence Region

While the mean of the signature function is the CDF of a desired object, the standard deviation of the signature function (square root of (4.6)) is usually much smaller than its mean (4.4). Therefore, the signature function of the moving object will allow us to define proper confidence regions, with a high probability p. It is possible to estimate the cdf with the signature function. The desired CDF mean and variance can be calculated using the object parameters from the target data in previous image frames of the video. Based on the confidence regions, we can determine if the tracked region in the current frame of the video has the same cdf as the moving object or not. Therefore, for each z and for a high confidence probability p, we can find a lower bound function $L(z)$ and an upper bound function $U(z)$ around the mean $F(z)$ such that

$$Pr\{L(z) \leq g(z, v_N) \leq U(z)\} = p \tag{4.7}$$

In statistics, the three-sigma rule, or empirical rule, states that for a normal distribution, almost all values lie within three standard deviations of the mean. For a better quality measure, the six sigma approach increases the standard deviation to 4.5 (equivalently p = 0.999997). Therefore, we can set the lower and upper bound functions as follows,

$$L(z) = F(z) - \lambda\sqrt{Var\{F(z)\}} \tag{4.8}$$

$$U(z) = F(z) + \lambda\sqrt{Var\{F(z)\}} \tag{4.9}$$

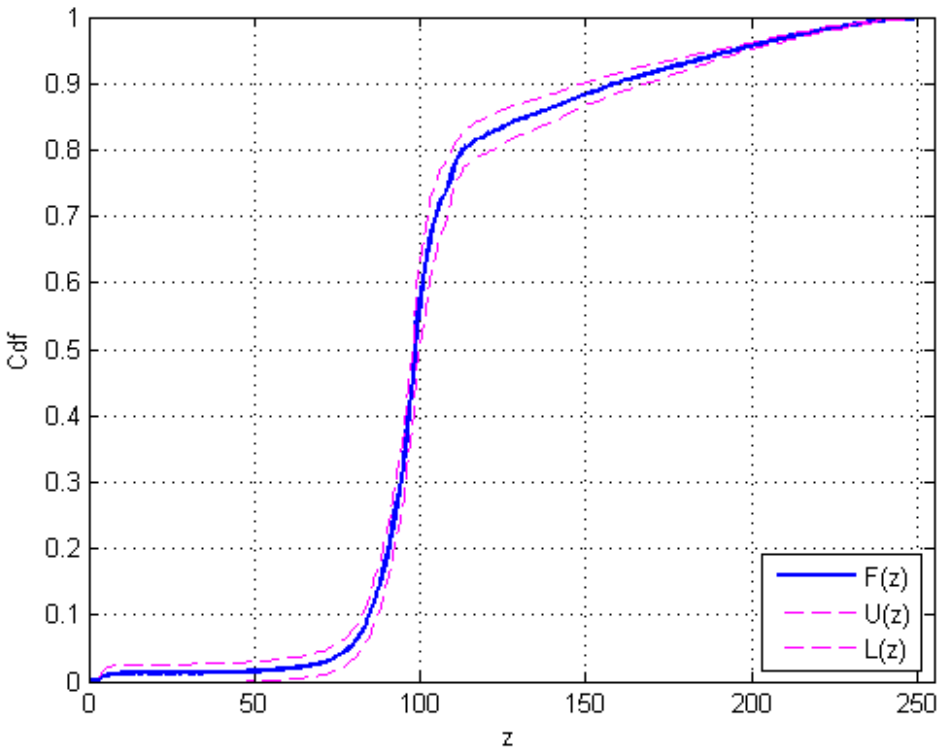
respectively with the Gaussianity assumption [1] . For example, Fig. 4.2a shows the bounds on $g(z, v_N)$ for the confidence probability p=0.9997 ($\lambda = 3$) for the object shown in Fig. 4.2b. The cdf $F(z)$ and its variance is empirically estimated from past image frames of the video.

The performance summary of the target loss detection method is given in Table 4.1.

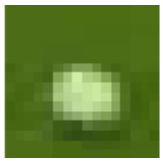
4.3 Experimental Results

As long as the cdf of the tracked region remains within upper and lower bounds we are confident with probability p that the object is properly tracked. When

[1]The signature function g in (4.3) can easily be approximated by a Gaussian distribution due to the central limit Theorem.



(a) Bounds on $g(z, v_N)$ for $\lambda = 3$



(b) Given 32x32 object

Figure 4.2: Upper and lower bounds on $g(z, v_N)$ for the confidence probability p=0.9997 and the corresponding object

the cdf of the tracked region exceeds the bounds, we conclude that the target is lost. In this section, the confidence interval method is applied to video sequences to find the frames in which trackers lose the target.

