Robust Watermarking for Scalable Video Sequence

Thesis submitted to the
Indian Institute of Technology, Guwahati
for the award of the degree

of

Doctor of Philosophy
in
Computer Science and Engineering

Submitted by
Nilkanta Sahu

Under the guidance of
Dr. Arijit Sur

Department of Computer Science and Engineering

Indian Institute of Technology Guwahati

2015

Dedicated to my family and friends

Whose blessings, love and support made my path to success

DECLARATION

I certify that

a. the work contained in this thesis is original and has been done by me under the guidance of my supervisor.

b. the work has not been submitted to any other Institute for any degree or diploma.

c. I have followed the guidelines provided by the Institute in preparing the thesis.

d. I have conformed to the norms and guidelines given in the Ethical Code of Conduct of the Institute.

e. whenever I have used materials (data, theoretical analysis, figures, and text) from other sources, I have given due credit to them by citing them in the text of the thesis and giving their details in the references. Further, I have taken permission from the copyright owners of the sources, whenever necessary.

Nilkanta Sahu

Copyright

Signature of Author.....

Nilkanta Sahu

Certificate

This is to certify that the project work entitled **"Robust Watermarking for Scalable Video Sequence"** being submitted to Department of Computer Science and Engineering, Indian Institute of Technology Guwahati by **Nilkanta Sahu**, in partial fulfillment for the award of the degree of Doctor of Philosophy in Computer Science and Engineering, is a bonafide work carried out by him under my supervision. To the best of my knowledge it has not been submitted elsewhere for award of degree

Date:

Place:

.....

Dr. Arijit Sur

Assistant Professor

Department of Computer Science and Engineering

IIT Guwahati

Acknowledgments

A great many people have contributed to production of this dissertation. I owe my gratitude to all those people who have made this possible.

I wish to express my deepest gratitude to my adviser, Dr. Arijit Sur. I have been fortunate to have an advisor who gave me the freedom to explore on my own, and at the same time the guidance to recover when my steps faltered. His patience, support constant motivation helped me overcome many crisis situations and finish this dissertation.

Besides my advisor, I would like to thank the rest of my thesis committee: Prof. S V Rao, Prof. P K Bora and Dr. Pinaki Mitra, for their insightful comments and encouragement. Their constructive criticism and suggestions helped me to widen my research from various perspectives.

I would also like to acknowledge the services and support of the Staff of Dept. of Computer Science and Engineering, IITG for providing access to valuable resources and extending all necessary support for the successful completion of my research work.

I am grateful to all my seniors, friends and juniors especially Shrinivasa sir, Ashok sir, Mamata di, Pravati, Shishendu, Mayank, Shashi, Shilpa, Basant, Sibaji, Satish, Rana, Anirban and many others for their unconditional help and support. You made my life at IIT Guwahati a memorable.

Most importantly, none of this would have been possible without the love and patience of my family. My family has been a constant source of love, concern, support and strength all these years.

Abstract

With the emergence of the scalable video coding (SVC), efficient and secure transmission of scalable video stream becomes an important research topic. In the recent literature, watermarking is regarded as an efficient tool for scalable video authentication. Primary motivation of this entire dissertation, is to develop robust watermarking solutions for different scalable adaptations like resolution, temporal and quality.

In the first part of this work, watermarking issues for resolution and quality scalability have been considered. It has been observed in the literature that there are two basic requirements for the scalable watermarking, firstly the watermark should be extracted from each of the scalable layers and secondly, reliability of the extracted watermark should be increased with the increase of the video quality layers i.e. achieving graceful improvement. In this context, an uncompressed domain watermarking scheme has been proposed to meet both the requirements while maintaining the decent visual quality of the watermarked video.

It is observed that the temporal adaptation is also a serious problem for designing robust scalable watermarking. In the next phase of this work, a robust algorithm has been devised where DCT based motion compensated temporal filtering is used to handle the temporal adaptation. A wavelet based spatial filtering is also used for embedding zone selection to achieve an acceptable visual quality.

Although, the proposed scheme against the resolution scalability outperforms recent existing schemes, its performance can be improved, especially when the resolution scaling is relatively large. In the third

phase of the work, a scale invariant feature transformation (SIFT) based image watermarking has been proposed which can easily be extended to the frame based video watermarking. The proposed scheme exploits the scale invariant property of the SIFT feature to devise a robust algorithm when the resolution scaling is relatively high.

It can be observed that the proposed algorithm against temporal adaptation, in the second part of the thesis, mostly outperforms the existing schemes, but it requires a location map for the extraction of the watermark which is an extra overhead. To take away this extra overhead, two schemes have been proposed in the final phase of this thesis, which require no location map for the watermark extraction. In the first scheme, a SIFT based watermarking algorithm is proposed which is invariant to the temporal scaling and performs well against temporal adaptation and any frame dropping and averaging attacks. In the second scheme, the frames of each temporal layer have been embedded with a different watermark which is generated by block DCT decomposition of a single watermark image to achieve graceful improvement in the successive enhancement layers.

Finally, the thesis concludes by briefly summarizing the works presented in the dissertation and explaining the possible future research directions.

Keywords : Watermarking, scale invariant watermarking, RST, content adaptation, SVC, MCDCT-TF, SIFT, visual saliency, wavelet, block DCT, base layer, enhancement layer.

Abbreviation

BIR	Bit Increase Rate
DCT	Discrete Cosine Transform
DFT	Discrete Fourier Transform
GOF	Group of Frames
GOP	Group of Pictures
GPE	Global Perceptual Error
HD	High Definition
HVS	Human Visual System
IMCDCT-TF	Inverse Motion Compensated DCT based Temporal Filtering
JSVM	Joint Scalable Video Model
MC	Motion Compensation
MCDCT-TF	Motion Compensated DCT based Temporal Filtering
MCTF	Motion Compensated Temporal Filtering
PSNR	Peak Signal to Noise Ratio
RST	Rotation Scaling Translation
SIFT	Scale Invariant Feature Transform
SSIM	Structural Similarity
SVC	Scalable Video Coding
VQM	Video Quality Metric

List of symbols

σ	Scale of Gaussian filter
α	Watermark strength
δ	Change in intensity
V	Input video/ Original video
V^w	Watermarked Video
I	Input Image/ original image
I^w	Watermarked Image
R	Residual Layer
Rw	Watermarked residual layer
D	SIFT descriptor of original image/frame
D'	SIFT descriptor of watermarked image/frame
D^w	Watermark descriptor
V_{th}	Visual quality threshold
$\|x\|$	modulus of x
$A \backslash B$	Set difference of A and B
W	Watermark signal
$H^{(x \to y)}, V^{(x \to y)}$	Motion Vector from frame x to frame y
$Lmap$	Location map

List of Figures

1.1 Use of scalable video 2
1.2 Different Scalability. 3
1.3 Video Watermarking 4
1.4 A basic Scalable Video Watermarking scenario 5
1.5 Block Diagram for Watermark Embedding [39] 11
1.6 Block Diagram for Watermark Extraction[39] 12
1.7 Block diagram of Meerwald's [44] watermark embedding method for two spatial layers. 13

2.1 Motion Compensated Temporal Filtering(Figure courtesy [54]). . . 20
2.2 DCT based MCTF (L, M,H are low middle and high frequency frames respectively) 21
2.3 Connected and unconnected pixel, (a)Fully connected pixels, (b)Unconnected pixels, (c)Partially connected pixel, (d)One to many connection . 23
2.4 RARE2012 flow chart [58] 27

3.1 Block DCT of video frame 35
3.2 Location Map based Technique for Spatial Coherency 37
3.3 Watermark Embedding Model 38
3.4 Watermark Extraction Model 44
3.5 PSNR comparison 50
3.6 Flicker Metric comparison 51
3.7 VQM comparison 52

3.8 SSIM comparison 53
3.9 Robustness comparison 58

4.1 Spatial Decomposition 62
4.2 DCT based Motion Compensated Temporal Filtering (MCDCT-TF) 62
4.3 Inverse motion compensation of the Location Map 64
4.4 Pixel categories and Location Map 65
4.5 Watermark Embedding Model 68
4.6 Frames after every step 68
4.7 Watermark Extraction Model 68
4.8 PSNR comparison 71
4.9 Flicker comparison 72
4.10 SSIM comparison 73
4.11 VQM comparison 74
4.12 Robustness at different Temporal layer 74
4.13 Robustness comparison 75

5.1 Lena Binary Image 81
5.2 Plot for finding best possible β 82
5.3 Robustness Comparison with scheme [24] (red) and [38](black) . 86
5.4 Matching ratio of watermark descriptor with original descriptor in previous scheme 90
5.5 Watermark Embedding Scheme 90
5.6 Visual Degradation Comparison between proposed scheme and existing schemes for Lena 92
5.7 Robustness Comparison between proposed scheme and existing schemes for Boy 93
5.8 Robustness Comparison between proposed scheme and existing schemes for Serano 93
5.9 Embedding of watermark for *baboon, barbara* and *lena*. Top row shows the original images. Middle row shows the watermarked images with patch. Bottom row shows the newly generated SIFT points due to insertion of patch. 94
5.10 Plot depicting variation of intensity with stability 97

5.11 Plot depicting variation of intensity with perceptual error 98
5.12 Plot depicting variation of stability with perceptual error 101
5.13 Variation of robustness with change in intensity of the patch by 20 102
5.14 Variation of GPE with change in intensity of the patch by 20 . . . 102

6.1 Side Plane and Embedding zone 104
6.2 Embedding Zones 105
6.3 Motion Map 106
6.4 New SIFT features for Smooth Area and Busy Area 108
6.5 Selection of blocks belonging to relatively smoother region 109
6.6 Zone Selection Procedure 109
6.7 Temporally adapted video and corresponding resizing for extraction 110
6.8 Robustness comparison with Chong's scheme [67] for City video . 114
6.9 PSNR comparison of the watermarked frames 115
6.10 SSIM comparison of the watermarked frames 116
6.11 Flicker comparison of the watermarked frames 116
6.12 Watermark Generation 118
6.13 DCT coefficients in zigzag scan 120
6.14 Watermark Embedding 121
6.15 Watermark Extraction 121
6.16 Graceful Improvement 122
6.17 Robustness comparison 123
6.18 PSNR comparison 124

List of Algorithms

3.1 Embedding Algorithm (V, α, W) 42
3.2 Residual Embedding $(R, \alpha, W, Lmap, MV)$ 43
3.3 Extraction Algorithm (V^w, α) 46
3.4 Residual Extraction $(Rw, \alpha, Lmap, MV)$ 47

4.1 $Embed_{bit}(A, B, wb, \delta)$ 64
4.2 Embedding Algorithm (V, α, W) 67
4.3 Extraction Algorithm (V^w) 69

