COMPRESSION ARTIFACT REDUCTION AND MOVING OBJECT TRACKING IN THE IMAGES AND VIDEOS

A THESIS

Submitted by

PURUSHOTHAMAN A

in partial fulfilment for the award of the degree

of

DOCTOR OF PHILOSOPHY

FACULTY OF ELECTRONIC AND COMMUNICATION ENGINEERING

ANNA UNIVERSITY OF TECHNOLOGY, COIMBATORE

COIMBATORE 641 047

JANUARY 2012

ANNA UNIVERSITY OF TECHNOLOGY
COIMBATORE 641 047

BONAFIDE CERTIFICATE

Certified that this thesis titled **“COMPRESSION ARTIFACT REDUCTION AND MOVING OBJECT TRACKING IN THE IMAGES AND VIDEOS”** is the bonafide work of **Mr.PURUSHOTHAMAN.A** who carried out the research under my supervision. Certified further, that to the best of my knowledge the work reported herein does not form part of any other project report or dissertation on the basis of which a degree or award was conferred on an earlier occasion on this or any other candidate.

SIGNATURE
Dr.R.RANGARAJAN
SUPERVISOR,
Principal,
Indus College of Engineering,
Kalimangalam, Alandurai,
Coimbatore-641 101,
Tamil Nadu,India.

ABSTRACT

Compression artifact reduction in the images and videos: Block artifact is one of the visually annoying problems that usually exist in low-bit-rate compression of images and videos. This thesis propose a simple but effective method to reduce block artifact based on pixel classification in spatial domain and frequency domain corrupted with impulses, Gaussian noises, artifacts (blocking, ringing, blur, etc.). In traditional methodology, linear filters are not effective in the removal of multiplicative noise, impulse noise and artifact (blocking, ringing, blur, etc.). And also it is large in amplitude; hence it dominates characterizations of the signals based on second-order statistics such as correlation and spectral analysis.

This thesis aims to design a non linear adaptive based algorithm, smooth region and edge region (low pass filter is performed for image smoothening) by using a binary edge map from the edge detection process, for removing the artifacts (blocking, ringing, blur, etc.) and to preserve edges and fine details in images and videos. This algorithm includes detection of corrupted pixels and the estimation of values for replacing the corrupted pixels. The main advantage of the proposed algorithm is the uncorrupted pixels are unaltered and produced in the output. The appropriate filter is used based on the variance of the filter window, for estimating the value for

replacing the corrupted value. This leads to reduced artifacts and high fine detail preservation for compressed images.

Moving object tracking in the images and videos: The moving object tracking in video pictures (Takashi morimoto et al 2005) has attracted a great deal of interest in computer vision. For **object recognition**, **navigation systems** and **surveillance systems** (Foresti 1998), object tracking is an indispensable first-step. A proposed a novel algorithm is proposed for object tracking in video pictures, based on **image segmentation** and **pattern matching** (Takashi morimoto et al 2005). With the use of image segmentation, one can detect all objects in the images no matter whether they are moving or not can be detected. Using image segmentation results of successive frames, pattern matching in a simple feature space for tracking of the objects is got. Consequently, the proposed algorithm can be applied to multiple moving and still objects even in the case of a moving camera. The algorithm is discussed in detail and is simulated.

The conventional approach to object tracking is based on the difference between the current image and the background image. However, algorithms based on the difference image cannot simultaneously detect still objects. They cannot be applied in the case of a moving camera. Algorithms including the camera motion information have been proposed previously, but, they still contain problems in separating the information from the background.

Real-time object tracking is recently becoming more important in the field of video analysis and processing. Applications like traffic-control,

user-computer interaction, online video processing and production and video surveillance need reliable and economically affordable video tracking tools. This thesis presents a new method for object tracking with improved efficiency by reducing the number of computations. The video frames are captured and smoothed by a low pass filter. Another group of algorithms deals with object tracking using adaptive particle filters and kalman filter. The captured frame in RGB (Red,Green and Blue) color space is transformed to YCbCr space. The Y component of YCbCr is subjected to Continuous Wavelet Transform (CWT) by dividing into 8×8 blocks, and then the Wavelet coefficients are extracted and are used for detecting moving objects. This method reduces the number of computations and computation time. In addition the noise in the signals are suppressed automatically. The thesis concludes with suggestions for future work.

ACKNOWLEDGEMENT

This thesis is by far the most significant accomplishment in my life and it would be impossible without the blessing of God.

I would like to extend my gratitude and my sincere thanks to my esteemed supervisor **Dr.R.Rangarajan,** Principal, Indus College of Engineering, Coimbatore-01., T.N.,India. He is not only a great researcher with deep vision but also most importantly a kind person. I sincerely thank for his exemplary guidance, continuous encouragement, diligent feedback and professional support.

I express my heartfelt thanks to the Doctoral Committee Members **Dr.A.Kandasawamy,** Professor and HOD/BME Department, PSG College of Technology, Coimbatore-, T.N.,India, **Dr.K.R.Shankar Kumar,** Professor, ECE Department, Sri Ramakrishna Engineering College, Coimbatore-22., T.N., India and **Dr.A.Rajeswari,** Professor and HOD/ECE Department, Coimbatore Institute of Technology, Coimbatore-, T.N.India. for their technical guidance and moral support.

I would like to thank my parents, who taught me the value of hard work, it's their blessings which always gave me courage to face all challenges and made my path easier.

I express my sincere thanks to The Management, The Director Academics of SNR charitable trust, The Principal, The Head of the department /ECE, of Sri Ramakrishna Engineering College for providing us with best facilities in the department.

Last but not the least I would like to thank all those who have helped me directly or indirectly for this thesis work.

PURUSHOTHAMAN A

TABLE OF CONTENTS

CHAPTER NO.	TITLE	PAGE NO.

CHAPTER NO.	TITLE	PAGE NO.

CHAPTER NO.	TITLE	PAGE NO.

CHAPTER NO.		TITLE	PAGE NO.

LIST OF TABLES

LIST OF FIGURES

FIGURE NO.	TITLE	PAGE NO.

FIGURE NO.	TITLE	PAGE NO.

LIST OF SYMBOLS AND ABBRIVATIONS

Symbols

θ	-	Angle of Orientation of the Edge
m_1	-	Average value of G_1
m_2	-	Average value of G_2
G_2	-	Background pixels
SQI,SQS	-	Correlation Coefficients
R_{xx}	-	Correlation Functions
C_{ij}	-	Current Block
*	-	Denotes complex conjugation
S_i	-	Difference between E_i and E_{i-1}
S_i	-	Difference between Fi and E_{i-1}
$\mathfrak{I}^{-1}$	-	Filtered Image
H(u,v)	-	Fourier Transform of the Filter Function h(x,y)
G(u,v)	-	Fourier Transform of the Filtered Image
F(u, υ)	-	Fourier Transform of the Input Image f (x, y)
f (t)	-	Function
G_x, G_y	-	Gradient
N	-	Indicates the Height of the Frame
E_i	-	Indicates the i^{th} Frame
E_i	-	Indicates the i^{th} Frame
M	-	Indicates the Width of Frame
Y	-	Indicates the Y Plane of the YCbCr Image
magn(x, y)	-	Magnitude Functions
MAXi	-	Maximum Possible Pixel Value of the Image
M*	-	Number of columns in the given image
N*	-	Number of rows in the given image
X^*_j	-	Original Image

G_1	-	Object pixels
E_{i-1}	-	Previous Frame
X_j	-	Reconstructed Image
R_{ij}	-	Reference Block
S	-	Scale
ψ^* (s,τ)	-	Set of Basis Function
N’	-	Size of the Image and Marco Block
f (x, y)*	-	Smooth the input image
T	-	Threshold
τ	-	Translation
I(i,j)	-	Wavelet at i^{th} row and j^{th} column
$\gamma(s,\tau)$	-	Wavelet transform Functions

Abbreviations

1D	-	One Dimensional
2D	-	Two Dimensional
3D	-	Three Dimensional
ATRAC	-	Adaptive Transform Acoustic Coding
AVI	-	Audio Video Interleaved
BLPF	-	Butterworth Low Pass Filter
BM	-	Block Matching
BMC	-	Block Motion Compensation
BMP	-	Bitmap
BWT	-	Burrows and Wheeler Transform
CD	-	Compact Disk
CD-ROM	-	Compact Disk- Read Only Memory
CG	-	Computer Generated
CIE	-	Commission Internationale de l’ Eclairage

CWT	-	Continuous Wavelet Transform
DBN	-	Dynamic Bayesian Networks
DCT	-	Discrete Cosine Transform
DIB	-	Device-Independent Bitmap
DS	-	Diamond Search
DTV	-	Digital Television
DVD	-	Digital Versatile Disc
EDP	-	Electronic Data Processing
EP	-	Electronic Photography
ES	-	Exhaustive Search
FLAC	-	Free Lossless Audio Codec
GA	-	Genetic Algorithm
GDI	-	Graphics Device Interface
GIF	-	Graphic Interchange Format
GIG	-	Gluten Intolerance Group
GLPF	-	Gaussian Low Pass Filter
GMI	-	Gradient Magnitude Intensities
HLS	-	Hue, Lightness and Saturation
HSV	-	Hue, Saturation and Value
HVS	-	Human Visual System
ILPF	-	Ideal Low Pass Filter
ITU	-	International Telecommunication Union
JPEG	-	Joint Photographic Experts Group
JPEG-LS	-	Lossless JPEG(JPEG-LS)
KLT	-	Karhunen-Loeve Transform
LDSP	-	Large Diamond Shape Pattern
LMI	-	Local Intensity Minimum
LZW	-	Lempel Ziv Welch
MAD	-	Mean Absolute Difference

MC	-	Motion Compensation
ME	-	Motion Estimation
MP3	-	Motion Picture Experts Group Layer 3
MPEG	-	Motion Picture Experts Group
MSE	-	Mean Square Error
MVD	-	Motion Vector Distribution
NTSC	-	National Television Standards Committee
NTSS	-	New Three Step Search
PAL	-	Phase Alternating Line
PCM	-	Pulse code Modulation
PDE	-	Partial Differential Equations-based
PDF	-	Probability Density Function
PNG	-	Portable Network Graphics
PPAM	-	Professional Photographers of Massachusetts
PPM	-	Prediction by Partial Matching
PS	-	Post Script
PSNR	-	Peak Signal to Noise Ratio
PT	-	Pan-Tilt
RBF	-	Radial Basis Function
RGB	-	Red, Green, Blue
ROI	-	Region-of-Interest
SBC	-	Subband Coding
SDSP	-	Small Diamond Shape Pattern
SPIHT	-	Set Partitioning In Hierarchical Trees
TIFF	-	Tagged Image File Format
TRL	-	Target Representation and Localization
TSS	-	Three Step Search
VHS	-	Video Home System
VIRIS	-	Video Reference Impairment System

WT	-	Wavelet Transform
WWC	-	World Wide Consortium
WWW	-	World Wide Web

CHAPTER 1

INTRODUCTION

1.1 MOTIVATION

Image and Video Compression: Today's needs for low bit rates in data compression are responsible for the Motion Picture Experts Group (MPEG) and International Telecommunication Union (ITU) family of video coding standards' success. However, it is important to realize that such standards involve lossy compression. This means a reconstructed image sequence contains degradations relative to the original, because information is discarded during compression. Low bit rate compression modes often introduce block edge, ringing, blurring, color bleeding, or mosquito noise effects. This thesis, proposes a Partial Differential Equations-based (PDE) method to reduce compression-related artifacts on MPEG-4 color image sequences. The main advantage of our approach is that both temporal and spatial artifacts removal are performed at the same time, by considering an image sequence as a 3D object. Although PDE-based methods for still images restoration are becoming quite popular, extensions to image sequences remain rare, especially in color. This is also the purpose of this thesis to introduce one.

Moving object tracking: The moving object tracking in video pictures has attracted a great deal of interest in computer vision. For **object recognition, navigation systems** and **surveillance systems**, object tracking is an indispensable first step. Object tracking has significance in real time environment because it enables several important applications such as **security and surveillance** (Foresti 1998) to recognize people and to provide better sense of security using visual information. In **Medical therapy,** to improve the quality of life for physical therapy patients and disabled people. In Retail space instrumentation, to analyze shopping behavior of customers, to enhance building and environment design, video abstraction to obtain **automatic annotation of videos**, to generate object based summaries, traffic management to analyze flow, to detect accidents, video editing to eliminate cumbersome human operator interaction and to design futuristic video effects.

1.2 LITERATURE REVIEW

1.2.1 Literature Review of Compression Artifact Reduction in the Images and Videos

The Discrete Cosine Transform (DCT) is the integrated part of the basic compression algorithm in Joint Photographic Experts Group (JPEG). The importance of this compression technique can be utilized due to its performance that matches the Karhunen-Loeve Transform (KLT), which is known to be optimal in the mean square error sense. Although this is the most popular compression approach, its main drawback is the “blocking effect.” Dividing the image into small blocks prior to coding causes discontinuities between adjacent blocks and affects the strong edges in the image. In this approach, correlation between adjacent blocks is not exploited. The quantization errors are generated in different blocks, reconstructed images exhibit severe artifacts at block boundaries for high compression ratios. The blocking effect in Joint Photographic Experts Group (JPEG) images can be characterized into five kinds as follows.

1. Stair case noise along the image edges
2. Grid noise in the monotone areas
3. Corner outliners in the corner points of the DCT blocks
4. Ringing artifacts near the strong edges
5. Corruption of edges across block boundaries. Due to the huge data requirements for multimedia, the attention is focused toward getting more compression and less visual defects

To remove the blocking effect, several deblocking techniques have been proposed in literature, as post process mechanisms after JPEG compression, depending on the perspective from which the deblocking problem is dealt with. Reeves and Lim (1984) have dealt with the easiest way of looking at this problem is to low-pass the blocky JPEG image. Crouse and Kannan (1998) have dealt with the approach which will reduce the effect of high frequency tendency but the image will be blurry and some details will be wiped out. Going further step in complexity and applying a simple nonlinear smoothing to the pixels will add another obstacle to the solution. Zhigang and Fu (1996) have dealt with more sophisticated approach which involves segmentation and smoothing that will reduce the ringing artifacts due to sharp variations. Sung-Wai et al (1996) have dealt with the importance of high-frequency details that may be overwhelmed with blocking noise; another restoration process that uses linear predictors to estimate the corresponding boundary pixels in the regions of interest is devised.

Yung-Kai et al (1996) have dealt with the classification of small local boundary regions according to their intensity distribution and to employ this information in designing proper predictors. Images with sharp variations cannot be easily configured with low order predictive filters. To improve the accuracy of the classified patterns, an iterative method for block removal using block classification and space frequency filtering is proposed. The features of the wavelet theory add another tool of exploration to the blocking problem; several ideas based on soft threshold approach in the wavelet domain are successfully implemented for deblocking JPEG coded images in Gopinath et al (1994). (Donoho 1995) The core of these techniques is to make use of Donoho's algorithm for denoising Gaussian noise and modifying the denoised wavelet coefficients to remove the effect of other types of noise. Direct application of Donoho's algorithm for removing blocking effect can be found from image and video. The further step has been adopted by presenting a simple and efficient denoising algorithm that exploits correlations among cross-scale wavelet coefficients to extract edge information and protect these edges during threshold operation in Xiong et al (1997).

The reconstructed images from highly compressed MPEG data have perceivable image degradation, such as blocking effect, ringing effect and corner outliners. At low bit-rate coding, blocking effect is visually more perceptible and annoying than others. The high amount of audio-visual data associated with typical multimedia services call for efficient data compression schemes in order to facilitate transmission and storage applications. Applications include video conferencing, video phones, remote monitoring and control, information based retrieval, video on demand and digital television.

There are various international video-compression standards that suit different requirements. Block-based motion estimation and a block-based Discrete Cosine Transform (DCT) are used in video compression standards like MPEG-1, MPEG-2, MPEG-4 and H.263. The standards use an 8 × 8 pixel block DCT for packing information into a few coefficients by utilizing the spatial correlation property of images. This block-based coding introduces blocking artifacts between block boundaries as the transform does not take into account the correlation between block boundaries and the block is independently coded. The blocking artifacts are the grid noise along the block boundaries in a relatively homogeneous area. The blocking artifacts are mainly resulted due to the quantization of DC and AC coefficients, which are nearer to DC coefficient. The other artifacts that are introduced due to truncation of high frequency coefficients by quantization are ringing effect around edges (due to Gibb's phenomenon), mosquito noise and corner outliners.

Motion compensated coding has been the most widely used video compression method amongst other compression schemes. The main difference of motion compensated coding from other transform and prediction based methods is its utilization of the motion characteristics in the images. While most other compression algorithms achieve compression by exploiting the redundancies only in spatial dimensions (in images), motion compensated coding exploits redundancies between two consecutive frames as well. In turn, motion compensated coding provides more efficient compression.

This project deals with the removal of compression artifacts in images and video by using an adaptive algorithm which improves Peak Signal to Noise Ratio (PSNR) and Mean square Error (MSE) when compared to the other techniques as mentioned in the literature.

1.2.2 Literature Review of Moving Object Tracking in the Images and Videos

There have been a number of surveys about object detection, classification, tracking and activity analysis in the literature, e.g. Aggarwal and Cai (1999), and Hwann-Tzong Chen et al (2001). The survey we present here covers only those works that are in the same context as our study.

However, for completeness, we also give brief information on some techniques which are used for similar tasks that are not covered in our study.

A generic smart video processing framework is shown in Figure 1.1 Although, some steps require interchange of information with other levels, this framework provides a good structure for the discussion throughout this survey.

Each application that benefits from smart video processing has different needs, thus requires different treatment. However, they have something in common moving objects. Thus, detecting regions that correspond to moving objects such as people and vehicles in video is the first basic step in almost every vision system since it provides a focus of attention and simplifies the processing on subsequent analysis steps.

Due to dynamic changes in natural scenes such as sudden illumination and weather changes, repetitive motions that cause clutter (tree leaves moving in blowing wind), motion detection is a difficult problem to process reliably. Frequently used techniques for moving object detection are given below,

- Background Subtraction
- Statistical Method
- Temporal Differencing
- Optical Flow
- Shadow and Light Change Detection

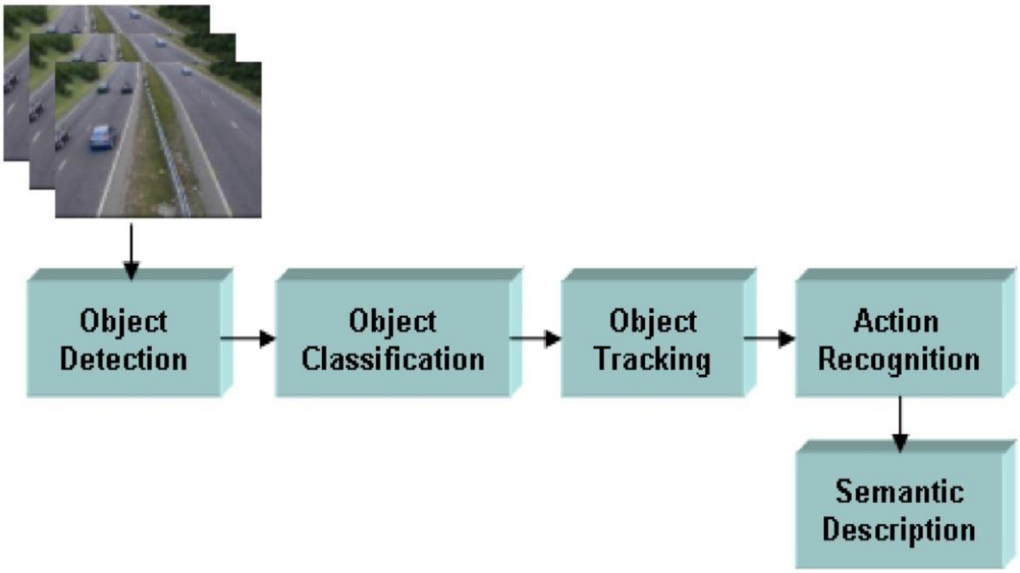


Figure 1.1 A generic framework for smart video processing algorithms

1.2.2.1 Background Subtraction

Background subtraction is particularly a commonly used technique for motion segmentation in static scenes (McIvor 2000). It attempts to detect moving regions by subtracting the current image pixel-by-pixel from a reference background image that is created by averaging images over time in an initialization period. The pixels where the difference is above a threshold are classified as foreground. After creating a foreground pixel map, some morphological post processing operations such as erosion, dilation and closing are performed to reduce the effects of noise and enhance the detected regions. The reference background is updated with new images over time to adapt a dynamic scene changes.

There are different approaches to this basic scheme of background subtraction in terms of foreground region detection, background maintenance and post processing.

Heikkila and Silven (1999) uses the simple version of this scheme where a pixel at location (x, y) in the current image It is marked as foreground if,

$$|I_t(x, y) - B_t(x, y)| > T \quad (1.1)$$

Equation (1.1) is satisfied where T is a predefined threshold. B_t background image is updated by the use of an Infinite Impulse Response (IIR) filter as follows

$$B_{t+1} = \alpha I_t + (1 - \alpha) B_t \quad (1.2)$$

The foreground pixel map creation is followed by morphological closing and the elimination of small-sized regions.

Although background subtraction techniques perform well at extracting most of the relevant pixels of moving regions even they stop, they are usually sensitive to dynamic changes when, for instance, stationary objects uncover the background (e.g. a parked car moves out of the parking lot) or sudden illumination changes occur.

1.2.2.2 Statistical Method

More advanced method that makes use of the statistical characteristics of individual pixels has been developed to overcome the shortcomings of basic background subtraction methods. These statistical methods are mainly inspired by the background subtraction methods in terms of keeping and dynamically updating statistics of the pixels that belong to the background image process. Foreground pixels are identified by comparing each pixel's statistics with that of the background model. This approach is becoming more popular due to its reliability in scenes that contain noise, illumination changes and shadow in Wang et al (2003).The W4 (Haritaoglu

1998) system uses a statistical background model where each pixel ‘I’ represented with its minimum (M) and maximum (N) intensity values and maximum intensity difference (D) between any consecutive frames observed during initial training period, where the scene contains no moving objects. A pixel in the current image is classified as foreground if it satisfies,

$$|M(x, y) - I_t(x, y)| > D(x, y) \text{ or } |N(x, y) - I_t(x, y)| > D(x, y) \qquad (1.3)$$

After thresholding, a single iteration of morphological erosion is applied to the detected foreground pixels to remove one-pixel thick noise. In order to grow the eroded regions to their original sizes, a sequence of erosion and dilation is performed on the foreground pixel map. Also, small-sized regions are eliminated after applying connected component labeling to find the regions. The statistics of the background pixels that belong to the non-moving regions of current image are updated with new image data. As another example of statistical methods, Stauffer and Grimson (1999) described an adaptive background mixture model for real-time tracking. In their work, every pixel is separately modeled by a mixture of Gaussians which are updated online by incoming image data. In order to detect a pixel which belongs to a foreground or background process, the gaussian distributions of the mixture model for that pixel is evaluated.

1.2.2.3 Temporal Differencing

Temporal differencing attempts to detect moving regions by making use of the pixel-by-pixel difference of consecutive frames (two or three) in a video sequence. This method is highly adaptive to dynamic scene changes, however, it generally fails in detecting whole relevant pixels of some types of moving objects. A sample object for inaccurate motion detection is shown in Figure 1.2.

The mono colored region of the human on the left hand side makes the temporal differencing algorithm to fail in extracting all pixels of the human's moving region. Also, this method fails to detect stopped objects in the scene. Additional methods need to be adopted in order to detect stopped objects for the success of higher level processing

(a) **(b)**

Figure 1.2 Temporal differencing sample (a) A sample scene with two moving objects (b) Temporal differencing fails to detect all moving pixels of the object on the left hand side since it is uniform colored. The detected moving regions are marked with red pixels.

Lipton et al (1998) presented a two-frame differencing scheme where the pixels that satisfy the following equation are marked as foreground.

$$|I_t(x, y) - I_{t-1}(x, y)| > T \quad (1.4)$$

In order to overcome shortcomings of two frame differencing in some cases, three frame differencing can be used in Wang et al (2003). For instance in Collins et al (2000) developed a hybrid method that combines three-frame differencing with an adaptive background subtraction model for their VSAM project. The hybrid algorithm successfully segments moving regions in video without the defects of temporal differencing and background subtraction.

1.2.2.4 Optical Flow

Optical flow methods make use of the flow vectors of moving objects over time to detect moving regions in an image. They can detect motion in video sequences even from a moving camera; however, most of the optical flow methods are computationally complex and cannot be used real-time without specialized hardware in Wang et al (2003).

1.2.2.5 Shadow and Light Change Detection

The algorithms described above for motion detection perform well in indoor and outdoor environments and have been used for real-time surveillance for years. However, without special care, most of these algorithms are susceptible to both local (e.g. shadows and highlights) and global illumination changes (e.g. sun being covered/uncovered by clouds). Shadows cause the motion detection methods fail in segmenting only the moving objects and make the upper levels such as object classification to perform inaccuracy. The proposed methods in the literature mostly use chromaticity in Horprasert et al (1999), Kaew TraKul Pong and Bowden (2002), McKenna et al (2000), Wang et al (2003), or stereo Haritaoglu et al (1998) information to cope with shadows and sudden light changes.

Horprasert et al (1999) present a novel background subtraction and shadow detection method. In their method, each pixel is represented by color model that separates brightness from the chromaticity component. A given pixel is classified into four different categories (background, shaded background or shadow, highlighted background and moving foreground object) by calculating the distortion of brightness and chromaticity between the background and the current image pixels. Like, the approach described McKenna et al (2000) in uses chromaticity and gradient information to cope with shadows. They make use of the observation that an area cast into shadow results in significant change in intensity without much change in chromaticity. They also use the gradient information in moving regions to ensure reliability of their method in ambiguous cases.

Wang et al (2003) The method presented adopts a shadow detection scheme which depends on two heuristics: a) pixel intensity values within shadow regions tend to decrease in most cases when compared to the background image b) the intensity reduction rate changes smoothly between neighboring pixels and most shadow edges do not have strong edges. An efficient method to deal with shadows is using stereo as presented in W4S Haritaoglu et al (1998) system.

In W4S, stereo image is generated by an inexpensive real-time device called SVM which uses two or more images to calculate a range image by using simple stereo image geometry. With the help of the range information provided by SVM, W4S is able to cope with shadows, sudden illumination changes and complex occlusion cases. In some systems, a global light change is detected by counting the number of foreground pixels and if the total number exceeds some threshold (e.g. 50% of the total image size), the system is reset to adapt to the sudden illumination change in Owens and Hunter (2002), Xuhui Zhou et al (2003).

1.2.3 Object Classification

Moving regions detected in video may correspond to different objects in real-world such as pedestrians, vehicles, clutter, etc. It is very important to recognize the type of a detected object in order to track it reliably and analyze its activities correctly. Currently, there are two major approaches towards moving object classification which are shape-based and motion-based methods by Wang et al (2003). Shape-based methods make use of the objects' 2D spatial information whereas motion-based methods use temporally tracked features of objects for the classification solution.

1.2.3.1 Shape-based Classification

Common features used in shape-based classification schemes are the rectangle area, silhouette and gradient of detected object regions. The approach presented in Qiang Chen et al (2010) makes use of the objects' silhouette contour length and area information to classify detected objects into three groups: human, vehicle and other. The method depends on the assumption that humans are, in generally, smaller than vehicles and have complex shapes. Dispersedness is used as the classification metric and it is defined in terms of object's area and contour length (perimeter) as follows

$$\text{Dispersedness} = \frac{\text{perimenter}^2}{\text{area}} \tag{1.5}$$

Classification is performed at each frame and tracking results are used to improve temporal classification consistency.

The classification method developed in Collins et al (2000) uses view dependent visual features of detected objects to train a neural network classifier to recognize four classes: human, human group, vehicle and clutter. The inputs to the neural network are the dispersedness, area and aspect ratio of the object region and the camera zoom magnification. Like the previous method, classification is performed at each frame and results are kept in a histogram to improve temporal consistency of classification.

Saptharishi et al (2000) propose a classification scheme which uses a logistic linear neural network trained with Differential Learning to recognize two classes: vehicle and people.

Papageorgiou et al (1998) reported a method that makes use of the Support Vector Machine classification trained by wavelet transformed object features (edges) in video images from a sample pedestrian database. This method used to recognize moving regions that respond to humans. Another classification method is proposed in Brodsky et al (2002), uses a Radial Basis Function (RBF) classifier which has a similar architecture like a three-layer.

1.2.3.2 Motion-based Classification

Some of the methods in the literature use only temporal motion features of objects in order to recognize their classes in Cutler et al (2000), (Lipton 1998) and Wixson et al (1998). Generally, they are used to distinguish non-rigid objects (e.g. human) from rigid objects (e.g. vehicles).

Cutler et al (2000), the method proposed is based on the temporal self-similarity of a moving object. As an object that exhibits periodic motion evolves, its self-similarity measure also shows a periodic motion. The method exploits this clue to categorize moving objects using periodicity.

Optical flow analysis is also useful to distinguish rigid and non-rigid objects. Lipton et al (1999) proposed a method that makes use of the local optical flow analysis of the detected object regions. It is expected that non-rigid objects such as humans will present high average residual flow whereas rigid objects such as vehicles will present little residual flow. Also, the residual flow generated by human motion will have a periodicity. By using cue, human motion, thus humans, can be distinguished from other objects such as vehicles.

1.2.3.3 Fire Detection

The number of papers that discuss fire detection using video are very few in computer vision literature. Most of the proposed methods exploit the color and motion features of fire. Healey et al (1993) and Heijden (1996) use a model which is based only on color characteristics of fire. Obviously this method generates false alarms due to fire colored regions. An improved approach which makes use of motion information as well as the color property is presented in Philips et al (2000).

Liu et al (2004) presented a method that defines spectral, spatial and temporal models of fire to detect its presence in video. The spectral model is represented in terms of fire pixel color probability density. The spatial model describes the spatial structure of a fire region and the temporal model captures the changes in the spatial structure over time.

1.2.3.4 Object Tracking

Tracking is a significant and difficult problem that arouses interest among computer vision researchers. The objective of tracking is to establish correspondence of objects and object parts between consecutive frames of video. It is a significant task in most of the surveillance applications, since it provides cohesive temporal data about moving objects which are used both to enhance lower level processing such as motion segmentation and to enable higher level data extraction such as activity analysis and behavior recognition. Tracking has been a difficult task to apply in congested situations due to inaccurate segmentation of objects. Common problems of erroneous segmentation are long shadows, partial and full occlusion of objects with each other and with stationary items in the scene. Thus, dealing with shadows at motion detection level and coping with occlusions both at segmentation level and at tracking level is important for robust tracking.

Tracking in video can be categorized according to the needs of the applications. It is used in or according to the methods used for its solution. Whole body tracking is generally adequate for outdoor video surveillance whereas objects part tracking is necessary for some indoor surveillance and higher level behavior understanding applications.

There are two common approaches in tracking objects as a whole (Amer 2003) one is based on correspondence matching and other one carries out explicit tracking by making use of position prediction or motion estimation. On the other hand, the methods that track parts of 3 objects (generally humans) employ model-based schemes to locate and track body parts. Some example models are stick figure cardboard model, 2D contour and 3D volumetric models combine motion estimation methods with correspondence matching to track objects. It is also able to track parts of people such as heads, hands, torso and feet by using the cardboard model which represents relative positions and sizes of body parts. It keeps appearance templates of individual objects to handle matching even in merge and split cases. Amer (2003) presents a non-linear voting based scheme for tracking objects as a whole. It integrates object features like size, shape, center of mass and motion by voting and decides final matching with object correspondence. This method can also detect object split and fusion and handle occlusions.

The algorithm incorporates size and positions of objects for seeding and maintaining a set of Kalman filters for motion estimation. Also, Extended Kalman filters are used for trajectory prediction and occlusion handling in the work of Rosales and Sclaroff (1998). As an example of model based body part tracking system, Wren et al (1997) Pfinder makes use of a multi-class statistical model of color and shape to track head and hands of people in real-time.

Moving object tracking is the process of locating a moving object in time using a camera as shown in figure 1.1. An algorithm analyses the video frames and outputs the location of moving targets within the video frame. The main difficulty in video tracking is to associate target locations in consecutive video frames, especially when the objects are moving fast relative to the frame rate. Here, video tracking systems usually employ a motion model which describes how the image of the target might change for different possible motions of the object to track.

Examples of simple motion models are

- To track planar objects, the motion model is a 2D transformation (affine transformation or homograph) of an image of the object.
- When the target is a rigid 3D object, the motion model defines its aspect depending on its 3D position and orientation.
- For video compression, key frames are divided into macro blocks. The motion model is a disruption of a key frame, where each macro block is translated by a motion vector given by the motion parameters.
- The image of deformable objects can be covered with a mesh; the motion of the object is defined by the position of the nodes of the mesh.

The role of the tracking algorithm is to analyze the video frames in order to estimate the motion parameters. These parameters characterize the location of the target.