The target loss detection results are summarized in Table 4.1 for 10 video clips. In video clip 1, the ball in a football match is tracked. The the tracker loses the ball in frame number 124 and jumps to the head of a player as shown in Fig. 4.3b. Target loss is detected in frame number 128. In this case the cdf of the region is clearly outside the bounds (Fig. 4.3a).

In Fig. 4.4, a recorded highway video sequence (video clip 2) is used as source to the object tracking algorithm. Although the used tracker does not completely

Table 4.1: Performance of the Target Loss Detection Method

	Tracking Method	Target	Object Loss at Frame	Object Loss Detected at Frame
Video 1	Covariance	Ball	124	128
Video 2	Covariance	Car	38	38
Video 3	Covariance	Head	438	443
Video 4	Mean-shift	Human	76	77
Video 5	Covariance	Ball	156	156
Video 6	Covariance	Chest	305	309
Video 7	Mean-shift	Player	178	178
Video 8	Mean-shift	Dog	91	91
Video 9	Mean-shift	Car	35	37
Video 10	Covariance	Man	282	289
Video 11	Covariance	Hand	75	78
Video 12	Mean-shift	Boy	115	119
Video 13	Covariance	Boat	93	93
Video 14	Covariance	Lion	27	33
Video 15	Mean-shift	Plane	61	61
Video 16	Covariance	Woman	267	269
Video 17	Mean-shift	Truck	88	89
Video 18	Mean-shift	Bicycle	57	62
Video 19	Covariance	Man	198	203

loses the track of the target, frame number 38 is flagged as target loss frame in Fig. 4.4b. The reason behind this decision is the shifted estimate target window causing differences in the features of the target region and the estimate target region as shown in Fig. 4.4a.

In video clip 3, the head of the man walking in a laboratory is given as target to the tracker (Fig. 4.5). The target is tracked successfully until frame 438. The target loss is detected in frame 443 as shown in Fig. 4.5b. The resulting cdf of the estimate target region features in Fig. 4.5a determines the target loss since it is out of the confidence bounds.

Two men walking across a street is observed in video clip 4 (Fig. 4.6). Tracked head of the man at left is lost by the tracker due to the noisy structure of the video sequence. In frame number 76, the tracker loses the target and the tracker estimates the target region incorrectly in frame 77 as shown in Fig. 4.6b. The corresponding cdfs of the target and the estimate are given in Fig. 4.6a. It is clearly observed that the estimate region cdf is out of the bounds.

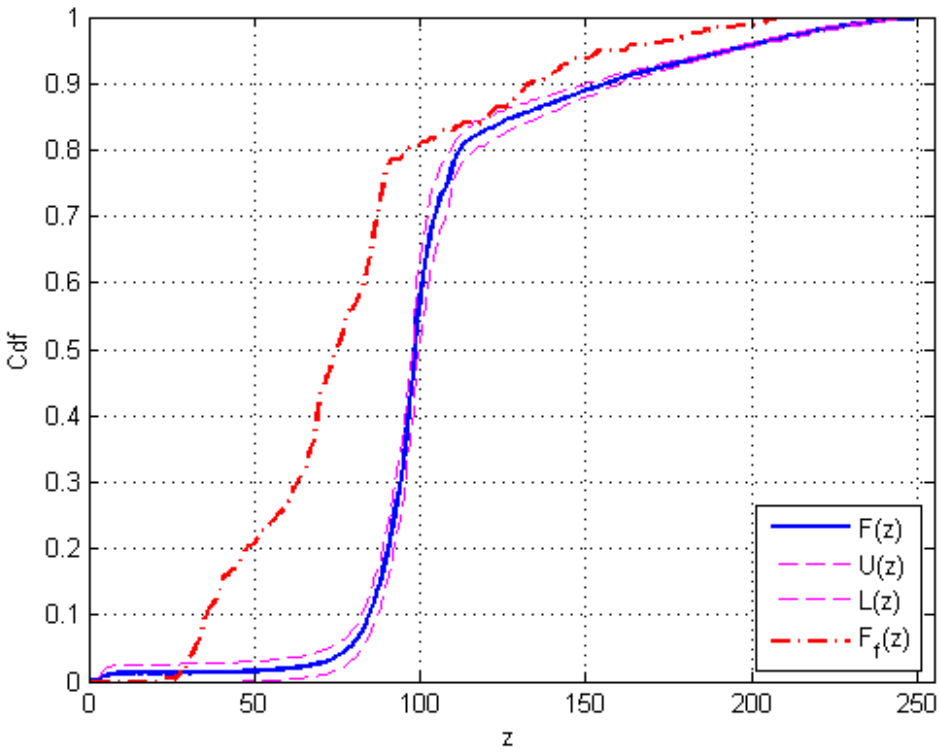
In video clip 6 in Table 4.1, the chest of a man walking into a building is tracked (Fig. 4.7). The tracker loses the target due to varying illumination conditions. The tracker starts to lose the target in frame number 305 and estimates the target region incorrectly as shown in Fig. 4.7b. The corresponding cdfs of the target and the estimate are given in Fig. 4.7a. Since the estimate region cdf is out of the bounds in frame number 309, it is concluded that the target is lost.