5.1 Watermark Zone selection 80
5.2 Watermark Embedding 83
5.3 Watermark Extraction & Authentication 84
5.4 Watermark Zone selection 88
5.5 Watermark Embedding 91

6.1 Watermark Embedding 111
6.2 Watermark Extraction & Authentication 112

List of Tables

2.1 Experimental Data set 31

3.1 Experimental Setup 48
3.2 Hamming distance of the extracted watermark from scalable CIF Bus video 54
3.3 Hamming distance of the extracted watermark from scalable CIF Akiyo video 55
3.4 Hamming distance of the extracted watermark from scalable HD Pedestrian area video 56
3.5 Hamming distance of the extracted watermark from scalable HD Sunflower video 57

4.1 Experimental Setup 70

5.1 GPE for Standard Images 85
5.2 Average Robustness for all the images in the dataset 86
5.3 Robustness for Standard images when scaled 87
5.4 GPE and Saliency for Standard Images 95
5.5 Median Robustness for all the images 95
5.6 Robustness for Standard images when scaled 96

6.1 Robustness against random frame dropping 113
6.2 Robustness against temporal scaling 113
6.3 Robustness against frame averaging 114

Contents

1 Introduction **1**

1.1 Digital Video Watermarking 2

1.1.1 Evaluation Parameters 4

1.1.2 Applications 5

1.2 Literature Survey 6

1.2.1 Scalable Image Watermarking 6

1.2.2 Scalable Video Watermarking 8

1.3 Motivation and Objectives 13

1.4 Contribution of the thesis 14

1.4.1 Watermarking against resolution and quality scalability . 14

1.4.2 Watermarking against temporal and quality scalability . . 14

1.4.3 Watermarking based on SIFT 15

1.4.4 Watermarking against temporal scalability 15

1.5 Thesis Organization 16

1.6 Summary 17

2 Research Background **19**

2.1 Motion Compensated Temporal Filtering (MCTF) 19

2.1.1 Motion Compensated DCT based Temporal Filtering (MCDCT-TF) 20

2.1.2 Inverse Motion Compensated DCT based Temporal Filtering (IMCDCT-TF) 22

2.1.3 Connected and unconnected pixels . . . 23
2.2 Scale Invariant Feature Transform . . . 24
2.3 Visual saliency model RARE2012 . . . 26
2.4 Evaluation Parameters . . . 28
2.4.1 Visual quality parameters . . . 28
2.4.2 Robustness parameter . . . 30
2.5 Experimental Dataset . . . 30
2.6 Summery . . . 30

3 Robust Video watermarking against Resolution and Quality Scalability 33
3.1 Introduction . . . 33
3.2 Background . . . 34
3.2.1 DC Frame . . . 35
3.2.2 Graceful Improvement of the Watermark . . . 35
3.3 Proposed Scheme . . . 37
3.3.1 Watermark Embedding Scheme . . . 37
3.3.2 Extraction Scheme . . . 41
3.3.3 Embedding Capacity . . . 45
3.4 Experimental Results . . . 48
3.4.1 Visual Quality . . . 48
3.4.2 Robustness . . . 49
3.5 Conclusion . . . 53

4 Robust Video watermarking against Temporal and Quality Scalability 59
4.1 Introduction . . . 59
4.2 Proposed Scheme . . . 60
4.2.1 Location Map . . . 61
4.2.2 Embedding Scheme . . . 63
4.2.3 Coefficient Selection . . . 64
4.2.4 Visual Quality Threshold . . . 68
4.2.5 Extraction Scheme . . . 69
4.3 Experimental Results . . . 69

4.3.1 Visual Quality 70
4.3.2 Robustness Comparison 72
4.3.3 Explanation 75
4.4 Conclusion 76

5 SIFT based Robust Image Watermarking against Resolution Scalability 77
5.1 Introduction 77
5.2 Proposed Scheme 78
5.2.1 Watermark Zone Selection 79
5.2.2 Derivation of Quality Parameter β 81
5.2.3 Watermark Embedding 82
5.2.4 Watermark Extraction & Authentication 84
5.2.5 Experimental Results 84
5.3 Improvement over the proposed scheme 87
5.3.1 Strength of Individual SIFT Feature 87
5.3.2 Modified Watermark Zone Selection 88
5.3.3 Watermark Embedding 89
5.3.4 Watermark Extraction 89
5.3.5 Experimental Result 92
5.4 Conclusion 99

6 Robust Video Watermarking against Temporal Scalability 103
6.1 SIFT based Video Watermarking Resilient to Temporal Scalability 104
6.1.1 Watermarking Zone Selection 105
6.1.2 Watermark Embedding 109
6.1.3 Watermark Detection & Authentication 110
6.1.4 Experimental Results 112
6.2 Robust video watermarking against Temporal Scalability 117
6.2.1 Proposed scheme 117
6.2.2 Watermark Generation 118
6.2.3 Watermark Embedding 119
6.2.4 Watermark Extraction 119
6.2.5 Graceful Improvement 119

6.2.6 Experimental Result 121
6.3 Conclusion 122

7 Conclusion and Future Works **125**
7.1 Watermarking against Resolution and Quality Scalability 125
7.2 Watermarking against temporal and quality scalability 126
7.3 Image watermarking based on SIFT against resolution scaling . . 126
7.4 Watermarking against Temporal Scalability 127
7.5 Future Research Scope 128

References **138**

Chapter 1

Introduction

The rapid growth in Internet technology and media communication started anew era of video broadcasting and transmission. In this new era, the heterogeneity among the end-using display devises increased considerably with respect to the display resolution, processing power, network bandwidth, etc. Depending on their computation power, display size or storage capacity, these devices have varying requirements in terms of video quality, frame rate, resolution, etc. It has been observed that achieving these scalable adaptation at the receiving side for a variety of end-using devices is a bit complicated process. Scalable video transmission provides a viable solution to this problem by performing these scalable adaptation at the multimedia servers rather than on the receiving end. A hypothetical scalable video transmission scenario is depicted in Fig. 1.1. A scalable video stream can be adapted to provide different types of resolutions, quality and spatio-temporal characteristics. A single video stream (high quality, high resolution and high frame rate) will be stored and each content consumer will be able to extract the best video representation for their applications or devices. Different types of video scalability are presented in Fig. 1.2. There are several strategies to achieve scalability : layered coding which is followed by MPEG-4 and its predecessor, embedded coding used by 3D subband coder, such as MC-EZBC and hybrid coding utilized by MPEG4-FGS and H.264/SVC [1].

The widespread and easy accesses to multimedia contents and the possibility to make unlimited copies without loss of considerable fidelity/quality have made

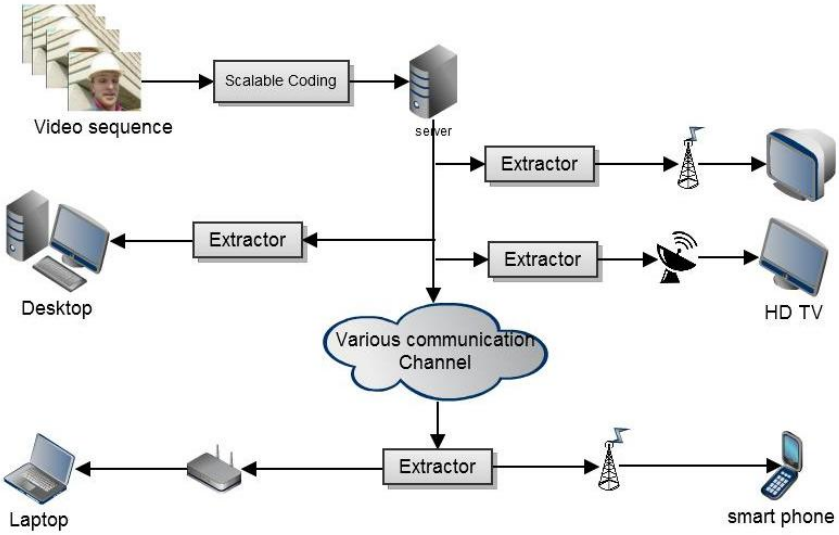


Figure 1.1: *Use of scalable video*

the digital rights management as an essential requirement for the efficient media transmission. Thus, assuring protection such as ownership as well as video content authentication became a challenging research problem especially when the scalable media is concerned. Encryption and cryptographic hashes are proposed to meet the solutions. But it is observed that the scalability property of the bit stream is lost [2, 3] if the video bit stream is encrypted with conventional cryptographic ciphers like AES [4]. Some schemes used multiple keys for multiple layers but it generally requires a complicated key management to meet the application scenario. Secret sharing based cryptographic solutions [5] also fails against resolution scaling. Watermarking has been used [6, 7, 8, 9, 10, 11] popularly in the last two decades for copyright protection and content authentication of multimedia content. In this research work, video watermarking is considered as the tool for ensuring secure video transmission.

1.1 Digital Video Watermarking

Digital video watermarking is a technique which inserts a digital signature (number sequence, binary sequence, logo, etc.) into the video stream which can be

(a) *Temporal Scalability*

(b) *Spatial Scalability*

(c) *SNR scalability*

Figure 1.2: *Different Scalability.*

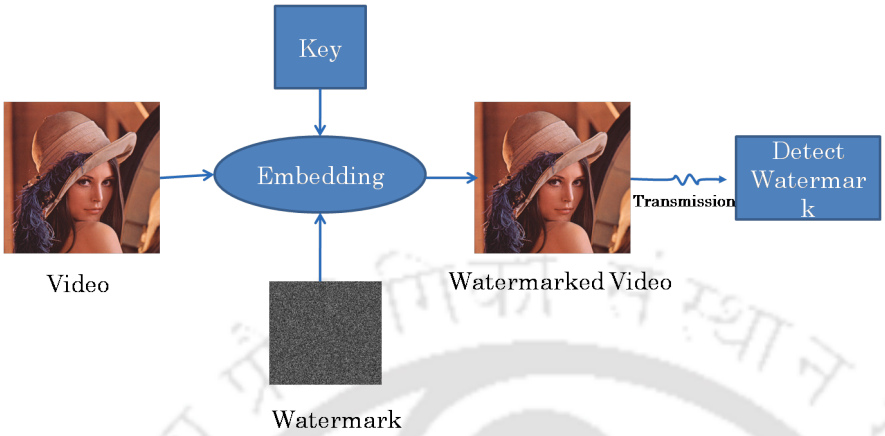


Figure 1.3: *Video Watermarking*

extracted or detected to authenticate the ownership of the video or the video itself. A fundamental video watermarking system is described in the Fig. 1.3 and the basic watermarking scenario for scalable video sequence is depicted in the Fig. 1.4.