Conventional approaches: There are two major components of a visual tracking system; Target Representation and Localization (TRL) and

Filtering and Data Association. TRL is mostly a bottom-up process. Typically the computational complexity for these algorithms is low. The following are some common Target Representation and Localization algorithms:

- **Blob tracking**: Segmentation of object interior (blob detection, block-based correlation)
- **Kernel-based tracking** (Mean-shift tracking): An iterative localization procedure based on the maximization of a similarity measure (Bhattacharyya coefficient).
- **Contour tracking**: Detection of object boundary.

Filtering and Data Association is mostly a top-down process, which involves incorporating prior information about the scene or object, dealing with object dynamics, and evaluation of different hypotheses. The computational complexity for these algorithms is usually much higher. The following are some common Filtering and Data Association algorithms;

- **Kalman filter**: An optimal recursive Bayesian filter for linear functions subjected to Gaussian noise.

- **Particle filter:** (Wen-Yan Chang 2008) Useful for sampling the underlying state-space distribution of non-linear and non-Gaussian processes.

Object tracking is an important task within the field of computer vision. The proliferation of high-powered computers and the increasing need for automated video analysis has generated a great deal of interest in object tracking algorithms. In general it is used mostly in surveillance systems. There are three key steps in video surveillance: Seol et al (2003) has detection of interesting moving object tracking of such objects from frame to frame, and analysis of object tracks to recognize their behavior. In this report we present a feature based object tracker which uses a Pan-Tilt (PT) camera to keep track of the target. Generally, the task is to keep the target at the center of the grabbed image. As the target starts moving in the real world, its position in the grabbed image is reported in subsequent frames through a feature based tracking algorithm, based on the work of Lukas, kanade and Tomasi. The image position error is processed by a proportional-integral controller and the camera is re-positioned accordingly to place the target in the pre-specified image region.

Several vehicle tracking and detecting methods have been proposed in the literature. The main, non-mutually exclusive categories identified are: motion based tracking region-based tracking, active contour-based tracking, feature-based tracking, model-based tracking and body part-based tracking. (Ashwani Aggarwal 2006) Uses background subtraction and motion estimation for tracking an object. DCT domain background subtraction in Y plane is used to locate candidate objects in subsequent I-frames after a user has marked an object of interest in the given frame. DCT domain histogram matching using Cb and Cr planes and motion vectors are used to select the target object from the set of candidate objects.

Karthik Hariharakrishnan (2005) proposes a tracking and detecting algorithm that predicts the object by predicting the object boundary using block motion vectors followed by updating the contour using occlusions/disclosing detection. Another group of algorithms deals with object tracking using adaptive particle filters in Xi Tao et al (2010); Zhuohua Duan et al (2009); kalman filter, Mehta et al (2010) and Xiaoqin Zhang et al (2008).

Mandal et al (1999) provides a critical review of the compressed domain indexing techniques. The method in Khansari et al (2008) proposes a tracking mechanism based on pixel features in the wavelet domain. Two-stage object tracking is performed by combining region-based method and contour-based method in Qiang Chen et al (2010).

The proposed scheme for vehicle objects tracking and detecting mainly deals with video surveillance from night vision applications where the camera is assumed to be fixed with a fairly wide angle of view. The object is tracked by our algorithm in subsequent frames till the object disappears from the field of view of the camera. It should be noted that, although the background is considered to be fixed. In this source it is robust in the presence of variations in the vehicle lighting conditions and extraneous movements.

Zhang et al (2003) a method which allows for the first time to perform automatically the detection and the tracking of microscopic objects directly from three dimensional image data is presented. Image multi-scale edge detection based on anti-symmetrical biorthogonal wavelet is given in detail by Zhang Chuang et al (2009).

Coded stepped-FMCW signal for mm-wave automotive collision warning radar (FLAR or AICC systems) an application is proposed in Zhang Jian-Hui et al (2001).Zhang et al (2010) is propose a fast drift-free solution based on feature point tracking using reference frames of high confidence over the temporal and spatial domains. Zhang Tao et al (2010) algorithm will preferably mitigate the large amount of calculation and solve shading interference existed in traditional image difference algorithm. Zhang Tao et al

(2010) novel method of moving target identification based on multi-features fusion is proposed.

The background subtraction detects object and approaches object contour with a rectangle which is also an object's tracking box. Furthermore, Camshift algorithm is used to make up the shortage of the background subtraction method in Zhang Qigui et al (2010).Zhang Ya et al (2010) in order to resolve the occlusion and temporal lost problem, a robust data association + filtering procedure is proposed. Zhao Ding-xuan et al (2010) method takes advantage of image analysis technology to detect the profile of specified moving object and compute its center. According to the center of moving object in the several previous frame images, predict the possible center in the next frame. Zhen Liang et al (2010) an approach to refine a region based attention model with eye tracking data. They are three main contributions. (1) A concept of fixation mask is proposed to describe the region saliency of an image by weighting the segmented regions using importance measures obtained in the Human Visual System (HVS) or computational models. (2) A Genetic Algorithm (GA) scheme for refining a region based attention model is proposed. (3) An evaluation method is developed to measure the correlation between the result from the computational model and that from the HVS in terms of fixation mask. A real time object tracking system based on multi-view cameras. In order to track a moving object, an active camera mounted on a pan/tilt platform controlled by static cameras is used in this system by Zhengya Xu et al (2009).

The weight parameter indicates the similarity between background feature and object feature and can reduce localization error of object tracking in Zhiqiang Wen et al (2007). Zhi-Qiang Wen et al (2006) Mean shift algorithm is recently widely used in tracking clustering, etc, however convergence of mean shift algorithm has not been rigorously proved. Zhiyuan Qiao et al (2011) proposed a rapid visualization method of GIS vector data on 3D terrain is proposed in this paper, including the organization of 3D vector data based on a quad-tree structure, the way to spatialize 2D vector data with terrain matching and methods to render a large amount of vector objects in a 3D scene. Zhou Hao et al (2010) Particle Swarm Optimization (PSO) was embedded into generic particle filter framework to achieve more robustness and flexibility. Zhou Xian cheng (et al (2008) solve the problem of poor anti-noise performance of the traditional Fuzzy C-Means (FCM) algorithm in image segmentation, a novel two-dimensional FCM clustering algorithm for image segmentation was proposed. Zhuohua Duan et al (2009) Particle filtering has proven very successful for non-linear and non-Gaussian estimation problems. The paper presents a method for occlusion detection and recovery for object tracking with adaptive particle filter. Firstly, object occlusion is detected with normalization factor. Secondly, adaptive transition function is employed to recovery from occlusion. Lastly, particle number is changed according to occlusion state.

A novel algorithm was proposed by Zhu Hong et al (2011) to eliminate noise correlations, using the subtraction values between the adjacent measurements which avoids higher-order matrix inversing, and with minimum hardware. Zhi et al (2010) has proposes a method of motion capture using joint skeleton tracking and surface estimation from the multi-view video sequence, which can capture the performance of a human or an animal. Zhu et al (2010) has proposed a model discussing the relationship between the objects and environments, and integrates it with Bayesian tracking. Further Particle filtering has been employed for solving the Bayesian tracking problem. Ziyou Xiong et al (2011) proposed algorithm enhancement algorithm identifies obscuration pixels by snow or rain in the near-field view and removes these pixels as snowflakes or rain drops; different from state-of-the-art methods, the algorithm in this step can detect snowflakes on foreground object and background, and choose different methods to fill in the removed regions.

1.3 COMPRESSION ARTIFACT REDUCTION AND MOVING OBJECT TRACKING IN THE IMAGES AND VIDEOS PROBLEM

Compression artifact reduction in the images and videos problem: compression artifact (or **artifact**) is the result of an aggressive data compression scheme applied to an image, audio, or video that discards some data that may be too complex to store in the available data-rate, or may have been incorrectly determined by an algorithm to be of little subjective importance, but is in fact objectionable to the user. Artifacts are often a result of the latent errors inherent in lossy data compression. Technically speaking; a compression artifact is a particular class of data error that is usually the consequence of quantization in lossy data compression. Where transform coding is used, they typically assume the form of one of the basic functions of the coder's transform space.

Compression artifacts occur in many common media such as Digital Versatile Disc (DVD), common computer file formats such as JPEG, MP3, or MPEG files, and Sony's Adaptive Transform Acoustic Coding (ATRAC) compression algorithm. Uncompressed media (such as on Laserdiscs, Audio Compact Disk (CD), and WAV files) or lossless compressed media (Free Lossless Audio Codec (FLAC), Portable Network Graphics (PNG), etc.) do not suffer from compression artifacts.

Artifacts (or **artifact**) are the result of an aggressive data compression scheme applied to an image, audio, or video that discards some data that may be too complex to store in the available data-rate, or may have been incorrectly determined by an algorithm to be of little subjective importance, but is in fact objectionable to the user. Artifacts are often a result of the latent errors inherent in lossy data compression. Technically speaking; a compression artifact is a particular class of data error that is usually the consequence of quantization in lossy data compression. Where transform coding is used, they typically assume the form of one of the basic functions of the coder's transform space.

Moving object segmentation in the images and videos problem: The task of video-object segmentation is to identify and separate the important objects in a video scene from the scene background. Clearly, to approach this problem, it is necessary to define what is exactly meant with important objects and how the correct object masks should look like. However, the practices that even an unambiguous definition of video objects are a fundamental problem. In the following, the involved definition problems are addressed and grouped into physical problems, being a consequence of the image formation, and semantic problems. The physical problems are as follows

- **Reflections:** The problem of handling reflections is actually similar to object shadows. However, reflections are more difficult, because the appearance of the reflected images depends on the physical properties of the reflecting surface and because the reflection is not necessarily attached to the object.

- **Occlusions:** The object shape can also change because of occlusions. It depends on the application whether the masks of occluded objects should be extended to their original shape.

- **Translucent objects:** Objects can appear partially translucent since they are made of translucent materials, or because thin structures like hair or cloth appear translucent. Moreover, pixels along object boundaries are always a mixture of foreground color and background color. To model the translucency, the segmentation algorithm has to compute an alpha-channel mask which identifies the translucency factor for each pixel instead of only computing a binary object mask. Accurate alpha-channel information cannot be obtained from a single image, but algorithms using heuristic approaches have been proposed.

(a) (b) (c)

Figure 1.3 (a) Input image (b) Background image (c) Segmented foreground

A part from the physical problems, there are semantic definition problems, like the following.

- **Objects of interest (foreground objects):** The first and obvious question of video segmentation is what parts of an image constitute the foreground objects as shown in the figure 1.3(c). This issue is already surprisingly difficult, since the intuitive human understanding of foreground objects is strongly depending on the scene context. Mostly, human intuition expects that this should be the main acting objects. For example, in sports broadcast, the players are usually considered foreground and the audience is considered background, even if the audience is moving. This distinction is on a very high semantic level, since it assumes knowledge about the meaning of the scene. A surveillance system in a sports stadium will be interested in other objects than a system for automatic analysis of the sports game.

- **Small background movements:** When taking a more detailed view on the last point, it can be observed that the distinction between foreground and background is in fact gradual as shown in figure 1.3 (c) and (b). The question is to what extent a background should change such that it is considered part of the foreground. For example, trees may occur in the background with leaves moving slightly in the wind, or there may be a clock on a wall at the back of the room.

- **Object-status change:** Objects can also change their classification over time. For example, most people would consider a car that drives along a street as an important object. But how to define the object status when the car stops and parks at the side of the street? Alternatively, the opposite case may occur that a car that was parked for a long time suddenly drives away. Note that it is practically impossible to separate all objects, including the static ones, into independent objects, since this would imply that all future actions would have to be predicted.

- **Multi-body objects:** Objects may be separated into several nonconnected regions in the image. One reason for this can be that an occluding object cuts the foreground object into pieces. Another complex example are objects that are really composed of several parts but still belonging together like flocking birds.

- **Hierarchical objects:** Additional to multi-body objects, there can also exist a hierarchical relationship between objects. One example is a car object that contains a driver object. When considering all of these problems simultaneously, it can only be concluded that a general-purpose segmentation of video objects is virtually impossible, since the definition of the expected output from the algorithm depends largely on the scene context and the application that we have in mind. However, despite all the mentioned problems, it is still possible to design algorithms that cover a multitude of specific applications and that work well in many practical cases.

1.4 PROPOSED APPROACH

Image segmentation and Pattern matching: The frames extracted from the video are segmented first, features of each object in the segmented image are extracted, pattern matching is done on the consecutive frames having the desired features in the hand, the motion vectors are calculated and mask is moved accordingly by Takashi Morimoto et al (2005).

Block Diagram

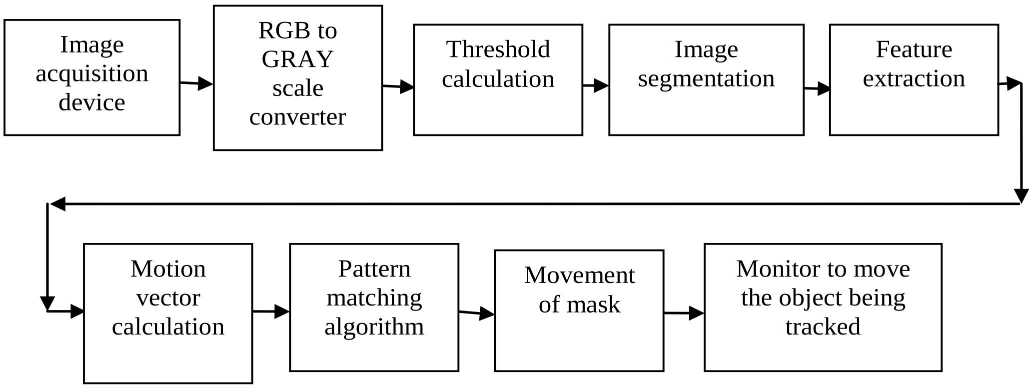


Figure 1.4 Block diagram of proposed method of object tracking

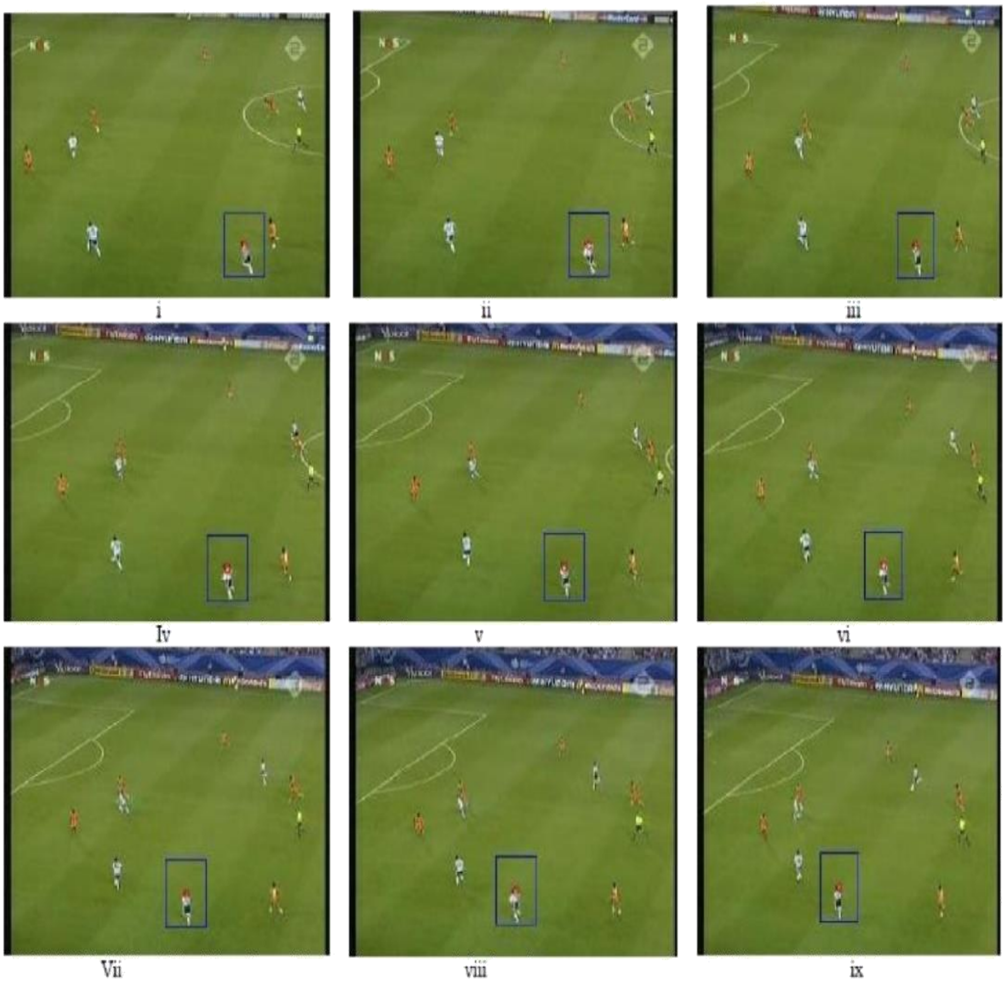


Figure 1.5 Illustration of moving object tracking in video sequences

1.5 ORGANIZATION OF THESIS

Chapter 1, discuss about the applications of compression artifacts occurs in many common media such as Digital Versatile Discs (DVDs), common computer file formats such as JPEG, MP3, or MPEG files, and Sony's Adaptive Transform Acoustic Coding (ATRAC) compression algorithm. Uncompressed media (such as Laserdiscs, Audio Compact Disks (CDs), and WAV files) or lossless compressed media (Free Lossless Audio Codec (FLAC) and Portable Network Graphics (PNG) etc.) do not suffer from compression artifacts. Removal of compression artifacts in images and video by using adaptive based algorithm which improves Peak Signal to Noise Ratio (PSNR) and Mean Square Error (MSE) compare to the other techniques as mentioned in the literature. And also Moving Object tracking, where it is used in real time environment, significance of object tracking in various fields like computer vision, robotics, artificial intelligent systems, surveillance systems, navigation systems, traffic monitoring systems in heavy crowded urban areas and it is also discussed about the conventional approaches for object tracking and proposed algorithm. The object tracking is shown in the figure 1.4.

Chapter 2, mainly concentrated on compression artifact images and video basic and several types of images and video artifacts techniques invented so far are briefly explained. And basic concepts in filter techniques, image and video format invented so far are briefly explained, the important aspects regarding the statistical measurement is also discussed.

Chapter 3, mainly concentrated on algorithm implementation in image sequence, algorithm steps for flow chat, wavelet compression, edge detection, low pass filter, adaptive median filter are briefly explained and the important aspects regarding the algorithm implemented in image simulation results is also discussed.

Chapter 4, mainly concentrated on algorithm implementation in video sequence, algorithm steps for flowchart, video frames, color processing, motion estimation and motion compensation, edge detection, low pass filter, adaptive median filter, temporal median filter are briefly explained and the important aspects regarding the algorithm implemented in video simulation results is also discussed.

In Chapter 5, the applications of moving object tracking, where it is used in real time environment significance of object tracking in various fields like computer vision, robotics, artificial intelligent systems, surveillance systems, navigation systems, traffic monitoring systems in heavy crowded urban areas and it is also discussed about the conventional approaches for object tracking and proposed algorithm.

All Simulation results are explained in Chapter 6, Compression artifacts reduction and moving object tracked images and videos sequences are shown, here different type of methods are used to remove artifacts and tracking the object and get the output result is given below.

1. Simulation Results for the proposed compression artifacts reduction image.
2. Simulation Results for the proposed compression artifacts reduction video.
3. Simulation Results for the proposed moving object tracking video

Conclusion and future recommendation are presented in Chapter 7.

CHAPTER 2

COMPRESSION ARTIFACT REDUCTION IN THE IMAGES AND VIDEOS

2.1 INTRODUCTION

A compression artifact (or artifact) is the result of an aggressive data compression scheme applied to an image, audio, or video that discards some data that may be too complex to store in the available data-rate, or may have been incorrectly determined by an algorithm to be of little subjective importance, but is in fact objectionable to the user. Artifacts are often a result of the latent errors inherent in lossy data compression. Technically speaking a compression artifact is a particular class of data error that is usually the consequence of quantization in lossy data compression. Where transform coding is used they typically assume the form of one of the basic functions of the coder's transform space.

Compression artifacts occur in many common media such as DVDs, common computer file formats such as JPEG, MP3, or MPEG files, and Sony's ATRAC compression algorithm. Uncompressed media (such as on Laserdiscs, Audio CDs, and WAV files) or lossless compressed media (FLAC and PNG, etc.) do not suffer from compression artifacts.

Compression is the process of representing information in a compact form. The goal of any compression technique is to reduce the bit rate for transmission and storage while maintaining the acceptable fidelity or data quality. Due to the increasing traffic caused by multimedia information and digitized form of representation of images and videos, image and video compression have become a necessity. Artifacts are defined as unwanted visible effects in the picture caused by disturbances and errors in the image, video compression, transmission or digital processing. An artifact arises due to inadequate acquisition, inadequate processing, and transmission error.

Nowadays transform-based compression is very popular in the still images and video like JPEG and MPEG. But these compressions are lossy compressions and in addition they are based on the method which divides each image into the blocks. In spite of these defect, usual decoder just perform inverse transform. Compressed video may possess a number of artifacts, both spatial and temporal. Spatial compression artifacts arise as a result of quantization of the transform-domain coefficients, and are often manifested as blocking and ringing artifacts. Temporal limitations in compressed video occur when the encoder, in an effort to reduce bandwidth, drops frames. The wavelet based compressions produce ringing and blur artifact.

A study of artifacts that creates major problem in images and video, and try to minimize the noise by using an non linear adaptive based algorithm. Generally noise removal can be achieved by using the linear filtering techniques like low pass; high pass filtering technique, but these filtering techniques introduce blurring effect in the picture. This give rise to the use of non-linear image processing techniques. The best-known and most widely used non-linear digital filters based on order statistics are Adaptive median filter.

2.2 TYPES OF ARTIFACT

Types of compression artifacts in images and video are as follows,

- Blocking Artifact.
- Ringing artifact.
- Blurring Artifact.
- Mosquito noise.

2.2.1 Blocking Artifact

Blocking in an image is caused by the quantization or truncation of the high frequency transform coefficients resulting from DCT-or-Wavelet-based coding. In spatial domain pixels across the block boundary are encoded with different set of basic functions. If the basis function overlaps between adjacent blocks, the blocking effect may be alleviated.

Figure 2.1 Blocking Artifact Image

2.2.2 Ringing Artifact

Ringing in images is caused by the quantization or truncation of the high frequency transform coefficients resulting from DCT-or-Wavelet-based coding. In spatial domain this causes ripples or oscillations around sharp edges or contours in the images and video. This is also known as Gibbs phenomenon.

Figure 2.2 Ringing Artifact Image

2.2.3 Blurring Artifact

Blur in an image is due to the attenuation of the high spatial frequencies. This commonly occurs during filtering or visual data compression.

Figure 2.3 Blurring Artifact Image

2.2.4 Mosquito Noise

Mosquito noise is most apparent around artificial or CG (Computer Generated) objects or scrolling credits (lettering) on a plain coloured background. It appears as some haziness and/or shimmering around high-frequency content (sharp transitions between foreground entities and the background or hard edges) and can sometimes be mistaken for ringing unfortunately; this peppered effect is also visible around more natural shapes like a human body. The Video Reference Impairment System (VIRIS) defines mosquito noise as follows: "Form of edge busyness distortion sometimes associated with movement, characterized by moving artifacts or blotchy noise patterns superimposed over the objects (resembling mosquito flying around a person's head and shoulders)."Mosquitoes" can also be found in other areas of an image. For instance, the presence of a very distinct texture or film grain at compression will also introduce mosquito noise. The result will be similar to random noise; the mosquitoes will seem to blend with the texture or the film grain and can look like original features of the picture.

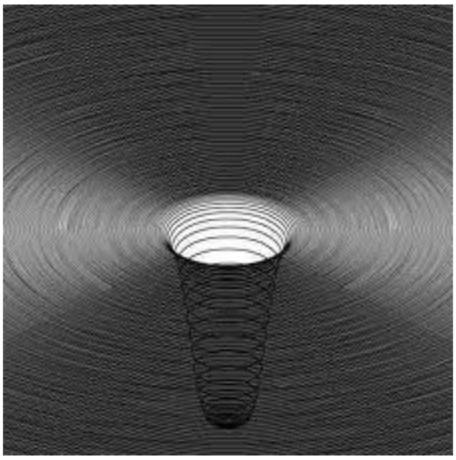


Figure 2.4 Mosquito Noise Image

2.3 FILTERING TECHNIQUES

Filtering is a technique for modifying or enhancing an image. Filter is used to emphasize certain features or remove other features, from an image. Image processing operations implemented with filtering include smoothing, sharpening, and edge enhancement. Filtering is a neighborhood operation, in which the value of any given pixel in the output image is determined by applying some algorithm to the values of the pixels in the neighborhood of the corresponding input pixel. A pixel's neighborhood is some set of pixels, defined by their locations relative to that pixel. There are two main types of filters used in image processing namely

- Linear filter
- Non-linear filter

2.3.1 Linear Filter

Linear filtering is filtering in which the value of an output pixel is a linear combination of the values of the pixels in the input pixel's neighborhood. A linear filter applies a linear operator to a time-varying input signal. Linear filters are very common in electronics and digital signal processing but they can also be found in mechanical engineering and other technologies. They are often used to eliminate unwanted frequencies from an input signal or to select a desired frequency among many others. Regardless of whether they are electronic, electrical, or mechanical, or what frequency ranges or timescales they work on, the mathematical theory of linear filters is universal. Linear filters, such as bandpass, highpass and lowpass.

2.3.2 Non-linear Filter

A variety of smoothing filters have been developed that are not linear. The main difference between linear and non-linear filters is that in linear filters an explicit mask is used, where as in non-linear filters the operation is based directly on the values of the pixels in the neighborhood under consideration. They do not explicitly use coefficients.

Non-linear filters locate and remove data that is recognized as noise. The algorithm is 'non- linear' because it looks at each data point and decides if that data is noise or valid signal. If the point is noise, it is simply removed and replaced by an estimate based on surrounding data points, and parts of the data that are not considered as noise are not modified at all.

2.4 IMAGE FORMATS

2.4.1 JPEG Format

In computing, JPEG is a commonly used method of compression for photographic images. The name JPEG stands for Joint Photographic Experts Group, the name of the committee that created the standard. The group was organized in 1986, issuing a standard in 1992, which was approved in 1994 as ISO 10918-1. JPEG is distinct from MPEG (Moving Picture Experts Group), which produces compression schemes for video. The JPEG standard specifies both the codec, which defines how an image is compressed into a stream of bytes and decompressed back into an image, and the file format used to contain that stream. The compression method is usually lossy compression, where some visual quality is lost in the process and cannot be restored, although there are variations on the standard baseline JPEG that are lossless. There is also an interlaced "Progressive JPEG" format, in which data is compressed in multiple passes of progressively higher detail. The large images that will be displayed while downloading over a slow connection, allowing a reasonable preview after receiving only a portion of the data. However, progressive JPEGs are not as widely supported.

2.4.2 BMP Format

The BMP file format, sometimes called bitmap or DIB file format (for device-independent bitmap), is an image file format used to store bitmap digital images, especially on Microsoft Windows and OS/2 operating systems.

Many graphical user interfaces use bitmaps in their built-in graphics subsystems; for example, the Microsoft Windows and OS/2 platforms' GDI subsystem, where the specific format used is the Windows and OS/2 bitmap file format, usually named with the file extension of .BMP or .DIB.

2.4.3 TIFF Format

Tagged Image File Format (abbreviated TIFF) is a file format for storing images, including photographs and line art. It is now under the control of Adobe. Originally created by the company Aldus for use with what was then called "desktop publishing", the TIFF format is widely supported by image-manipulation applications, by publishing and page layout applications, by scanning, faxing, word processing, optical character recognition and other applications. Adobe Systems, which acquired Aldus, now holds the copyright to the TIFF specification.

Since 1992, though several Aldus/Adobe technical notes have been published with minor extensions to the format, and several specifications, including TIFF/EP, have been based on the TIFF 6.0 specification.

2.4.4 PNG Format

Portable Network Graphics (PNG) is an extensible file format for the lossless, portable, well-compressed storage of raster images. PNG provides a patent-free replacement for GIF (Graphics Interchange Format) and can replace many common uses of RGB indexed and true color TIFF

(Tagged Image File Format). Indexed-color, grayscale and true color images are supported, plus an optical alpha channel for transparency information. Sample depths range from 1 to 16 bits. The PNG International Standard is designed to work well in online viewing applications, such as the World Wide Web (WWW) and is based on a specification developed by the World Wide Web Consortium (WWC).

2.4.5 GIF Format

CompuServe Information's Services introduced the Graphic Interchange Format (GIF) image format. GIFs were designed to be a platform independent format designed for slow modem speed transfers. GIFs are 8-bit compressed images that can be displayed without any loss of original graphics information. This process is known as losses compression. GIF files are limited to a palette of 256 colors and are considered one of the standard image formats found on the internet. A GIF file is encoded in binary. Binary files require the use of a software program to interpret their content. Transparent GIFs allow a background image color to be designated "see through" allowing the illusion of images that float over a web page's background color, texture, or graphics. Transparent GIFs may also be used to represent type fonts and blank spaces. GIF files can be in Gluten Intolerance Group (GIG) 87a format. The latter permits creation of an interleaved or animated image.

2.5 VIDEO FORMATS

2.5.1 AVI Format

'AVI' stands for "Audio Video Interleaved". AVI is one of the oldest formats in the computer world that was created by Microsoft.AVI, the proprietary format of Microsoft's "Video for windows" application, merely provides a framework for various compression algorithms. The first version of AVI supported a maximum resolution of 160 x 120 pixels with a refresh rate of 15 frames per second. The format attained widespread popularity, as the first video editing systems. However there were a number of restrictions leading from format incompatibility, because each manufacturer adapted the open AVI format according to own requirements. Despite its age and numerous problems, the AVI format is still used in semi-professional video editing cards. Many TV cards and graphic boards with a video input also use the AVI format. These are able to grab video clips at low resolutions (mostly 320 x 240 pixels.).File extension: avi.

2.5.2 MPEG Format

The abbreviation MPEG stands for "Motion Picture Experts Group". The MPEG formats are by far the most popular standards. In order to attain widespread use, the MPEG standard only specifies a data model for the compression of moving pictures and for audio signals. It means that it is not a definition of coding standard, and assumed to be only constraint on the data bit stream. In the MPEG remains platform independent. And various MPEG formats are:

1. **MPEG-1**: for most home user applications and business applications (image videos, documentation), the quality offered by MPEG-1 is adequate. It is file extensions are: mpg, mpeg.
2. **MPEG-2**: due to its high data rate compared with MPEG-1 and the increased requirement for memory space, MPEG-2 is currently only suitable for playback in home user filed. File extensions: m2v, mpeg2.
3. **MPEG-4**: something completely new in MPEG-4 is the organization of image contents into independent objects in order to be able to address or process them individually. It is

used for transmission over the Internet for example. Some manufactures transmit moving images to mobile phones. MPEG-4 is used to form a platform for this kind of data transfer.

4. **MPEG-7**: it is a standard to describe multimedia data and can be used independently of other MPEG standards. It aims at offering a comprehensive set of audio-visual descriptions, which will form the basic for applications enabling the needed quality access to content, which implies good storage solutions, high performance content identification, search and retrieval.

5. **MPEG-21**: The development of this format standard recently. It should be a standard that aims at creating a Multimedia Framework taking into consideration the different components involved in the delivery of content from creator to the user.

2.6 COMPRESSION TECHNIQUES

2.6.1 Lossless Compression Technique

Lossless compression methods may be categorized according to the type of data they are designed to compress. Some main types of targets for compression algorithms are text, executables, images, and sound. Whilst, in principle, any general-purpose lossless compression algorithm (general-purpose means that they can handle all binary input) can be used on any type of data, many are unable to achieve significant compression on data that is not of the form that they are designed to deal with. Sound data, for instance, cannot be compressed well with conventional text compression algorithms.

Most lossless compression programs use two different kinds of algorithms: one which generates a statistical model for the input data, and another which maps the input data to bit strings using this model in such a way that "probable" (e.g. frequently encountered) data will produce shorter output than "improbable" data. Often, only the former algorithm is named, while the second is implied (through common use, standardization ect.) or unspecified.

Statistical modeling algorithms for text (or text-like binary data such as executables) include:

- Burrows-Wheeler transform (BWT; block sorting preprocessing that makes compression more efficient)
- LZ77 (used by deflate)
- Lempel-Ziv-Welch (LZW)
- Prediction by Partial Matching (PPM)
- Lossless JPEG(JPEG-LS)
- Set Partitioning In Hierarchical Trees (SPIHT)
- Electronic Data Processing (EDP)
- Professional Photographers of Massachusetts (PPAM)

Encoding algorithms to produce bit sequences are:

- Huffman coding (also used by Deflate)
- Arithmetic coding

Many of these methods are implemented in open-source and proprietary tools, particularly LZW and its variants. Some algorithms are patented in the USA and other countries and their legal usage requires licensing by the patent holder. Because of patents on certain kinds of LZW compression, some open source activists encouraged people to avoid using the Graphics Interchange Format (GIF) for compressing image files in favor of Portable Network Graphics PNG, which combines the LZ77-based deflate algorithm with a selection of domain-specific prediction filters. However, the patents on LZW have now expired.