In Fig. 4.8, a hand moving over a complex background is given as target to the tracking algorithm (video clip 11). Due to the complex color structure of the background, the tracker loses the target at frame number 75 in Fig. 4.8b. The cdfs corresponding to the target and estimate at frame number 78 is given in Fig. 4.8a. The estimate region cdf shown in 4.8a indicates a target loss.

The probability of error in Table 4.1 is p=0.982 which is close to the 3 sigma probability. Estimated probability verifies the confidence measure theory. The small discrepancy of 0.9997-0.982=0.0177 is probably due to the fact that the upper and lower bounds on cdf are also estimated from the data.

4.4 Summary

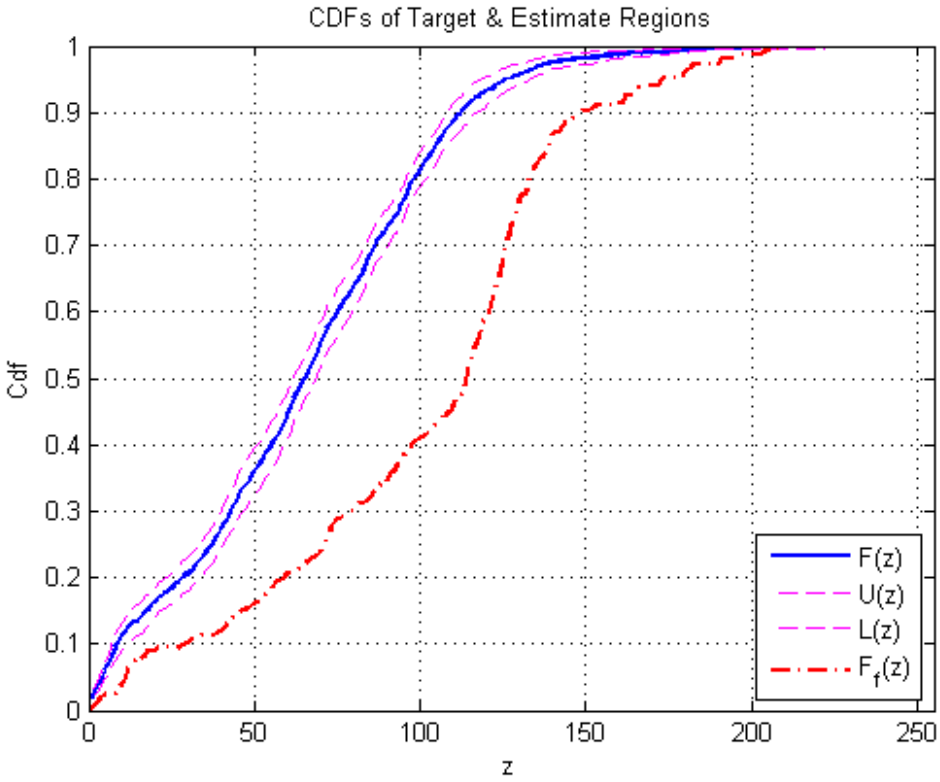
In this chapter, a statistical method based on confidence interval functions is developed for target loss detection. Upper and lower bound functions on the cumulative distribution function (cdf) of the target feature vector are estimated for a given confidence level. Whenever the estimated cdf of the detected region exceeds the bounds, the target is regarded as being no longer tracked by the tracking algorithm. The method is applied to various video sequences in which the target is lost at some frame and it is observed that the confidence interval based target loss detection method detects the target losses successfully. In addition, the method is also practical since it is applicable to most tracking algorithms using features of the target image region.