1.1.1 Evaluation Parameters

There are many parameters to evaluate the efficiency of a watermarking system. These parameters are often mutually conflicting. For example, the visual quality (imperceptibility) may be reduced to increase the robustness of the scheme. Few important parameters are described below:

Robustness : The robustness of a watermarking scheme is defined as how efficiently the watermarking scheme withstands against intentional and unintentional attacks.

Imperceptibility : Imperceptibility implies that the watermark should not be perceptually noticeable in the watermarked video.

Payload : Payload measures the number of bits or the size of the watermark which is embedded to the cover media.

Blindness : A watermarking scheme is called blind if the original content is not required at the time of watermark extraction.

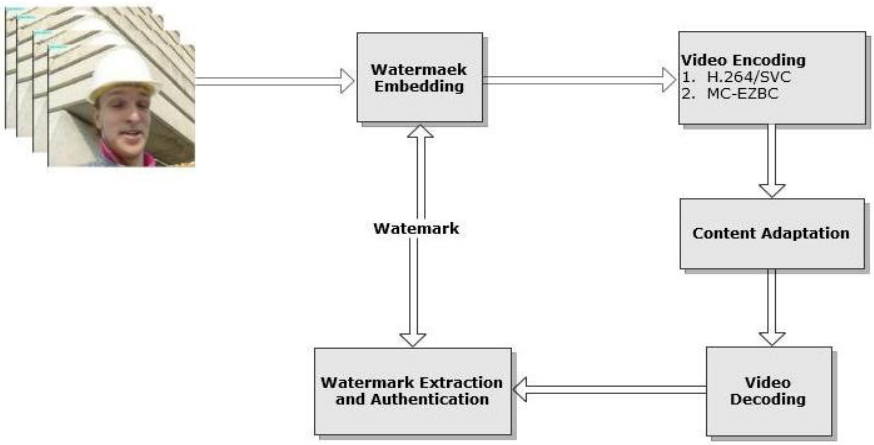


Figure 1.4: *A basic Scalable Video Watermarking scenario*

Bit Increase Rate : The video bit rate may get increased due to embedding. An efficient watermarking scheme embeds watermark in such a way that bit increase rate (BIR) should not be increased considerably.

1.1.2 Applications

Despite the widespread applications of digital watermarking as content or copyright authentication tool, it may be used for several purposes. Some of them are narrated below:

Ownership Authentication / Copyright Protection [9] : As video content is a valuable commodity, ownership or copyright of the content must be protected. Watermarking resolves copyright issues of digital media by using copyright data as watermark information.

Video Authentication [12] : Alteration to a video content can be done easily and such alteration is often very difficult to detect. Watermarking can be used to verify the authenticity of the content by identifying the possible video tampering or forgery.

Traitor Tracing[13] : Watermark can also be used to trace the source of pirated video content to stop the unauthorized content distribution.

Broadcast Monitoring [14] : Watermark can also be used for managing video broadcasting by putting a unique watermark in each video clip and assessing broadcasts by an automated monitoring station.

Medical Application [15] : A mix up in X-rays and MRI scans of two patients may be disastrous and must be avoided. Visible watermark can be used to identify the patient accurately by embedding some digital signature [16].

1.2 Literature Survey

It is observed in the literature that few video watermarking schemes are reported as a direct extension of the existing image watermarking scheme [17, 18]. More specifically, frame by frame video watermarking may be achieved by applying existing image watermarking scheme. Although, there exists few limitations of such extension, scalable image watermarking may be a good starting point to analyze the merits and demerits of state of the art scalable video watermarking.

1.2.1 Scalable Image Watermarking

Piper et al. mentioned explicitly about scalable watermarking (for image) in [19]. They used different coefficient selection methods for the spread-spectrum embedding proposed by Cox [6] and have evaluated the robustness of the scheme against quality and resolution scalability. Later, Piper [20] discussed that the spatial resolution and quality scalable watermarking can be achieved by exploiting the characteristic of Human Visual System (HVS). In another work, Seo et. al. [21] evaluated a scalable image watermarking scheme for protecting distant learning content and proposed a watermark embedding technique using wavelet based image coding.

Content-based image watermarking schemes are generally used to resist the geometric attacks [22, 23, 24]. There are a variety of watermarking techniques that aim to be robust against a specific subset of geometric attacks. In some schemes [25, 26], watermark is embedded in such domains (e.g. Fourier-Mellin, Radon) which are invariant to the geometric attacks. In, [27], a reference pattern

or template is embedded which can be used to synchronize the watermark during extraction whereas in [28], exhaustive search has been performed.

Feature point-based approach is one of the most promising classes of image watermarking. Kutter et al. [29] argued that a watermark is more robust if the embedding is performed using these feature points as they can be viewed as containing second-order information of the image. In this direction, Bas et al. [22] proposed a scheme where delaunay triangulation is computed on the set of feature points which are robust to the geometric distortion. The watermark is then embedded into the resulting triangles and detection is done using the correlation properties on the different triangles. Drawback of this method is that the extraction process may not accurately extract the watermark especially when the extracted feature points from the original and distorted images are not matched as the sets of triangles generated during watermark insertion and detection are different.

Scale Invariant Feature Transform (SIFT) [30] is an image descriptor for image based matching developed by David Lowe. SIFT features have been used in many applications like multi view matching [31, 32], object recognition [33], object classification [34, 35], robotics [36] etc. It is also being used for robust image watermarking against geometric attacks [24, 37, 38]. Miyaki et al. [37] proposed a RST invariant object based watermarking scheme where SIFT features are used for object matching. In the detection scheme, the object region is first detected by feature matching. The transformation parameters are then calculated, and the message is detected. Though the method produces quite promising results but it is a type of informed watermarking. The register file has to be shared between the sender and receiver which may not be always desirable. In another SIFT based work, Kim et al. [24] have inserted watermark into the circular patches generated by the SIFT. The detection ratio of the method varies from 60% to 90% depending upon the intensity of the attack. It is observed that under strong distortions due to attenuation and cropping, this additive watermarking method sometimes fails to accurately detect the watermark. More recently, Jing et al. [38] used SIFT points to form a convex hull. The SIFT points are then optimally triangulated. The watermark is embedded into the circles centered around the

centroid of each triangle. This method also fails to sustain the watermark when the image is scaled down considerably.

It the watermark observed in the literature that above mentioned image watermarking schemes are often directly used for frame by frame video watermarking. But there are certain limitations for the frame by frame embedding. Firstly, it can create flickering artifacts [17, 39], secondly these schemes are generally vulnerable against collusion attacks [17, 40, 41].

1.2.2 Scalable Video Watermarking

It is discussed in the previous subsection that the scalable video transmission is an emerging field of research in recent times. But, literature reveals that relatively less attention has been paid to the scalable video watermarking in comparison with the general video watermarking. Lu [42] possibly the first characterized the scalable watermarking and argued that the watermark should be detected at every resolution and quality layers. Piper [20] mentioned *graceful improvement* as another important property of the scalable watermarking where the watermark detection should become more and more accurate with the improvement of the video quality (with addition of enhancement layers).

Challenges of the Scalable Video Watermarking

The main problem of scalable watermarking is that the bit-budget for a scalable sub-stream is not known a-priory as main bit stream can be truncated at any spatio-temporal bit truncation point. In scalable watermark, it is required to protect the base layer as well as the enhancement layers which generally causes substantial bit increase for the watermarked video. Keeping low bit increase rate (BIR) for scalable video watermarking is a real challenging task [43, 44]. Since, different scalable parameters like resolution, frame rate, quality etc. are different in nature, assuring combined watermarking security to all of them sometimes requires conflicting demands. Thus achieving combined scalable watermarking is a difficult task [20]. It is also observed that the statistical distribution of the transform domain coefficients of the base layer is substantially different than that of enhancement layer. It makes the multi-channel detection more complicated

for incremental detection performance. Finally, watermarking zone selection becomes challenging in presence of the inter layer prediction structure of the scalable coding.

Literature on Scalable Video Watermarking

Alattar et al. [45] proposed a compressed domain watermarking scheme for MPEG-4 against RST attack. They used a synchronization template to resist the RST attacks. In [41], Jung et al. proposed a RST invariant watermarking scheme where the content adaptive watermark signal is embedded in the Discrete Fourier Transform (DFT) domain of the video stream. Authors have used log polar projection to detect the watermark. The problem of extending this work for scalable video is that it may not withstand quality and temporal adaptation although it achieves desired robustness if only resolution scaling is considered. Moreover, visual artifacts may be generated due to logarithmic mapping during watermark embedding. Chang et al. [46] have combined encryption and watermarking to realize layered access control to a temporally scalable M-JPEG stream. They have encrypted enhancement layer and embedded the key needed to decrypt it in the base layer. So that, the key receives stronger error-protection than the content. Wang et al. [43] proposed a blind watermarking scheme for MPEG-4 where watermark embedded into FGS bit planes for authentication of enhancement layer. One bit is embedded by forcing the number of non-zero bits T_j per bit plane j and block to even or odd depending on the watermark. In another recent scheme, Y. Wang and A. Pearmain [47] have proposed a blind scale-invariant watermarking scheme for MPEG-2. In this scheme, authors embedded the watermark in a single frame (middle frame) of the GOP (Group of Picture) of 3 frames. It is observed that the scheme is vulnerable against type I collusion attack [40, 41] as the watermark can easily be estimated by comparing watermarked frame and the two adjacent non-watermarked frame. The scheme may also be vulnerable against frame dropping attack as the watermark for an entire GOP has been lost if the single watermarked frame is dropped or replaced. The temporal artifacts may be caused by watermark embedding, as the scheme is not using motion compensated embedding. Moreover, since the watermark can

only be extracted from the base layer, base layer computation is always required for the watermark extraction from any of the enhancement layers.