Many of the lossless compression techniques used for text also work reasonably well for indexed images, but there are other techniques that do not work for typical text that are useful for some images (particularly simple bitmaps), and other techniques that take advantage of the specific characteristics of images (such as the common phenomenon of contiguous 2-D areas of similar tones, and the fact that color images usually have a preponderance to a limited range of colors out of those representable in the color space).

2.6.2 Lossy compression Technique

A lossy data compression method is one where compressing data and then decompressing it retrieves data that may well be different from the original, but is "close enough" to be useful in some way. Lossy data compression is used frequently on the Internet and especially in streaming media and telephony applications. These methods are typically referred to as codecs in this context. Most lossy data compression formats suffer from generation loss: repeatedly compressing and decompressing the file will cause it to progressively lose quality. This is in contrast with lossless data compression.

Types of lossy compression

There are two basic lossy compression schemes:

- In lossy transform codecs, samples of picture or sound are taken, chopped into small segments, transformed into a new basis space, and quantized. The resulting quantized values are then entropy coded.

- In lossy predictive codecs, previous and/or subsequent decoded data is used to predict the current sound sample or image frame. The error between the predicted data and the real data, together with any extra information needed to reproduce the prediction, is then quantized and coded.

2.7 STATISTICAL MEASUREMENT

All parameters used for the performance of compression artifact reduction in the images and videos are evaluated by Mean Square Error (MSE) and Peak Signal to Noise Ratio (PSNR) methods.

2.7.1 Mean Square Error (MSE)

MSE indicates average error of the pixels throughout the image. A definition of a higher MSE does not indicate that the denoised image suffers more errors instead it refers to a greater difference between the original and denoised image. That means significant artifact reduction. The formula for the MSE calculation is given by equation (2.1).

$$MSE = \frac{1}{N}\sum_{j=0}^{N-1} (X_j - X_j^*)^2 \tag{2.1}$$

where X_j is the reconstructed image

X^*_j is the original image

N’ is the size of the image

2.7.2 Peak Signal-to-Noise Ratio (PSNR)

Video quality can be measured with formal metrics like PSNR or with subjective video quality using expert observation. The phase peak signal-to-noise ratio, often abbreviated PSNR, is an engineering term for the ratio between the maximum possible power of a signal and the power of corrupting noise that affects the fidelity of its representation. Because many signals have a very wide dynamic range, PSNR is usually expressed in terms of the logarithmic decibel scale. The PSNR is most commonly used as a measure of quality of reconstruction in image compression etc. It is most easily defined via the mean squared error (MSE) which for two m×n monochrome images I and K where one of the images is considered a noisy approximation of the other is defined as;

$$MSE = \frac{1}{mm}\sum_{j=0}^{m-1}\sum_{j=0}^{n-1}\| I(i,j) - K(i,j) \|^2 \qquad (2.2)$$

The PSNR is define

$$PSNR = 10\log_{10}\left(\frac{MAX_i^2}{MSE}\right) = 20\log_{10}\left(\frac{MAX}{\sqrt{MSE}}\right) \qquad (2.3)$$

Here, MAXi is the maximum possible pixel value of the image. When the pixels are represented using 8 bits per sample. Generally, when samples are represented using linear Pulse Code Modulation (PCM) with B bits per sample, MAXi is 2^B-1(where ^ stands for the power function).For color images with three RGB values per pixel, the definition of PSNR is the same except the MSE is the sum over all squared value differences divided by image size and by three.

CHAPTER 3

COMPRESSION ARTIFACT REDUCTION IN THE IMAGES

3.1 INTRODUCTION

An image is an array, or a matrix, of square pixels (picture elements) arranged in columns and rows.

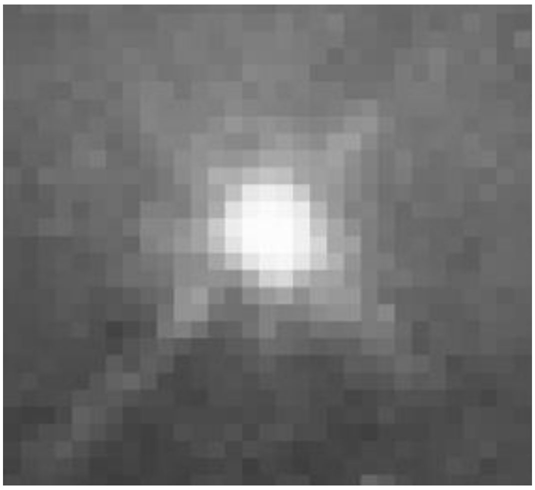


Figure 3.1 Grayscale Image (8-bit)

An image an array or a matrix of pixels arranged in columns and rows. In a (8-bit) grayscale image each picture element has an assigned intensity that ranges from 0 to 255. A grayscale image is what people normally call a black and white image, but the name emphasizes that such an image will also include many shades of grey as shown in the Figure 3.1.

Each pixel has a value from 0 (black) to 255 (white). The possible range of the pixel values depend on the color depth of the image as shown in the Figure 3.2, the 8 bit = 256 tones or grayscales. A normal grayscale image has 8 bit color depth = 256 grayscales. A “true color” image has 24 bit color depth = 8 x 8 x 8 bits = 256 x 256 x 256 colors = 16 million colors.

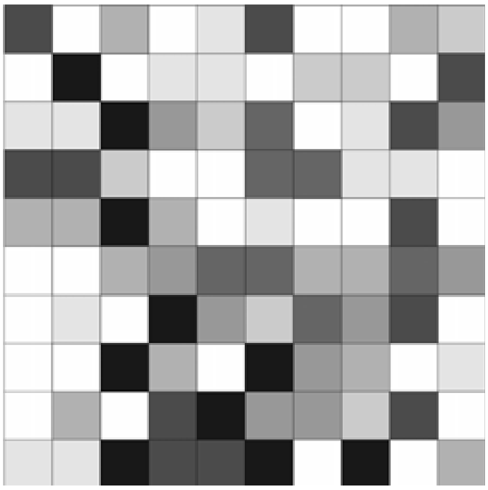


Figure 3.2 Grayscale Image (8-bit each pixel has a value from 0 (black) to 255 (white))

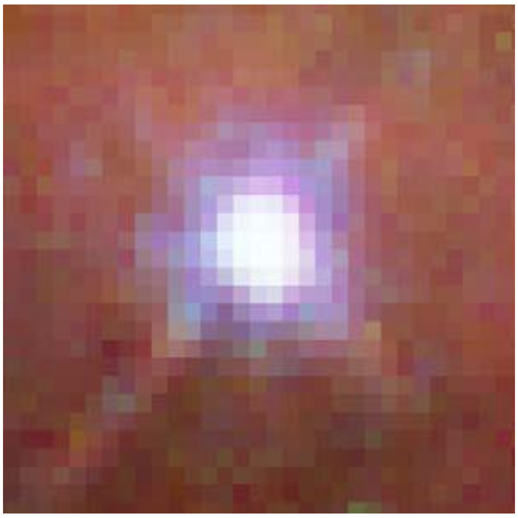


Figure 3.3 True Color Image (Assembled from three grayscale Images colored red, green and blue)

A true-color image assembled from three grayscale images colored red, green and blue as shown in the figure 3.3. Such an image may contain up to 16 million different colors. Some grayscale images have more grayscales, for instance 16 bit = 65536 grayscales. In principle three grayscale images can be combined to form an image with 281,474,976,710,656 grayscales.

1. There are two general groups of ‘images’: vector graphics (or line art) and bitmaps (pixel-based or ‘images’). Some of the most common file formats are
2. **GIF:** An 8-bit (256 color), non-destructively compressed bitmap format. Mostly used, and several sub-standards one of which is the animated GIF.
3. **JPEG:** A very efficient (much information per byte) and destructively compressed 24 bit (16 million colors) bitmap format. Widely used, especially for web and Internet (bandwidth-limited).
4. **TIFF:** The standard 24 bit publication bitmap format. Compresses non-destructively with, for instance, Lempel-Ziv-Welch (LZW) compression.
5. **PS**: Postscript, a standard vector format and numerous sub-standards and can be difficult to transport across platforms and operating systems.
6. **PSD:** A dedicated Photoshop format that keeps all the information in an image including all the layers.

3.2 FLOWCHART OF THE PROPOSED ALGORITHM

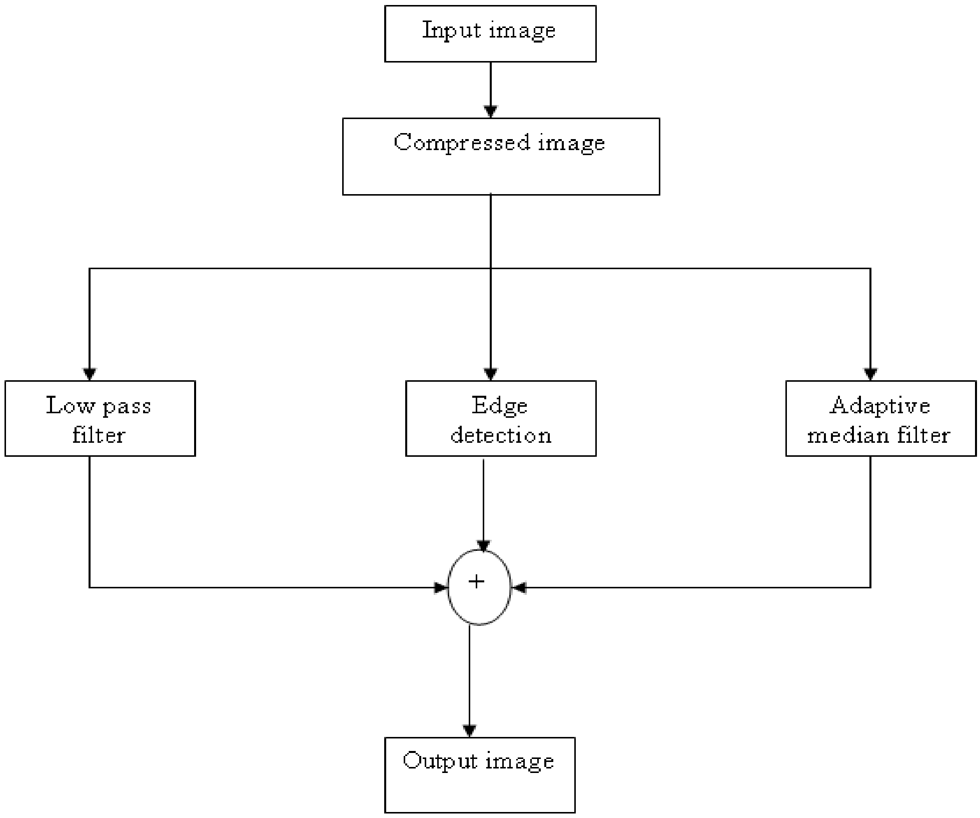


Figure 3.4 Flowchart of the proposed algorithm

3.3 COMPRESSION ARTIFACT REDUCTION IN THE IMAGES ALGORITHM

- **Input image:** The entire process, the original artifact image.
- **Compressed image:** The original artifacts image is compressed.
- **Edge detection:** Edge detection is performed using Sobel edge detector.
- **Low pass filter:** Low pass filter is performed for image smoothening.

- **Adaptive median filter:** The output pixel of adaptive median filter is better when compared to the output pixel of other filter. It is used to replace the corrupted pixel.
- **Output image:** After completing the entire process, the processed image is finally converted back into original image.

3.4 WAVELET BASED COMPRESSION

Wavelets are functions defined over a finite interval and having an average value of zero. The basic idea of the wavelet transform is to represent any arbitrary function as a superposition of a set of such wavelets or basis functions. These basis functions or baby wavelets are obtained from a single prototype wavelet called the mother wavelet, by dilations or contractions (scaling) and translations (shifts). The Discrete Wavelet Transform of a finite length signal x(n) having N components, for example, is expressed by an N x N matrix.

The wavelet transform is defined as follows. The transformed signal is a function of two variables ‘τ’ and ‘s’, the translation and scale parameters, respectively. ‘τ’ is the transforming function, and called as mother wavelet.

The wavelet transform is formally written as:

$$\gamma(s,\tau) = \int f(t)\ \psi^{*}s,\tau\ (t)dt \tag{3.1}$$

where $*$ = Denotes complex conjugation

$f(t)$ = Function

$\psi^{*}s,\tau$ = Set of Basis Function

s = Scale

τ = Translation

The advantages of JPEG compression schemes based on DCT namely simplicity, satisfactory performance, and availability of special purpose hardware for implementation. Since the input image needs to be “blocked," correlation across the block boundaries is not eliminated. This results in noticeable and annoying “blocking artifacts" particularly at low bit rates. The wavelet transform attempt to solve this problem by using smoothly overlapping blocks. Although blocking effects are reduced in wavelet transform compressed images, increased computational complexity of such algorithms do not justify wide replacement of DCT by WT. Wavelet transform is an efficient approach to reduce spatial redundancies without the annoying blocking artifacts at low bit rates.

Figure 3.5 Original image

Figure 3.6 Wavelet compressed image

Figure 3.7 Artifact image

Over the past several years, the wavelet transform has gained widespread acceptance in signal processing in general and in image compression research in particular. In many applications wavelet-based schemes (also referred as subband coding) outperform other coding schemes like the one based on DCT. Since there is no need to block the input image and its basis functions have variable length, wavelet coding schemes at higher compression avoid blocking artifacts. Wavelet-based coding is more robust under transmission and decoding errors, and also facilitates progressive transmission of images. Wavelet coding schemes are especially suitable for applications where scalability and tolerable degradation are important.

The fundamental concept behind SubBand Coding (SBC) is to split up the frequency band of a signal (image in our case) and then to code each subband using a coder and bit rate accurately matched to the statistics of the band. SBC has been used extensively first in speech coding and later in image coding because of its inherent advantages namely variable bit assignment among the subbands as well as coding error confinement within the subbands. Over the years, there have been many efforts leading to improved and efficient design of filter banks and subband coding techniques. Since 1990, methods very similar and closely related to subband coding have been proposed by various researchers under the name of Wavelet Coding (WC) using filters specifically designed for this purpose.

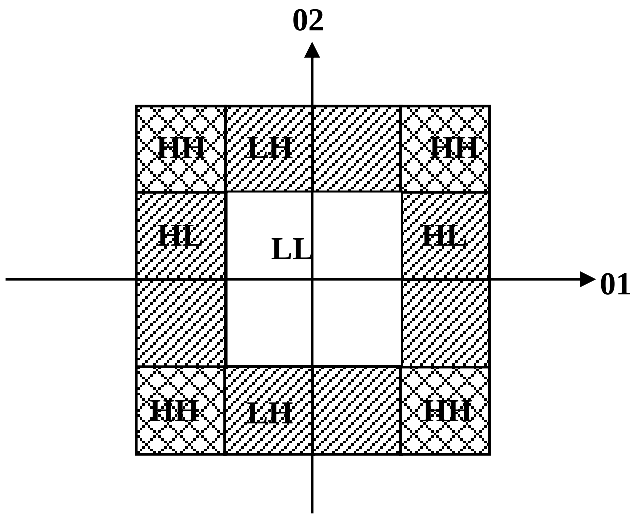


Figure 3.8 Subband Coding (SBC)

Such filters must meet additional and often conflicting requirements. It can be include short impulse response of the analysis filters to preserve the localization of image features as well as to have fast computation, short impulse response of the synthesis filters to prevent spreading of artifacts (ringing around edges) resulting from quantization errors, and linear phase of both types of filters since nonlinear phase introduce unpleasant waveform distortions around edges. The wavelet transform based compression produces the ringing artifacts around edges. My proposed algorithm removes the wavelet based compression artifacts effectively.

Blurring and ringing (edge) artifact calculation:

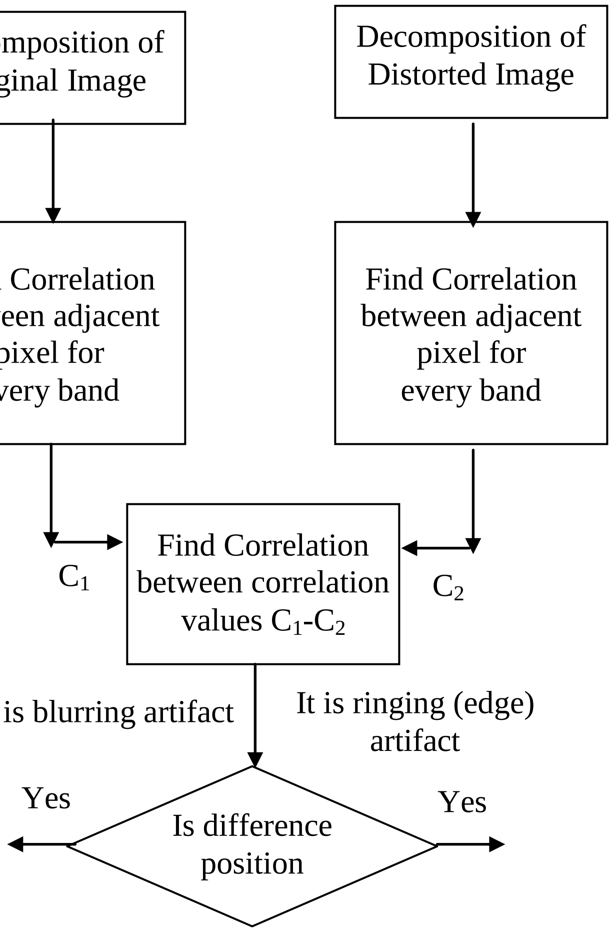


Figure 3.9 Flowchart for blur and ringing artifact measurement

$$I = \frac{1}{N^* \times M^*} \sum_{i=0}^{n-x} \sum_{j=01}^{m-y} I(i, j) \qquad (3.2)$$

where I(i,j) = Wavelet at i^{th} row and j^{th} column

N* = Number of rows

M* = Number of columns in the given image

$$S = \frac{1}{N^* \times M^*} \sum_{i=0}^{n-x} \sum_{j=01}^{m-y} I(i+x, j+y) \qquad (3.3)$$

$x = 0, y = 1$

$$SQI = \sqrt{\sum_{i=0}^{n-x} \sum_{j=01}^{m-y} I(i,j) - I)^2} \qquad (3.4)$$

$$SQS = \sqrt{\sum_{i=0}^{n-x} \sum_{j=01}^{m-y} I(i+x, j+y) - S)^2} \qquad (3.5)$$

Correlation coefficient

$$R_{xx} = \frac{\sum_{i=0}^{n-x} \sum_{j=01}^{m-y} I(i,j) - I)^2 \times \sum_{j=01}^{m-y} (I(i+x, j+y) - S)^2}{SQI \times SQS} \qquad (3.6)$$

If difference, C_1-C_2 is positive it is treated as ringing (edge) artifact otherwise it is blur. Difference C_1-C_2 is found for all bands positive are added together which for all bands positive values are added together which gives total ringing artifact while addition of all negative values gives overall blur.

On account of sensitivity of the eye being different for different spatial frequencies. In this thesis introduce as 2 for resolution level 3, 1.514 for resolution level 2 and 1for resolution level1.

Total ringing (edge) artifact value = 2*(Ringing artifact at resolution level 3)
+1.514(Ringing artifact at resolution level
2)+1(Ringing artifact at resolution level 1)

Total blur artifact value = 2*(Ringing artifact at resolution level
3)+1.514(Ringing artifact at resolution level
2)+1(Ringing artifact at resolution level 1)

3.4.1 Types of Wavelet

There are three types of wavelet such as;

- Haar Wavelet
- Daubechies Wavelet
- Coiflets

3.4.1.1 Haar Wavelet

A Haar wavelet is the simplest type of wavelet. The Haar serves as prototype for all other wavelet transforms.

It can be described as a step function f(x) with

$$f(x) = \begin{cases} 1 & 0 \le r < 1/2, \\ -1 & 1/2 \le r < 1, \\ 0 & \text{otherwise.} \end{cases} \tag{3.7}$$

The 2×2 Haar matrix that is associated with the Haar wavelet is

$$H2 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \tag{3.8}$$

3.4.1.2 Daubechies Wavelet

The Daubechies Wavelet transforms produce averages and differences using just a few more values from the signal. This slight change provides a tremendous improvement in the capabilities of the new transform. The first level Daubechies transform is same as the Haar transform.

3.4.1.3 Coiflets

These wavelets are designed for the purpose of maintaining a close match between the trend values and the original signal values. The coif let wavelets produce a zero fluctuation value whenever a signal is linear over its support.

3.5 EDGE DETECTION

Edge detection is a terminology image processing and computer vision, particularly in the areas of feature detection and feature extraction, The aim of **e**dge detection refers to the process of identifying and locating sharp discontinuities in an image. The discontinuities are abrupt changes in pixel intensity which characterize boundaries of objects in a scene. Classical methods of edge detection involve convolving the image with an operator (a 2-D filter), which is constructed to be sensitive to large gradients in the image while returning values of zero in uniform regions. There is an extremely large number of edge detection operators is available, and some types of edges to be sensitive the detection operator include,

- **Edge orientation:** The geometry of the operator determines a characteristic direction. It is most sensitive to edges. Operators can be optimized to look for horizontal, vertical and diagonal edges.
- **Noise environment:** Edge detection is difficult in noisy images, since both the noise and edges contain high-frequency content. Attempts to reduce the noise result in blurred and distorted edges. Operators used on noisy images are typically larger in scope, so they can be concluded average enough data to discount localized noisy pixels. This results in less accurate localization of the detected edges.

- **Edge structure:** Edges involved a step change in intensity. Effects such as refraction or poor focus the results can be assumed by the objects with boundaries a gradual change in intensity. Newer wavelet-based techniques actually characterize the nature of the transition for each edge in order to distinguish, for example, edges associated with hair from edges associated with a face. There are many ways to perform edge detection. However, the majority of different methods may be grouped into two categories;

- **Gradient:** The gradient method detects the edges by looking for the maximum and minimum in the first derivative of the image.

- **Laplacian:** The Laplacian method searches for zero crossings in the second derivative of the image to find edges. An edge has the one-dimensional shape of a ramp and calculating the derivative of the image can highlight and its location. So the following signal, with an edge shows the jump in intensity;

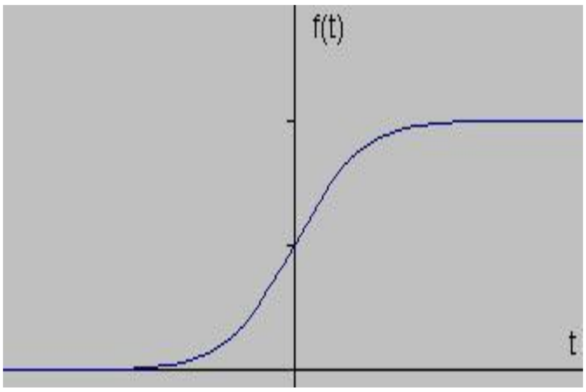


Figure 3.10 Edge intensity

If we take the gradient of this signal (which, in one dimension, is just the first derivative with respect to t) we get the following:

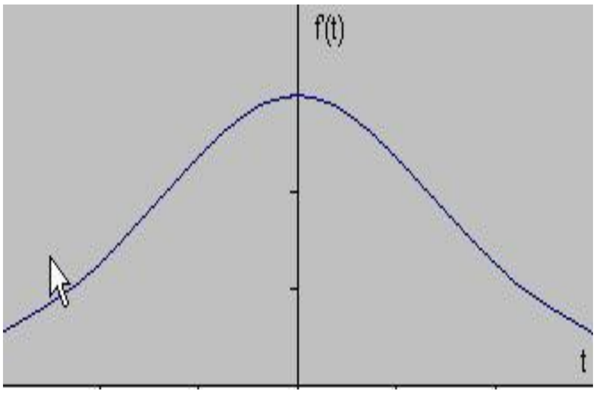


Figure 3.11 Gradient of the signal

Clearly, indicated that the derivative shows a maximum located at the center of the edge in the original signal. This method of locating edge is characteristic of the “gradient filter” family of edge detection filters and includes the sobel method. A pixel location is declared as an edge location if the value of the gradient exceeds some threshold. As mentioned before, edges will have higher pixel intensity values than those surrounding it. So once a threshold is set, you can compare the gradient value to the threshold value and detect an edge whenever the threshold is exceeded. Furthermore, when the first derivative is at a maximum, the second derivative is zero. As a result, another alternative to find the location of an edge is to locate the zeros in the second derivative. These methods are known as laplacian and the second derivative of the signal is shown below;

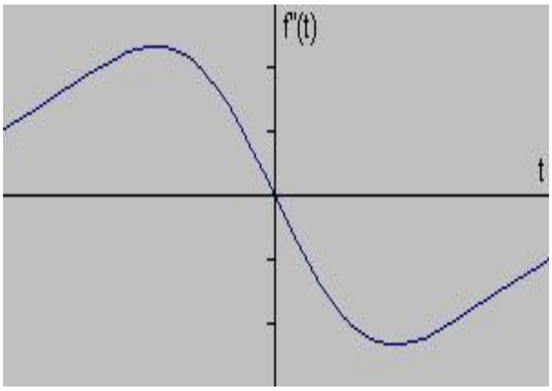


Figure 3.12 Laplacian of the signal

3.5.1 Sobel Operator

The sobel operator is used in our algorithm. Technically, it is a discrete differentiation operator, computing an approximation of the gradient of the image intensity function. At each point in the image, the result of the sobel operator is either the corresponding gradient vector or the norm of this vector. The sobel operator is based on convolving the image with a small, separable and integer valued filter in horizontal and vertical direction. Therefore, relatively inexpensive in terms of computations and the gradient approximation, which it produces, is relatively crude in particular for high frequency variations in the image.

A popular gradient magnitude computation is the sobel operator. Based on the one-dimensional analysis; the theory can be carried over to two-dimensions as long as there is an accurate approximation to calculate the derivative of a two-dimensional image. The sobel operator performs a 2-D spatial gradient measurement on an image. Typically it is used to find the approximate absolute gradient magnitude at each point in an input grayscale image. The sobel edge detector uses a pair of 3x3 convolution masks, one estimating the gradient in the x-direction (columns) and the other estimating the gradient in the y-direction (rows). A convolution mask is usually much smaller than the actual image. Sobel operator can produce thick edges and used for image sharpening. The operator consists of a pair of 3×3 convolution kernels as shown in the figure 3.13. One kernel is simply the other rotated by 90°.

-1	0	+1
-2	0	+2
-1	0	+1

G_x

+1	+2	+1
0	0	0
-1	-2	-1

G_y

Figure 3.13 Sobel Operator

These kernels are designed to respond maximum to the edges running vertically and horizontally relative to the pixel grid, one kernel for each of the two perpendicular orientations. The kernels can be applied separately to the input image, to produce separate measurements of the gradient component in each orientation (call these G_x and G_y). These can be combined together to find out the absolute magnitude of the gradient at each point and the orientation of each gradient. The gradient magnitude is given by;

$$|G| = \sqrt{G_x^2 + G_y^2} \tag{3.9}$$

Typically, an approximate magnitude is computed using

$$|G| = |G_x| + |G_y| \tag{3.10}$$

The angle of orientation of the edge (relative to the pixel grid) giving rise to the spatial gradient is given by

$$\theta = \tan^{-1}\left(\frac{G_y}{G_x}\right) \tag{3.11}$$

Steps in edge detection

(1) Smooth the input image (f (x, y)* = f (x, y) * G(x, y))

(2) $f_x^* = f(x, y)^* \times G_x(x, y)$

(3) $f_y^* = f(x, y)^* \times G_y(x, y)$

(4) magn(x, y) = $| f_x^* | + | f_y^* |$

(5) dir(x, y) = $\tan^{-1}(f_y^* / f_x^*)$

(6) If magn(x, y) > T, then possible edge point.

3.6 LOW PASS FILTER

Low pass filter is a filter that attenuates high frequencies and retains low frequencies unchanged. Smoothing filters in the spatial domain since high frequencies are blocked.

Gaussian filters are important in many signal processing, image processing and communication applications. These filters are characterized by narrow bandwidths, sharp cutoffs, and low overshoots. A key feature of Gaussian filters means Fourier transform of a Gaussian. So the filter has the same response shape in both spatial and frequency domains.

The form of a Gaussian low pass filter in two-dimensions is given by

$$H(u,v) = e^{-D^2(u,v)/2\sigma^2} \tag{3.12}$$

Where D(u,v) is the distance from the origin in the frequency plane (Equation (3.12)) The parameter σ measures the spread or dispersion of the Gaussian curve as shown in the figure 3.14. Larger the value of σ, larger the cutoff frequency and milder filtering section.

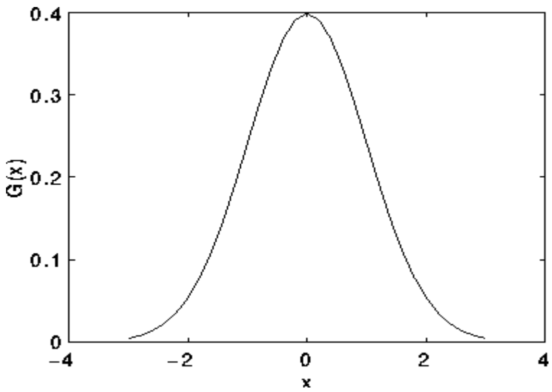


Figure 3.14 1-D Gaussian distribution with mean 0 and σ = 1

When letting σ = r_0, which leads a more familiar form as previous discussion.

$$H(u,v) = e^{-D^2(u,v)/2r_0^2} \tag{3.13}$$

When $D(u, v) = r_0$, the filter is down to 0.607 of its maximum value of 1.

A perspective plot, image display and radial cross section of a GLPF function are shown in the figure 3.15 to 3.17.

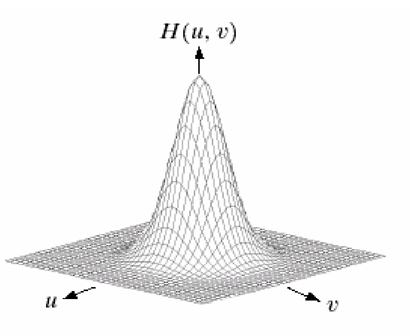


Figure 3.15 Perspective plot of a GLPF transfer function

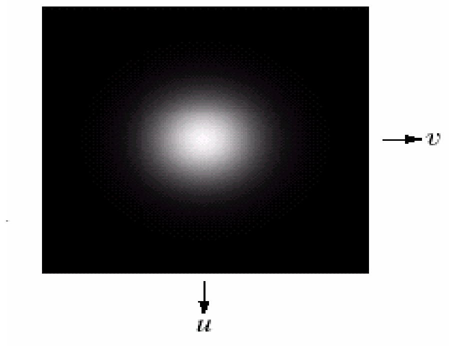


Figure 3.16 Filter displayed as a image

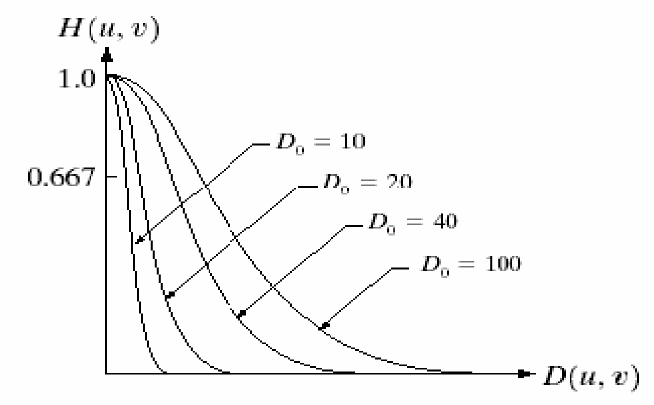


Figure 3.17 Filter radial cross sections for various values of $D_0 = r_0$

Basic steps of filtering in the frequency domain

1 Multiply the input image $f(x, y)$ by $(-1)^{(x+y)}$ to center the transform.

2 Indicated as following equation

$$\tau\,[f(x, y)\,(-1)^{(x+y)}] = F(u - M/2, \upsilon - N/2).$$

3 Compute $F(u, \upsilon)$, the DFT of the input image from (1).

4 Multiply $F(u, \upsilon)$ by a *filter* function $H(u, \upsilon)$.

5 Compute the inverse DFT of the result in (3).

6 Obtain the real part (better take the magnitude) of the result in (4).

7 Multiply the result in (5) by (-1) $^{(x + y)}$.

In step 2, the Two-Dimensional DFT:

$$F(u,v) = \frac{1}{MN}\sum_{X=0}^{M-1}\sum_{y=0}^{N-1} f(x,y)e^{-j2\pi\left(\frac{ux}{M}+\frac{vy}{N}\right)} \tag{3.14}$$

And it's inverse

$$f(x,y) = \frac{1}{MN}\sum_{u=0}^{M-1}\sum_{v=0}^{N-1} F(u,v)e^{j2\pi\left(\frac{ux}{M}+\frac{vy}{N}\right)} \tag{3.15}$$

$$G(u,v) = F(u,v)H(u,v) \tag{3.16}$$

where; F(u, υ) and H(u, υ) denote the Fourier transform of the input image f (x, y), and the filter function h(x, y), respectively. G (u, υ) is the Fourier Transform of the filtered image, which is the multiplication of two-dimensional functions H and F on an element-by-element. The important point to keep in mind is that the filtering process is based on modifying the transform of an image (frequency) in some way via a filter function, and then taking the inverse of the result to obtain the filtered image;

$$\text{Filtered Image} = \Im^{-1}[G\ (u,\ \upsilon)] \tag{3.17}$$

As mentioned earlier, the Gaussian has the same shape in the spatial and Fourier domains. Therefore does not incur the ringing effect in the spatial domain of the filtered image. The advantages over Ideal Low Pass Filter (ILPF) and Butterworth Low Pass Filter (BLPF), especially in some situations where any type of artifact is not acceptable, such as medical image.