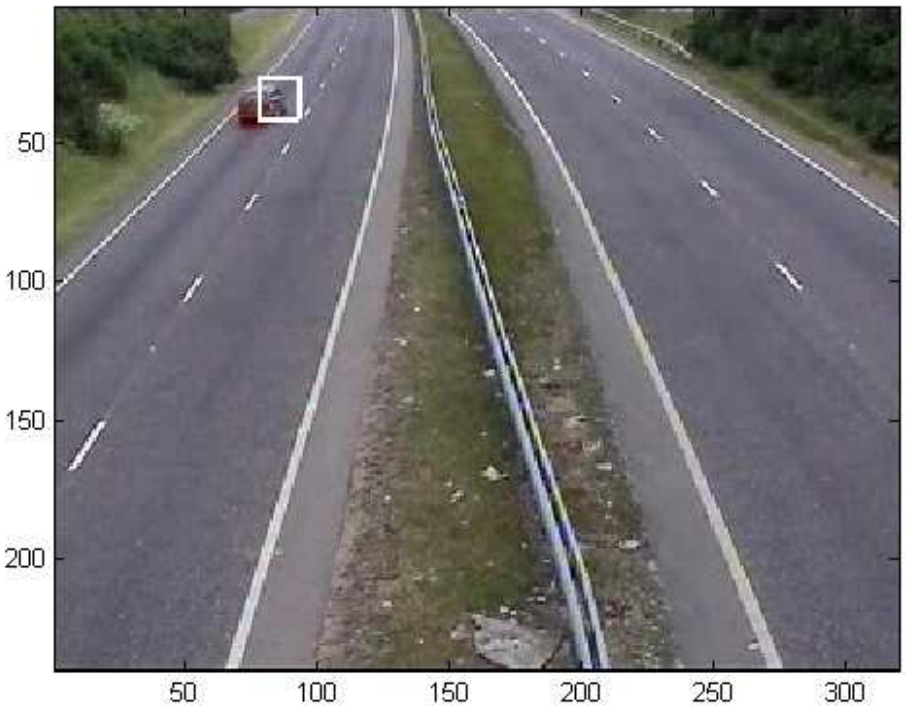
(a) Cdfs of the target and estimate regions

(b) Falsely estimated 32x32 target region

Figure 4.3: Tracker loses the ball and determines the head of the player as the object