A RADON transformation based RST invariant watermarking scheme [48] has been proposed by Liu and Zhao. In this scheme, authors have used temporal DFT and embedded the watermark using RADON coefficients. Objectionable visual artifacts in the watermarked video may be caused due to embedding by altering the RADON coefficients. Since the embedded watermark in the base layer is same as in the different enhancement layers, the cross-layer collusion attack [40, 41] may be mounted over the watermarked video. Moreover, temporal motion may degrade the embedded watermark. In schemes [49, 50], 3D wavelet coefficients are used for watermark embedding. Since the temporal motion is not considered in these schemes, they may suffer from the flickering artifact due to embedding and may produce visually degraded watermarked video [39]. However, these algorithms are not designed for recent scalable video coding techniques, and they did not introduce the concept of adaptation detection into the scalable watermark model. In last few years, after standardization of SVC [51] in 2007, very few works are published in the literature on watermarking of scalable video content. Some significant works are described as follows:

Watermarking based on Temporal Filtering

It is observed in the previous section that flickering artifact is caused due to frame-by-frame watermarking. It is found in the literature [39, 40] that the motion compensation in the temporal direction during embedding generally reduces these artifacts. On the other hand, temporal filtering can be used to resist the frame dropping or frame averaging kind of attacks. Considering these facts, there exists a group of scalable watermarking schemes which use motion compensated temporal filtering (MCTF) [52, 53, 54] to find the suitable embedding zone for watermarking [39, 40]. In this subsection, MCTF based schemes and their pros and cons are described. P. Vinod and P. K. Bora [40] used MCTF structure for video watermarking to resist the collusion attacks. First they segmented the video into different scenes. Then every scene is temporally decomposed using wavelet based MCTF. Then Watermark is embedded along the motion trajectory of low pass frames. Bhowmik et al. [39] proposed a motion compensated

spatio-temporal sub-band decomposition scheme, based on the modified MCTF for video watermarking. They have used a 2D+t+2D decomposition framework where they decomposed the video sequence in different temporal and spatial levels and choose the best subband for embedding. Temporal decomposition is done using Modified MCTF. Embedding distortion is evaluated using MSE and flicker difference metric. The proposed sub-band decomposition also has low computational cost as MCTF is performed only on sub-bands where the watermark is embedded. Authors have proposed two approaches, one is blind and another one is non-blind. In the non-blind approach additive watermarking is used where coefficients are increased or decreased according to the Equation 1.1. Block diagram of the embedding and extraction technique for the blind technique is shown in Fig. 1.5 and Fig. 1.6 respectively.

$$C'_{s,t}[m,n] = C_{s,t}[m,n] + \alpha C_{s,t}[m,n]W \tag{1.1}$$

where $C_{s,t}[m,n]$ and $C'_{s,t}[m,n]$ is $[m,n]th$ is original and corresponding watermarked coefficient respectively, α is watermark strength and W is watermark signal. Authors have evaluated it against only quality scalability attack which

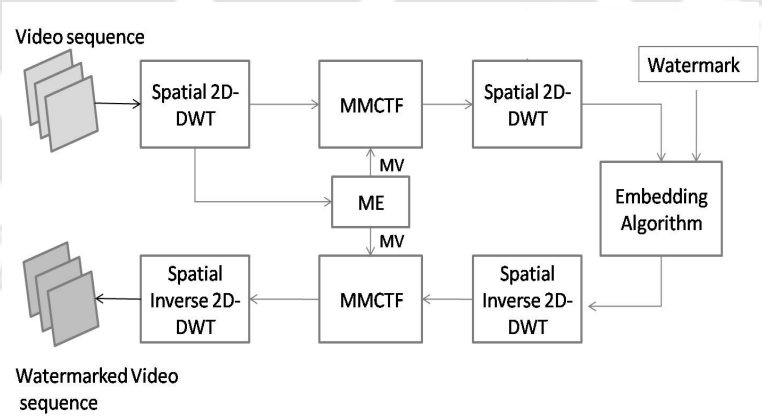


Figure 1.5: *Block Diagram for Watermark Embedding [39]*

may not be useful in a real world situation where a video will be scaled in any dimension according to the requirement. Visual quality, Bit Increase Rate are (BIR) not evaluated for watermarked video. They haven't considered any zone selection or coefficient selection method where more research can be done.

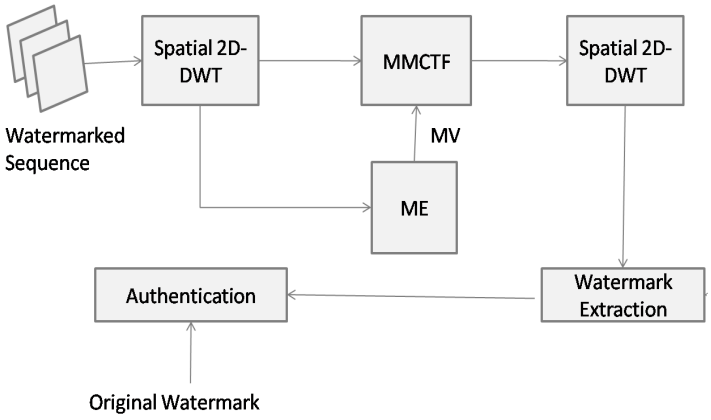


Figure 1.6: *Block Diagram for Watermark Extraction[39]*

Watermarking in Compressed Domain

Meerwald et al. [44] proposed a compressed domain watermarking for H.264/SVC. They extended a framework for robust watermarking of H.264-encoded video proposed by Noorkami et al. [55] to scalable video coding (SVC). The main objective of this work is to handle spatial (resolution) scaling. Authors have shown that the watermark embedding in the base layer of the video is insufficient to protect the decoded video at higher enhancement layers as the embedded watermark may get faded in higher resolution layers. Moreover, the bit rate of the enhancement layer may get increased. To solve this problem, they up-sampled the base layer watermark signal and embedded in the enhancement layer.
At first, watermark is embedded in the base layer stream using the Eqn. 1.2 proposed in [55].

$$R_k^w = R_k + S_k.W_k \tag{1.2}$$

Where W_k is watermark signal, S_k is location matrix and R_k and R_k^w is base layer residual block and corresponding watermarked block. Then upsampled watermark signal is embedded using the Eqn. 1.3.

$$R_k{}^{Ew} = R_k{}^{E} + W_k{}^{E} \tag{1.3}$$

where $W_k{}^{E}$ is upsampled watermark signal, $R_k{}^{E}$ is the enhancement layer residual block . Proposed Watermarking structure is shown in Fig. 1.7. Problem with this approach is that the higher resolution video

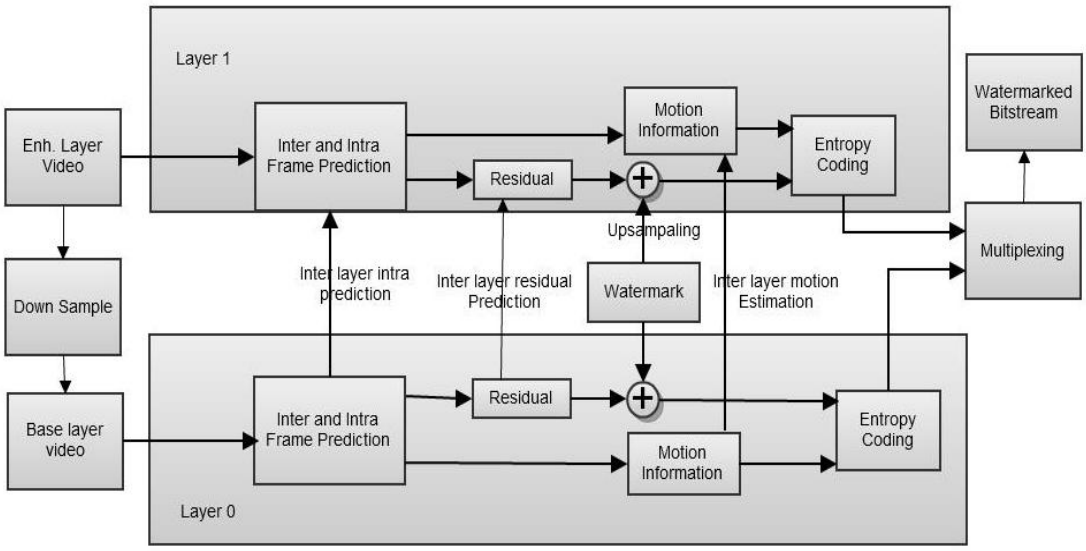


Figure 1.7: *Block diagram of Meerwald's [44] watermark embedding method for two spatial layers.*

must be down-sampled to the base layer for the watermark extraction. Thus, it doesn't follow the convention given in [19]. It is also observed that the BIR for the enhancement layer is bit high and the scheme becomes too complex for more than two enhancement layers.

1.3 Motivation and Objectives

From the above literature, it is observed that most of the schemes are not performing well for relatively high degree of resolution scaling. The temporal adaptation is also not properly taken care off and the temporal synchronization for heterogeneous motion, collusion resistant watermarking etc. issues should be properly handled. The suitable zone selection for maintaining decent a visual quality of the watermarked video during scalable watermarking is also an important issue and should be more explored. Finally, SIFT based schemes look very promising and how SIFT features are more efficiently used for watermarking may be another very interesting study. Motivated by these issues, main objective of this work is to enhance the robustness of the watermarking scheme against the content adap-

tation attacks while maintaining the decent visual quality of the watermarked video. This has been carried out by

1. Proposing robust watermarking against the resolution, temporal and quality adaptation attacks.
2. Maintaining a decent visual quality of the watermarked video by proposing suitable zone selection methods for watermarking.
3. Proposing robust watermarking schemes by exploiting SIFT features against temporal and resolution adaptation.

1.4 Contribution of the thesis

1.4.1 Watermarking against resolution and quality scalability

In the first work, a watermarking scheme against resolution and quality scalability is proposed where the base layer embedding is done on the DC frame which is generated by accumulating DC values of non-overlapping blocks for every frame in the input video sequence. DC frame sequence is up-sampled and subtracted from the original video sequence to generate the residual frame sequence. Then Discrete Cosine Transform (DCT) based temporal filtering is applied on DC as well as residual frame sequence. Watermark is embedded in the low pass DC frames and up sampled watermark is embedded in the low pass residual frames to achieve the graceful improvement of watermark signal in successive enhancement layers. It is experimentally shown that the proposed scheme performs well against resolution and quality adaptation and outperforms existing related schemes.

1.4.2 Watermarking against temporal and quality scalability

In the next phase, a blind scalable video watermarking scheme is proposed, which is robust against quality and temporal scalability. In the proposed scheme, DCT based temporal filtering and wavelet based spatial filtering is used for selecting

watermark embedding zone. Temporal filtering is used on GOP to exploit the correlation among frames. Watermark is embedded in the low pass frames. In this work, location map is used to accurately describe the embedding locations for efficient and blind watermark extraction.

1.4.3 Watermarking based on SIFT

In this work, a novel image watermarking scheme is proposed which is robust to the resolution scaling. In the proposed method, SIFT features are used which are invariant to scaling. Most of the existing SIFT based watermarking algorithms, fail to retain the watermark when the image is heavily scaled down. In the proposed scheme, a context coherent object or patch is inserted in the image such that it generates strong SIFT features. These newly generated SIFT features are themselves used as the watermark. Since the SIFT features are invariant to scaling, these features can be extracted from any image resolution with high probability.

1.4.4 Watermarking against temporal scalability

In the final phase of the work, two watermarking schemes are proposed against temporal scalability. In the first work, SIFT features are used to handle the temporal scalability. In this work, a cubic region of the video is modified to generate new SIFT feature, which are stored in database as the watermark. Modification is done in a low motion area of a randomly selected frame set as the embedding in high motion area often creates flickering artifacts. To resist temporal scaling and frame dropping attack, SIFT features are extracted from the side plane. In the proposed scheme, low motion and high texture zones of selected n consecutive frames are chosen as the embedding location. Effectiveness of the scheme is experimentally justified against the temporal adaptation and frame dropping attacks.