3.7 ADAPTIVE MEDIAN FILTER (AMF)

The adaptive median filtering has been applied widely as an advanced method compared with standard median filtering. The Adaptive Median Filter performs spatial processing to determine which pixels in an image have been affected by impulse noise. The Adaptive Median Filter classifies pixels as noise by comparing each pixel in the image to its surrounding neighbor pixels. The size of the neighborhood is adjustable, as well as the threshold for the comparison. A pixel that is different from a majority of its neighbors, as well as being not structurally aligned with those pixels to which it is similar, is labeled as impulse noise. These noise pixels are then replaced by the median pixel value of the pixels in the neighborhood that have passed the noise labeling test.

The adaptive median filter removes the impulse noise from the image and reduces distortion in the image. It can be used to handle the filtering operation of an image corrupted with impulse noise of probability greater than 0.2. Adaptive median filters also smoothness and gives a much better output image compare than the other filters. Our filter is based on the algorithm.

The adaptive filter works on a rectangular region Sxy. The adaptive median filter changes the size of Sxy during the filtering operation. The output of the filter is a single value which replaces the current pixel value at (x, y), the point on Sxy is centered at that time. The following notation is adapted from the book Machine-Vision-Digital-Image-Processing-Fundamentals-Louis-J-Galbiati and is reintroduced here;

Purpose

1. Remove impulse noise
2. Smoothing of other noise

3. Reduce distortion, like excessive thinning or thickening of object boundaries

How it works?

- Adaptive median filter changes size of Sxy (the size of the neighborhood) during operation.

Notation

Zmin = minimum gray level value in Sxy

Zmax = maximum gray level value in Sxy

Zmed = median of gray levels in Sxy

Zxy = gray level at coordinates (x, y)

Smax = maximum allowed size of Sxy

Algorithm

Level A: A1 = Zmed - Zmin
A2 = Zmed – Zmax

if A1 > 0 AND A2 < 0, go to level B
else increase the window size
if window size < Smax, repeat level A
else output Zxy

Level B: B1 = Zxy - Zmin
B2 = Zxy - Zmax
if B1 > 0 AND B2 < 0, output Zxy
else output Zmed

Explanation

Level A: IF Zmin < Zmed < Zmax, then

Zmed is not an impulse

(1) go to level B to test if Zxy is an impulse ...

ELSE

Zmed is an impulse

(1) the size of the window is increased and

(2) level A is repeated until ...

(a) Zmed is not an impulse and go to level B or

(b) Smax reached: output is Zxy

Level B: IF Zmin < Zxy < Zmax, then

Zxy is not an impulse

(1) output is Zxy (distortion reduced)

ELSE

either Zxy = Zmin or Zxy = Zmax

(2) output is Zmed (standard median filter)

Zmed is not an impulse (from level A)

The key to understanding the mechanics of this algorithm is to keep in mind that it has three main purposes:

- To remove salt and pepper (impulse noise) noise.
- To provide smoothing of other noise that may not be impulsive.
- To reduce distortion such as excessive thinning or thickening of objects boundaries.
- The algorithm to be “impulsive like” noise components, even if these are not the lowest and highest possible pixel value in the image.

3.8 SIMULATION RESULTS FOR THE PROPOSED ARTIFACT REDUCTION IN THE IMAGES

The image sequence is separated into compressed. Each compressed image is processed by “Adaptive based Algorithm.” These images are motion estimated to find the motion vectors and get the output artifact removed image sequence.

Image File

Table 3.1 Types of image file

S.No	Name of the Image Sequence	Format	Size
1	Lenna.jpeg	JPEG	256X256
2	Lenna.bmp	BMP	512x512
3	Pict1679.tiff	TIFF	416x316
4	purple flower.png	PNG	512x512
5	xray-skeleton.gif	GIF	500x500

Image file characteristics

`Shown image file have different characteristics, such as fast motion, moderate motion and slow motion as below Table 3.2.

Table 3.2 Image file characteristics

S.No	Name of the Image Sequence	Format	Characteristics
1	Lenna.jpeg	JPEG	Fast
2	Lenna.bmp	BMP	Fast
3	Pict1679.tiff	TIFF	Moderate
4	purple flower.png	PNG	Fast
5	xray-skeleton.gif	GIF	Moderate

(a)

(b)

(c)

(d)

(e)

(f)

Figure 3.18 Leena.JPEG (a) Original image (b) Compressed image (c) Edge detection (d) Low pass filtered image (e) AMF filtered image (f) Output of the proposed algorithm

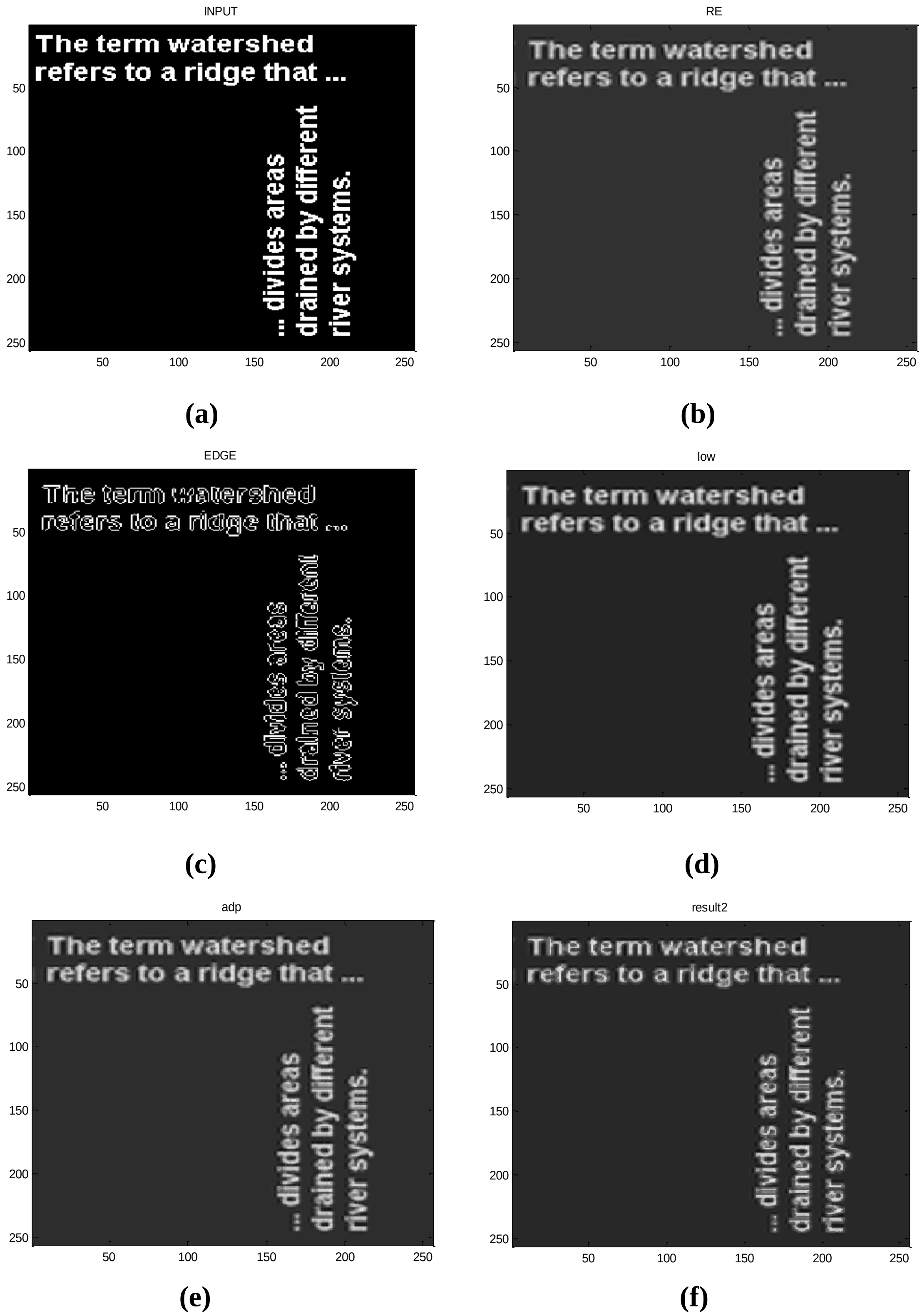


Figure 3.19 Text.PNG (a) Original image (b) Compressed image (c) Edge detection (d) Low pass filtered image (e) AMF filtered image (f) Output of the proposed algorithm

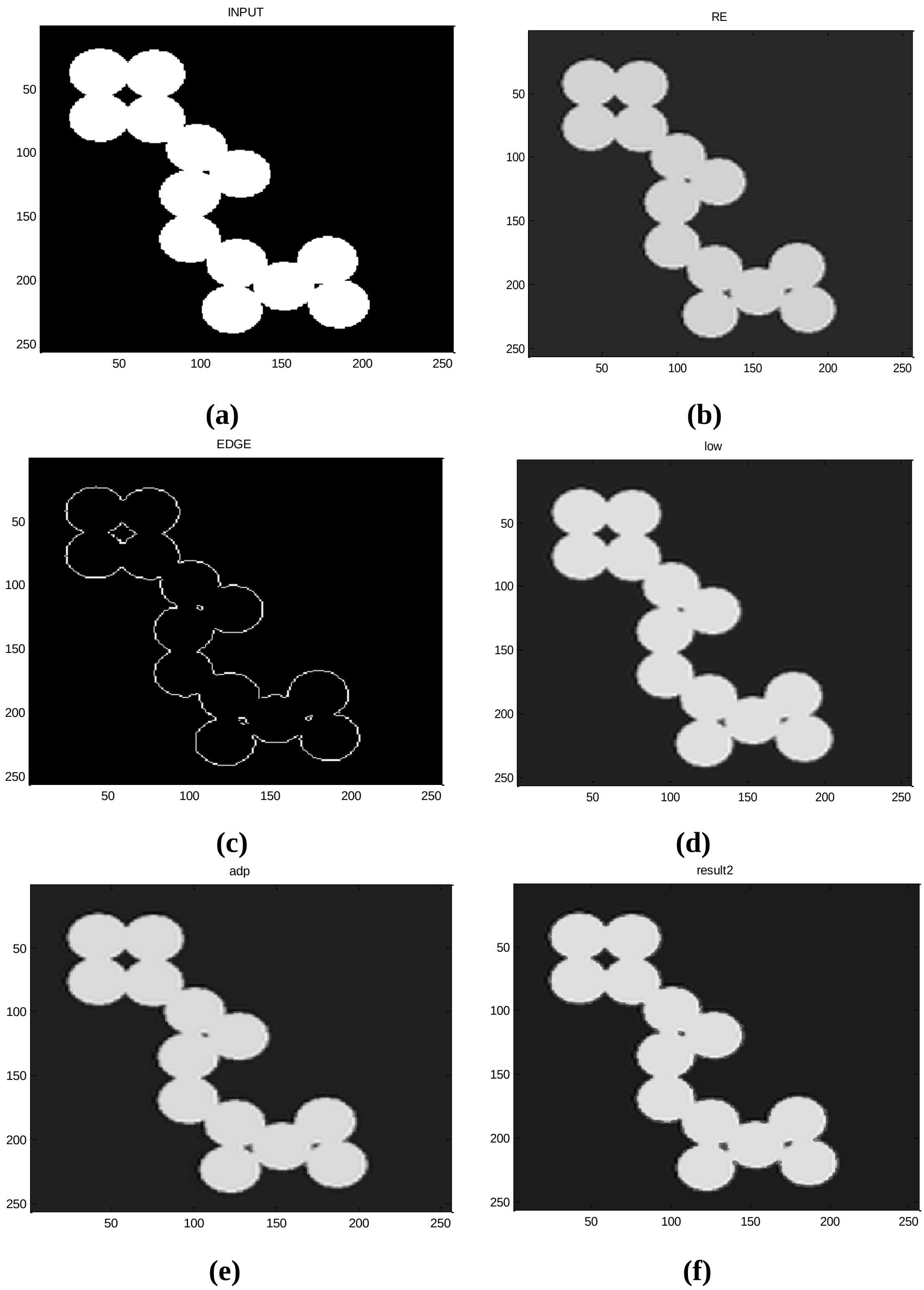


Figure 3.20 Circles.PNG (a) Original image (b) Compressed image (c) Edge detection (d) Low pass filtered image (e) AMF filtered image (f) Output of the proposed algorithm

Table 3.3 Comparison of proposed algorithm performance with adaptive median filter algorithm

S.No.	Name of the image sequence	MSE		PSNR	
		MF	AMF	MF	AMF
1	Text.PNG	5.377	1.499	43.523	51.176
2	Circles.PNG	3.843	1.852	37.943	45.488
3	Lenna.JPEG	6.241	1.992	42..423	55.174
4	Lenna.BMP	4.981	1.432	32.781	41.355
5	Pict1679.TIFF	3.853	1.243	45.423	58.432
6	x-ray-skeleton.GIF	2.981	0.984	33.425	46.543

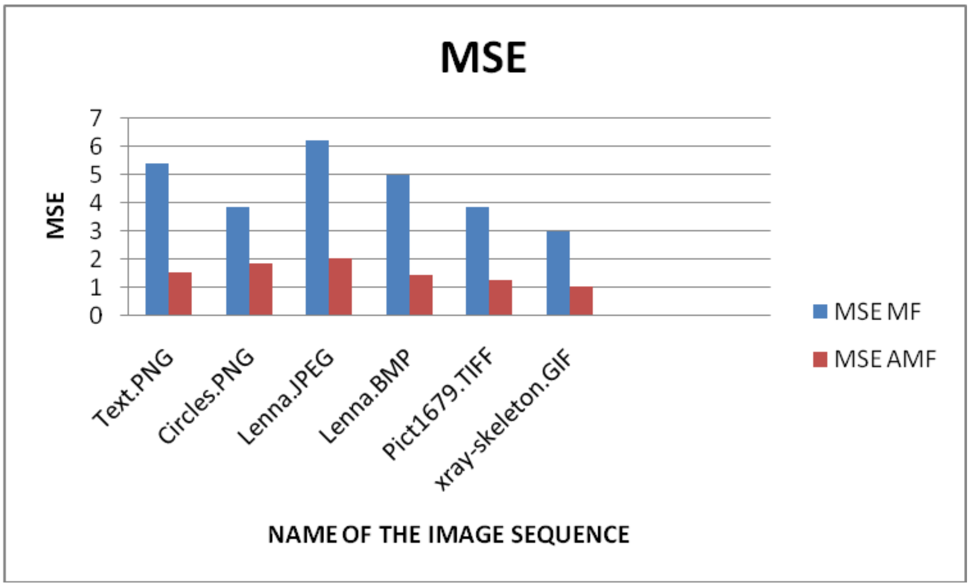


Figure 3.21 MSE for the Image Sequence

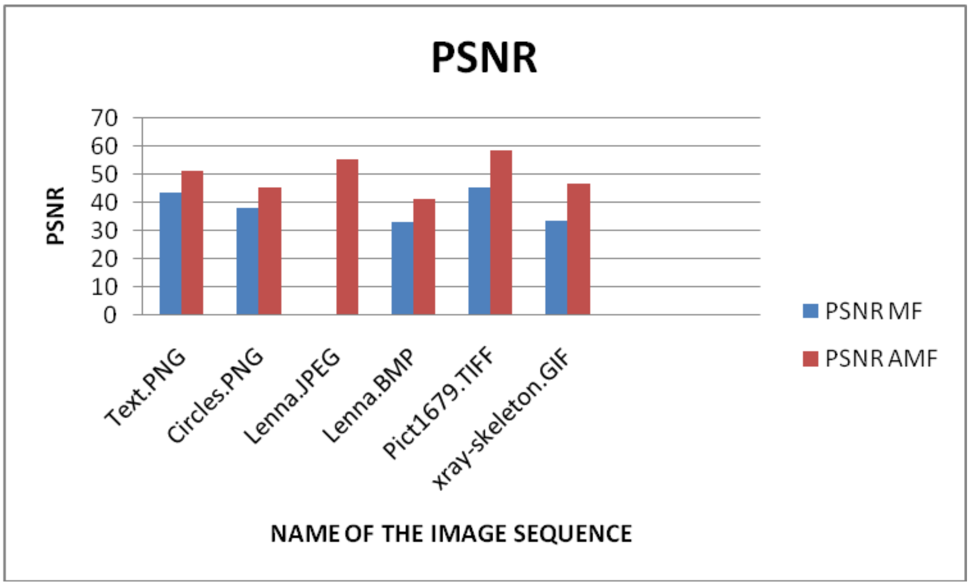


Figure 3.22 PSNR for the Image Sequence

Comparison Results

Saving percentages of all selected algorithms are shown in Table 3.4. Lowest saving percentage is given by JPEG (DCT based) and relatively best values are given by scalar quantized algorithm. Average values are given by all the three SPIHT algorithms, but the values of Adaptive method are higher than the two static methods. The differences do not exceed more than 2%.

Table 3.4 Saving percentages of all selected algorithms

	JPEG(DCT)	Scalar Quantized	SPIHT	Adaptive
Text.PNG	38.24	37.42	36.06	39.21
Circles.PNG	43.78	38.32	37.06	39.32
Lenna.JPEG	30.70	32.60	31.99	35.88
Lenna.BMP	47.98	41.70	40.91	44.15
Pict1679.TIFF	69.03	41.94	40.82	42.53
xray-skeleton.GIF	44.35	41.07	40.72	42.11

In order to compare the performance of the selected algorithms the compression and decompression times, and compressed image are compared. Figure 3.23 shows the compression times 6 images for all the algorithms.

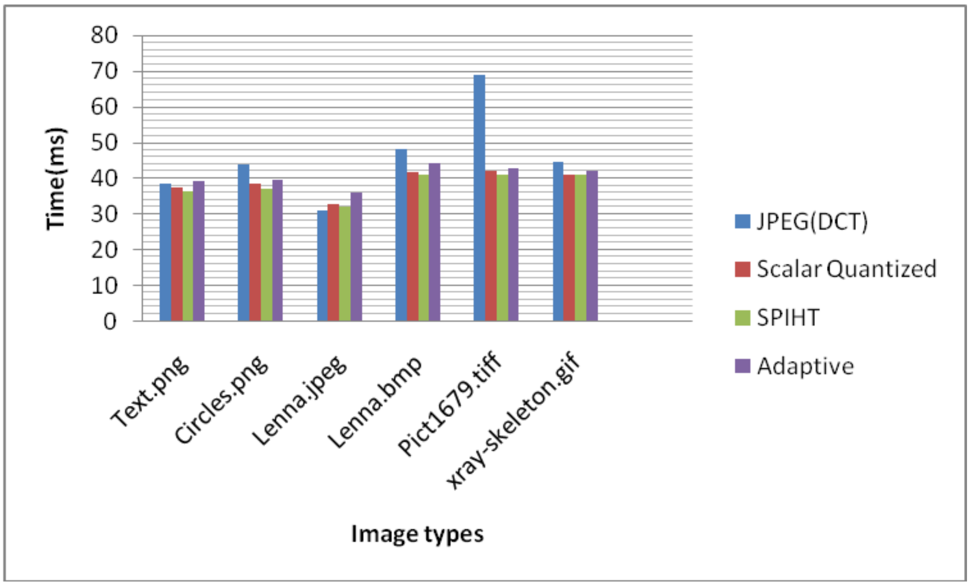


Figure 3.23 Compression times

3.9 CONCLUSION

Image processing has become an important task and found usage in many fields. In this thesis, images and videos degraded by artifact noises are taken in account and removed using adaptive median filter. The proposed algorithm produces better edge preservation, fine detail preservation and also significantly reduces the artifacts when compared to the median filter. Exhaustive search block matching algorithm produces better PSNR for motion estimation in motion pictures, but in case of large search windows, the computational expenses are large compared to the other motion estimation techniques. The computational expense is less for Diamond search estimation. The MSE and PSNR calculated at compressed image sequences.

CHAPTER 4

COMPRESSION ARTIFACT REDUCTION IN THE VIDEOS

4.1 INTRODUCTION

Video is the technology of electronically capturing, recording, processing, storing, transmitting and reconstructing a sequence of still images representing scenes in motion. Video technology was first developed for television systems, but further developed in many formats to allow for consumer video recording. Video can also be viewed through the Internet as video clips or streaming media clips on computer monitors. The video commonly refer to as several storage formats for moving eye pictures such as digital video formats, including DVD, QuickTime and MPEG-4 and analog videotapes, including Video Home System (VHS) and Betamax. Video can be recorded and transmitted in various physical media: in magnetic tape when recorded as Phase Alternating Line *(*PAL) or National Television Standards Committee (NTSC) electric signals by video cameras, or in MPEG-4 or DV digital media. Quality of video used essentially depends on the capturing method and storage. Digital Television (DTV) is a relatively recent format with higher quality than earlier television formats and standard for television video. Video can be interlaced or progressive. Video sequence is a equal three-dimensional array of data in the vertical, horizontal and temporal dimensions. There sequence is a collection of frames and display the fixed time intervals. A video frame is a picture made up of 2D discrete grid of pixels. These frames converted to images for image processing.

4.2 FLOWCHART OF THE PROPOSED ALGORITHM

A new algorithm has been proposed to remove artifacts from a noise artifact video source. The video sequence is first converted into frames and frames into images. Edge detection using a sobel edge detector is performed. The low pass filter applied to the video sequence.

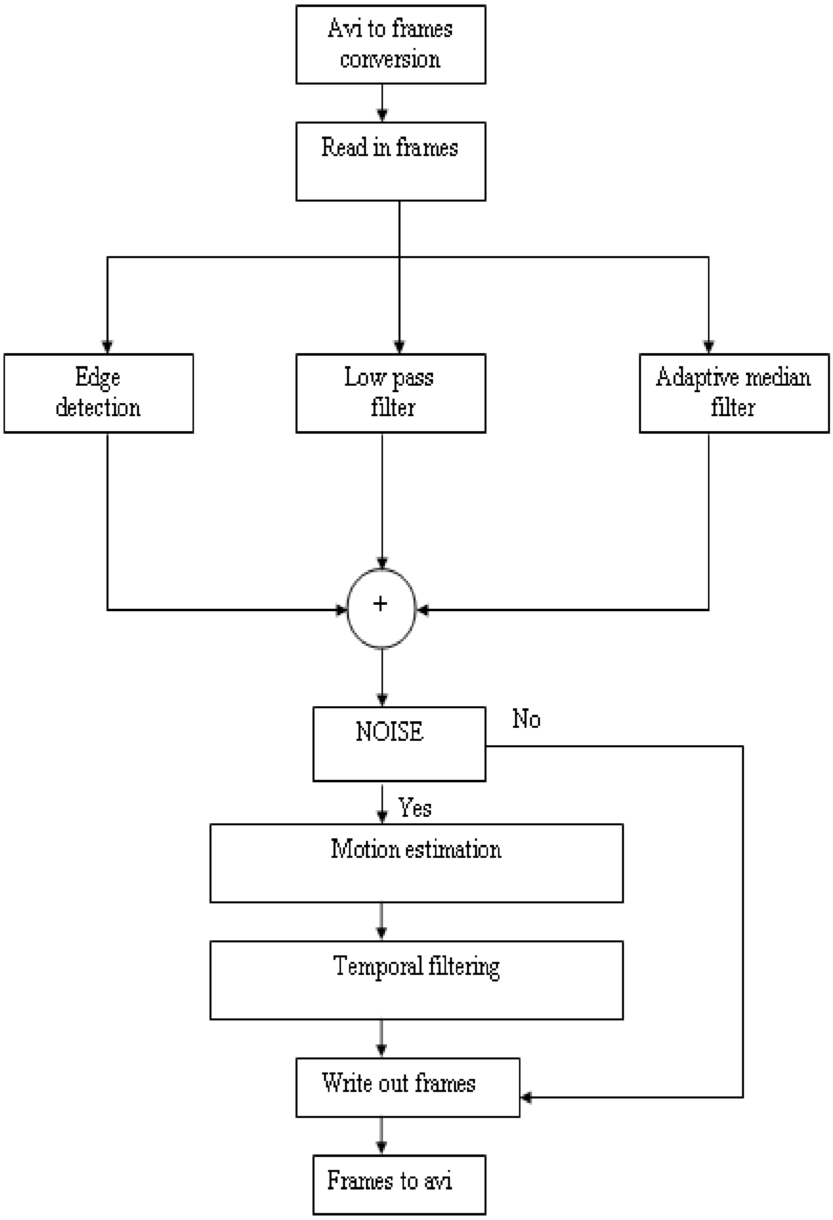


Figure 4.1 Flowchart of the proposed algorithm

Next the adaptive median filter is applied to the video sequence and compares the results. A block-matching algorithm based on Diamond search is used for motion estimation and finally a temporal median filter is applied to obtain the restored video sequence.

4.3 COMPRESSION ARTIFACT REDUCTION IN THE VIDEOS ALGORITHM

- **Video to frames**: The noisy video sequence containing artifact is converted into avi format, which is an uncompressed format and frames are extracted from the video
- **Frames:** Frames are used for further processing.
- **Edge detection:** Edge detection is performed using Sobel edge detector.
- **Low pass filter:** Low pass filter is performed for image smoothening.
- **Adaptive median filter:** The output pixel of adaptive median filter is better when compared to the output pixel of other filter. It is used to replace the corrupted pixel.
- **Motion estimation & motion compensation:** It is performed for all the frames. In motion estimation and vectors are determined by considering two consecutive frames. These frames are divided into several macro blocks. A block in the reference frame is compared with in the current frame and thus motion vectors are obtained. A block matching used for Diamond searching. The motion vectors and compensation is performed which improves the quality of the picture.
- **Temporal median filtering:** They are the original frame, processed frame and the motion compensated frame. The output frame reduces the noise.
- **Frames to movie:** After completing the entire process, the processed frames are finally converted back into original movie.

4.4 COLOR PROCESSING

RGB color image is an M x N x 3 array of color pixels, where as each color pixel is a triplet corresponding to the red, green and blue components of an RGB image at a specific spatial location. RGB image may be viewed as a “stack” of three gray-scale images, when fed into the red, green, and blue inputs of a color monitor and produce a color image of the screen. The data class of the component images determines their range of values. The RGB image is a class of double and range of value is [0, 1]. Similar, before can [0, 255] and [0, 65535] the number of bits used to represent the pixel values of the component images determines the bit depth of an RGB image. For example, if each component image is 8-bit image and corresponding RGB image is said to be 24 bits deep. Generally, the number of bits in all the component images is the same.

4.4.1 Extraction of three Components

These three components will combine in proper proportions to form all the possible colors, which are required to form the real color image. The three components (red, green, blue) are extracted from the color frame that is to be processed. Thus a single color frame is converted into three primary color component frames. These frames are processed using the proposed filter and thus noise in each component is removed. For the color video component colors are motion estimated and compensated individually. Then the three processed frames are concatenated to form the resultant color image.

4.5 MOTION ESTIMATION

Motion estimation techniques, to track scratch on the frame, gave a good method to sniff out line scratches. The key role in prediction and interpolation play successful estimate of a motion vectors. For fast motion prediction commonly used technique is Block Matching (BM) motion estimator.

The motion vector is obtained by minimizing a cost function measuring the mismatch between a block and each predictor candidate. The Motion Estimation (ME) is essential tool for determining of motion trajectories. It gives us motion vector of each pixel or block of pixels. The more accurate is the motion estimation and better image sequence restoration we can reach. Due to motion of objects in scene (i.e. corresponding regions in an image sequence) the same region don't occur in the same place in previous frame as in current one. Hence for proper function of the proposed scratch removal algorithm is favorable to estimate future motion in observed image sequence as accurate as possible. The motion typically encountered in an image sequence has three possible forms: translation, rotation and zoom of objects in the scene.

Regions in image are moving along their motion trajectories through the image sequence. So it is necessary to find similar areas in neighboring frames and to treat them together. Solution is to find motion of the regions in each frame and then restore missing information in some frame by filtering along motion trajectories.

4.5.1 Block Matching Method

Block matching (BM) is very popular & robust type of methods for motion estimation, which are widelsy adopted in various coding standards, such as MPEG-1, MPEG-2 and MPEG-4. It is used for fast motion estimation of the blocks, into which usually evaluated image is divided.

The underlying supposition behind motion estimation is that the patterns corresponding to objects and background in a frame of video sequence move within the frame to form corresponding objects on the subsequent frame. The idea behind block matching is to divide the current frame into a matrix of ‘macro blocks’ that are then compared with corresponding block and its adjacent neighbors in the previous frame to create a vector. The macro block will one location to another location andcalculated for all the macro blocks comprising a frame and constitutes the motion estimated in the current frame. The search area for a good macro block match is constrained up to ‘p’; pixels on all fours sides of the corresponding macro block in previous frame. So ‘p’ is called as search parameter. Larger motions require a larger ‘p’ and the larger the search parameters are the more computational and expensive. Usually, the macro block is taken as a square of side 16 pixels, and the search parameter ‘p’ is 7 pixels.

4.5.2 Mean Absolute Difference (MAD)

Mean Absolute Difference (MAD) is computed for all pixel values between the blocks in the present frame and the previous frame. It is blocks matching method.

$$\mathrm{MAD} = \frac{1}{N^2}\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}\left|C_{ij} - R_{ij}\right| \tag{4.1}$$

Where, C_{ij} = Pixels in the Current Block

R_{ij} = Pixels in the Reference Block

N = Macro block Size

Finally which gives the lowest MAD value is chosen as the motion vector.

4.5.3 Types of Block Matching Method

There are several search strategies used for block matching techniques such as;

- Exhaustive Search (ES)
- Three Step Search (TSS)
- New Three Step Search (NTSS)
- Diamond Search (DS)

4.5.3.1 Exhaustive Search (ES)

Exhaustive Search (ES) algorithm is known as full search and it is the most computational and expensive block matching algorithm. This algorithm calculates the cost function at each possible location in the search window. As a result of which it finds the best possible match amongst any block matching algorithm. The disadvantage of ES is the larger search window gets the more computations.

4.5.3.2 Three Step Search (TSS)

This algorithm was introduced by Koga et al in 1981. TSS is a popular fast algorithm. It became very popular because of its simplicity and also robust and near optimal performance. It searches for the best motion vectors in a course to fine search pattern. It is a block based search technique and uses a maximum of three steps. It is mainly used for ME in low bit rate video compression applications such as videophone and video conferencing.

The TSS algorithm is a simple and effective method used for ME. When compared with Exhaustive Block Matching Algorithm (EBMA) it has a saving factor of greater than a 100. It has a fixed number of three search steps and a maximum and minimum number of search points of 25[4] The disadvantage of this technique is that it uses a uniformly allocated checking point pattern in the first step which makes it inefficient for the estimation of small motion.

1. An initial step size is picked. Eight blocks at a distance of step size from the centre (One center point and eight points on the boundary of the search square) are compared in the first step and thereafter only 8 points are searched.
2. At the start of a new step the search point center is moved from the best matching point from the previous step. Step size is reduced by half after each step.
3. At the end of the search the step size is one pel.

Advantages

- Simplicity and also robust and near optimal performance.
- Best motion vectors in a course to fine search pattern.

Disadvantages

- Inefficient for the estimation of small motion.

4.5.3.3 New Three Step Search (NTSS)

NTSS algorithm is a modification of the well known TSS algorithm, which is widely used as the motion estimation technique in some low bit-rate video compression algorithms. However, as mentioned in the previous section, TSS algorithm allocates uniform search points in the first step. Therefore, it is not very efficient in detecting small motions appearing in stationary or quasi blocks. Since most real world motion sequences are smooth and gentle, detecting small motions becomes an important issue. One solution to this could be to make the search adaptable to the uncertainty and motion scale. This is achieved by employing a center-biased checking point pattern in the first step. The increase in complexity due to the addition of points-in the first step is compensated by having a half-way stop technique. This makes the complexity of NTSS comparable to that of TSS.