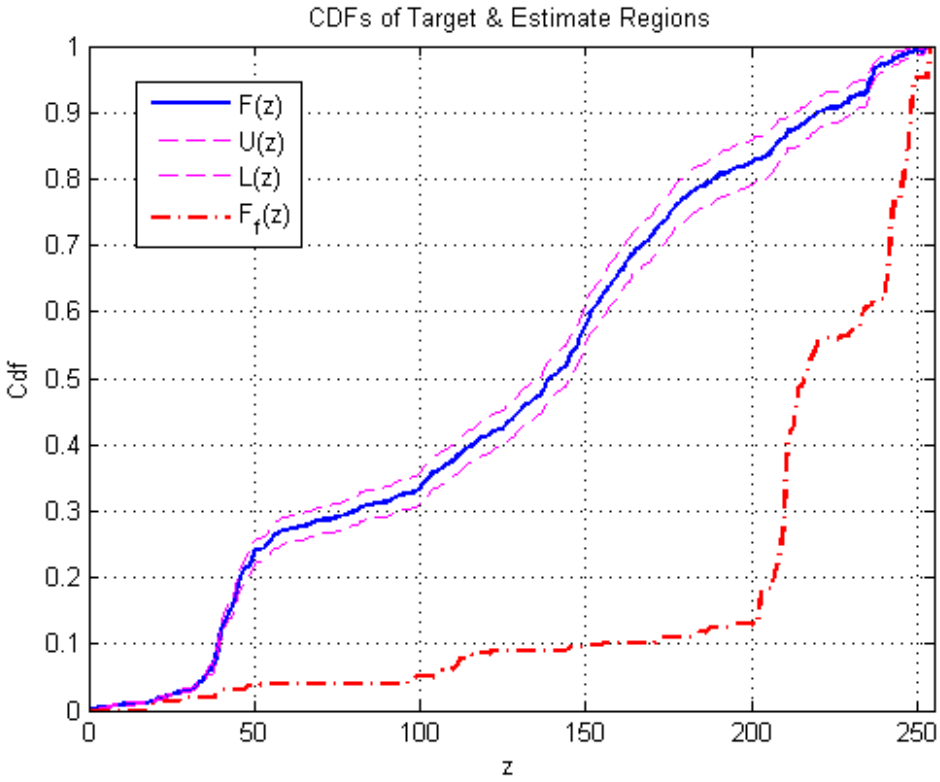


(a) Cdfs of the target and estimate regions

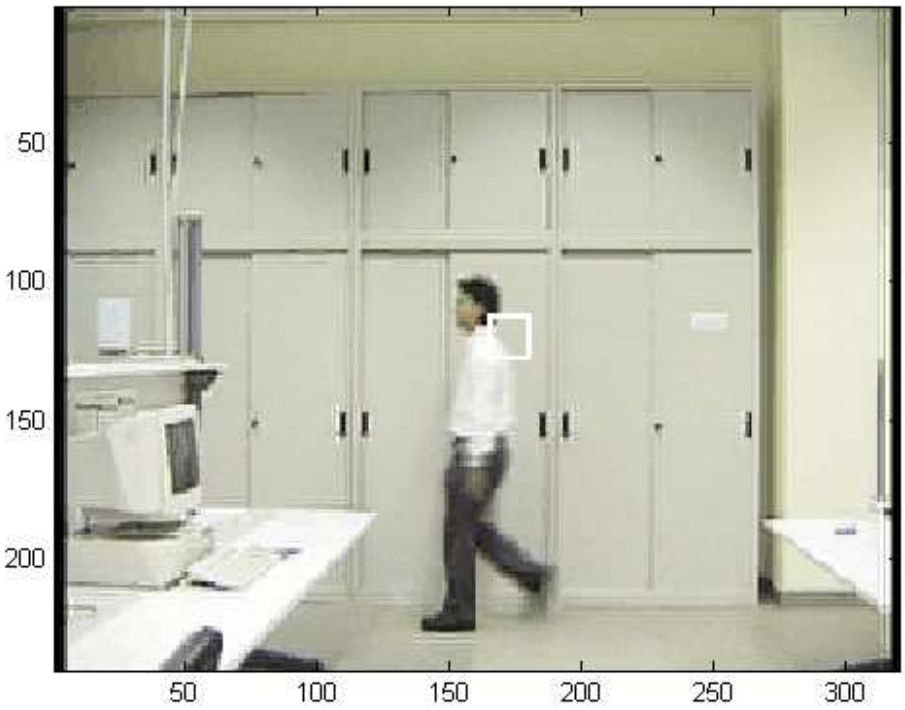


(b) False 16x16 target region estimate

Figure 4.4: Highway video sequence is used as source to the object tracking algorithm