In the second work, each temporal layer has been separately embedded with a different watermark which is generated by DCT domain decomposition of a single watermark image to ensure the graceful improvement of the extracted watermark along successive higher layers. A zigzag sequence of the block DCT coefficients

of the watermark image is partitioned into non-overlapping sets and each set is embedded separately into different temporal layers. The base layer is embedded with the first set of DCT coefficient (which includes DC coefficient of each block) and successive layers are embedded with successive non-overlapping coefficient sets. The coefficients of each set are chosen in such a fashion that uniform energy distribution across all temporal layers can be maintained.

1.5 Thesis Organization

This PhD dissertation consists of seven chapters. The first chapter consists of a brief introduction of scalable image and video watermarking, a brief literature survey, research motivation and objectives of the thesis, contribution of the thesis and the organization of the thesis.

- Chapter 2 describes the background of the research which includes some preliminary concepts like MCTF, SIFT etc., evaluation metrics, experimental data set etc. which are used in later chapters.
- In the 3rd chapter, a robust watermarking algorithm is presented against the resolution and quality scaling where watermark is first embedded in base layer and then the up-sampled watermark is embedded in the enhancement layer to achieve the graceful improvement.
- Chapter 4 introduces a MCDCT-TF based robust watermarking against temporal and quality scalability. Watermark is embedded in the temporal low pas frames and a location map is used for the watermark extraction.
- In chapter 5, SIFT based scale invariant image watermarking is presented. In this scheme, intensity of an image patch is changed to generate new SIFT features, which is stored as the watermark.
- Chapter 6 describes two different schemes against temporal scalability and temporal attacks. In the first work, SIFT features of the side planes of the video sequence are used for watermarking to handle temporal scalability. In the second work, different temporal layer frames are embedded with different watermarks.

- The chapter 7 briefly summarizes the PhD dissertation and suggests future research directions.

1.6 Summary

In this introductory chapter, at first the domain of the research is defined. Then a brief literature survey is presented and the corresponding limitation have been identified. Based on these limitations, motivation and the objective of the research work is formulated. Finally, the brief description of the contributions and the thesis organization are presented.

Chapter 2

Research Background

In this chapter, a brief overview of mathematical preliminaries and theoretical foundations relevant to the topics of interests are presented. This includes a discussion on MCTF, SIFT and RARE2012. In addition, the different evaluation parameters for the proposed algorithms and corresponding data set used for experimentations are also described.

2.1 Motion Compensated Temporal Filtering (MCTF)

MCTF [54, 52, 53] is as its name suggests, the low pass filtering of the input frames along the motion direction. It is used to remove temporal correlation within the sequence [54]. Input frames need to be aligned along the motion trajectories for better de-correlation. A basic MCTF based on Haar wavelet transformation is described in Fig. 2.1.

In Fig. 2.1, each odd frame is predicted from previous even frame using operator P. Predicted frame then subtracted from the original frame to obtain the residual error frame (high-pass temporal frame) or H-frame. Low-pass frame are generated by adding the residual information to the reference frame by the U (Update) operator, which performs an additional Motion Compensation (MC) stage using the reversed motion field. The Haar wavelet is a short filter and provides limited de-correlation. Longer length filter can make better use of correlation in

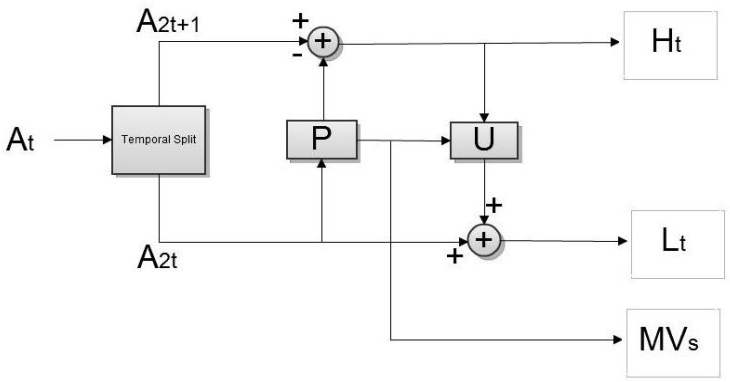


Figure 2.1: *Motion Compensated Temporal Filtering(Figure courtesy [54]).*

the temporal domain. A DCT based temporal filtering which uses a longer filter is described in the next sub-section.

2.1.1 Motion Compensated DCT based Temporal Filtering (MCDCT-TF)

MCTF have been used for video watermarking in [39, 40]. While these watermarking schemes used 2-tap Haar filter for MCTF, use of longer length filter can make better use of correlation in the temporal domain. Atta et al. used a DCT based temporal filtering (MCDCT-TF) in [56] for scalable video encoding. They have used size of Group of Frame (GOF) as 9. First temporal filtering is done on sub-GOF of 3 frames. For next level of temporal decomposition, they used 2 low pass frames from first level decomposition. In our work, a variant of MCDCT-TF is used for scalable watermarking. In the used MCDCT-TF, one low pass frame is used for the next level of filtering to avoid overlapping of watermark information. In this work, video sequence is divided into group of N frames, which are again subdivided into group of K frames. After applying $K \times 1$ temporal DCT, a new sequence of $\frac{N}{K}$ low pass frames are formed. By recursive decomposition of the video sequence we finally get one low pass frame. Procedure of MCDTC-TF for GOFs of 3 frames is explained in Fig.2.2, where L,M and H are low, middle and high frequency frames respectively. For the simplicity, all equations are written with the above assumption (GOF=9, sub-GOF=3).

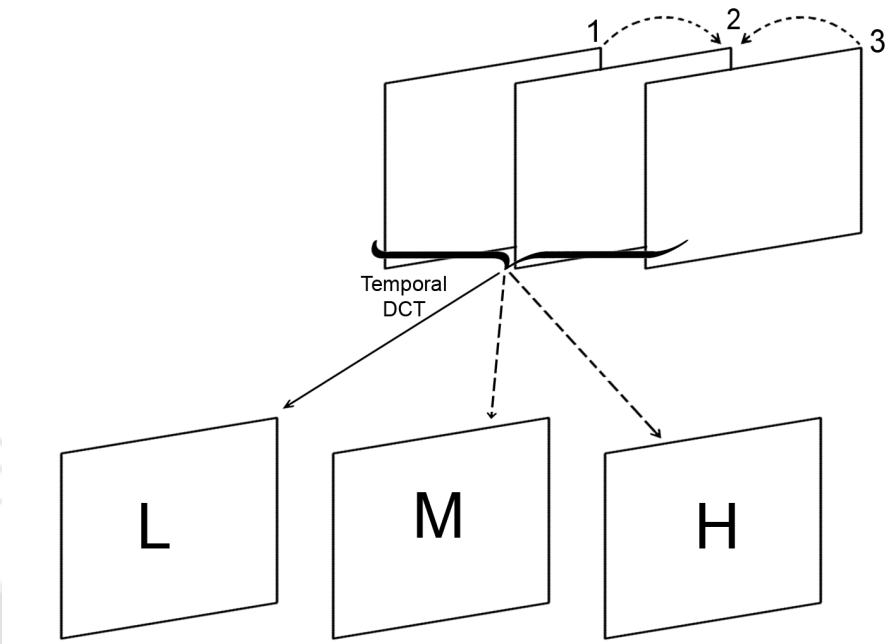


Figure 2.2: *DCT based MCTF (L, M,H are low middle and high frequency frames respectively)*

The predicted frames can be given as Eqn. 2.1 and 2.2.

- From first frame to second frame

$$I^{1}_{3t+2}[m,n] = I_{3t+1}[m + H^{1->2}, n + V^{1->2}] \tag{2.1}$$

- From third frame to second frame

$$I^{3}_{3t+2}[m,n] = I_{3t+3}[m + H^{3->2}, n + V^{3->2}] \tag{2.2}$$

where I_{3t+1}, I_{3t+2} *and* I_{3t+3} are 3 frames in sub-GOF. $(H^{1->2}, V^{1->2})$, $(H^{3->2}, V^{3->2})$ are motion vectors of I_{3t+2} with respect to I_{3t+1} and I_{3t+3} respectively. After the motion compensation, frames are aligned along the motion trajectories. Now (3×1) temporal DCT (refer to Eqn. 2.3) is done pixel by pixel. Temporal DCT results three frequency level frames

- Low frequency high energy frame L
- Middle level frequency frame M
- High frequency and low energy frame H

$$\begin{bmatrix} L[m,n] \\ M[m,n] \\ H[m,n] \end{bmatrix} = A \times \begin{bmatrix} I^1_{3t+2}[m,n] \\ I_{3t+2}[m,n] \\ I^3_{3t+2}[m,n] \end{bmatrix} \tag{2.3}$$

where A= $\begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{6}} & -\sqrt{\frac{2}{3}} & \frac{1}{\sqrt{6}} \end{bmatrix}$ $is \quad 3 \times 1 DCT kernel.$

2.1.2 Inverse Motion Compensated DCT based Temporal Filtering (IMCDCT-TF)

For the inverse transform, 3×1 inverse DCT is done on L, M and H as in equation 2.4. Then inverse motion compensation is done on result using Eqn. 2.5 and Eqn. 2.6

$$\begin{bmatrix} I^1_{3t+2}[m,n] \\ I_{3t+2}[m,n] \\ I^3_{3t+2}[m,n] \end{bmatrix} = A^T \times \begin{bmatrix} L[m,n] \\ M[m,n] \\ H[m,n] \end{bmatrix} \tag{2.4}$$

where A^T is transpose of matrix A

$$\begin{aligned} I_{3t+1}[m,n] &= I^1_{3t+2}[m - H^{1->2}, n - V^{1->2}], \quad for\, fully\, connected\, pixels \\ &= I_{3t+1}[m,n] \qquad\qquad\qquad , \quad for\, all\, other \end{aligned} \tag{2.5}$$

$$\begin{aligned} I_{3t+3}[m,n] &= I^3_{3t+2}[m - H^{3->2}, n - V^{3->2}], \quad for\, fully\, connected\, pixels \\ &= I_{3t+3}[m,n] \qquad\qquad\qquad , \quad for\, all\, other \end{aligned} \tag{2.6}$$

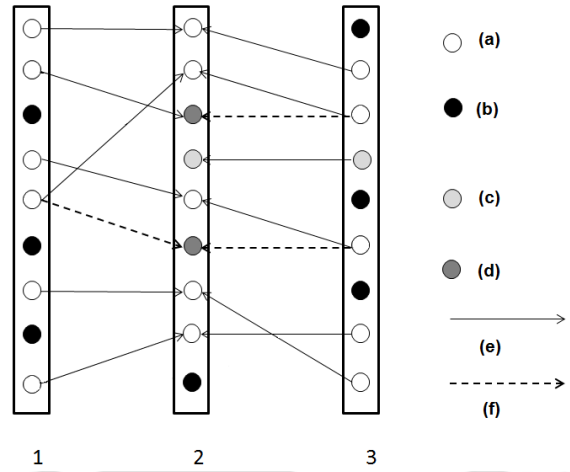


Figure 2.3: *Connected and unconnected pixel, (a)Fully connected pixels, (b)Unconnected pixels, (c)Partially connected pixel, (d)One to many connection*

2.1.3 Connected and unconnected pixels

If the video frames are filtered along all motion trajectories, then some pixels are filtered more than once and some pixels are not filtered at all. To avoid this problem, Ohm [57] categorized pixels into two classes "connected" and "unconnected" by their estimated motion vectors. In [39], the problem of unconnected pixels in the MC temporal filtering phase was considered between every pair of input frames. Different strategy is required when MCDCT-TF is applied to three frames. In [56], pixels are categorized in three different categories. The basic mechanism of treating unconnected pixels with our DCT temporal analysis of three frames is illustrated in Fig. 2.3. During MCDCT-TF we get three kind pixels,

1. **Fully connected pixels :** In this case, the pixels of the second frame have a connection from first and third frame. White pixels in motion compensated frames in Fig 2.3 are connected pixels.