The NTSS algorithm is as follows. As mentioned above, NTSS differs from TSS by full search assuming a center biased checking point pattern in the first step and TSS incorporating a half-way stop technique for stationary or quasi stationary blocks. Details are as follows.

In addition to the original checking points used in TSS, eight extra points are added, which are the eight extra neighbors of the search window (center-bias).

A half-way stop technique is used for stationary and quasi stationary block in order to fast identify and then estimate the motions for these blocks.

- If the minimum BDM in the first step occurs at the search window center, stop search (first-step-stop).
- If the minimum BDM point in the first step is one of the eight neighbors of the window center, the search in the second step will be performed only for eight neighboring points of the minimum and then stop search (second-step-stop).

Advantages

- Complexity of NTSS comparable to TSS.
- Highest PSNR compare to other algorithms and nearest to Full Search algorithm.

- To overcome the TSS disadvantages.
- Elapse time less compare to Full Search.

Disadvantages

- Processing time is complex compared to TSS and DS.

4.5.3.4 Diamond Search (DS)

The search point pattern is diamond, and there is no limit and the number of steps in the algorithm. The DS uses two different types of fixed patterns such as Large Diamond Search Pattern (LDSP) of 9 points and Small Diamond Search Pattern (SDSP) of 5 points as shown in the figure 4.2.

(a) LDSP (b) SDSP

Figure 4.2 Search Patterns

The LDSP is repeatedly used in the searching procedure until the step in which the minimum block distortion (MBD) point occurs at the center of the diamond. The search pattern is then switched to SDSP. The point yielding the MBD among the 5 checking points in the SDSP provides the motion vector of the best matching block.

The DS algorithm is summarized as follows.

1. The initial LDSP is centered at the origin of the search window, and the 9 checking points of LDSP are tested. If the

MBD point calculated is located at the center position, go to Step 3; otherwise, go to Step 2.

2. The MBD point found in the previous search step is re-positioned as the center point to form a new LDSP. If the new MBD point obtained is located at the center position, go to Step 3; otherwise, recursively repeat this step.

3. Switch the search pattern h m LDSP to SDSP. The MBD point found in this step is the final solution of the motion vector which points to the best matching block.

Advantages

- Elapse time shorter.
- Compress video sequence in high value.
- Reduce the complexity of the motion estimation

Disadvantages

- Does not restrict the number of search steps.
- No efficient hardware implementation

4.5.3.5 Diamond Search Procedure

These figure shows an example path to motion vector (-4, -2) in five search steps four times of LDSP and one time of SDSP.

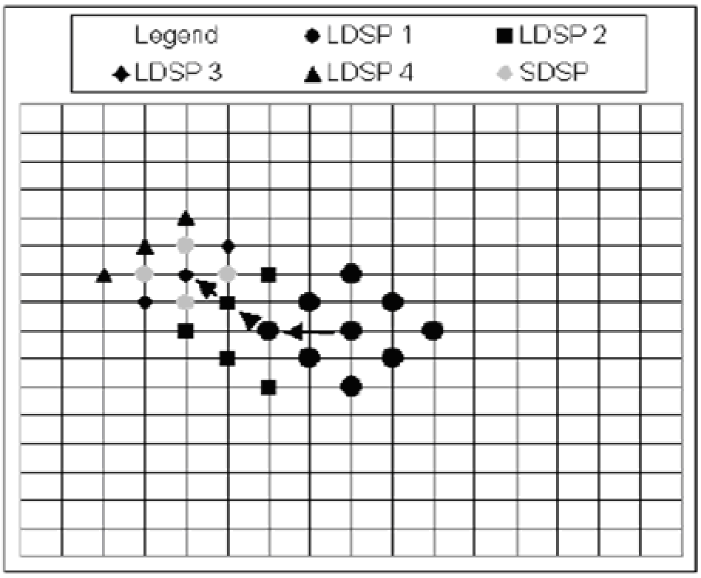


Figure 4.3 Diamond Search (DS)

4.5.4 Motion Compensation

Motion compensation is a way of describing the difference between consecutive frames in terms of the where each section of the former frame has moved to the crucial thing in image sequences is motion. If the motion is insignificant, then any motion non-compensated method of filtering can be applied. However, if the noise is significant and the most often unsatisfactory results. In order to increase the quality of frames and motion compensated filters are used.

4.5.4.1 Principle of Motion Compensation

If an image sequence shows moving objects, then their motion within the scene can be measured, and the information is used to predict the content of frames later in the sequence. Each image frame is divided into a fixed number of square blocks. For each block in the frame, a search is made in the reference frame over an area of the image that allows for the maximum translation that the coder can use. The searches is for the best matching block gives the least prediction error, usually minimizing either mean square difference, or mean absolute difference which is easier to compute, such as absolute difference.

A good match during the search means that a good prediction can be made, but the improvement in prediction must outweigh of the cost of transmitting the motion vector. Whole block has undergone the same translation, and the block should not overlap objects in the image that have different degrees of motion. Including the background. The choice of block-size to use for motion compensation is always a compromise, smaller and more numerous blocks can better represent complex motion than fewer large ones and reduces the work and transmission cost and greater cost for the motion information itself. The choice of block-size can be affected not only by motion vector but also by the other scene characteristics such as, texture and inter-frame noise.

4.5.4.2 Block Motion Compensation

In BMC, the frames are partitioned in blocks of pixels (e.g. macro blocks of 24×24 pixels). Each block is predicted from a block of equal size in the reference frame. The blocks are not transformed in any way apart from being shifted to the position of the predicted block. There are shift represented by motion vector. The motion vectors are the parameters of motion model and have to be encoded into the bit-stream. As the motion vectors are not always independent (e.g. if two neighboring blocks belong to the same moving object) and they are usually encoded differentially to save bit-rate. This means that the difference of the motion vector and the neighboring motion vector(s).

4.5.5 Temporal Median Filtering

The temporal median filter smoothes out sharp transitions in intensity at each pixel position, it not only denoise the whole frame and remove blotches, but also helps in stabilizing the fluctuations. There mostly is beneficial for both visually and motion estimation performance in the next stage. Temporal median filtering removes the temporal noise in the form of small dots and streaks found in some videos.

In this approach, dirt is viewed as temporal impulse (single-frame incident) and treated by inter-frame processing by taking into account at least three consecutive frames. A pixel was fluffed as dirt if the corresponding absolute differences between the current frame and each of the previous next frames are high. Generally, motion compensation can be essential component of a dirt detection algorithm. Motion estimation fails and may be generated unpredictable results. A temporal filtering may be regarded as a useful tool and complement to motion compensated approaches. Following steps are performed

1. Three frames are used such as
   - Original Frame
   - Processed Frame
   - Motion Compensated frames.
2. Using adaptive based algorithm
3. Removed artifact and median is determined finally, resultant before frame.

4.6 SIMULATION RESULTS FOR THE PROPOSED ARTIFACT REDUCTION IN THE VIDEOS

The video sequence is separated into frames. Each frame is processed by “Adaptive based Algorithm”. These frames are motion estimated and compensated to find the motion vectors. The motion compensated frames are subjected to temporal filtering and the output for artifact removed frames. The sample frames are following processing sequences such as

Video file

Table 4.1 Types of video file

S.No	Name of the Video Sequence	Format	Size	Frames
1	Akiyo.avi	QCIF	144x176x0	100
2	Claire.avi	QCIF	144x176x0	100
3	Coastguard.avi	QCIF	144x176x0	100
4	Foreman.avi	QCIF	144x176x0	100
5	Salesman.avi	QCIF	144x176x0	100
6	Tennis.avi	QCIF	144x176x0	100
7	Vipcolor Segmentation.avi	QCIF	144x176x0	100
8	Vipmen.avi	QCIF	144x176x0	100

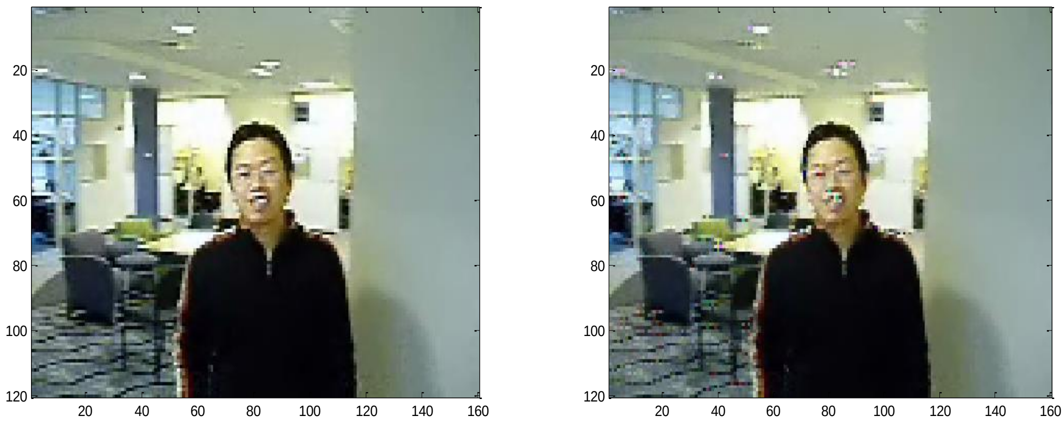
Video File Characteristics

Shown video file have different characteristics, such as fast motion, moderate motion and slow motion as shown in the table 4.2.

Table 4.2 Video file characteristics

S.No	Name of the Test Video Sequence	Characteristics
1	Akiyo.avi	Slow
2	Claire.avi	Slow
3	Coastguard.avi	Moderate
4	Foreman.avi	Fast
5	Salesman.avi	Moderate
6	Tennis.avi	Fast
7	Vipcolor Segmentation.avi	Fast
8	Vipmen.avi	Fast

Sequence 1



(a) **(b)**

Figure 4.4 Frame no 26 (a) Before processing (b) After processing

(a) **(b)**

Figure 4.5 Frame no 70 (a) Before processing (b) After processing

Sequence 2

(a) **(b)**

Figure 4.6 Frame no 101 (a) Before processing (b) After processing

(a) **(b)**

Figure 4.7 Frame no 185 (a) Before processing (b) After processing

Table 4.3 Comparison of proposed algorithm performance with adaptive median filter algorithm

S.No.	Name of the video sequence	MSE		PSNR	
		MF	AMF	MF	AMF
1	Akiyo.avi	15.152	4.563	30.102	33.521
2	Claire.avi	15.532	2.105	29.521	34.926
3	Coastguard.avi	16.521	5.305	31.515	34.105
4	Foreman.avi	14.825	4.865	31.926	35.690
5	Salesman.avi	19.522	5.582	32.584	37.087
6	Tennis.avi	20.523	5.521	33.921	38.542
7	Vipcolor Segmentation.Avi	21.1045	6.6863	34.9228	39.7360
8	Vipmen.Avi	3.6708	2.5734	42.9369	45.8479

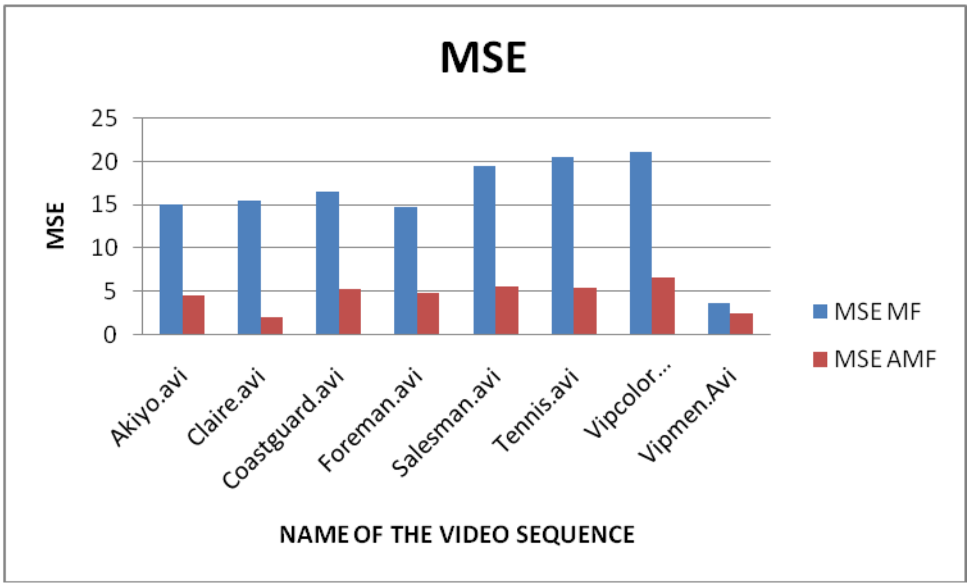


Figure 4.8 MSE for the Video Sequence

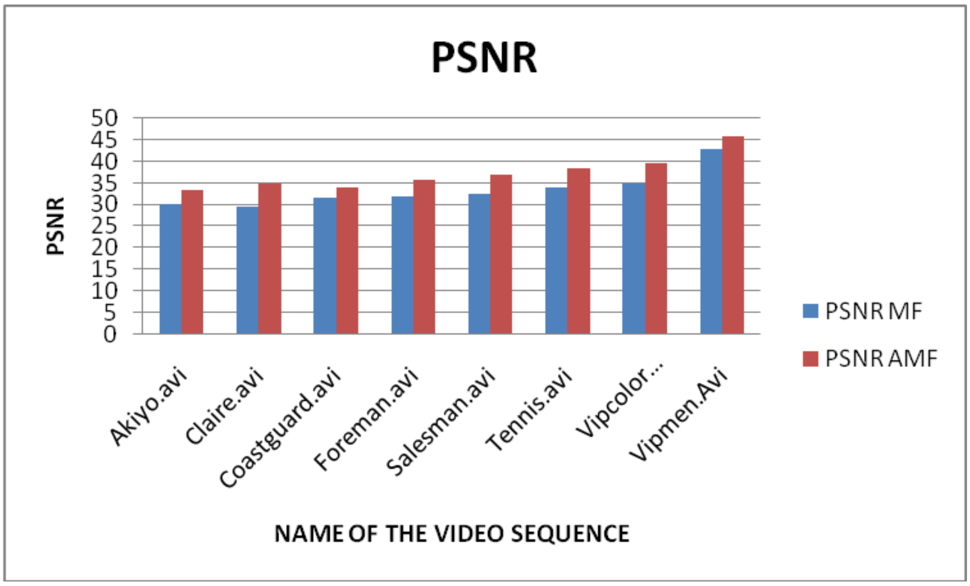


Figure 4.9 PSNR for the Video Sequence

Comparison Results

Saving percentages of all selected algorithms are shown in Table 4.4. Lowest saving percentage is given by MPEG (DCT based) and relatively best values are given by scalar quantized algorithm. Average values are given by all the three SPIHT algorithms, but the values of Adaptive method are higher than the two static methods.

Table 4.4 Saving percentages of all selected algorithms

	MPEG(DCT)	Scalar Quantized	SPIHT	Adaptive
Akiyo.avi	38.24	37.42	36.06	39.21
Claire.avi	43.78	38.32	37.06	39.32
Coastguard.avi	30.70	32.60	31.99	35.88
Foreman.avi	47.98	41.70	40.91	44.15
Salesman.avi	69.03	41.94	40.82	42.53
Tennis.avi	44.35	41.07	40.72	42.11
Vipcolor Segmentation.Avi	45.63	42.32	39.23	37.45
Vipmen.Avi	52.12	45.34	39.25	35.24

In order to compare the performance of the selected algorithms the compression and decompression times, and compressed image are compared. Figure 4.10 shows the compression times 8 videos for all the algorithms.

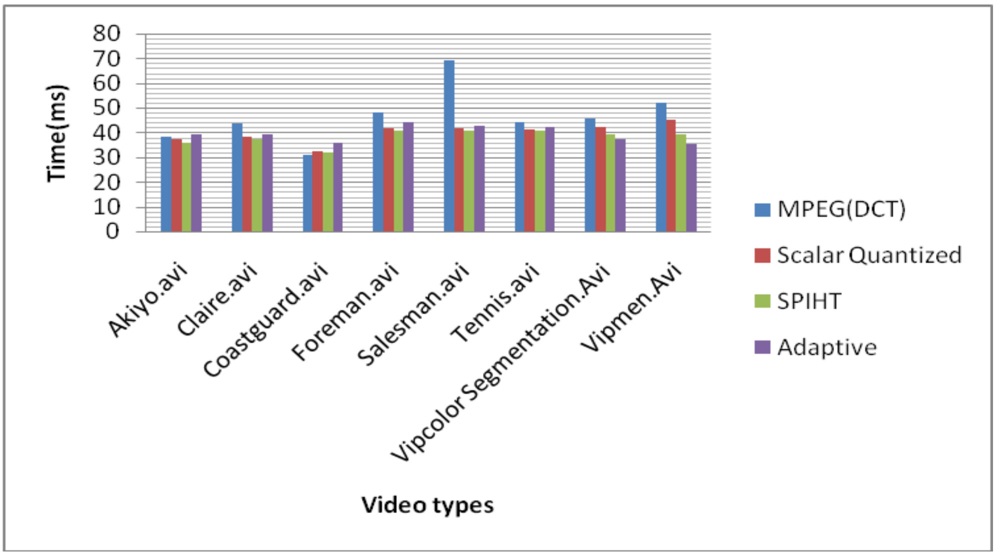


Figure 4.10 Compression times

4.7 CONCLUSION

Video processing has become an important task and found usage in many fields. It can be concluded that the images and video degraded by artifact noises are taken in account and removed using adaptive median filter. The proposed algorithm produces better edge preservation, fine detail preservation and also significantly reduces the artifacts when compared to the median filter. Exhaustive search block matching algorithm produces better PSNR for motion estimation in motion pictures, but in case of large search windows, the computational expenses are large compared to the other motion estimation techniques. The computational expense is less for Diamond search estimation. The MSE and PSNR calculated at compressed video sequence.

CHAPTER 5

MOVING OBJECT TRACKING IN THE IMAGES AND VIDEOS

5.1 INTRODUCTION

In designing automated systems for the interpretation and manipulation of image data, system developers often need to perform software imaging operations is also called as segmentation .Figure 5.1 (a), (b), (c) and (d) extract information about the structure of objects and to separate and discern various parameters of interest within the data.

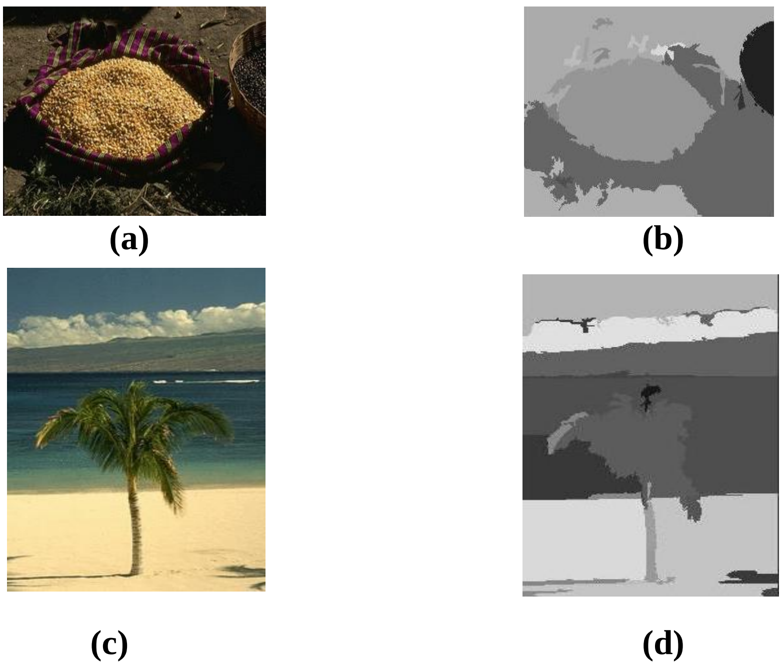


Figure 5.1 Moving object tracking in the images (a) Input image (b) Segmented image (c) Input image (d) Segmented image

5.1.1 Moving Object Tracking in the Images

Measurements or attributes of these objects is known as features and that can be used for defect inspection, quality control and clinical qualitative analysis. Accordingly, common vision processes deal with the identification of discrete objects within an image. Such processes transform single-pixel representations of the image data into geometric descriptors representing groups of pixel elements. These descriptors, known as objects and to the form of points, lines, regions, polygons and other unique representations.

Segmentation techniques are divided into two basic categories: edge-based and region-based. Edge-based segmentation is primarily used to look for image discontinuities. Generally these techniques are applied for the changes of gray-level intensity occur in the images. The assumption is that changes occur in the data at the boundary between objects of interest. The output of edge-segmentation schemes can be x and y gradient. These two image are found to be one is X direction and other one is Y direction.

1. Gradient strength and direction
2. Binary edge map
3. Edge representation.

In contrast, region-based segmentation is used to look for similarities between adjacent pixels. The pixels that possess similar attributes are grouped into unique regions. The assumption is made that each region represents one object of interest. Using gray-level intensity is the most common means of assigning similarity, but many other possibilities exist, such as variance, color, and multispectral features.

Most commercial vision systems are use the region-based segmentation schemes based on pixel-intensity values. These segmentation techniques assume that the objects of interest possess uniform shading, significant and constant gray-level change occurs between the objects of interest and the background. However, in many vision applications, these assumptions have proven erroneous. Therefore, these techniques are considered fragile and commonly require controlled conditions and human supervision. The effects of uneven sample illumination, shadowing, partial occlusion, clutter, noise and subtle object-to-background changes can all contribute to errors in basic segmentation processes. Generally, results false segmentations of the background, partial segmentations of the objects of interest, clumping of objects and inadequate segmentations. Errors in the segmentation of the data can also result in the calculation of erroneous features. Therefore, it is essential that the segmentation method chosen support the final processing goals of the vision system. Thresholding (Otsu 1979) is perhaps the most common segmentation technique and the most basic region-segmentation technique.

The technique separates pixels into background and foreground (object of interest) classes based upon their similarities in gray-level intensity. To implement these techniques, a Threshold (T) value is chosen. Every pixel in the image is then compared to the T value. Each pixel is given a region label of "0" (background) if the pixel value is less than or equal to T or "1" (foreground) if greater than T. These form of region segmentation results in a binary image, and each region is either white (1) or black (0). Many variations exist within the general concept of segmentation by thresholding, which will be discussed in future column. Image segmentation is one of the fundamental problems in image processing and computer vision.

Segmentation is also one of the first steps in many image analysis tasks. Image understanding systems such as face or object recognition often assume that the objects of interests are well as segmented.

Different visual cues, such as color, texture, and motion, helps in the achieving segmentation. Segmentation is also goal dependent and subjective. However, generic criteria applied to a large variety of images and that can be adapted for specific applications.

5.1.2 Various Types of Image Segmentation Techniques

5.1.2.1 Clustering Method

The K-means algorithm is an iterative technique that can be used to partition an image into *K* clusters. The basic algorithm is

1. Pick *K* cluster centers, either randomly or based on some heuristic
2. Assign each pixel in the image to the cluster that minimizes the variance between the pixel and the cluster center
3. Re-compute the cluster centers by averaging all of the pixels in the cluster
4. Repeat steps 2 and 3 until convergence is attained (i.e. no pixels change clusters)

In this case, variance is the squared or absolute difference between a pixel and a cluster center. The difference is typically based on pixel color, intensity, texture, and location, or a weighted combination of these factors. *K* can be selected manually, randomly or by a heuristic. This algorithm is guaranteed to converge, but return to the optimal solution. The quality of the solution depends on the initial set of clusters and the value of *K*. In statistics and machine learning, the k-means algorithm is clustering algorithm to partition n objects into k clusters, where as k < n. It is similar to the expectation-maximization algorithm for mixtures of Gaussians in that they both attempt to find the centers of natural clusters in the data. The model requires that the object attributes correspond to elements of a vector 14 space. The objective it tries to achieve is to minimize total intra-cluster variance, or, the squared error function. The k-means clustering was invented in 1956. The most common form of the algorithm uses an iterative refinement heuristic known as Lloyd's algorithm. Lloyd's algorithm starts by partitioning the input points into k initial sets, either at random or using some heuristic data. It can be calculate the mean point, or centroid of each set. It constructs a new partition by associating each point with the closest centroid.

The centroids are recalculated for the new clusters, and algorithm repeated by alternate application of these two steps until convergence, the points no longer switch clusters (or) alternatively centroids are no longer changed.

5.1.2.2 Histogram-Based Method

Tsai and Chen (1992), have histogram-based methods are very efficient when compared to other image segmentation methods because they are typically requiring only one pass through the pixels. In this technique, a histogram is computed from all of the pixels in the image peaks and valleys histogram are used to locate the clusters in the image. Color or intensity can be used as the measure. A refinement of their technique is to recursively apply the histogram-seeking method to clusters in the image in order to divide them into smaller clusters. This is repeated with smaller and smaller clusters until no more clusters are formed. One of disadvantages of the histogram-seeking method is that it is can be difficult to identify significant peaks and valleys in the image. In this methods of image classification distance metric and integrated region matching are familiar.

5.1.2.3 Region Growing Method

The first region growing method was the seeded region growing method. These methods take a set of seeds as input along with the image. The seeds mark each of the segmented objects. The regions are iteratively grown by comparing all unallocated neighboring pixels to the regions. The difference between a pixel's intensity value and the region's mean, δ, is used as a measure of similarity. The pixel with the smallest difference measured and allocated to the respective region. These processes continue until all pixels are allocated to a region. Seeded region growing requires seeds as additional input. The segmentation results are dependent on the choice of seeds. Noise in the image can cause the seeds to be poorly placed. Unseeded region growing is a modified algorithm that doesn't require explicit seeds.

5.1.2.4 Edge Detection Method

Edge detection is a well-developed field on its own within image processing. Region boundaries and edges are closely related, since there is often a sharp adjustment in intensity at the region boundaries. Edge detection techniques have been used as the base of another segmentation technique. The edges identified by edge detection are often disconnected. To segment an object from an image however, one needs closed region boundaries. Discontinuities are bridged if the distance between the two edges is within some predetermined threshold. One such method is the edge linking method, proposed.

5.1.2.5 Watershed Transformation

The watershed transformation considers the gradient magnitude of an image as a topographic surface. Pixels having the highest Gradient Magnitude Intensities (GMIs) correspond to watershed lines, which represent the region boundaries. Water placed on any pixel enclosed by a common watershed line flows downhill to a common Local Intensity Minimum (LMI). Pixels draining to a common minimum form a catch basin, which represents a segment.

5.1.2.6 Graph Partitioning Method

Graphs can be effectively used for image segmentation. Usually a pixel or a group of pixels are vertices and edges define the similarity among the neighborhood pixels. Some popular algorithms of are category are random walker, minimum mean cut, minimum spanning tree- based algorithm and normalized cut. The normalized cuts method was first proposed in Shi and Malik (1997).In this method, the image being segmented is modeled as a weighted, undirected graph. Each pixel is a node in the graph and an edge is formed between every pair of pixels. The weight of an edge is a measure of the similarity between the pixels. The image is partitioned into disjoint sets (segments) by removing the edges connecting the segments. The optimal partitioning of the graph is the one that minimizes the weights of the edges that were removed (the cut). Shi's algorithm seeks to minimize the normalized cut, which is the ratio of the cut to all of the edges in the set.

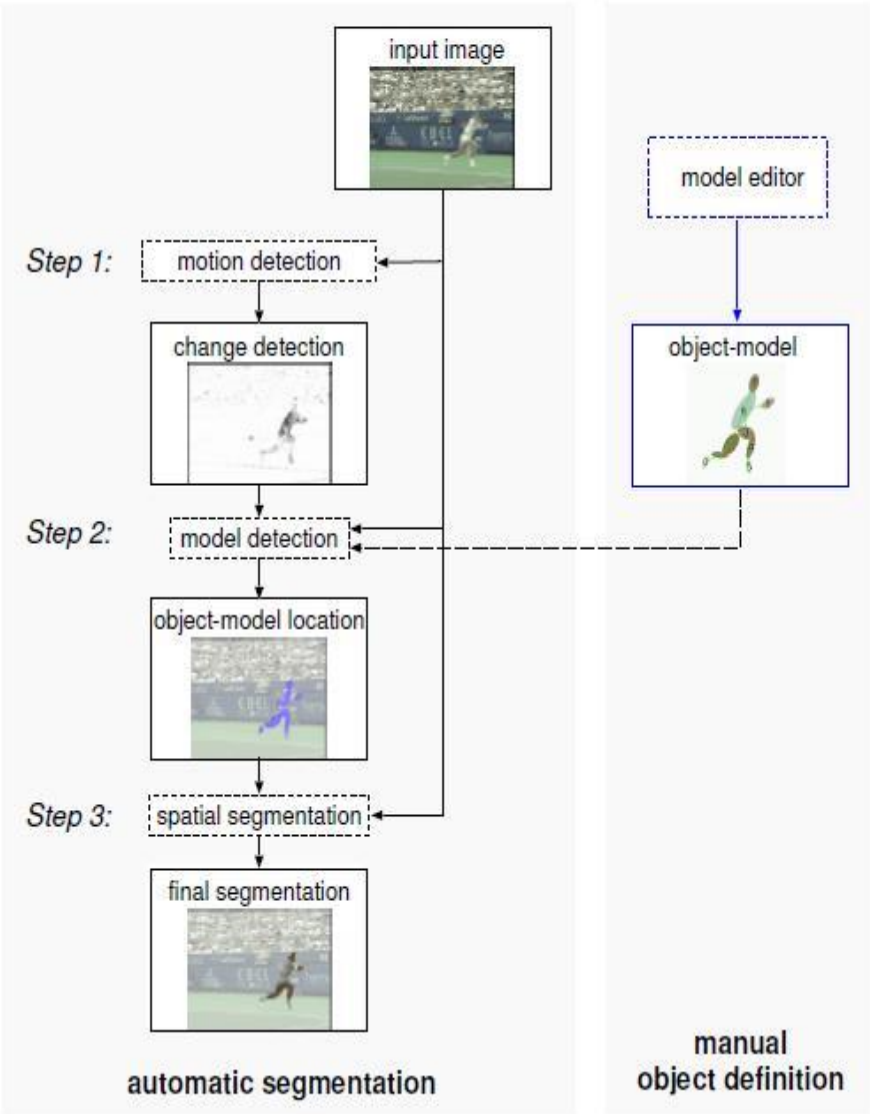


Figure 5.2 Image segmentation using graph partitioning method

5.1.2.7 Model Based Segmentation

The central assumption of such an approach is that structures of interest/organs have a repetitive form of geometry. Therefore, one can seek for a probabilistic model towards explaining the variation of the shape of the organ and then segmenting an image imposes constraints using this model as prior. Such a task involves

1. Registration of the training examples to a common pose
2. Probabilistic representation of the variation of the registered samples.
3. Statistical inference between the model and image.
4. State of the art methods in the literature for knowledge-based segmentation involve active shape, and appearance models, active contours, deformable templates and level-set based methods.

5.1.3 Global Thresholding Approach

5.1.3.1 Thresholding

Thresholding is one of the most powerful and important tools for image segmentation. The segmented image obtained from thresholding has the advantages of smaller storage space, fast processing speed and ease in manipulation compared with gray level image which usually contains 256 levels. The thresholding techniques, which can be divided into bi-level and multilevel category. In bi-level thresholding (Figure 5.3 (a)), a threshold is determined to segment the image into two brightness regions which correspond to background and object. Several methods have been proposed to automatically select the threshold. Otsu et.al formulates the threshold selection problem as a discriminate analysis where the gray level histogram of image is divided into two groups and the threshold is determined when the variance between the two groups is the maximum. Even in the case of unimodal histogram images, that is, the histogram of a gray level image does not have two obvious peaks, Otsu‘s method can still provide satisfactory result. Therefore, it is referred to as one of the most powerful methods for bi-level thresholding. In multilevel thresholding, as shown in the figure 5.3 (b), more than one threshold will be determined to segment the image into certain brightness regions which correspond to one background and several objects.

The selection of a threshold will affect both accuracy and efficiency of the subsequent analysis of the segmented image. The principal assumption behind the approach is that the object and the background can be distinguished by comparing their gray level values with a suitably selected threshold value. If background lighting is arranged so as to be fairly uniform, and the object is rather flat that can be silhouetted against a contrasting background, segmentation can be achieved simply by thresholding the image at a particular intensity level.

The simplicity and speed of the thresholding algorithm make it one of the most widely used algorithms in automated systems ranging from medical applications to industrial manufacturing. The binarized image is especially suitable as the input for hardware implementation of template matching through correlation and moment based recognition. Besides the application of thresholding in image segmentation, it is also used in various classification problems in pattern recognition. Suppose that the gray-level histogram, as shown in the Figure 5.3 (a), corresponds to an image, $f(x; y)$, composed of light objects on a dark background, in such a way that object and background pixels have gray levels grouped into two dominant modes. One obvious way to extract the objects from the background is to select a threshold T that separates these modes. Then any $(x; y)$ for which $f(x; y) > T$ is called an object point; otherwise, the point is called as background point. (Figure 5.3 (b)) shows a slightly more general case of this approach, where three dominant modes characterize the image histogram (for example, two types of light objects on a dark background).