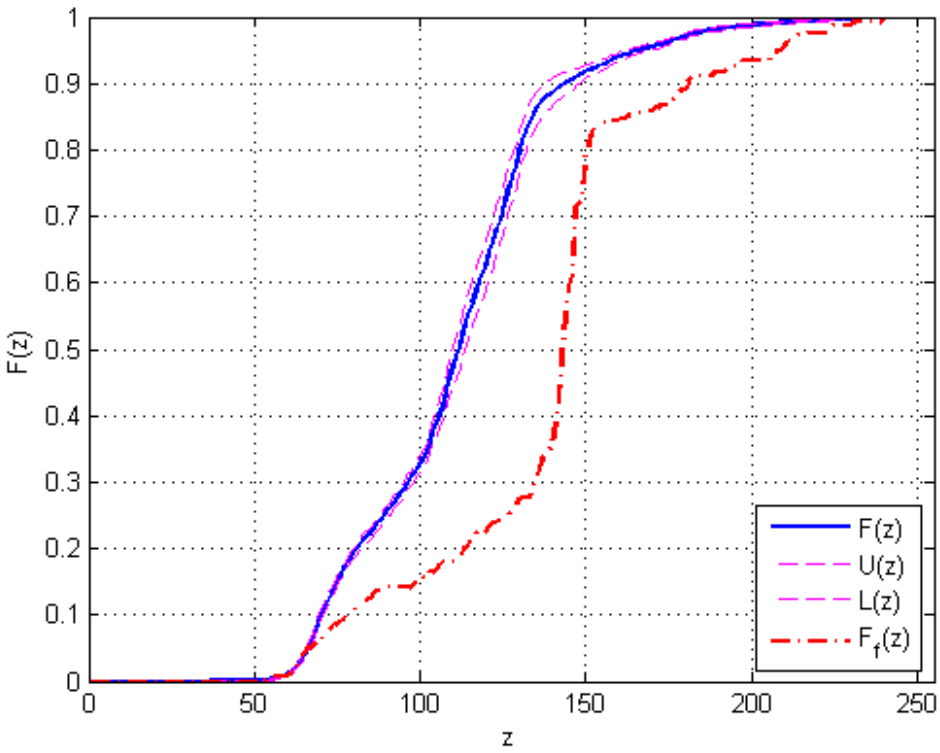


(a) Cdfs of the target and estimate regions



(b) False 15x15 target region estimate

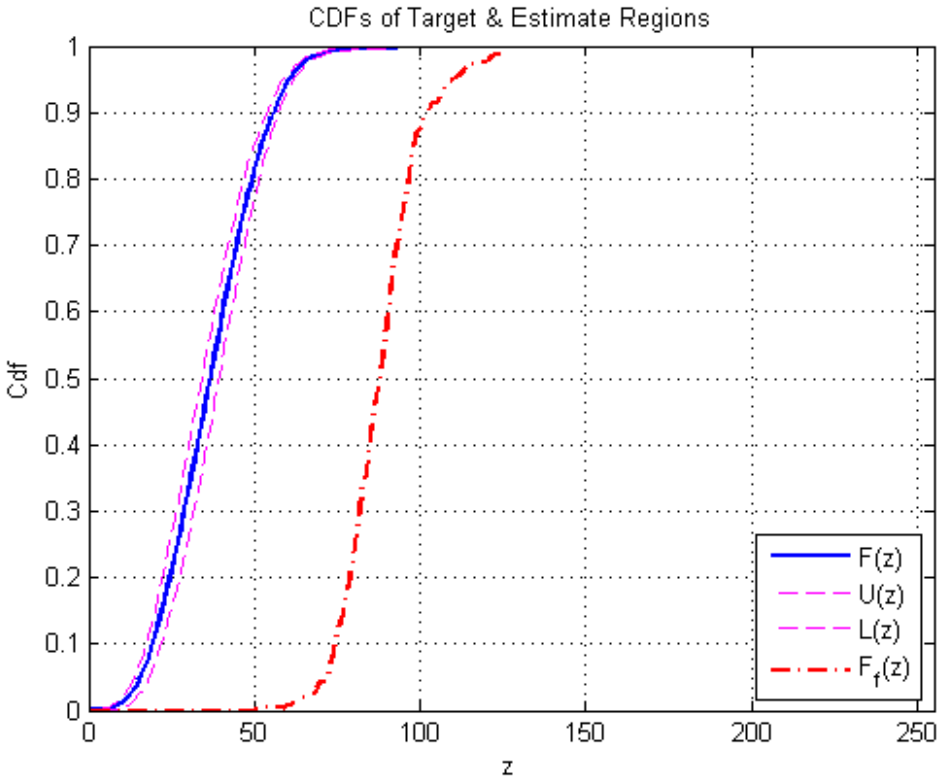
Figure 4.5: A man walking into a building is lost by the tracker due to varying illumination conditions



(a) Cdfs of the target and estimate regions

(b) False 20x20 target region estimate

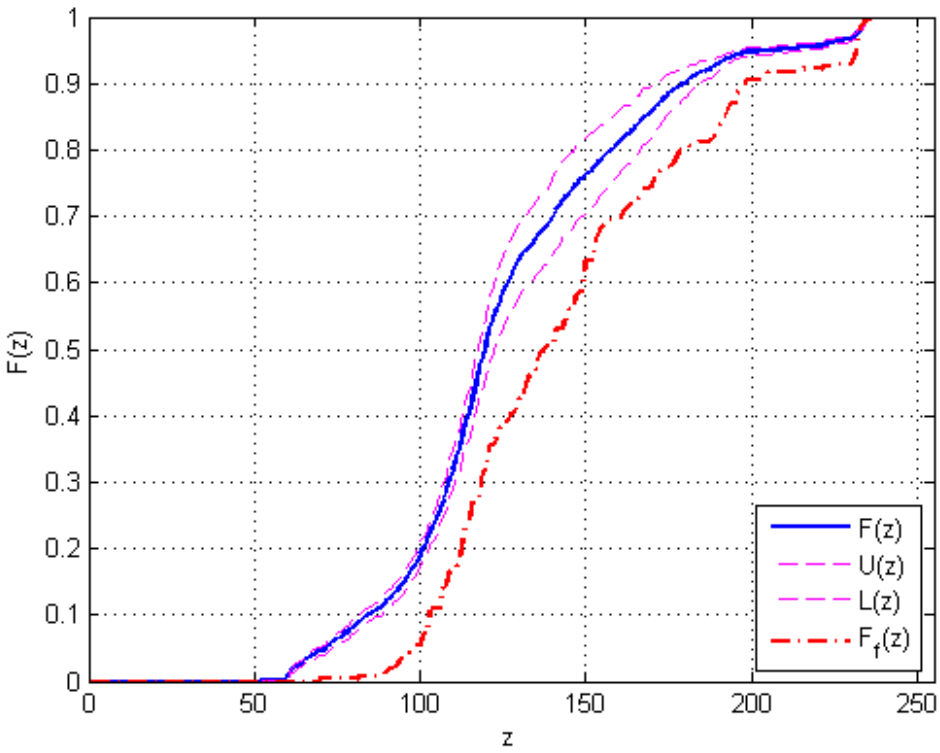
Figure 4.6: A man walking into a building is lost by the tracker due to varying illumination conditions