2. **Unconnected pixels :** These types of pixels occur only in first referee and third referee frames. Black pixels in motion compensated frames in Fig 2.3 are unconnected pixels.

3. **Partially connected pixel :** When the pixel in the second frame is connected to the pixel in the first frame but unconnected to the pixel in the third frame or when the pixel in the second frame is connected to the pixel in the third frame but unconnected with the pixel in the first frame, those pixels are considered as partially connected. Grey pixels in Fig 2.3 are partially connected pixels.

There are pixels in the reference frames (1st or 3rd) which are connected to more than one pixels in 2^{nd} frame, any one of them are selected during motion compensation.

2.2 Scale Invariant Feature Transform

It is observed in the literature that SIFT based watermarking performs well against RST attacks. The SIFT [30] algorithm extracts distinctive features of local image patches and is proved to be invariant to image scaling and rotation. SIFT descriptors are robust to noise and changes in illumination, distortion and viewpoint. These local invariant features are highly distinctive and are matched with a high probability against large image distortions. The SIFT descriptor extracts features and their properties, such as the location (x, y), the scale $(\sigma$)and the orientation (θ).

Four major steps for finding SIFT descriptors are (1) Scale-space extrema detection; (2) Keypoint localization; (3) Orientation assignment; (4) Key point descriptor.

1. **Scale-space extrema detection**
   Given an input image I(x,y), the scale space of image I can be defined as Difference of Gaussian (DOG) as in Eqn. 2.7 in [30].

$$\begin{aligned} DOG(x, y, \sigma) &= [G(x, y, k\sigma) - G(x, y, \sigma)] * I(x, y) \\ &= L(x, y, k\sigma) - l(x, y, \sigma) \end{aligned} \tag{2.7}$$

   where $*$ is the *convolution* operation in x and y directions, and

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} \exp^{-\frac{x^2+y^2}{\sigma^2}} \tag{2.8}$$

is the Gaussian kernel and σ denotes the standard deviation of the Gaussian kernel. The scale-space extrema are detected from the points (x, y, σ) in scale-space at which the scale-normalized Laplacian assumes local extrema with respect to space and scale. In a discrete setting, such comparisons are usually made in relation to all neighbors of a point in a $3 \times 3 \times 3$ neighborhood over space and scale.

2. **Key point localization:**
   Many unstable keypoint candidates are detected in scale-space extrema detection. In this step, only stable keypoints are retained and unstable points are filtered out. The points with low contrast or located along the edges are rejected. Details are given in [30].

3. **Orientation assignment:**
   All the retained key-points in the previous step are assigned to one or more orientations based on local image gradient direction to make it rotation invariant. The key-point orientation is calculated from an orientation histogram of local gradients from the closest smoothed image $L(x, y, \sigma)$. For each image sample L(x, y) at the key-points scale σ, the gradient magnitude $mag(x, y)$ and orientation $\theta(x, y)$ is computed using pixel differences, using Eqn. 2.9 and Eqn. 2.10 respectively.

$$mag(x, y) = \sqrt{L_1^2 + L_2^2} \tag{2.9}$$

$$\theta(x, y) = \arctan(L_2/L_1) \tag{2.10}$$

   where,
   $L_1 = L(x + 1, y, \sigma) - L(x - 1, y, \sigma)$, and
   $L_2 = L(x, y + 1, \sigma) - L(x, y - 1, \sigma)$

   Then a orientation histogram is formed and the peak of this histogram is selected as the direction of that feature.

4. **Keypoint Descriptors:**
   In this part, a distinct key-point descriptor is computed for each key-point to make it invariant to illumination change, 3D view point etc. First the gradient magnitude and orientation at each image sample point in a region around the key-point location is computed in a 16×16 neighborhood. Then a weighted histogram of these samples for each of 16 4×4 sub-region is formed. A 128 bit descriptor is formed by concatenating 16 such histograms.

Descriptor matching: To match these computed key point descriptors with another set of descriptors, the nearest keypoint i.e. one with having minimum Euclidean Distance has been found. To reduce the ambiguous matches, only those matches are selected for which the ratio between distances to the nearest and second nearest point is less than 0.8.

2.3 Visual saliency model RARE2012

In chapter 5, a saliency map is used for embedding zone selection. Brief description of the saliency model named RARE2012 [58] is given in this section. In [58] authors used multi-scale rarity to select the information which attracts human attention. Saliency of a region is calculated in three major steps. First low-level color features and medium-level orientation features are extracted by principal component analysis and gabor filter. By convolving with Gabor filter at 8 different orientation, 8 maps (map_i) are generated. Each map then alloted an efficiency coefficient using Eqn. 2.11. Maps are sorted according to their efficiency and each map is weighted according to their rank using Eqn 2.12 as mentioned in [58].

$$EC_i^2 = (max_i - mean_i)^2 \tag{2.11}$$

$$\forall i \in [1, N], \quad \begin{cases} if\ \frac{EC_i}{EC_n} \geq T & M_i = \frac{i}{N} \times map_i \\ else & M_i = 0 \end{cases} \tag{2.12}$$

Then, a multi-scale rarity mechanism is applied on the combination of the maps. Finally, rarity maps are fused into a single final saliency map. Flow chart of the algorithm is given in Fig. 2.4. Authors claimed that the model out performs other recent attention models.

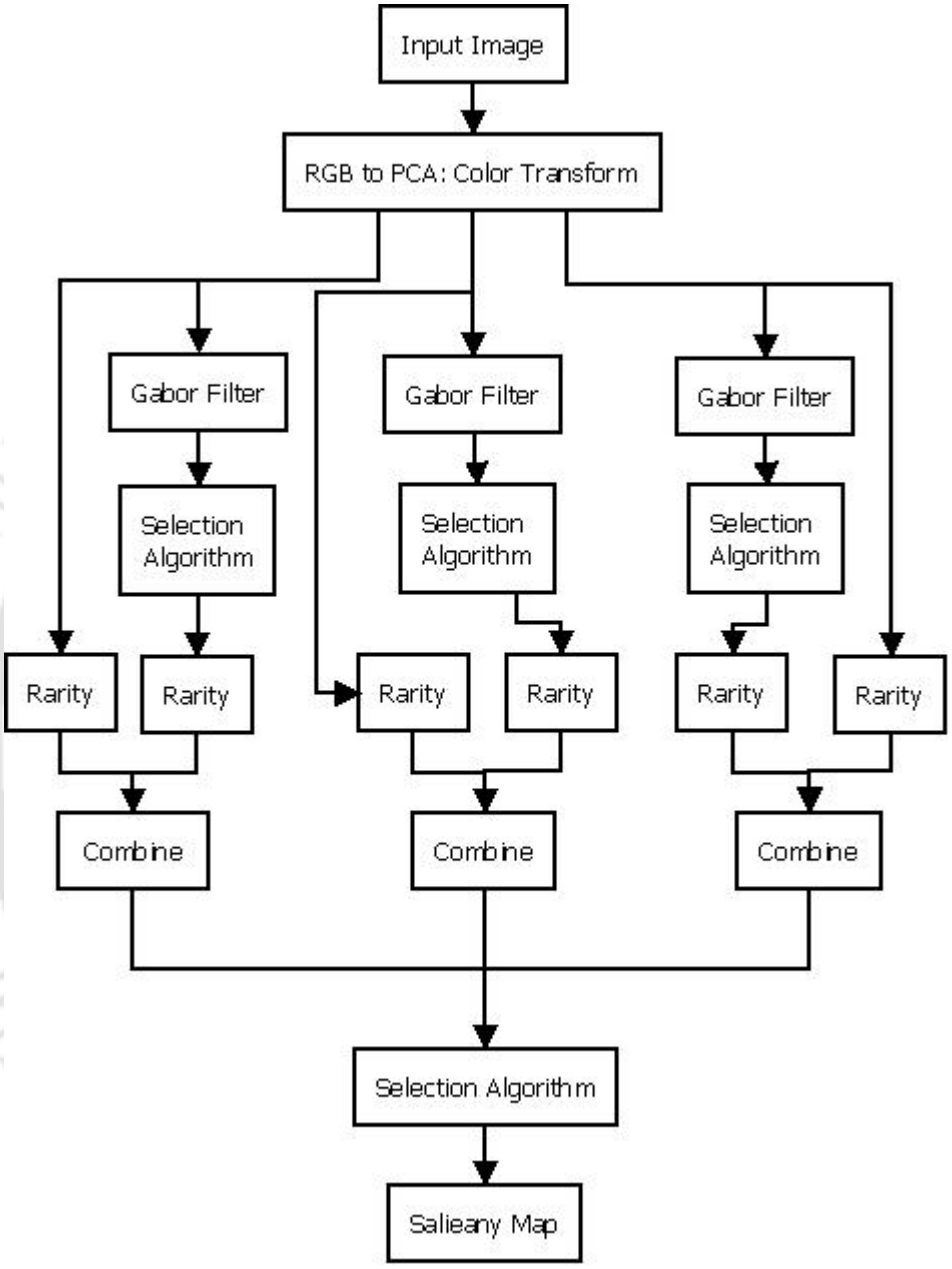


Figure 2.4: *RARE2012 flow chart [58]*

2.4 Evaluation Parameters

2.4.1 Visual quality parameters

In this work, Peak Signal to Noise Ration (PSNR), Structural Similarity (SSIM) [59], flicker metric [60] and Video Quality Metric (VQM) [61] are measured using MSU video quality measurement tool [62] to quantify the distortion due to watermark embedding. The evaluation parameters are described below.