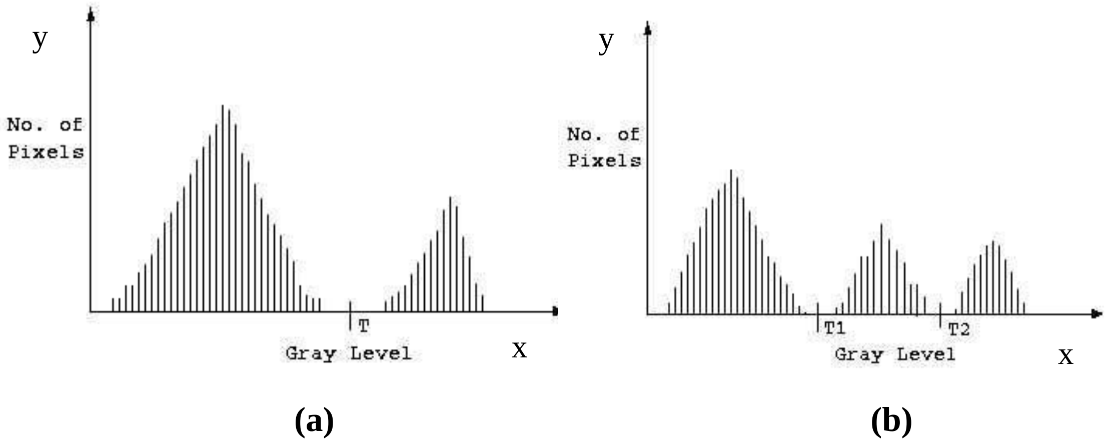


Figure 5.3 Gray-level histograms that can be partitioned by (a) A single threshold (b) Multiple thresholds

Here, multilevel thresholding classifies a point (x; y) as belonging to one object class if T1 < f(x; y) T2, to the other object class if f(x; y) > T2, and to the background if f(x; y) T1. Based on the preceding discussion, thresholding may be viewed as an operation that involves tests against a function T of the form

$$T = T[x; y; p(x; y); f(x; y)] \tag{5.1}$$

where f(x; y) is the gray level of point (x; y) and p(x; y) denotes some local property of these point-for example, the average gray level of a neighborhood centered on (x; y). A threshold image g(x; y) is defined as;

$$g(x,y) = \begin{cases} 0 - \text{if} \ - f(x,y) \le T \\ 1 - \text{if} - f(x,y) > T \end{cases} \tag{5.2}$$

Thus, pixels labeled 1 (or any other convenient gray level) correspond to objects, whereas pixels labeled 0 (or any other gray level not assigned to objects) correspond to the background. When T depends only on f(x; y) (that is, only on gray-level values) the threshold is called global. If T depends on f(x; y) and p(x; y), the threshold is called local. If, in addition, T depends on the spatial coordinates x and y, the threshold is called dynamic or adaptive.

One simple technique for finding a suitable threshold arises in situation where the proportion of the background that is occupied by objects is substantially constant in a variety of conditions. The technique that is most frequently employed for determining thresholds involves analyzing the histogram of intensity levels in the gray image.

Methods subjected to the following major difficulties

1. The valley may be so broad that it is difficult to locate a significant minimum.
2. May be a number of minimum because of the type of detail in the image, and selecting the most significant one will be difficult.
3. Noise within the valley may inhibit location of the optimum position.
4. There may be no clearly visible valley in the distribution because noise may be excessive or because the background lighting may vary appreciably over the image.
5. Either of the major peaks in the histogram (usually that dye to the background) may be much larger than the older, and then bias the position of the minimum.
6. The histogram may be inherently multimodal, making it difficult to determine the relevant thresholding level.

5.1.3.2 Otsu's Global Thresholding Method

This method is a nonparametric and unsupervised method of automatic threshold selection for image segmentation. An optimal threshold is calculated by the discriminant criterion, namely, so as to maximize the between-class variance or to minimize the within-class variance. The method is very simple, utilizing only the zeroth and first order cumulative moments of the gray level histogram

Let the pixels of a given image represented in L gray levels [1 2……. L]. The number of pixels at level i is denoted by n_i and the total number of pixels by $N = n_1+n_2+\ldots\ldots+n_L$. For simplification, the gray-level histogram is normalized and regarded as a probability distribution.

$$P_i = \frac{n_i}{N}, P_i \geq 0, \sum_{i=1}^{L} P_i = 1 \qquad (5.3)$$

To emphasize the partitioned windows technique, only Otsu's thresholding method is considered among many other techniques. This method can be stated as follows: For a given image f(x,y) with m gray levels 0, 1, . . . ,m-1, let the threshold be j, where 0< j< m-1. Then, all pixels in image f(x,y) can be divided into two groups: group A with gray level values of pixels less than or equal to j; and group B with values greater than j. Also, let (ω_1(j), M_1(j)), (Eqn.5.4,5.5) (ω_2(j), M_2(j)) (Eqn.5.6,5.7) be the number of pixels and the average gray level value in group A and group B, respectively. Then

$$\omega_i(j) = \sum_{i=n}^{j} n_i, 0 \leq j \leq m-1 \qquad (5.4)$$

$$M_1(j) = \frac{\sum_{i=n}^{j} (i, n_i)}{\omega_1(j)} 0 \le j \le m-1 \tag{5.5}$$

$$\omega_2(j) = \sum_{i=j+1}^{j=m-1} n_i, 0 \le j \le m-1 \tag{5.6}$$

$$M_2(j) = \frac{\sum_{i=j+1}^{j=m-1} (i, n_i)}{\omega_2(j)}, 0 \le j \le m-1 \tag{5.7}$$

where n_i is the number of pixels with gray level value *i*. Expressing the average gray level value M_T of all the pixels in image *f(x,y)* as

$$M_T = \frac{\omega_1(j) M_1(j) + \omega_2(j) M_2(j)}{\omega_1(j) + \omega_2(j)}, 0 \le j \le m-1 \tag{5.8}$$

The variance between the two groups, denoted as $\sigma_B^2(j)$ is

$$\sigma_B^2(j) = \omega_1(j)(M_1(j) - M_T)^2 + \omega_2(j)(M_2(j) - M_T)^2 \tag{5.9}$$

$$\sigma_B^2(j) = \frac{\omega_1(j)\omega_2(j)(M_1(j) - M_2(j))^2}{\omega_1(j) + \omega_2(j)} \tag{5.10}$$

For j ranging from ‘0’ to ‘m-1’, calculate each $\sigma_B^2(j)$ Using above Equation (5.9, 5.10), and the value j corresponding to the greatest $\sigma_B^2(j)$ is the resulting threshold.

5.1.4 Moving Object Tracking in the Video

Object tracking and detecting has been an area of intensive research in the field of video analysis and processing. Visual content in a video can be modeled as a hierarchy of abstractions. At the lowest level are the raw pixels with color information leading to lines, curves, edges, corners and regions. At the highest level are the human level concepts involving one or more objects and relationships among them. The first step in high level video processing is to identify the objects present in a scene. The next step is to see how these detected objects move with respect to each other. The above two problems combined, can be termed as “Object Tracking” and also detected to night vision “Object Tracking”

Applications like human-computer interaction, security, surveillance, video communication, compression, augmented reality, traffic control, medical image and video editing need reliable and economically affordable video tracking tools. Video tracking can be a time consuming process due to the amount of data that is contained in video scenes.

The two major sources of information in video that can be used to track objects are visual features (such as color, texture, shape and light beam) and motion information. Extraction of these two types of information can be done either in the pixel domain or in the compressed domain and also spatial or in the frequency domain.

Section 5.2 describes the overview of the related research, Section 5.5 explains the algorithm for the proposed approach, Section 5.6 includes experimental results to demonstrate the proposed method. Application and Conclusions have been included in the last section. Several vehicle tracking and detecting methods have been proposed in the literature. The main, non-mutually exclusive categories identified are: motion based tracking region-based tracking, active contour-based tracking, feature-based tracking, model-based tracking and body part-based tracking.

Karthik Hariharakrishnan et al (2005) proposes a tracking and detecting algorithm that predicts the object by predicting the object boundary using block motion vectors followed by updating the contour using occlusions and disclosing detection.

Ashwani Aggarwal et al (2006) uses background subtraction and motion estimation for tracking an object. DCT domain background subtraction in Y plane is used to locate candidate objects in subsequent I-frames after a user has marked an object of interest in the given frame. DCT domain histogram matching using Cb and Cr planes and motion vectors are used to select the target object from the set of candidate objects.

Another group of algorithms deals with object tracking using adaptive particle filters in Zhuohua Duan et al (2009) and Xi Tao et al (2010), kalman filter in Xiaoqin Zhang et al (2008) and Mehta et al (2010).Literature, Mandal et al (1999) provides a critical review of the compressed domain indexing techniques. Khansari et al (2008) proposes a tracking mechanism based on pixel features in the wavelet domain. Two-stage object tracking is performed by combining region-based method and contour-based method in Qiang Chen et al (2010).

5.2 OVERVIEW OF THE PROPOSED MOVING OBJECT TRACKING IN THE VIDEOS METHOD

The proposed scheme for vehicle objects tracking and detecting mainly deals with video surveillance applications where the camera is assumed to be fixed with a fairly wide angle of view. The object is tracked by our algorithm in subsequent frames till the object disappears from the field of view of the camera. It should be noted that, although the background is considered to be fixed. In these source is robust in the presence of variations in the vehicle lighting conditions and extraneous movements.

Block Diagram

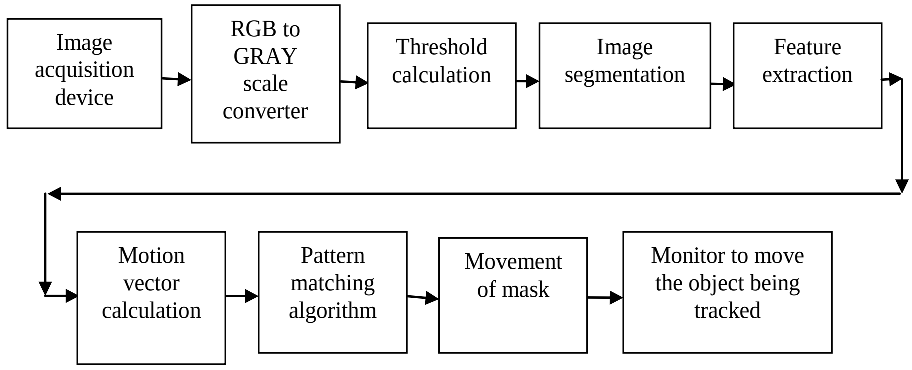


Figure 5.4 Block diagram of proposed method of object tracking

The Figure 5.4 shows the overview of the proposed method of object tracking.

1. **Image acquisition device:** Capture the frame from the video clip.
2. **RGB to GRAY scale converter:** The captured RGB space color is transformed to Y_{CbCr} space color.
3. **Threshold calculation**: The threshold value is calculated from the captured frames.
4. **Image segmentation:** The frames extracted from the video are segmented.
5. **Feature extraction:** Features of each object in the segmented frames are extracted.
6. **Motion vector calculation:** The motion vector is calculated from the feature extracted frame.
7. **Pattern matching algorithm:** Pattern matching is done on the consecutive frames.

8. **Movement of mask:** The targeted frame is masked.
9. **Monitor to move the object being tracked:** After completing the entire process, the processed to move the object is being tracked.

5.3 COLOR MODELS

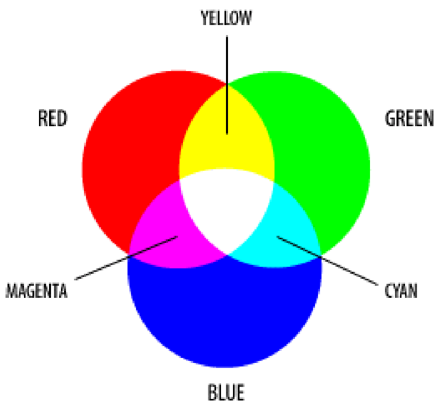
The purpose of a color model is to facilitate the specification of colors in some standard generally accepted way. In essence, a color model is a specification of a 3-D coordinate system and a subspace within that system where each color is represented by a single point.

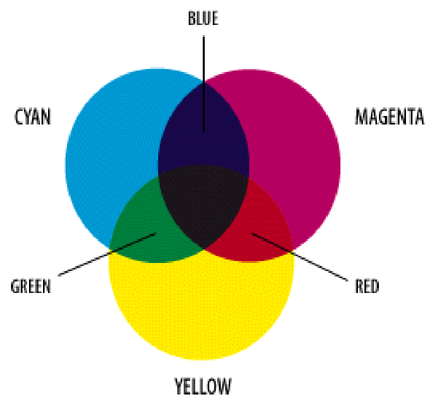
Each industry that uses color employs the most suitable color model. For example, the RGB color model is used in computer graphics, YUV or YCbCr are used in video systems, PhotoYCC is used in PhotoCD production. Transferring color information from one industry to another requires transformation from one set of values to another. Intel IPP provides a wide number of functions to convert different color spaces to RGB and vice versa.

5.3.1 RGB Color Model

In the RGB model, each color appears as a combination of red, green, and blue. This model is called as additive, and the colors also called as primary colors. The primary colors can be added to produce the secondary colors of light (see Figure 5.5 "Primary and Secondary Colors for RGB and CMYK Models") - magenta (red plus blue), cyan (green plus blue), and yellow (red plus green). The combination of red, green, and blue at full intensities makes white.

Primary and Secondary Colors for RGB and CMYK Models





(a) **(b)**

Figure 5.5 Color Models (a) RGB Model (b) CMYK Model

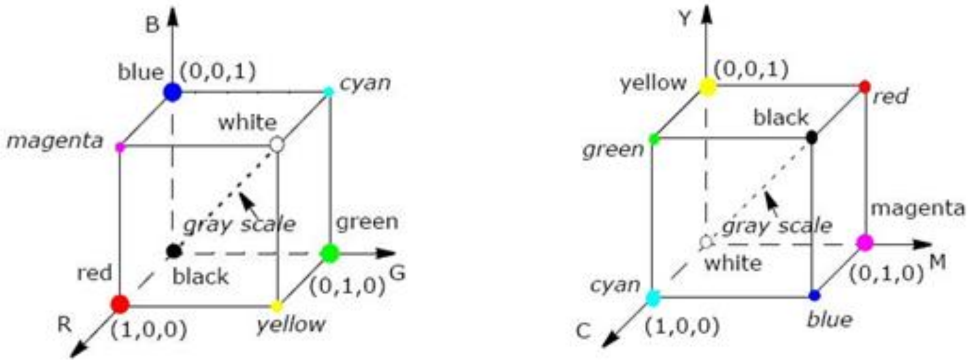
The color subspace of interest is a cube shown in the figure "RGB and CMY Color Models" (RGB values are normalized to 0.1), in which RGB values are at three corners; cyan, magenta, and yellow are the three other corners, black is at their origin; and white is at the corner farthest from the origin.

The gray scale extends from black to white along the diagonal joining these two points. The colors are the points on or inside the cube, defined by vectors extending from the origin.

Thus, images in the RGB color model consist of three independent image planes, one for each primary color. As a rule, the Intel IPP color conversion functions operate with non-linear gamma-corrected images R'G'B'. The importance of the RGB color model is that it relates very closely to the way that the human eye perceives color. RGB is a basic color model for computer graphics because color displays use red, green and blue to create the desired color. Therefore, the choice of the RGB color space simplifies the architecture and design of the system. Besides, a system designed using the

RGB color space and advantage of a large number of existing software routines, because this color space has been around for a number of years.

5.3.2 RGB and CMY Color Model



(a) **(b)**

Figure 5.6 Color Models (a) RGB Model (b) CMY Model

However, RGB is not very efficient when dealing with real-world images. To generate any color within the RGB color cube, all three RGB components need to be of equal pixel depth and display resolution. Also, any modification of the image requires modification of all three planes. The figure 5.6(a) and (b) shows RGB and CMY model.

5.4 CONTINUOUS WAVELET TRANSFORM

The Continuous Wavelet Transform (CWT) helps separate the image into parts (or spectral sub-bands) of differing importance The CWT is similar to the discrete wavelets transform: it transforms a signal or image from the spatial domain to the frequency domain.

Wavelets are mathematical functions that divide data into different frequency components, and then study each component with a resolution matched to its scale. The fundamental idea behind wavelets is to analyze signal according to scale rather than frequency. The scale is defined as a frequency inverse. Wavelets have advantages over traditional Fourier methods in analyzing physical situations where as signal contains discontinuities and sharp spikes.

Wavelet techniques can divide a complicated function into several simpler ones and separately. This property, along with fast wavelet algorithms which are comparable in efficiency to the FFT algorithm makes the wavelet techniques very attractive in analysis and synthesis problems different types of wavelets have been used as tools to solve problems in signal analysis, image analysis, medical diagnostics, geophysical signal processing, statistical analysis, pattern recognition, and many others.

The one-dimensional forward Continuous Wavelets Transform (1-D CWT) of *N* samples is formulated by

In one dimension

$f(x) \in L_2(R),$

$$\mathrm{cwt}(s,b) = \frac{1}{\sqrt{s}} \int f(x)\, \psi\left(\frac{x-b}{s}\right) dx \quad \text{(Time domain)} \qquad (5.11)$$

$$\mathrm{CWT}(s,\omega) = \sqrt{s}F(\omega)\phi(s\omega) \quad \text{(Frequency domain)} \qquad (5.12)$$

The general equation for a 2D Continuous Wavelets Transform (*N* by *M* image) CWT is defined by the following equation

In two dimension

$f(x,y) \in L_2(R),$

$$\mathrm{cwt}(s,a,b) = \frac{1}{\sqrt{s}\int\int f(x,y)\,\psi\left(\frac{x-a}{s},\frac{y-b}{s}\right)dxdy}$$

(Time domain) (5.13)

$$\mathrm{CWT}(s,\omega_1,\omega_2) = \sqrt{s}\,F(\omega_1,\omega_2)\,\phi(s\omega_1,s\omega_2)$$

(Frequency domain) (5.14)

The basic Operation of the CWT is as follows

1. Take a wavelet and compare it to a section at the start of the original signal.
2. Calculate a number C, that represents how closely correlated the wavelet is with this section of the signal. The large the number C is in absolute value the more the similarity. CWT coefficients are calculated with an inner product. Inner Products for more information on how inner products measure similarity.

If the signal energy and the wavelets energy are equal to one, C may be interpreted as a correlation coefficient. In generally, the signal embryos not equal one and the CWT coefficients are not directly interpretable as correlation coefficients.

As described in the CWT coefficients explicitly depend on the analyzing wavelets. Therefore, the CWT coefficients are different when you compute the CWT for the same signal using different wavelets.

3. Shift the wavelet to the right and repeat steps 1and step 2 until you have covered the whole signal.
4. Scale the wavelet and repeat step 1 through 3.
5. Repeat step 1 through 4 for all scales.

5.5 ALGORITHM IMPLEMENTATION OF THE MOVING OBJECT TRACKING IN THE VIDEOS

The procedure for tracking and handling occlusions can be summarized in the following steps:

1. Capture the frames from the video clip
2. Apply low pass filter
3. The captured RGB space color frame is transformed to Y_{CbCr} space
4. Divide the Y plane of the image into 8x8 blocks.
5. For each block find CWT coefficients
6. Extract CWT coefficients which results in M/8 x N/8 matrix E_i
7. Find the difference $S_i = E_{i-1} - E_i$
8. Object area is where $S_i > T$

where Y - Indicates the Y Plane of the YCbCr Image

M -Indicates the Width of Frame

N - Indicates the Height of the Frame

E_i - Indicates the i^{th} Frame

E_{i-1} Previous Frame

S_i is Difference between E_i and E_{i-1}

T is Indicates the Threshold Value

Which is experimentally selected as 4 for the test video

Threshold value calculation:

1. An initial threshold (T) is chosen; this can be done randomly or according to any other method desired.

2. The image is segmented into object and background pixels as described above, creating two sets:

    1. $G_1 = \{f(m,n):f(m,n)>T\}$ (object pixels)

    2. $G_2 = \{f(m,n):f(m,n)\leq T\}$ (background pixels) (note, f(m,n) is the value of the pixel located in the m^{th} column, n^{th} row)

3. The average of each set is computed.

    1. m_1= average value of G_1

    2. m_2= average value of G_2

4. A new threshold is created that is the average of $m_1$and m_2

    1. $T = (m_1+m_2)/2$

5. Go back to step two, now using the new threshold computed in step four, keep repeating until the new threshold matches the one before it.

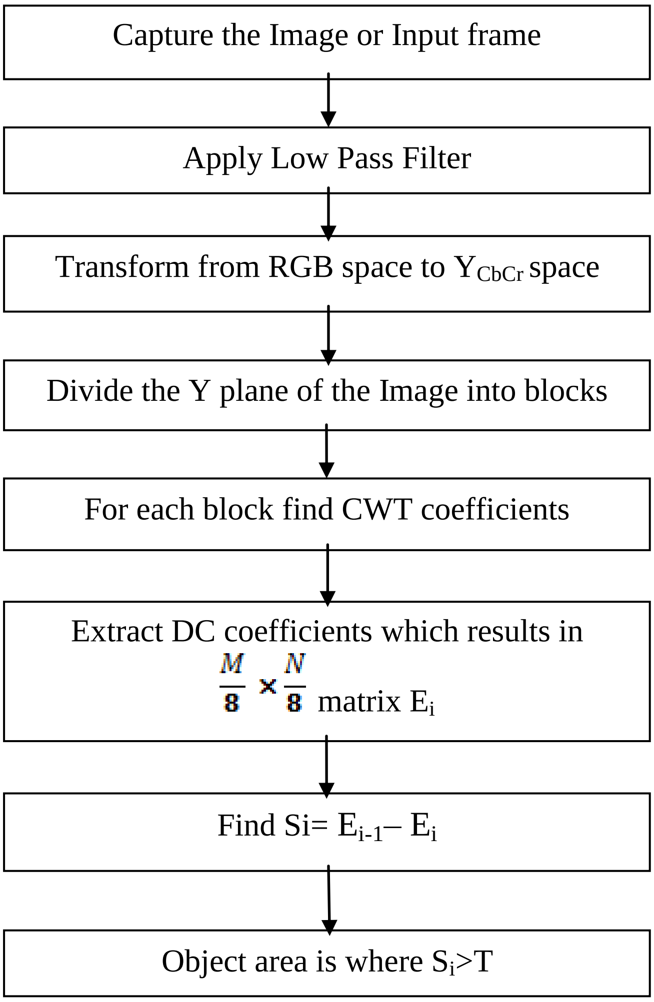


Figure 5.7 Flowchart of the moving object tracking method

5.6 SIMULATION RESULTS FOR THE PROPOSED MOVING OBJECT TRACKING IN THE VIDEOS

The algorithm described in chapter is implemented and the results are observed and analyzed. It is detected that most of the frames in the location of the moving object is detected successfully. Various kinds of moving objects were used for testing, including car, bus, human and different environment. Object tracking can be done in a number of situations. For example the tracking results of a person is shown in the figure 5.9.

Figure 5.9 (a) shows the moving object in rectangular region in Y plane of YCbCr frame converted from the original frame, Figure 5.9 (b) shows the CW coefficient image, Figure 5.9 (c) shows the regions of moving object, Figure 5.9 (d) shows the binary image of detected region of the moving object. Figures 5.11 and 5.12 shows the sample video frames for outdoor and indoor.

As seen in the figure frame wise results are shown and even though the background is considered to be fixed, our system is robust in the presence of variations in the night vision lighting conditions and extraneous movements.

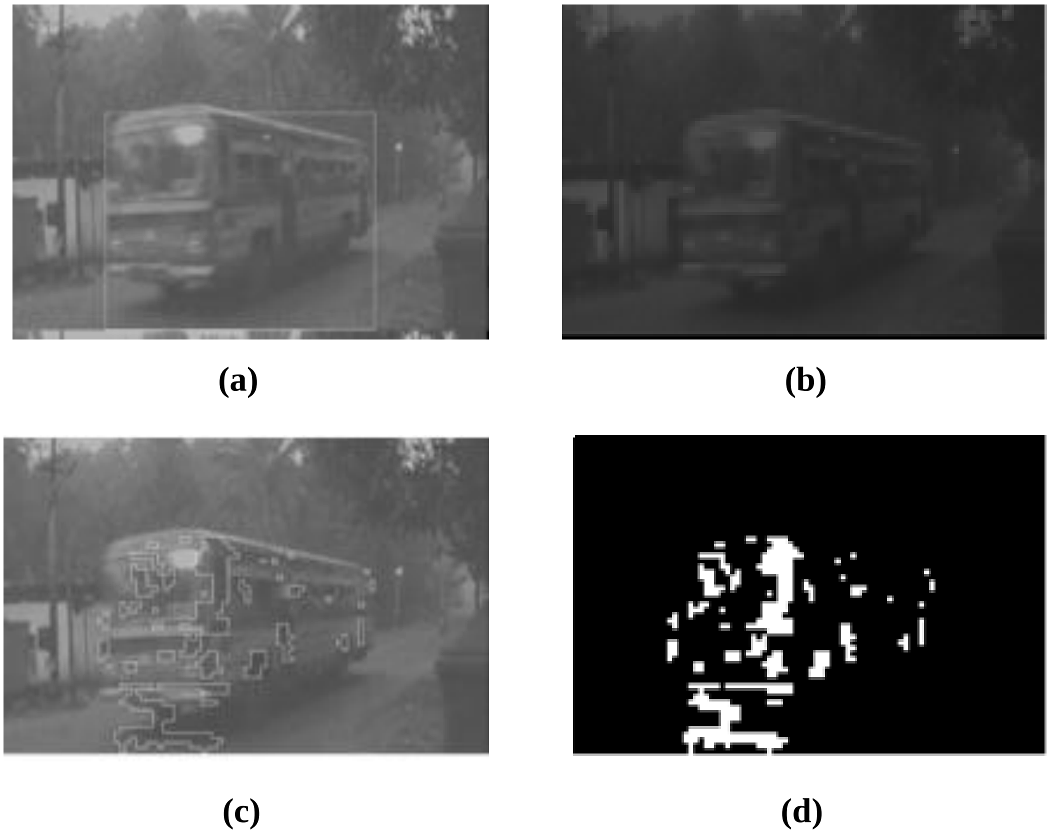
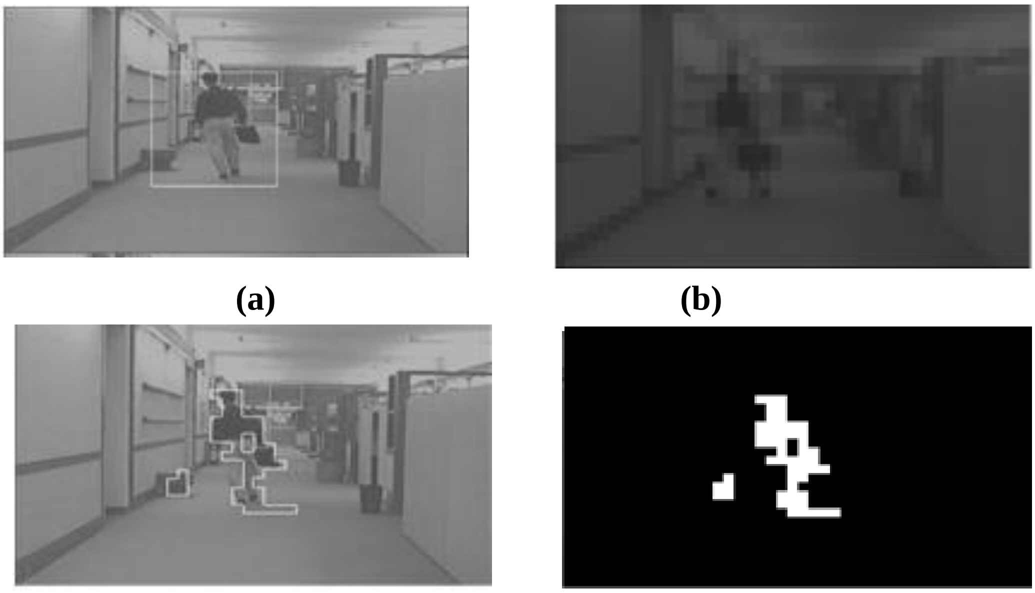


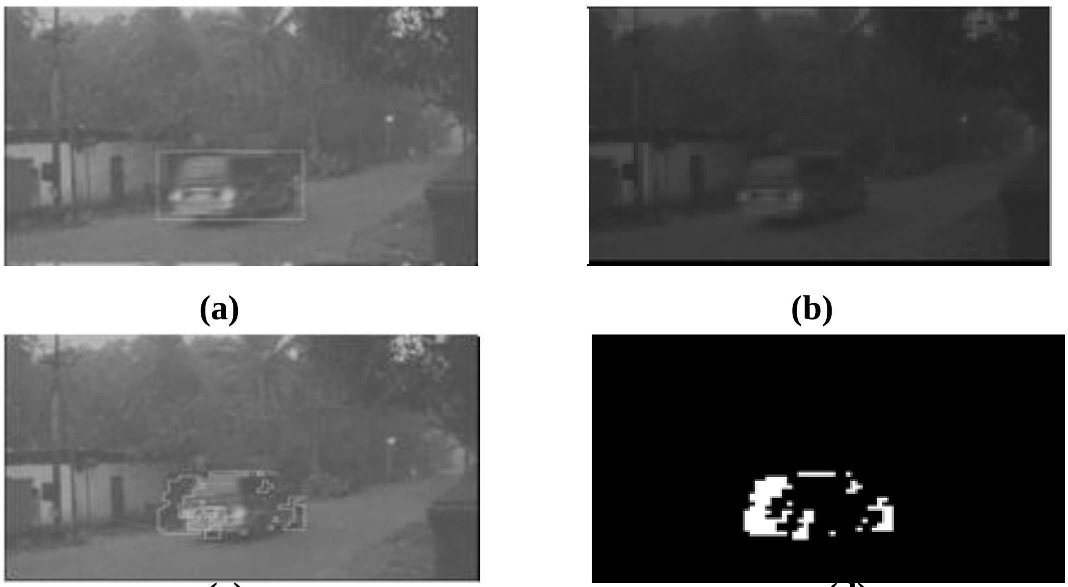
Figure 5.8 Tracking a bus in outdoor (Frame no 101) (a) Moving object in region (b) DC coefficient image (c) Regions of moving object (d) Binary image of detected region



(a) (b)

(c) (d)

Figure 5.9 Tracking a person (Frame no 76) (a) Moving object in rectangular region (b) DC coefficient image (c) Regions of moving object (d) Binary image of detected region



(a) (b)

(c) (d)

Figure 5.10 Tracking a car in outdoor (Frame no 11) (a) Moving object in rectangular region (b) CWT coefficient image (c) Regions of moving object (d) Binary image of detected region

Figure 5.11 Sample video frames ((1 to 7 and 11 to 16)) Outdoor

Figure 5.12 Sample video frames (1 to 5 and 11 to 15) Indoor

5.7 APPLICATION

Medical Imaging

1. Locate tumors and other pathologies
2. Measure tissue volumes
3. Computer guided surgery
4. Diagnosis
5. Treatment planning
6. Study of anatomical structure
7. Locate objects in satellite image (roads, forests, etc.)
8. Face recognition
9. Fingerprint recognition and Traffic control systems

5.8 CONCLUSION

The Video tracking can be a time consuming process due to the amount of data that is contained in video. A new method for object tracking with improved efficiency by reducing the number of computations is presented. Tracking and detecting is performed in the compressed domain using the CWT coefficients. In addition the noises in the signals are suppressed automatically. Since the CW coefficient is only considered the spatial and frequently. The experimental result carried out show the algorithm is very efficient. The drawback of the proposed method is our system failed to handle occlusions. Our future enhancements include handling occlusions multiple object tracking and detecting in different environments.

CHAPTER 6

SIMULATION RESULTS AND DISCUSSION

6.1 INTRODUCTION

All Simulation results are explained in this chapter, Compression Artifacts Reduction and Moving Object tracked images and video sequences are shown, here we are used different type of method used to remove artifacts and tracking the object and get the output result is given below.

1. Simulation results for the proposed compression artifacts reduction in the images.
2. Simulation results for the proposed compression artifact reduction in the videos.
3. Simulation results for the proposed moving object tracking video scene

6.2 SIMULATION RESULTS FOR THE PROPOSED ARTIFACT REDUCTION IN THE IMAGES

The image sequence is separated into compressed. Each compressed image sequence is processed by “Adaptive based Algorithm.” These images are motion estimated to find the motion vectors and get the output artifact removed image sequence. They are following the sequence of processing.