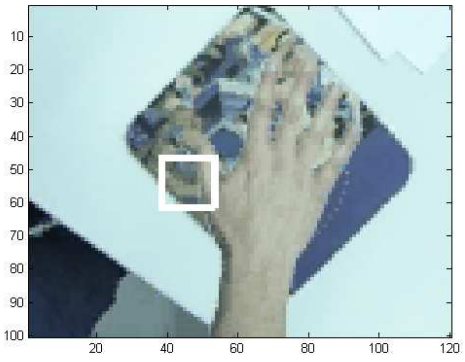
(a) Cdfs of the target and estimate regions

(b) False 25x25 target region estimate

Figure 4.7: A man walking into a building is lost by the tracker due to varying illumination conditions



(a) Cdfs of the target and estimate regions



(b) False 16x16 target region estimate

Figure 4.8: A hand moving over a complex background is lost by the tracker

Chapter 5

Conclusions

In this thesis, novel methods are proposed for recent research areas in video processing such as moving shadow detection, robust object tracking under illumination variations and target loss detection.

Cepstral domain analysis based shadow detection method is proposed as a contribution in the first part. The proposed method is composed of two steps. Moving object detection is followed by cepstral analysis for moving shadow detection. Cepstral domain analysis steps are based on the shadow regions retaining the underlying color and texture of the background region. The possible shadow blocks are determined in the first step. In the second step a pixel based decision mechanism is used to determine the shadow boundaries. The success of the proposed method is verified by the benchmark test results. Accurate shadow and object pixel segmentation in all video sequences is obtained. Quantitative measures show the improvements over previous approaches.

In the second part, an object tracking method based on cepstrum analysis is proposed for robust object tracking under varying illumination conditions. The proposed object tracking method is a combination of the covariance tracking method and the usage of 2D-cepstral features of the target region. The 2D-cepstrum is used for its property of retaining the underlying color and texture information under light-intensity variations. The method is applied to video sequences having varying light intensity. It is observed that the proposed method produces better results than the covariance tracking method. In addition, 2D-cepstral features of the target region is introduced to computationally efficient co-difference method and similar performance improvements are obtained.

Finally, a statistical method based on confidence interval functions is proposed for target loss detection in video. Upper and lower bound functions on the cumulative distribution function (cdf) of the target feature vector are calculated for a given confidence level. The target is regarded as being no longer tracked by the tracking algorithm when the estimated cdf of the detected region exceeds the calculated bounds. The method is applied to different video sequences in which the target is lost at some point. It is observed that the confidence interval based target loss detection method is successful in detecting the target losses. The proposed method is applicable to most tracking algorithms using features of the target image region.

Bibliography

[1] Y.-B. Lee, B.-J. You, and S.-W. Lee, "A real-time color-based object tracking robust to irregular illumination variations," in *Proceedings of IEEE International Conference on Robotics and Automation*, vol. 2, 2001.

[2] Y. Raja, S. McKenna, and S. Gong, "Tracking and segmenting people in varying lightning conditions using colour," in *Proceedings of 3rd International Conference on Automatic Face and Gesture Recognition*, 1998, pp. 228–233.

[3] B. U. Toreyin and A. E. Cetin, "Shadow detection using 2d cepstrum," in *Proceedings of SPIE, the International Society for Optical Engineering.*

[4] T. Horprasert, D. Harwood, and L. Davis, "A statistical approach for real-time robust background subtraction and shadow detection," in *Proceedings of the 7th IEEE International Conference on Computer Vision, Frame Rate Workshop (ICCV '99).*

[5] R. Cucchiara, C. Grana, M. Piccardi, and A. Prati, "Detecting objects, shadows and ghosts in video streams by exploiting color and motion information," in *Proceedings of the 11th International Conference on Image Analysis and Processing.*

[6] R. Cucchiara, C. Grana, M. Piccardi, A. Prati, and S. Sirotti, "Improving shadow suppression in moving object detection with hsv color information," in *Proceedings of IEEE International Conference on Intelligent Transportation Systems.*

[7] B. Chen, "Indoor and outdoor people detection and shadow suppression by exploiting hsv color information," in *Proceedings of the The Fourth International Conference on Computer and Information Technology.*

[8] C. Jiang and M. O. Ward, "Shadow identification," in *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition.*

[9] J. Stauder, R. Mech, and J. Ostermann, "Detection of moving cast shadows for object segmentation," *IEEE Transactions on Multimedia*, vol. 1, no. 1, pp. 65–76, 1999.

[10] K. Onoguchi, "Shadow elimination method for moving object detection," in *Proceedings of Fourteenth International Conference on Pattern Recognition*.

[11] Y. Sonoda and T. Ogata, "Seperation of moving objects and their shadows, and their application to tracking on the loci in the monitoring images," in *Proceedings of Fourth International Conference on Signal Processing*.

[12] A. Prati, I. Mikic, C. Grana, and M. M. Trivedi, "Shadow detection algorithms for traffic flow analysis: A comparative study," in *Proceedings of IEEE International Conference on Intelligent Transportation Systems*, 2001, pp. 340–345.