PSNR: PSNR is the ratio between the maximum possible value (power) of a signal and the power of distorting noise that affects the quality of the signal. PSNR is calculated using Eqn.2.13.

$$PSNR = 10 \times \log_{10} \frac{Peak^2 \times M \times N}{\sum_{i=1}^{M} \sum_{j=1}^{N} \left(I(i,j) - I^w(i,j)\right)^2} \tag{2.13}$$

where $Peak$ is maximum possible intensity, M and N are width and height of the frame/image respectively, I is the original image and I^w is modified image.
SSIM: SSIM index is an image quality metric. It is function of three components eg. luminance similarity ($l(x,y)$), contrast similarity ($c(x,y)$) and structural similarity ($S(x,y)$) as given in Eqn. 2.14

$$SSIM(x,y) = [l(x,y)]^{\alpha}.[c(x,y)]^{\gamma}.[s(x,y)]^{\eta} \tag{2.14}$$

where α, β and η are parameters used to adjust the relative importance of the three components. In [62] SSIM is calculated for each frame. Detail description of the algorithm is given in [59].

Temporal Flicker: To determine the temporal flickering the average brightness value is first calculated for each of the frames. The flickering metric is calculated as modulus of difference between average brightness values of previous and current frames [61]. In this paper, flicker difference of original video and watermarked video is calculated.

VQM [61]: VQM is a DCT based video quality metric. To calculate VQM each DCT coefficients are converted to local contrast (LC) using Eqn. 2.15

$$LC(i,j) = DCT(i,j) * Power(DC/1024, 0.65)/DC \tag{2.15}$$

where DC is the DC component of the block, 1024 is the mean DCT value for 8 bit image, 0.65 is the best parameter for fitting psychophysics data as claimed by the author of the refereed paper. Then LC coefficients are converted to just-noticeable differences (JND) by multiplying each DCT coefficient by its corresponding entry in the spatial contrast sensitivity function (SCSF) matrix.

Then weighted pooling of mean and maximum distortion is done using equation 2.16

$$\begin{aligned} &\text{Mean\_Dist} = 1000 * \text{mean}(\text{mean}(\text{abs}(\text{diff}))) \\ &\text{Max\_dist} = 1000 * \text{maximum}(\text{maximum}(\text{abs}(\text{diff}))) \\ &\text{VQM} = (\text{Mean\_dist} + 0.005 * \text{Max\_dist}) \end{aligned} \tag{2.16}$$

Maximum distortion weight parameter 0.005 is chosed based on several primitive psychophysics experiments. Parameter 1000 is the standardization ratio.

In MSU (video quality measurement tool [62]) a modified version of [61] is implemented.

Watson Metric: Watson metric [63] is a DCT based perceptual error measure. The quantization errors for each coefficient in each block are scaled by the corresponding visual sensitivities of each DCT basis function in the block. The visual sensitivities are determined by three factors: contrast sensitivity, luminance masking and contrast masking. Initially luminance threshold (t_{ij}) for each DCT basis function is computed as a function of mean luminance of the display by taking contrast sensitivity into account. Then t_{ij} is adjusted by taking the approximation of luminance masking i.e. the local mean luminance within the image. The masked threshold m_{ij} is then computed by considering the contrast masking. From the masked threshold m_{ijk} and quantization error e_{ijk}, the perceptual error in each frequency of each block is given by:

$$d_{ijk} = \frac{e_{ijk}}{m_{ijk}} \tag{2.17}$$

The total perceptual error is then given by:

$$d(i,i') = \frac{1}{N^2}[\sum_{i,j}(\sum_{k} d_{ijk}^{\beta_s})^{\frac{\beta_f}{\beta_s}}]^{\frac{1}{\beta_f}} \tag{2.18}$$

Watson recommends $\beta_s = \beta_f = 4$. In our scheme, total perceptual error $d(i,i')$ is calculated by taking i as our original image and i' as modified image. The details are given in [63].

2.4.2 Robustness parameter

To measure the robustness of the watermarking scheme, *Hamming* distance of the extracted watermark and the original watermark is calculated. The Hamming distance (H) between the original watermark W and the extracted watermark W' is calculated using Eqn.2.19.

$$H = \frac{1}{L}\sum_{i=1}^{L} W_i \oplus W_i' \tag{2.19}$$

where L is length/size of the watermark signal.

2.5 Experimental Dataset

The data set (video sequences and images) used for the experimentation covers a wide range of different video and images. Th data set are enlisted in Table 2.1.

2.6 Summery

In this chapter few background concepts eg. MCTF, SIFT etc. as well as few evaluation parameter are explained. These concepts are in different phase of the watermarking schemes proposed in the later chapters. The evaluation parameters described are used to measure the efficiency of those scheme. Along with that dataset used for experimentation are also summarized in this chapter.

Video Sequence Used	*Bus, Foreman, Crew, Coast Guard, Mobile, Akiyo, City, Hall, Mother Daughter, Sunflower , Pedestrian area , News*
Image Data sets	Complex Scene Saliency Dataset (CSSD) Extended Complex Scene Saliency Dataset (ECSSD) [64], Caltech 256 dataset [65], LabelMe-12-50k dataset [66], few other standard images eg. *lena, cameraman, baboon* etc.
Video Resolution	1080p(Full HD), CIF, 4CIF
Watermark signal	32×32 and 64×64 binary image

Table 2.1: *Experimental Data set*

Chapter 3

Robust Video watermarking against Resolution and Quality Scalability

3.1 Introduction

As mentioned in introduction chapter, there are two main characteristics of the scalable video watermarking. Firstly, watermark should be extractable from each layer and secondly, robustness of the watermark should be increased with increase of enhancement layers. It has been observed in the literature (refer to Sec. 1.2) that existing schemes against resolution scaling [43, 44] fail to achieve both of the above mentioned requirements for scalable watermarking.

In this chapter, a video watermarking algorithm is proposed which is robust against spatial and quality adaptation attacks. In the proposed scheme, watermark can be extracted from all the layers as well as the scheme has achieved the graceful improvement in the successive higher layers for both resolution and quality scaling.

To handle the spatial and quality adaptation attack, in the proposed work, watermark is embedded by altering the coefficients of the motion compensated temporal filtered DC frames which are generated by accumulating DC values of non-overlapping blocks of original video frames. The proposed scheme is described in subsequent sections.

3.2 Background

In the proposed scheme, watermark is embedded in the DCT based motion compensated coefficients of the DC frames. DC frames are generated by accumulating DC values of non-overlapping blocks of given size from the original frame. The main motivation of the proposed scheme is to prevent the degradation of the embedded watermark signal when a controlled resolution scaling is performed on the watermarked video frame to achieve the resolution scalability. DC frame based embedding helps to handle this controlled resolution scaling. A detail discussion of the DC frame is presented in the Sec. 3.2.1.

Another important motivation of the proposed scheme is to achieve the graceful improvement of the watermark signal with higher enhancement layers. The most essential requirement to achieve this improvement is to maintain the spatial synchronization of watermark embedding locations between every successive layers. In other words, in every enhancement layer, two watermark signals are added (one is up-sampled from previous layer and other is embedded in the residual layer of the corresponding enhancement layer) and they should be co-located. Intuitively, this spatial coherency achieves the graceful improvement of the embedded watermark signal in successive enhancement layers. The concept is elaborated in Sec. 3.2.2.

In this work, motion compensated DCT based temporal filtering (MCDCT-TF) is done on the DC frames and low pass frames are selected for embedding. Embedding in the low pass frames, spreads the embedded watermark into all frames such that frame dropping and collusion attacks can be resisted. During the inverse MCDCT-TF, additive noise due to embedding affects the all frame in a GOP. Thus, the watermark is spread over all the frames in that GOP. Motion coherent embedding helps to reduce temporal flickering. MCDCT-TF is discussed in Sec. 2.1.1. During MCDCT-TF, only temporally connected pixels are considered for temporal filtering and for watermark embedding. The concept of connected-unconnected pixels is narrated in Sec. 2.1.3.

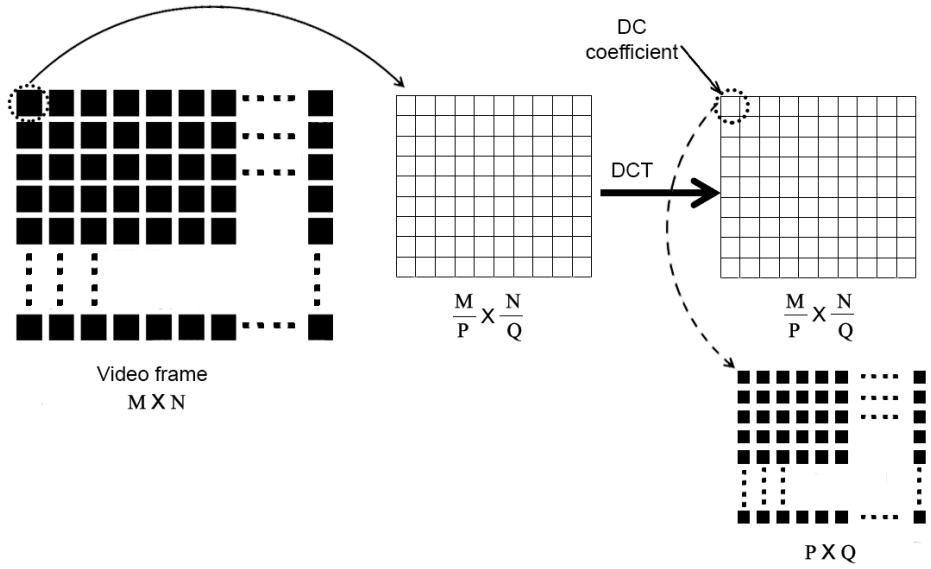


Figure 3.1: *Block DCT of video frame*

3.2.1 DC Frame

The concept of DC frame based watermarking is first introduced by Y. Wang et. al. [47] where DC frame is generated by accumulating the DC value (after 2D block DCT transform) of non overlapping 8×8 blocks within a video frame. Embedding in the DC frame results in the spreading of the watermark signal during the up-sampling process for generating enhancement layer frames. In the proposed scheme, the size of the DC frame is fixed and the non-overlapping block size are determined based on the size of full resolution video frame size.

The DC frame formation in this work is depicted in the Fig. 3.1. Since the DC frame size $(P \times Q)$ is fixed, the video frame $(M \times N)$ is divided into $(\frac{M}{P} \times \frac{N}{Q})$ non-overlapping blocks. The DC frame is generated by accumulating all DC coefficients of such $(\frac{M}{P} \times \frac{N}{Q})$ non-overlapping blocks after 2D block DCT.