(a)

(b)

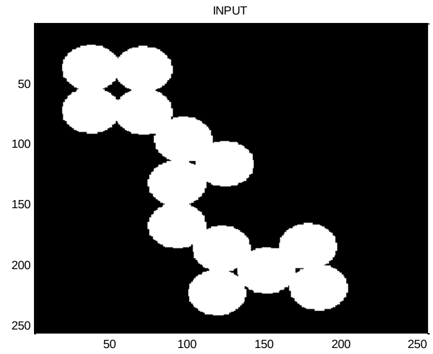
(c)

(d)

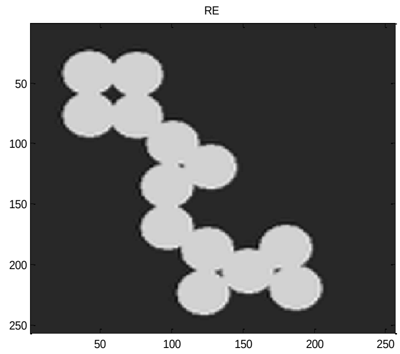
(e)

(f)

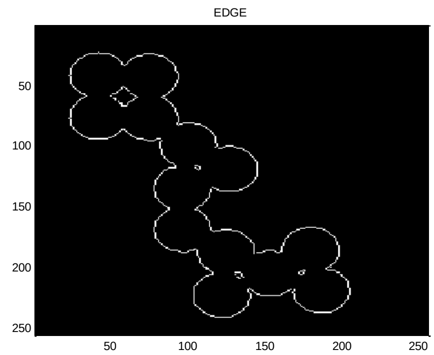
Figure 6.1 Leena.JPEG (a) Original image (b) Compressed image (c) Edge detection (d) Low pass filtered image (e) AMF filtered image (f) Output of the proposed algorithm



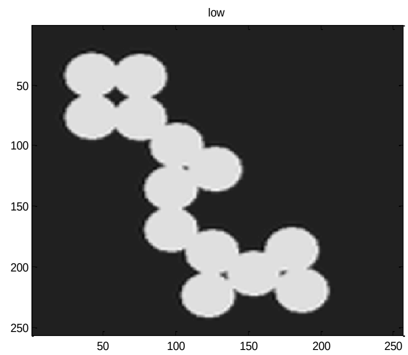
(a)



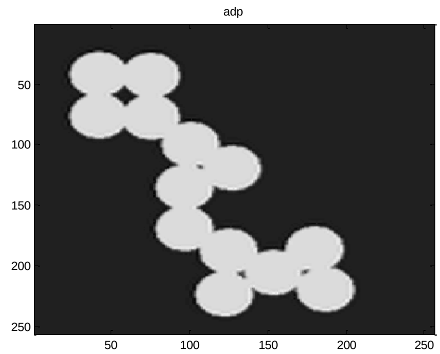
(b)



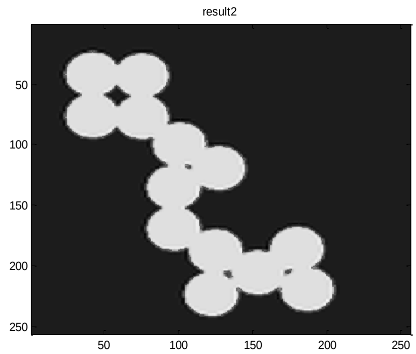
(c)



(d)



(e)



(f)

Figure 6.2 Circles.PNG (a) Original image (b) Compressed image (c) Edge detection (d) Low pass filtered image (e) AMF filtered image (f) Output of the proposed algorithm

6.3 SIMULATION RESULTS FOR THE PROPOSED ARTIFACT REDUCTION IN THE VIDEOS

The video sequence is separated into frames. Each frame is processed by “Adaptive based Algorithm.” These frames are motion estimated to find the motion vectors. The frames are then motion compensated using these vectors. The motion compensated frames are subjected to temporal filtering to get the output artifact removed frames. These following sample frames of processed sequences.

Sequence 1

(a) **(b)**

Figure 6.3 Frame no 26 (a) Before processing (b) After processing

Sequence 2

(a) **(b)**

Figure 6.4 Frame no 70 (a) Before processing (b) After processing

Sequence 3

(a) **(b)**

Figure 6.5 Frame no 101(a) Before processing (b) After processing

(a) **(b)**

Figure 6.6 Frame no 185 (a) Before processing (b) After processing

6.4 SIMULATION RESULTS FOR THE PROPOSED MOVING OBJECT TRACKING IN THE VIDEOS

In this chapter all the simulation results for all the moving object tracking techniques are explained. Initially the moving objects in video images are tracked based on image segmentation and pattern matching techniques. The Object tracking is done in three ways, those are

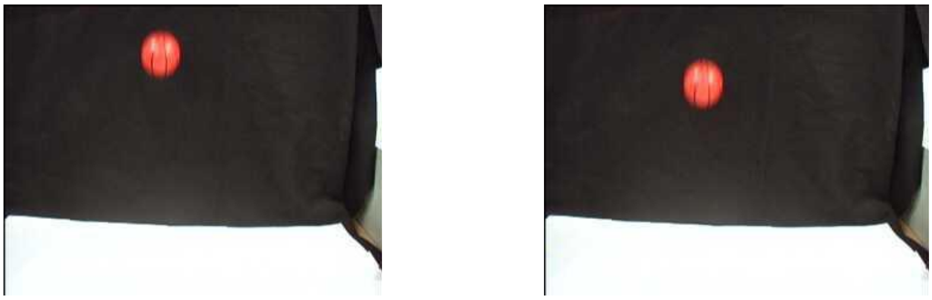
1. Histogram based thresholding by manual threshold selection and Pattern matching.

2. Segmentation by Otsu‘s global thresholding and Pattern matching.
3. Moving object tracking using YCbCr space color segmentation.

After the segmentation, it is possible to detect all the objects or connected regions in the image, no matter whether they are moving or not. Later we extract the features of the object patterns which we are interested to track. The segmentation is done for all the consecutive video images in the sequence and then sent to feature extraction process. Then feature matching (pattern matching) is done with next frame, among all the patterns the pattern matches with our desired object only and subject to the condition that desired object only satisfies minimum distance search in the feature space. i.e. after segmentation we find the distances of an object i in t^{th} frame to all the objects $(j=1,2,3….n)$in the $(t+1)^{th}$ frame. let us consider a video sequence in which 3 objects are in motion, if the frame rate is high and all the frames in the sequence are taken in to account, then one object in a frame gives minimum distance with the same object in next frame. So the object is tracked in the next frame based on two conditions one is pattern matching and the second is minimum distance search in the feature space. Once the pattern matching is over, then the motion vectors are calculated for each object in the image. But in real life nonlinear motion dominates linear motion. In video sequences the movement of the objects may be either linear or non linear, if the movement is linear then estimated positions (adding distance in x,y directions to previous centroid) in the next frame becomes the real positions, whereas nonlinear case in addition to this, distance to real position will have to be added. Then we can move the region of interest with the pattern we want to track. Here, the 3 techniques have been repeated for ball video sequence and tennis video sequence. Each image in ball video sequence is of size 320 × 240 and tennis video sequence is of size 352 × 220. In the ball video sequence the frames from 9 to 35 are taken, whereas in tennis video sequence the frames from 21 to 35 are taken. The ROI (region of interested) for ball video sequence is of size 50 ×50, and it is 16 × 16 for tennis video sequence.

The simulation results of histogram based tracking are shown below

6.4.1 Histogram based Tracking and Pattern Matching

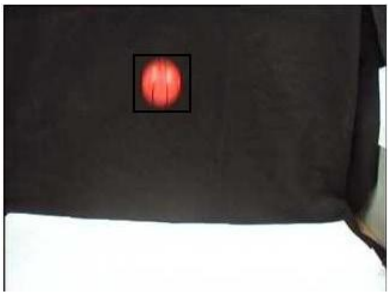


(a) (b)

Figure 6.7 (a) Frame 10 of ball video sequence (b) Frame 11 of ball video sequence

(a) (b)

Figure 6.8 (a) Segmented image of frame 10 of ball video sequence (b) Segmented image of frame 11 of ball video sequence



(a)

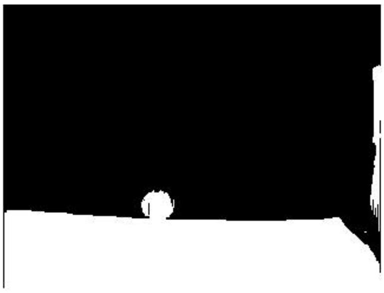
(b)

Figure 6.9 (a) Ball in frame 10 of ball video sequence is tracked (b) Ball in frame 11 of ball video Sequence is tracked

6.4.1.1 Drawback of Histogram based Thresholding

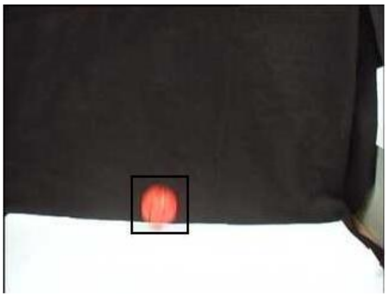
The disadvantage of histogram based thresholding is the pattern in the segmented image will not be clear. So in the segmented image shown in the figure 6.10(a), there are three patterns ball, floor and side wall. In the figure 6.10(b) also the same three patterns exists but with some variation. So may it happen sometimes that the pattern in the previous frame may not be recognized in next frame because of segmentation fault and in histogram based segmentation the threshold is selected manually, if the proper threshold is not selected then we cannot differentiate foreground and background objects. So the object tracking for desired object is not possible.

The disadvantage of this method is, it cannot be used for complex image sequences where the background is changed or the camera is moving. It can be used with video sequences with stationary camera and with less tracking efficiency.




(a) **(b)**

Figure 6.10 (a) Segmented image of frame 27 of ball video sequence (b) Segmented image of frame 28 of ball video sequence




(a) **(b)**

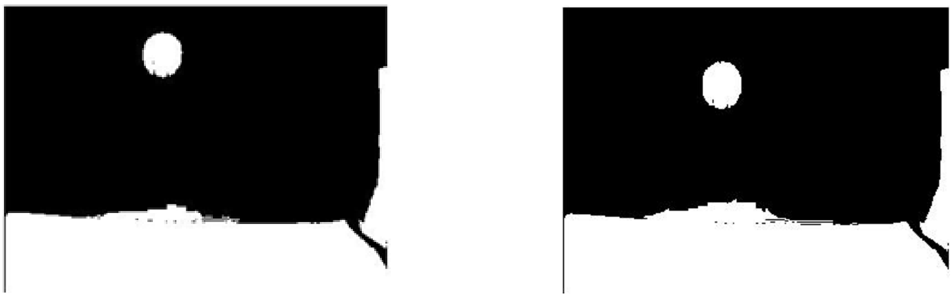
Figure 6.11 (a) Pattern matched object in frame 27 of ball video sequence (b) Pattern matching fails in frame 28 of ball video sequence

In the above figure 6.10 (a) the 3 patterns in the image are clearly visible in frame 27, but in figure 6.10 (b) the ball, wall patterns are not clearly visible, pattern matching fails here and the same is repeated for frames 52 and 57 also.

6.4.2 Tracking using Otsu's Global Thresholding

The second approach for segmentation is Otsu‘s global thresholding. In this approach the segmentation accuracy increases, but the difficulty arises when the illumination changes from one frame to another frame in video sequence as shown in the figure 6.7 (a) and 6.7 (b). It is found that there is non uniform illumination present in the image. This approach is better than histogram based segmentation.

Ball video sequence



(a) **(b)**

Figure 6.12 (a) Segmented image of frame 10 of ball video sequence (b) Segmented image of frame 11 of ball video sequence

(a) **(b)**

Figure 6.13 (a) Pattern matched object in frame 10 of ball video sequence (b) Pattern matched object in frame 11 of ball video sequence

Table 6.1 Extracted features for ball video sequence

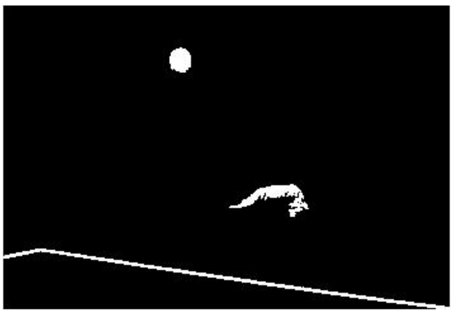
Object	Area	width	Height	cenx	ceny	mv-x	mv-y
(10,1)	20988	66	318	207	161	0	0
(10,2)	961	33	31	22	134	20	1
(10,3)	3020	151	21	128	310	0	0
(11,1)	20988	66	318	207	161	0	0
(11,2)	954	37	31	22	142	44	1
(11,3)	3029	151	21	128	310	0	0

Table 6.2 Euclidean distance between successive frames for ball video

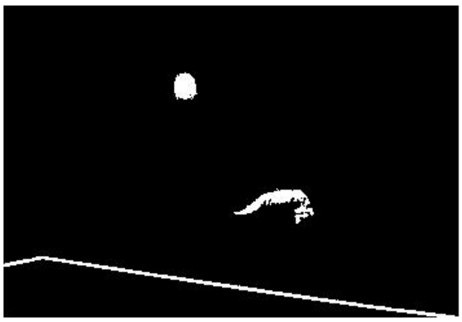
(Frame, Object)	(10,1)	(10,2)	(10,3)
(11,1)	**0**	176.3589	168.6476
(11,2)	186.9599	**20.0250**	205.4556
(11,3)	168.6476	196.7867	**0**

Tennis video sequence

In tennis video sequence, there are 2 moving objects in the sequence, one is ball and the other is hand of the player, In segmented image three patterns appears as the connected regions.



(a)



(b)

Figure 6.14 (a) Segmented image of frame 22 of tennis video sequence (b) Segmented image of frame 23 of tennis video sequence

Features are extracted for these three patterns and matched with all the patterns in the next frame, and the ROI is moved accordingly. In this case, the camera is moved in a long shot, appearance of the images in the video sequence changes, though the algorithm works. But the segmentation results are less accurate when compared to FCM-PSO method. But this method is faster than FCM-PSO, can be used for real time applications. The efficiency of the tracker is less when compared to FCM-PSO technique. Each frame in the video is of size 352 × 220, the mask regions are 24×24 for ball and 35 × 35 for hand. In this video the ball and the hand are moving so the motion vectors of these two objects are finite, whereas the edge of the table being the third pattern in the segmented image is stationary in all frames so motion vectors for this object is always zero. If there exists any stationary objects in the image then nonzero minima in all distances have to be considered, if that is not the case then minimum distance search in the feature space is to be done (Santhosh Kumar Kadarla 2009).

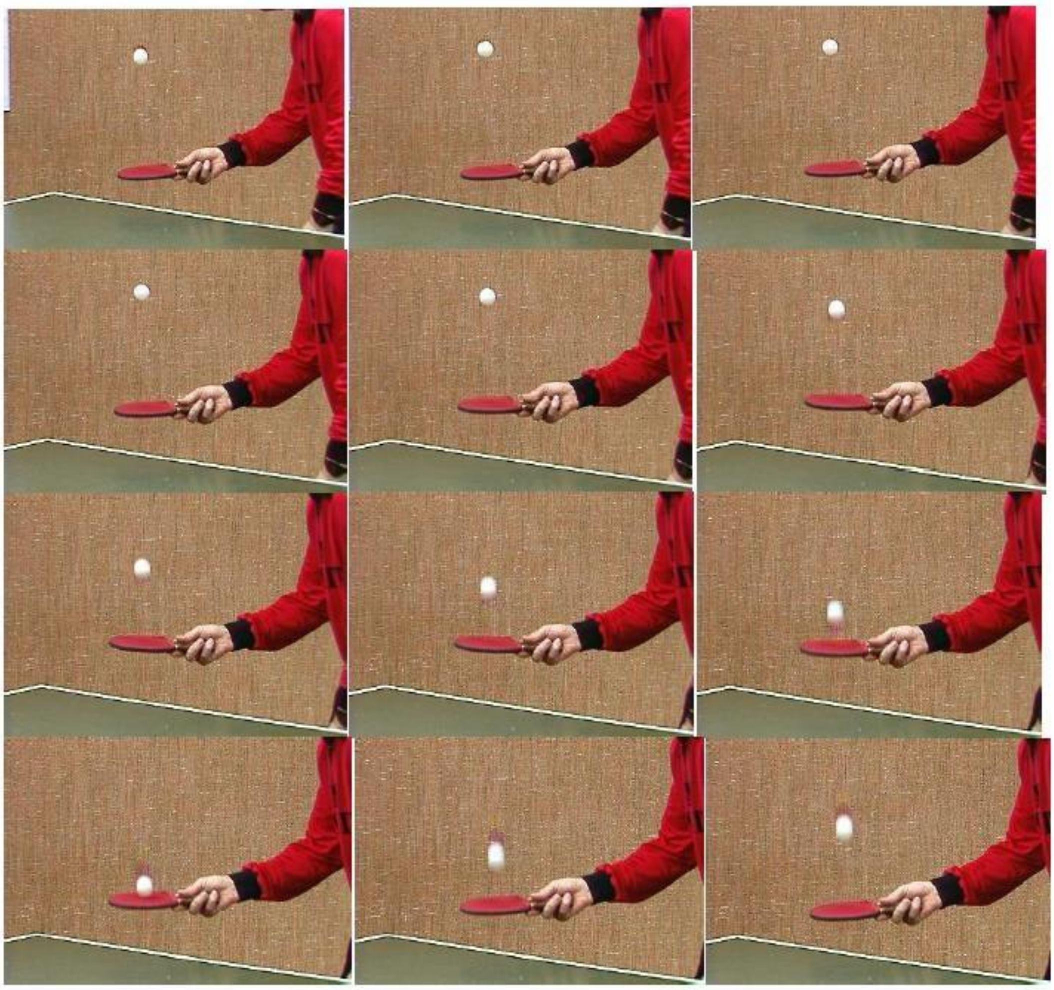


Figure 6.15 Frames 22 to 34 of tennis video sequence

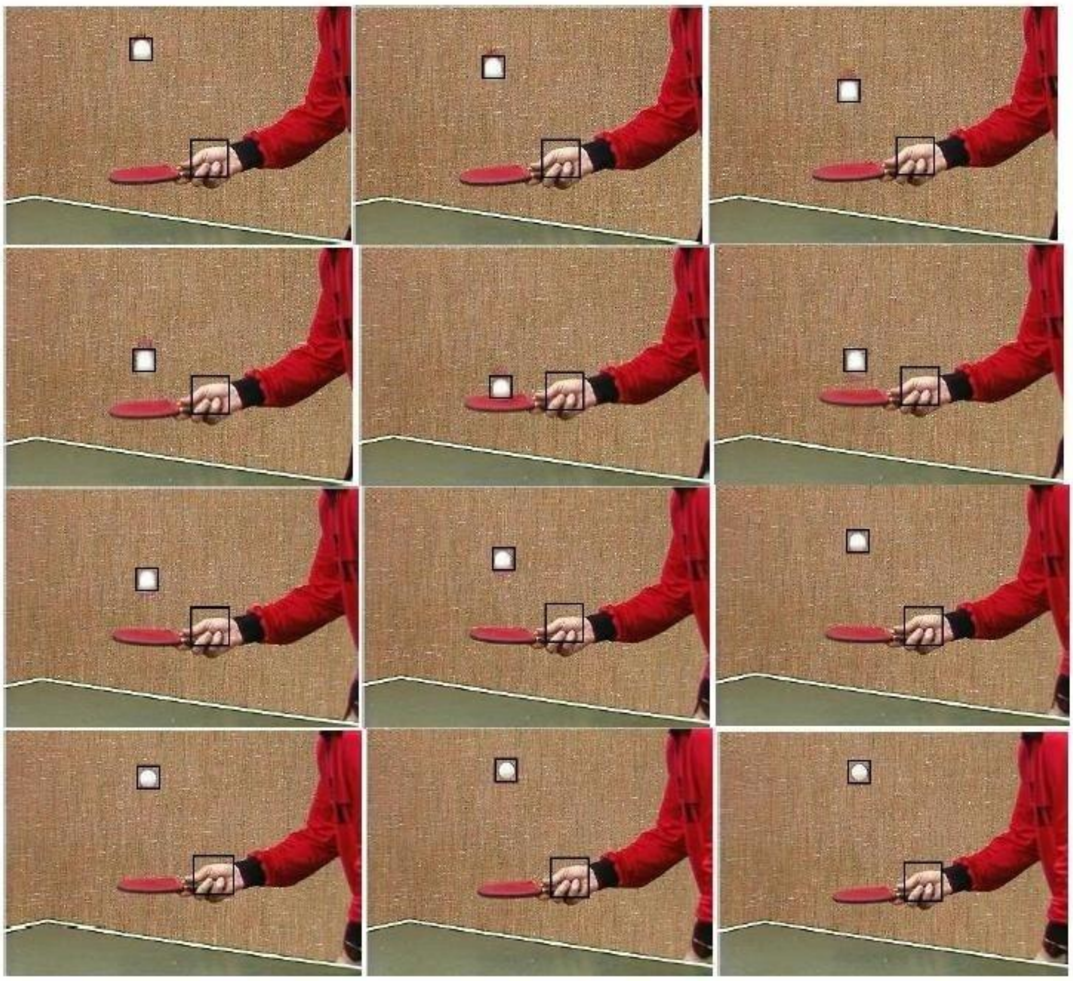


Figure 6.16 The objects tracked (ball, hand) images (frame 22 to 34) of tennis video sequence

Table 6.3 Extracted features for tennis video sequence

Frame,Object)	Area	Width	Height	Cenx	Ceny	mv-x	mv-y
(22,1)	16497	47	351	217	177	2	1
(22,2)	304	19	16	44	140	14	0
(22,3)	1550	25	62	155	210	0	0
(23,1)	16497	47	351	217	177	1	0
(23,2)	340	20	17	62	142	32	2
(23,3)	1586	61	26	54	210	0	0

Table 6.4 Euclidean distance between successive frames for tennis video

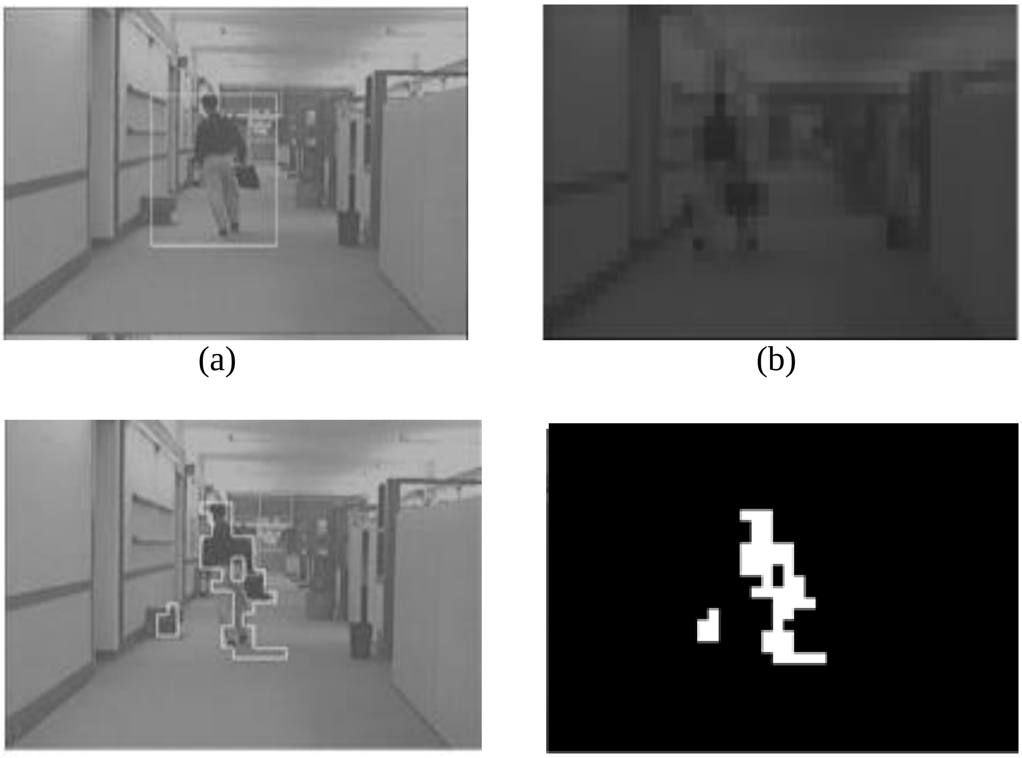
(Frame, Object)	**(23,1)**	**(23,2)**	**(23,3)**
(22,1)	**0**	158.9025	71.196
(22,3)	70.2353	115.2085	**1**

6.4.3 Moving Object Tracking using YCbCr Space Color Segmentation

This method is more robust, accurate and the efficiency of the tracker is high in all the three methods because the segmentation accuracy is dependent on the no of clusters, which is known apriori. If the clusters are more than the segmented image consists of all the finer details of the image so that several patterns exists in the segmentation result and takes some time to match the pattern from one frame to another frame. In this case, the tracker does not mislead, because the segmentation is so clear and it is very easy to perform pattern matching.

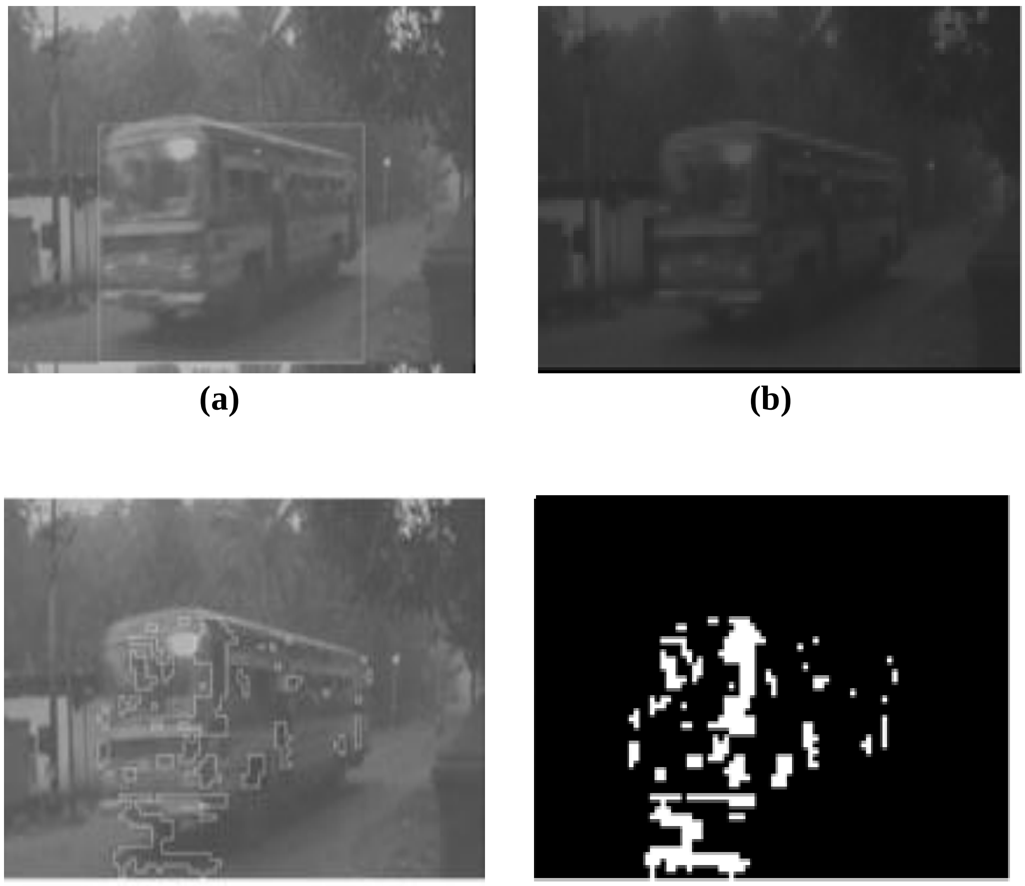
The algorithm described in this chapter is implemented and the results are observed and analyzed. It is detected that most of the frames in the location of the moving object is detected successfully. Various kinds of moving objects were used for testing, including car, bus, human and different environment.

Object tracking can be done in a number of situations. For example the tracking results of a person are shown in the figure 6.17. Figure 6.17 (a) shows the moving object in rectangular region in Y plane of YCbCr frame converted from the original frame, Figure 6.17 (b) shows the CW coefficient image, Figure 6.17 (c) shows the regions of moving object, Figure 6.17 (d) shows the binary image of detected region of the moving object. Figure 6.18 and figures 6.19 shows the sample video tracking frames outdoor. The sample video frames are shown in the figure 6.20 and figure 6.21.



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

Figure 6.17 Tracking a person in indoor (a) Moving object in rectangular region (b) DC coefficient image (c) Regions of moving object (d) Binary image of detected region

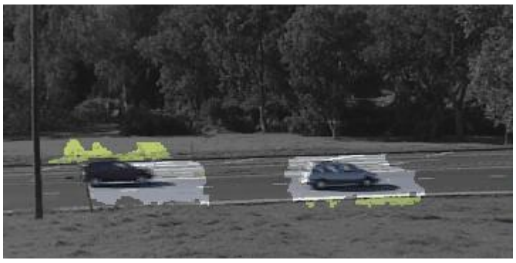


(a) **(b)**

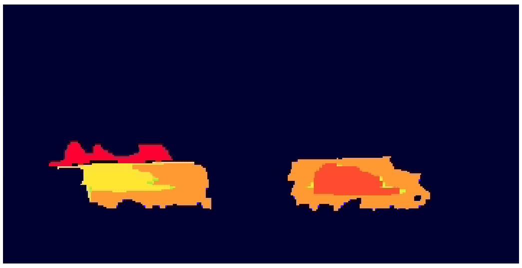
(c) **(d)**

Figure 6.18 Tracking a bus in outdoor (a) Moving object in rectangular region (b) DC coefficient image (c) Regions of moving object (d) Binary image of detected region

(a)



(b)



(c)

Figure 6.19 Tracking a car in outdoor (a) Moving object in rectangular region (b) CWT coefficient image object (c) Binary image of detected region

Figure 6.20 Sample video frames (1 to 5 and 11 to 15) Outdoor

Figure 6.21 Sample video frames (1 to 5 and 11 to 15) Indoor

CHAPTER 7

CONCLUSION

7.1 CONCLUSION

Image and video processing have become an important task and found its usage in many fields. It can be concluded that images and video degraded by artifact noises are taken in account and removed using adaptive median filter. The proposed algorithm produces better edge preservation, fine detail preservation and also significantly reduces the artifacts when compared to the median filter. Exhaustive search block matching algorithm produces better PSNR for motion estimation in motion pictures, but in case of large search windows, the computational expenses are large compared to the other motion estimation techniques. The computational expense is less for diamond search estimation. Our future enhancements include handling occlusions and multiple artifacts removed from image and video in different environments.

The Video tracking can be a time consuming process due to the amount of data contained in video. A new method for object tracking with improved efficiency by reducing the number of computations is presented. Tracking and detecting is performed in the compressed domain using the CWT coefficients. In addition the noises in the signals are suppressed automatically. Since the CW coefficient is only considered for the spatial and frequently. The experimental result carried out show that the algorithm is very efficient.

7.2 FUTURE RECOMMENDATION

The goal of thesis is to review the state of the art tracking methods, classify them into different categories and identify new trends. Object tracking, In general, is a challenging problem. Difficulties in tracking objects can arise due to abrupt object motion, changing appearance patterns of the object and the scene, nonrigid object structures, object to object and object to scene occlusions, and camera motion. Tracking is usually performed in the context of higher-level applications that require the location and/or shape of the object in every frame. Typically, assumptions are made to constrain the tracking problem in the context of a particular application. In these surveys, they are categorizing the tracking methods on the basis of the object and motion representations used, provide detailed descriptions of representative methods in each category, and examine their pros and cons. Moreover, the important issues related to tracking include the use of appropriate image features, selection of motion models, and detection of objects.

Significant progress has been made in object tracking during the last few years. Several robust trackers have been developed which can track objects in real time in simple scenarios. However, it is clear from the papers reviewed in these survey that the assumptions used to make the tracking problem tractable (for example, smoothness of motion, minimal amount of occlusion, illumination constancy, high contrast with respect to background) are violated in many realistic scenarios and therefore a limit tracker's usefulness in applications like automated surveillance, human computer interaction, video retrieval, traffic monitoring and vehicle navigation. Thus, tracking and associated problems of feature selection, object representation, dynamic shape, and motion estimation are very active areas of research and new solutions are continuously being proposed. One challenge in tracking is to develop algorithms for tracking objects in unconstrained videos, for example, videos obtained from broadcast news networks or home videos.

These videos are noisy, compressed, unstructured, and typically contain edited clips acquired by moving cameras from multiple views. Another related video domains of formal and informal meetings. These videos usually contain multiple people in a small field of view. Thus, there is severe occlusion, and people are only partially visible. One interesting solution in this context is to employ audio in addition to video for object tracking. There are some methods being developed for estimating the pointed location of audio source, for example, a person's mouth, based on four or six microphones. These audio-based localization of the speaker provides additional information which then can be used in conjunction with a video-based tracker to solve problems like severe occlusion.