[13] A. Prati, C. Mikic, M. M. Trivedi, and R. Cucchiara, "Detecting moving shadows: Algorithms and evaluation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 7, pp. 918–923, 2003.

[14] A. Prati, R. Cucchiara, C. Mikic, and M. M. Trivedi, "Analysis and detection of shadows in video streams: a comparitive evaluation," in *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*, 2001, pp. 571–576.

[15] R. Colins, A. Lipton, and T. Kanade, "A system for video surveillance and monitoring," Robotics Institute, Carnegie Mellon University, Tech. Rep., May 2000, cMU-RI-TR-00-12.

[16] C. Stauffer and W. E. Grimson, "Adaptive background mixture models for real-time tracking," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 1999, pp. 246–252.

[17] I. Haritaoglu, D. Harwood, and L. Davis, "W4: Real-time surveillence of people and their activities," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 809–830, 2000.

[18] S. Furui, "Cepstral analysis technique for automatic speaker verification," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 29, no. 2, pp. 254–272, 1981.

[19] S. R. Oppenheim, A.V., "From frequency to quefrency: A history of the cepstrum," IEEE Signal Processing Magazine, 2004.

[20] "Video surveillance online repository," http://www.openvisor.org.

[21] A. Joshi, S. Atev, O. Masoud, and N. Papanikopoulos, "Moving shadow detection with low- and mid-level reasoning," in *Proceedings of IEEE International Conference on Robotics and Automation*, 2007, pp. 4827– 4832.

[22] F. Porikli, O. Tuzel, and P. Meer, "Covariance tracking using model update based on means on riemannian manifolds," in *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*, 2006, pp. 728–735.

[23] H. Tuna, I. Onaran, and A. E. Cetin, "Image description using a multiplier-less operator," *IEEE Signal Processing Letters*, vol. 16, no. 9, pp. 751–753, 2009.

[24] C. Stauffer and W. Grimson, "Learning patterns of activity using real-time tracking," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 747–757, 2000.

[25] A. Elgammal, D. Harwood, and L. Davis, "Non-parametric model for background subtraction," in *Proceedings of 6th European Conference on Computer Vision*, 2000, pp. 751–767.

[26] Y. Wu, J. Lin, and T. Huang, "Capturing natural hand articulation," in *Proceedings of IEEE International Conference on Computer Vision*, vol. II, 2001.

[27] M. Cascia, S. Sclaroff, and V. Athitsos, "Fast, reliable head tracking under varying illumination: An approach based on registration of texture-mapped 3d models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 4, pp. 322–336, 2000.

[28] J. Yang and A. Waibel, "A real-time face tracker," in *Proceedings of IEEE Workshop on Applications of Computer Vision*, 1996, pp. 142–147.

[29] F. Moreno-Noguer, J. Andrade-Cetto, and A. Sanfeliu, "Fusion of color and shape for object tracking under varying illumination," in *Proceedings of IEEE Iberian Conference on Pattern Recognition and Image Analysis*, 2003, pp. 580–588.

[30] C. Shen, X. Lin, and Y. Shi, "Moving object tracking under varying illumination conditions," *IEEE Pattern Recognition Letters*, vol. 27, no. 14, pp. 1632–1643, 2006.

[31] S. Jabri, Z. Duric, H. Wechsler, and A. Rosenfeld, "Detection and location of people in video images using adaptive fusion of color and edge information," in *Proceedings of IEEE International Conference on Pattern Recognition*, vol. 4, 2000.

[32] L. Li and M. Leung, "Integrating intensity and texture differences for robust change detection," *IEEE Transactions on on Image Processing*, vol. 11, no. 2, pp. 105–112, 2002.

[33] T. Horprasert, D. Harwood, and L. Davis, "A statistical approach for real-time robust background subtraction and shadow detection," in *Proceedings of 7th IEEE International Conference on Computer Vision Frame-Rate Workshop*, 1999, pp. 1–19.

[34] F. Porikli, O. Tuzel, and P. Meer, "Covariance tracking using model update based on lie algebra," in *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*, 2006, pp. 728–735.

[35] A. Yilmaz, O. Javed, and M. Shah, "Object tracking: A survey," ACM Comput. Surv. 38, 2006.

[36] C. Chang and R. Ansari, "Kernel particle filter for visual tracking," *IEEE Signal Processing Letters*, vol. 12, no. 3, pp. 242–245, 2005.

[37] S. Beheshti, M. Hashemi, X.-P. Zhang, and N. Nikvand, "Noise invalidation denoising," http://arxiv.org/abs/1006.4801.