Generally, an important issue that has been neglected in the development of tracking algorithms is integration of contextual information. For example, in a vehicle tracking application, the location of vehicles should be constrained to paths on the ground as opposed to vertial walls or the sky. Recent work in the area of object recognition Torralba (2003), Kumar and Hebert (2003) has shown that exploiting contextual information is helpful in recognition. In addition, advances in classifiers (Friedman et al 2000) and (Tipping 2001) have made accurate detection of scene context possible, for example, manmade structures, paths of movement and class of objects, etc. A tracker that takes advantage of contextual information to incorporate general constraints on the shape and motion of objects will usually perform better than one that does not exploit this information. These tracker designed to give the best average performance in a variety of scenarios can be less accurate for a particular scene than a tracker that is attuned (by exploiting context) to the characteristics of that scene. The used for particular feature set for tracking can also greatly affect the performance. Generally, the features that best discriminate between multiple objects and, between the object and background are also best for tracking the object.

Many tracking algorithms use a weighted combination of multiple features assuming that a combination of preselected features will be discriminative. A wide range of feature selection algorithms have been investigated in the machine learning and pattern recognition communities. However, these algorithms require offline training information about the target and/or background. Moreover, as the object appearance or background varies and the discriminative features also vary. Thus, there is a need for online selection of discriminative features. Some work has been done in these areas for online selection of individual features (Stern and Efros 2002). However, the problem of efficient online estimation of discriminative feature sets remains unresolved. One promising direction to achieve this goal is the use of the online boosting methods (Oza 2002) for feature selection. In a similar vein, most tracking algorithms use pre specified models for object representation. The capability to learn object models online will greatly increase the applicability of a tracker. Motion-based segmentation (Vidal and Ma 2004), (Black and Anandan 1996), (Wang and Adelson 1994), multibody factorization (Costeira and Kanade 1998) and (Gear 1998) methods have been used to learn models for multiple objects moving in a scene. However, these approaches assume rigid body motion. Unsupervised learning of object models for multiple nonrigid moving objects from a single camera remains an unsolved problem.

One interesting direction that has largely been unexplored is the use of semi supervised learning techniques for modeling objects. These techniques (Levin et al 2003), (Blum and Mitchell 1998), transductive SVMs (Joachims 1999), Moreover, they can not only learn nonrigid shapes and/or appearance, but they can also encode the knowledge of the background in the form of negative training data. Probabilistic state-space methods including Kalman Filters (Bar-Shalom and Foreman 1988), Dynamic Bayesian Networks (DBNs) and (Jensen 2001) have been extensively used to estimate object motion parameters. Among these methods, DBNs are probably the most general method for representation of conditional dependencies between multiple variables and/or image observations. They also provide a principled framework for fusing information from different sources. However, there is a need for more efficient solutions for inference before DBNs are more commonly used in tracking applications. Overall, we believe that additional sources of information, in particular prior and contextual information should be exploited whenever possible to attune the tracker to the particular scenario used. A principled approach to integrate these disparate sources of information will result in a general tracker that can be employed with success in a variety of applications. Very Large Scale Integration (VLSI) implementation of the proposed algorithm is possible.

The drawback of the proposed method is that system failed to handle occlusions. The future enhancements include handling occlusions and multiple object tracking and detecting in different environments.

REFERENCES

1. Aggarwal, J.K. and Cai, Q. “Human motion analysis” A review, Computer Vision and Image Understanding, Vol.73, No.3, pp.428-440, 1999.

2. Yilmaz, A. Omar Javed and Mubarak Shah, “Object tracking: A survey”, ACM Computing Surveys, Vol. 38, No. 4, Article 13, pp.1-45, 2006.

3. Amer, A. “Voting-based simultaneous tracking of multiple video objects”, In Proc. SPIE Int. Symposium on Electronic Imaging, Santa Clara, USA, pp.500-511, 2003.

4. Levin, A., Viola, P. and Freund, Y. “Unsupervised improvement of visual detectors using co-training”, In IEEE International Conference on Computer Vision (ICCV), pp.626-633, 2003.

5. Ashwani Aggarwal, Susmit Biswas, Sandeep Singh, Shamik Sural and Majumdar, A.K. “Object tracking using background subtraction and motion estimation in MPEG videos,” Springer-Verlag Berlin heidelberg, ACCV 2006, LNCS 3852, 2006.

6. Bar-shalom, Y. and Foreman, T. “Tracking and Data Association”, Academic Press Inc, 1988.

7. Black, M. and Anandan, P. “The robust estimation of multiple motions: Parametric and piecewisesmooth flow fields”, Computer Vision Image Understand., Vo.63, No. 1, pp.75-104, 1996.

8. Blum, A. and Mitchell, T. “Combining labeled and unlabeled data with co-training”, In 11th Annual Conference on Computational Learning Theory, pp.92-100, 1998.

9. Borko Furht, Joshua Greenberg and Raymond Westwater, “Motion Estimation Algorithms for Video Compression. Massachusetts”, Kluwer Academic Publishers, Chapter.2 and Chapter.3, 1997.

10. Brodsky, T. “Visual Surveillance in Retail Stores and in the Home”, Video-Based Surveillance Systems. Kluwer Academic Publishers, Boston. Chapter 4, pp.5161, 2002.

11. Cleary, J. and Witten, I. “Data compression using adaptive coding and partial string matching”, IEEE Transactions on Communications, 32(4), pp.396-402, April 1984.

12. Cleary, J. Teahan, W. and Witten, I. “Unbounded length contexts for PPM”, IEEE Data Compression Conference, 1995.

13. Cleary, J. and Teahan, W. “Unbounded length contexts for PPM”, The Computer Journal, 40(2/3), pp.67-75, February 1997.

14. Costeira, J. and Kanade, T. “A multibody factorization method for motion analysis”, International Journal of Computer Vision, Vol. 29, No. 3, pp.159-180, 1998.

15. Cox, I.J. “A review of statistical data association techniques for motion correspondence”, International Journal of Computer Vision, Vol.10, No.1, pp.53-66, 1993.

16. Crouse, C.J. and Kannan, M.R. “Simple algorithm for removing blocking artifacts in block-transform coded images”, IEEE Signal Processing Lett., Vol.5, pp.33-35, 1998.

17. Cutler, R. and Davis, L.S. “Robust real-time periodic motion detection, analysis and applications”, In IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.22, No.8, pp.781-796, 2000.

18. Donald Adjeroh, Timothy Bell and Amar Mukherjee, “The Burrows-Wheeler Transform:: Data Compression, Suffix Arrays, and Pattern Matching”, 2008.

19. Donoho, D.L. “De-Noising by Soft-Thresholding”, in IEEE Transactions on Information Theory, Vol.41, No.3, pp.613-627, 1995.

20. Foresti, G.L. “A real-time system for video surveillance of unattended outdoor environments”, IEEE Trans. Circuits and Systems for Vid. Tech., Vol.8, No.6, pp.697–704, 1998.

21. Friedman, J. Hastie, T. and Tibshirani, R. “Additive logistic regression: A statistical view of boosting,” annals of statistics, Ann. Stat., Vol.38, No. 2, pp.337-374, 2000.

22. Gavrila, D.M. “The analysis of human motion and its application for visual surveillance”, In Proc. of the 2nd IEEE International Workshop on Visual Surveillance, Fort Collins, U.S.A, pp.3-5,1999.

23. Gear, C.W. “Multibody grouping from motion images”, International Journal of Computer Vision, Vol. 29, No.2, pp.133-150, 1998.

24. Gonzalez .R and Woods .R “Digital Image Processing”, Addison-Wesley Publishing Company, pp.414-428, 1992.

25. Gopinath, R.A. Guo, H. Lang, M. and Odegard, J.E. “Wavelet-based post-processing of low bit rate transform coded”, in IEEE Int. Conf. Image Processing, 1994.

26. Haritaoglu, I. Harwood, D. and Davis, L.S. W4: “A real time system for detecting and tracking people”. In Computer Vision and Pattern Recognition, Vol.22, No.8, 1998.

27. Haritaoglu, I. “A Real Time System for Detection and Tracking of People and Recognizing Their Activities”. PhD thesis, University of Maryland at College Park, 1998.

28. Healey, G. Slater, D. Lin, T. Drda, B. and Goedeke, D. “A system for real-time fire detection”, Computer Vision and Pattern Recognition, pp.605-606, 1993.

29. Heijden, F. Image Based Measurement Systems: “Object Recognition and Parameter Estimation”, Wiley, 1996.

30. Heikkila, J. and Silven, O. “A real-time system for monitoring of cyclists and pedestrians”. In Proc. of Second IEEE Workshop on Visual Surveillance, Fort Collins, Colorado, pp.74-81,1999.

31. Horprasert, T. Harwood, D. and Davis, L.S. “ A statistical approach for real time robust background subtraction and shadow detection”, In Proc. of IEEE Frame Rate Workshop, Kerkyra, Greece, pp.1-19, 1999.

32. Hsung, T. and Lun, D.P. “Application of singularity detection for the deblocking of JPEG decoded images,” IEEE Trans. Circuits Syst. II, Vol.45, No.5, pp.640-644, 1998.

33. Hsung, T.C., Lun, D.P. and Siu, W. “A deblocking technique for block-transform compressed image using wavelet transform modulus maxima,” IEEE Trans. Image Processing, Vol.7, No.10, pp.1488-1496, 1998.

34. Hwann-Tzong Chen, Horng-Horng Lin and Tyng-Luh Liu, “Multi-object tracking using dynamical graph matching”, In Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp.210-217,2001.

35. “Information Technology – Generic Coding of Audio-Visual Objects – Part 2: Visual,”MPEG-4 standard, ISO/IEC/ JTC 1/SC29/WG 11 N 2688, Seoul, March 1999.

36. Jensen, F.V. “Bayesian Networks and Decision Graphs”, Springer, 2001.

37. Joachims, T. “Transductive inference for text classification using support vector machines”, In International Conference on Machine Learning, pp.200-209, 1999.

38. Ju, S. Black, M. and Yaccob, Y. “Cardboard people: a parameterized model of articulated image motion”, In Proc. of the IEEE International Conference on Automatic Face and Gesture Recognition, pp.38-44, 1996.

39. Junqiu wang and yasushi yagi, “Integrating colour and shape texture features for adaptive real time object tracking”, IEEE transactions on image processing, Vol.17, No.2, pp.235-240, 2008.

40. KaewTraKulPong, P. and Bowden, R. “An Improved Adaptive Background Mixture Model for Real-time Tracking with Shadow Detection”, Video-Based Surveillance Systems. Kluwer Academic Publishers, Boston, chapter 11, pp.135-144,2002.

41. Karthik Hariharakrishnan and Dan Schonfeld, “Fast object tracking using adaptive block matching”, IEEE Trans. Multimedia, Vol.7, No.5, pp.853-859, 2005.

42. Kaur, L. Chauhan, R. C. and Saxena, S. C. "Performance improvement of the SPIHT coder based on statistics of medical ultrasound images in the wavelet domain",pp.297-301,2005.

43. Kennedy, J. and Eberhart, R.C. "Particle Swarm Optimization", in Proceeding of IEEE International Conference on Neural Networks, IEEE Press, Perth, WA, pp.1942-1948,1995.

44. Kennedy, J., Eberhart, R.C. and Shi, Y. "Swarm intelligence", San Francisco, Morgan Kaufmann Publishers, 2001.

45. Khansari, M. Rabiee, H. R., Asadi, M. and Ghanbari1, M. "Object tracking in crowded video scenes based on the undecimated wavelet features and texture," EURASIP Journal on Advances in Signal Processing, Vol.8, 2008.

46. Kumar, S. and Hebert, M. "Discriminative random fields: A discriminative framework for contextual interaction in classification", In IEEE International Conference on Computer Vision (ICCV), pp.1150-1157, 2003.

47. Li Ying-hong, Pang Yi-gui, Li Zheng-xi and Liu Ya-li, "An intelligent tracking technology based on kalman and mean shift algorithm", IEEE Second Intl. Conf. on Computer Modeling and Simulation 2010, Vol.1, pp.107-109, 2010.

48. Lipton H. Fujiyoshi and Patil, R. "Moving target classification and tracking from real-time video", In Proc. of Workshop Applications of Computer Vision, pp.129-136, 1998.

49. Lipton, A.J. "Local application of optic flow to analyses rigid versus non-rigid motion", Technical Report CMU-RI-TR-99-13, Robotics Institute, Carnegie Mellon University, Pittsburgh, 1999.

50. Liu, C.B. and Ahuja, N. "Vision based fire detection", In IEEE International Conference on Pattern Recognition, Cambridge, UK,2004.

51. Loncaric, S. "A survey of shape analysis techniques. Pattern Recognition", Vol.31, No.8, pp983-1001, 1998.

52. Mahamed G.H. Omran, Ayed Salman and Andries P. Engelbrecht, "Dynamic clustering using particle swarm optimization with application in image segmentation", Pattern Analysis Applications, Vol.6, No.8, pp. 332-344, 2006.

53. Mahesh Saptharishi, John B. Hamshire II and Pradeep Khosla, "Agent-based moving object correspondence using differential discriminative diagnosis", In Proc. Of Computer Vision and Pattern Recognition, pp.652-658, 2000.

54. Mahidizadeh, E. and Tavakkoli moghaddam, R. "A hybrid Fuzzy clustering PSO algorithm for a clustering supplier problem", Proceedings of IEEE IEEM, Vol.7, pp.1466-1470, 2007.

55. Mandal, M.K., Chandni Goyal, M.C. Srivastava and R.C.Jain, "A critical evaluation of image and video indexing techniques in the compressed domain", Image and Vision Computing Journal, Vol.17, No.7, pp. 513-529, 1999.

56. McIvor, A.M. "Background subtraction techniques", In Proc. of Image and Vision Computing, Auckland, New Zealand, pp.55-60,2000.

57. McKenna, S.J., Jabri, S., Duric, Z. and Wechsler, H., "Tracking interacting people", In Proc. of International Conference on Automatic Face and Gesture Recognition, pp.348-353,2000.

58. Md. Zahidul Islam, Chi-Min Oh and Chil-Woo Lee, "Video Based Moving Object Tracking by Particle Filter", International Journal of Signal Processing, Image Processing and Pattern, Vol.2, No.1, 2009.

59. Nam Chul Kim, N. C. Jang, I. H. Kim, D. H. and Hong, W. H. "Reduction of blocking artifact in block-coded images using wavelet transform", IEEE Trans. Circuits Syst. Video Technology, Vol.8, No.3, pp.253-257,1998.

60. Oleg Michailovich and Allen Tannenbaum IEEE, "Segmentation of Tracking Sequences Using Dynamically Updated Adaptive Learning", IEEE Transactions on Image Processing, Vol.17, No.12, pp.2403-2412,2008.

61. Otsu, N. "A threshold selection method from gray-level histogram", IEEE Transactions on System Man Cybernetics, Vol.9, No.1, pp. 62-66, 1979.

62. Owens, J. and Hunter, A. “A fast model-free morphology-based object tracking algorithm”, In Proc. of British Machine Vision Conference, Cardiff, UK, pp.767-776, 2002.

63. Oza, N.C. “Online ensemble learning”, PhD Thesis, University of California, Berkeley, 2002.

64. Papageorgiou, C. Evgeniou, T. and Poggio, T. “A trainable pedestrian detection system”, In Proc. of IEEE Int. Conf. on Intelligent Vehicles, Germany, pp.241-246, 1998.

65. Phillips III, W. Shah, M. and Da Vitoria Lobo, N. “Real time object detection and tracking: Histogram matching and kalman filter approach”, IEEE Trans., Vol.5, pp.796-801, 2010.

66. Phillips III, W. Shah, M. and Da Vitoria Lobo, N. “Flame recognition in Video”, In Fifth IEEE Workshop on Applications of Computer Vision, pp.224-229, 2000.

67. Qiang Chen, Quan-Sen Sun, Pheng Ann Heng and De-Shen Xia, “Two-stage object tracking method based on kernel and active contour”, IEEE Trans. Circuits and Systems for Video Technology, Vol.20, No.4, pp.605-609, 2010.

68. Qingming Haung, Wen Gao and Wenjian Cai, “Thresholding technique with adaptive window selection for uneven lighting image”, Elsevier, Pattern recognition letters, Vol.26, No.6, pp.801-808, 2005.

69. Reeves, H.C. and Lim, J.S. “Reduction of blocky effect coding”, Opt. Eng, Vol.23, pp.34-37, 1984.

70. Collins, R.T., Lipton, A.J., Takeo Kanade, Hironobu Fujiyoshi, David Duggins and Yanghai Tsin “A system for video surveillance and monitoring”, VSAM final report. Technical report CMU-RI-TR-00-12, Robotics Institute, Carnegie Mellon University, 2000.

71. Rosales, R. and Sclaroff, S. “Improved tracking of multiple humans with trajectory prediction and occlusion modeling”. In Proc. of IEEE CVPR Workshop on the Interpretation of Visual Motion, Santa Barbara CA, 1998.

72. Santhosh Kumar Kadarla, “Object tracking in video image based on image segmentation and pattern matching”, ME Thesis, National Institute of Technology, Rourkela, 2009.

73. Saptharishi, M. Hampshire II, J.B. and Khosla, P. "Agent-based moving object correspondence using differential discriminative diagnosis", In Proc. Of Computer Vision and Pattern Recognition, pp.652-658, 2000.

74. Seol, S.W. "An automatic detection and tracking system of moving objects using double differential based motion estimation", Proc. of Int. Tech. Conf. Circ./Syst.,Comput. and Comms. (ITC-CSCC2003), pp.260-263, 2003.

75. Shajeena, J. and Ramar, K. "A Novel Way of Tracking Moving Objects in Video Scenes", IEEE Emerging Trends in Electrical and Computer Technology (ICETECT), International Conference, pp.805-810, 2011.

76. Stauffer, C. and Grimson, W.E.L. " Adaptive background mixture models for realtime tracking", In Proc. of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol.2 - 252 Vol. 2, pp.246-252,1999.

77. Stauffer, C. and Grimson, W.E.L. "Learning patterns of activity using real-time tracking", IEEE Pattern Recognition and Machine Intelligence, Vol.22, No.8, pp.747-757,2000.

78. Stern, H. and Defros, B. "Adaptive color space switching for face tracking in multi-colored lighting environments", In IEEE International Conference on Automatic Face and Gesture Recognition, pp.0249, 2002.

79. Sung-Wai, H. Yuk-Hee, C. and Wan-Chi, S. "A Practical Real-Time Post-Processing Technique for Block Effect Elimination", Image Processing, Proceedings, International Conference on, Vol.2, pp.21-24, 1996.

80. Takashi Morimoto, Osama Kiriyama and Youmei Harada, "Object tracking in video images based on image segmentation and pattern matching", IEEE conference proceedings, Vol.5, pp.3215-3218, 2005.

81. Tipping, M.E. "Sparse bayesian learning and the relevance vector machine", Journal of Mach. Learn. Resear., Vol.1, No.1,pp. 211-244, 2001.

82. Toreyin, B.U. “Moving object detection and tracking in wavelet compressed video”, Master’s thesis, Bilkent University, Department of Electrical and Electronics Engineering, 2003.

83. Torralba, A. “Contextual priming for object detection”. International Journal Computer Vision, Vol. 53, No. 2, pp.169191, 2003.

84. Tsai, D.M. and Chen, Y.H. “A fast histogram-clustering approach for multilevel thresholding”, Pattern Recognition Letters, Vol. 13, No. 4, pp. 245-252, 1992.

85. “Video Coding for Low Bit Rate Communication”, H.263 Standard, ITU-T recommendation H.263, 1998.

86. Vidal, R. and MA, Y. “A unified algebraic approach to 2-d and 3-d motion segmentation”, In European Conference on Computer Vision (ECCV), pp.1-15, 2004.

87. Wang, J. and Adelson, E. “Representing moving images with layers”, IEEE Image Process., Vol. 3, No. 5, pp.625-638, 1994.

88. Wang, L., Hu, W. and Tan, T. “Recent developments in human motion analysis, Pattern Recognition”, pp.585-601, 2003.

89. Wei Chen and Kangling Fang, “A hybrid clustering approach using particle swarm optimization for image segmentation”, ICALIP 2008, pp.1365-1368, 2008.

90. Wei Qu and Dan Schonfeld, “Real-Time Decentralized Articulated Motion Analysis and Object Tracking from Videos”, IEEE Transactions on Image Processing, Vol.16, No.8, pp.2129-2138, 2007.

91. Weina Wang, Yunjie Zhang, Yi Li and Xiaona Zhang, “The global fuzzy C means clustering algorithm”, Proceedings of 6th world congress on intelligent control and automation, June 21-23, dalian, china, pp.3604-3607, 2006.

92. Wen-Yan Chang, Chu-Song Chen and Yong-Dian Jian, “Visual Tracking in High-Dimensional State Space by Appearance-Guided Particle Filtering”, IEEE Transactions on Image Processing, Vol.17, No.7, pp.1154-1168, 2008.

93. Wixson, L. and Selinger, A. “Classifying moving objects as rigid or non-rigid”, In Proc. of DARPA Image Understanding Workshop, pp.341-358, 1998.

94. Wren, C.R., Azarbayejani, A., Darrell, T.J. and Pentland, A.P. “Pfinder: Real-time tracking of the human body”, IEEE Pattern Recognition and Machine Intelligence, Vol.19, No.7, pp.780-785, 1997.

95. XI Tao, Zhang Shengxiu and Yan shiyuan, “A robust visual tracking approach with adaptive particle filtering”, IEEE Second Intl. Conf. on Communication Software and Networks, Vol.50, No.7, pp.549-553, 2010.

96. Xiaoli Tang, Sungdae Cho and Pearlman, W.A. “Comparison of 3D set partitioning methods in hyperspectral image compression featuring an improved 3D-SPIHT”, Center for Image Process. Res., Rensselaer Polytech.Inst.,Troy,NY,USA Data Compression Conference, Proceedings.pp.25-27,2003.

97. Xiaoqin Zhang, Weiming Hu, Zixiang Zhao, Yan-guo Wang, Xi Li and Qingdi Wei, “SVD based kalman particle filter for robust visual tracking”, IEEE Trans., pp.1-4, 2008.

98. Xiong, Z., Orchard, M.T. and Zhang, Y. “A deblocking algorithm for JPEG compressed images using over complete wavelet representations”, IEEE Trans. Circuits Syst. Video Technology, Vol.7, No.2, pp.433-437, 1997.

99. Xu, M. and Ellis, T. “Colour-Invariant Motion Detection under Fast Illumination Changes”, Video-Based Surveillance Systems, Kluwer Academic Publishers, Boston, chapter 8, pp.101-111,2002.

100. Yuhua Zheng and Yan Meng, “Adaptive Object Tracking using Particle Swarm Optimization”, Proceedings of the 2007 IEEE International Symposium on Computational Intelligence in Robotics and Automation, Jacksonville, FL, USA, June 20-23,2007, pp.43-48,2007.

101. Yuk-Hee Chan, Wan-Chi Siu and Sung-Wai Hong, “Practical post processing technique for real-time block-based coding system”, IEEE Trans. Circuits, Syst., Video Technology, Vol.8, No.1, pp.4-8, 1998.

102. Yung-Kai, L., Jin, L. and Kuo, C.C.J. "Image enhancement for low-bit-rate JPEG and MPEG coding via post processing", Proc. SPIE, Vol.2727, No.3, pp.1484-1494, 1996.

103. Zhang Chuang, Chi Jian-nan, Zhang Han, Liu Yang and Yan Yan-tao, "Approach of moving objects detection in active video surveillance", Decision and Control, held jointly with the 28th Chinese Control Conference, CDC/CCC 2009. Proceedings of the 48th IEEE Conference on 2009, pp. 3130 -3136, 2009.

104. Zhang Jian-Hui, Liu Guo-Sui, Gu Hong and Su Wei-Min, "A novel transmit signal based on high range-resolution concept for FLAR or AICC system applications", Radar Processing, CIE International Conference on, pp.8184,2001.

105. Zhang Qigui and Liu Xiaoming, "Recognition and tracking of hovering object", Signal Processing Systems (ICSPS),2nd International Conference on, Vol.1pp.680 -683,2010.

106. Zhang Tao, Gui, Jun, Liu and Lijun, "The Application of Kalman Filtering in the System of Image Difference Moving Objects Tracking", Computational and Information Sciences (ICCIS), International Conference on, pp.83-86,2010.

107. Zhang Tao, Wang Xiao-yi, Liu Zai-wan and Lian Xiao-feng, "Study on moving-objects identification based on evidence theory", Intelligent Control and Automation (WCICA), 2010 8th World Congress on pp. 2610 -2613, 2010.

108. Zhang Ya, Dang Xiaoyan, Wang Wei, Wang Zhuo and Wang Zhihua, "A Robust Multiple Object Tracking Algorithm under Highly Occlusion", Computer Graphics, Imaging and Visualization (CGIV), Seventh International Conference on, pp.5-8, 2010.

109. Zhang, B., Genovesio, A. and Olivo-Marin, J.C. "Tracking of multiple fluorescent biological objects in three dimensional video microscopy", International Conference Image Processing, Vol.1, pp. 1105-1108, 2003.

110. Zhang, Q., Cai, Q., Sankaranarayanan, A., Zhang, Z. and Liu, Z., "Real time head pose tracking from multiple cameras with a generic model", Computer Vision and Pattern Recognition Workshops (CVPRW), IEEE Computer Society Conference on, pp.25-32, 2010.

111. Zhao Ding-xuan and Cui Yu-xin, “Research on the Method of Tracking and Measuring Moving Object Based on Machine Vision”, Computational Intelligence and Software Engineering (CISE), International Conference on, pp.1-4,2010.

112. Zhen Liang, Hong Fu, Zheru Chi and Dagan Feng, “Refining a region based attention model using eye tracking data”, Image Processing (ICIP), 17th IEEE International Conference on, pp. 1105, 2010.

113. Zhengya Xu, Hong Ren Wu and Xinghuo Yu, “Interest points based object tracking with controlled cameras”, Industrial Technology, ICIT 2009, IEEE International Conference on, pp.1-6, 2009.

114. Zhi-Qiang Wen and Zi-Xing Cai, “Mean Shift Algorithm and its Application in Tracking of Objects”, Machine Learning and Cybernetics, International Conference on, pp. 4024 – 4028, 2006.

115. Zhiqiang Wen and Zixing Cai, “A Robust Object Tracking Approach using Mean Shift”, IEEE Natural Computation, ICNC 2007. Third International Conference on, pp.170 -174, 2007.

116. Zhiyuan Qiao, Jingnong Weng, Zhengwei Sui, Heng Cai and Xuzhao Zhang, “A rapid visualization method of vector data over 3D terrain, Geoinformatics”, 19th International Conference on, pp.1-5, 2011.

117. Zhi-Hong Yang, Hui-Xiang Xu, Su Yu and Zhi-Chao Li, "A Method of Motion Capture from the Multi-view Video Sequence,"IEEE International Conference on Advanced Computer Theory and Engineering (ICACTE), Vol.4, V4-492-V4-495, 2010.

118. Ziv, J. and Lempel, A. “A universal algorithm for sequential data compression”, IEEE Transactions on Information Theory, 23(3), pp. 337–343, May 1977.

119. Ziv, J. and Lempel, A. “Compression of individual sequences via variable-rate coding”, IEEE Transactions on Information Theory, 24(5), pp. 530– 536, September 1978.

120. Ziyou Xiong, Zhen Jia, Hongcheng Wang, Caballero.R, Jianwei Zhao and Finn.A, “A Two-step approach to see-through bad weather for surveillance video quality enhancement” IEEE International Conference on Robotics and Automation (ICRA),pp. 5309 – 5314,2011.

121. Zhou Xian Cheng, Shen Quan-Tai and Liu Li Mei, “New two dimensional fuzzy C means clustering algorithm for image segmentation”, Springer link, Vol.15, pp.882-887, 2008.

122. Zhou, Hao, Xuejie Zhang, Haiyan Li and Jidong Li, “Video object tracking based on swarm optimized particle filter”, Industrial Mechatronics and Automation (ICIMA), 2nd International Conference on, Vol.2, pp.702-706, 2010.

123. Zhuohua Duan, Zixing Cai and Jinxia Yu, “Occlusion detection and recovery in video object tracking based on adaptive particle filters”, IEEE Trans. China, pp.466-469, 2009.

124. Zhu Hong , Guixia Guan and Minhua Wu, “An efficient estimation approach for moving object tracking with correlated measuring noises” Artificial Intelligence, Management Science and Electronic Commerce (AIMSEC),2nd International Conference on Coll. of Inf. Eng., Capital Normal Univ., Beijing, China,pp.955 – 958, 8-10 Aug. 2011.

125. Zhu.J, Yuanwei Lao and Zheng, Y.F, “Object Tracking in Structured Environments for Video Surveillance Applications”Circuits and Systems for Video Technology, IEEE Transactions on, pp.223-235, Feb. 2010.

LIST OF PUBLICATIONS

International Journals

1. **Purushothaman, A.,** Rangarajan, R., Shankar Kumar, K.R. and Kandaswamy, A. "Compression Block Artifact Reduction in Low Bit Rate Image and Video" **European Journal of Scientific Research (EJSR),** Volume 64, Issue 2,pp. 244-253, 2011.

2. **Purushothaman, A.,** Rangarajan, R., Shankar Kumar, K.R. and Kandaswamy, A. "Compressed Novel Way of Tracking Moving Objects in Image and Video Scenes" **European Journal of Scientific Research (EJSR),** Volume 64.Issue 3,pp.353-360, 2011.

3. **Purushothaman, A.,** Rangarajan, R., Shankar Kumar, K.R. and Kandaswamy, A. "Compressed Video Based distance traffic analysis: Applications to vehicle detecting and tracking and counting" **European Journal of Scientific Research (EJSR)** [Accepted for Publication], 2011.

4. **Purushothaman, A.,** Rangarajan, R., Shankar Kumar, K.R. and Kandaswamy, A. "Block Artifact and Edge Reduction in low bit rate Image and video coding" **ICGST International Journal on Graphics, Vision and Image Processing (GVIP)**[Accepted for Publication], 2011.

5. **Purushothaman, A.,** Rangarajan, R. and Shankar Kumar, K.R. "Compression Block Artifact Reduction in Low Bit Rate Images and Video Coding" **(CSC Journal) International Journal of Image Processing**(**IJIP-446),** Volume 5, Issue 5. [Accepted for Publication], 2011.

6. **Purushothaman, A.,** Rangarajan, R. and Shankar Kumar, K.R. "Low Bit Rate Video Coding Algorithms-A Survey" **(CSC Journal) International Journal of Image Processing (IJIP-448),** Volume 5, Issue 5. [Accepted for Publication], 2011.

7. **Purushothaman, A.,** Rangarajan, R. and Shankar Kumar, K.R. "Compressed artifacts reduction novel way moving Object Tracking in Image and Video Scenes using DWT" **EURASIP Journal on Image and Video Processing** [Under Review], 2011.

International Conference

8. **Purushothaman, A.,** Rangarajan, R. and Shankar Kumar, K.R. "Block Artifact Reducation in low bit rate Image and video coding" **Proceedings of the International Conference on Emerging Trends in Computing (ICETC-2011)** 17-18,pp.43-49. March,2011.

9. **Purushothaman, A.,** Rangarajan, R. and Shankar Kumar, K.R. "Compression Artifact Removal Image and Video Coding using Motion Estimation Technique" **Proceedings of the International Conference on Emerging Trends in Computing ICETC-2011)** 17-18, pp.50-59. March, 2011.

10. **Purushothaman, A.** and Shankar Kumar, K.R. "Compression Artifacts Removal in low Bit Rate Image and Video Coding Algorithms" **International Conference on Advances in Information, Commincation Technology and VLSI Design (ICAICV),** id (ICA092), pp.74-75.

11. **Purushothaman, A.** and Shankar Kumar, K.R. "Compression Block Artifacts Removal in Low Bit Rate Image and Video Coding" **International Conference on Advances in Information, Communication Technology and VLSI Design(ICAICV),** id (ICA093),pp.76.

National Conference

12. **Purushothaman, A.,** Rangarajan, R. and Shankar Kumar, K.R. "Compression Artifact Reducation in low bit rate Image and video coding" **National Conference on Emerging Trends in Communication and Signal Processing (NCETCSP-2011),** id: DIP11, pp 34, 2011.

CURRICULUM VITAE

PURUSHOTHAMAN A received the B.E. degree from the Department of Electronics and Communication Engineering, Anna University, Chennai, Tamil Nadu, India, in 2006, and the M.E.(Communication System) degree from the Department of Electronics and Communication Engineering, Anna University, Chennai, Tamil Nadu, India, in 2008. He is currently pursuing the Ph.D. degree from the Department of Electronics and Communication Engineering, Anna University of Technology, Coimbatore, Tamil Nadu, India. His areas of Image and Video Processing, Digital Signal Processing, interest include Mathematics. He has published 5 papers in reputed National / International Journals and more than 6 papers in National/International Conferences. He has obtained a patent on "Compression Artifact Reduction and Moving Object Tracking in the Images and Videos".