3.2.2 Graceful Improvement of the Watermark

Graceful improvement [20] means, the improvement in the quality of the extracted watermark signal along with the increase video quality (addition of successive

enhancement layer with respect to the different scalable parameters). Intuitively, there are two main problems associated with this DC frame based embedding to achieve graceful improvement. Firstly, since enhancement layers are predicted from base layer, the embedded watermark signal in the base layer is also up-sampled in higher layers and there is a chance of watermark signal degradation when residual component are added at different higher layers. Secondly, as a consequence of first case, in every successive higher layers, watermark signal are getting degraded due to this error propagation to the watermark signal. So, the grace full up-gradation of the watermark signal is not possible, rather there is a chance for continuous degradation of the watermark signal towards the higher layers. Moreover, the enhancement layer residual component (to be added in each enhancement layer with the up-sampled version from the lower layer) is not secured.

As a counter-measure, an up-sampled watermark can be separately added to the residual components of each enhancement layer which is to be added to the up-sampled version from the previous lower layer [44]. The main challenge in this proposed technique is that two watermark signal (one which is getting up-sampled with last lower layer of the original signal and the other which is embedded in residual component of the current enhancement layer) should be spatially collocated such that addition of two watermark signal should up-grade their signal strength rather degrade it. A location map ($Lmap$) is used in the proposed scheme to maintain this spatial coherency when adding two watermark signal in any enhancement layer. The proposed location map ($Lmap$) based technique for achieving said spatial coherency is depicted in Fig. 3.2. It is observed in the Fig. 3.2 that the location map ($Lmap$) is obtained during the embedding in the DC frame. The location map ($Lmap$) which is a binary matrix is up-sampled (according to the resolution ratio of the base layer (DC frame) and corresponding enhancement layer) to get the required location map ($Lmap^U$) for a particular enhancement layer. The up-sampled location map ($Lmap^U$) is used during the embedding in the residual frame for that particular enhancement layer to achieve spatially co-located watermark embedding.

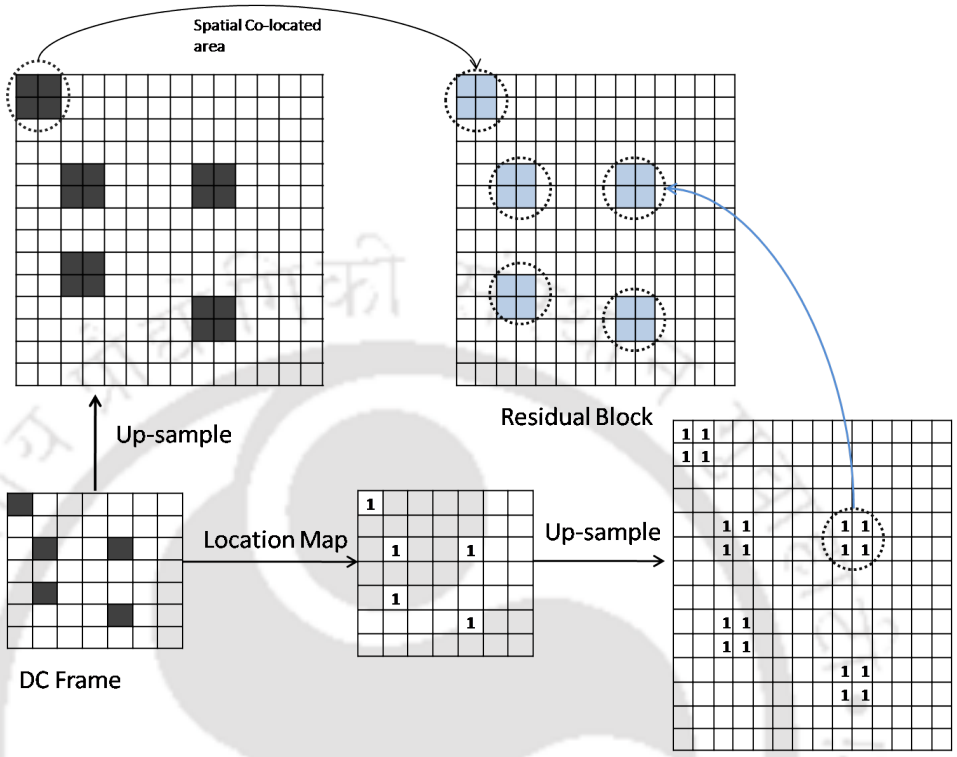


Figure 3.2: *Location Map based Technique for Spatial Coherency*

3.3 Proposed Scheme

3.3.1 Watermark Embedding Scheme

In this subsection, proposed watermarking embedding scheme has been described. The embedding scheme has three modules. Firstly, in Sec. 3.3.1, embedding zone selection is discussed using spatio-temporal filtering with the help of the robustness as well as the visual quality thresholds. In Sec. 3.3.1 and Sec. 3.3.1, base layer embedding and enhancement layer embedding schemes are presented respectively. The block diagram for the overall watermark embedding scheme is depicted in the Fig. 3.3.

Watermark Zone Selection

In this subsection, block based zone selection is described to embed the watermark signal. Each frame of the video sequence I is subjected to 2D Block DCT

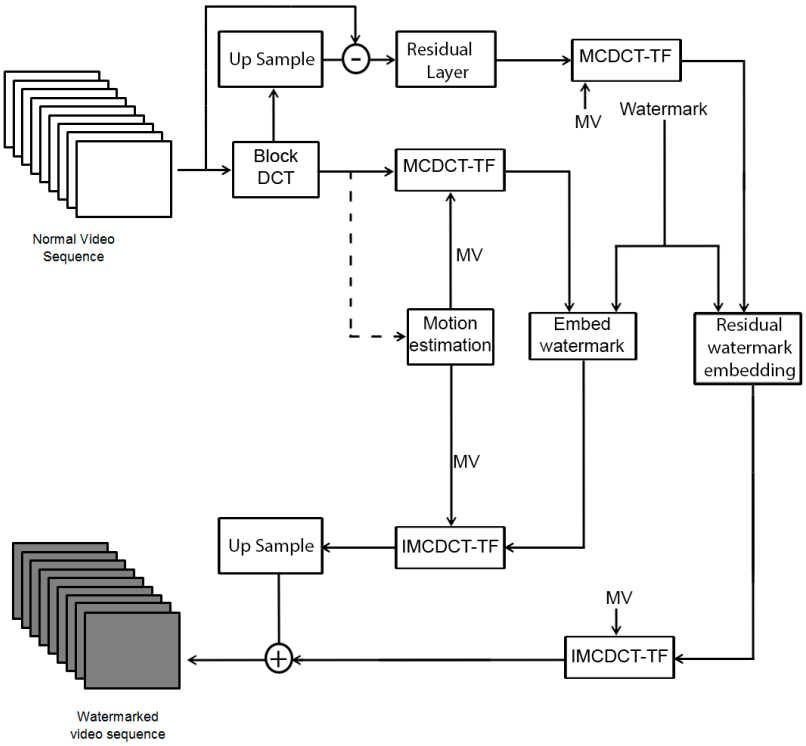


Figure 3.3: *Watermark Embedding Model*

transform as described in Fig. 3.1 and the DC values of the non overlapping blocks of size $(\frac{M}{P} \times \frac{N}{Q})$ (Sec. 3.2.1) are accumulated to obtain the DC frame sequence C which is considered as the base layer.

The base layer frames i.e. DC frames (C) are up-sampled to obtain the predicted enhancement layer frame sequence (say E) using Eqn. 3.1

$$E_i = \uparrow C_i \tag{3.1}$$

where i is the frame index. Up-sampled DC frames is subtracted from the original enhancement layer video frames to get the residual frame sequence R using Eqn. 3.2.

$$R_i = V_i - E_i \tag{3.2}$$

MCDCT-TF has been employed (as described in Fig. 2.2) on the extracted base layer as well as residual of enhancement layers. The motion compensation has been done using Eqn. 2.1, 2.2 and the connected and unconnected pixel regions are identified as shown in Fig. 2.3. Using the Eqn. 2.3, the DCT based temporal filtering has been done to get the low pass coefficients C_t from motion compensated base layer and Rt from motion compensated residual layer.

In the proposed watermarking scheme, the watermark is embedded in base layer (DC frame) by modifying the DC values. Three consecutive coefficient in low pass frame (namely C_{t1}, C_{t2}, C_{t3}) are used for embedding one watermark bit. It may sometime happen that the absolute difference between C_{t2} and the average of C_{t1} and C_{t3} (i.e. $|C_{t2} - (C_{t1} + C_{t3})/2|$) is relatively high. For this case, proposed embedding scheme adds relatively higher noise which may cause flickering artifacts [60]. As a countermeasure, an adaptive threshold [*say Visual Quality Threshold* (V_{th})] is incorporated which is described in Eqn. 3.3 to select the suitable coefficients for watermark embedding such that the embedding noise will be under an acceptable limit. The value of the threshold is $(C_{t1} + C_{t3}) * 2\alpha$, (where α is the *robustness threshold*) which is used to increase the embedding strength of the watermark to make it robust against the content adaptation attack. Generally the value of α is taken very close to 0 to avoid the objectionable artifacts. In this scheme, α value is taken as 0.01 for experimentation.

$$\left.\begin{array}{ll} if & \left|(\frac{C_{t1}+C_{t3}}{2} - C_{t2})\right| \leqslant V_{th} \quad then\ the\ coefficients \\ & \qquad\qquad\qquad\qquad\qquad trio\ are\ selected\ for\ embedding \\ else & the\ coefficients\ trio\ are\ rejected \end{array}\right\} \tag{3.3}$$

where $|\ |$ represents absolute value, C_{t1}, C_{t1} and C_{t3} are three consecutive coefficient in low pass frames. In the proposed scheme, if the coefficients do not satisfy the visual threshold then the middle value is increased (or decreased) to a certain limit to mark the corresponding set of coefficients unsuitable for the embedding such that at the time of extraction, decoder can identify the set easily as non-embedded set. A location map ($Lmap$) is derived during the time of base layer (DC frame) embedding which is used to select embedding coefficients during any enhancement layer embedding (refer to Sec. 3.2.2).

Base Layer Embedding

In the proposed scheme, the watermark is embedded into the motion compensated low pass DC frames as shown in Fig. 2.2. The proposed blind watermark embedding scheme is depicted in Fig.3.3. A set of 3 consecutive coefficients [Ct_1, Ct_2, Ct_3] which satisfies the visual quality threshold (refer to Eqn. 3.3) is selected for embedding using Eqn. 3.4.

$$Ct_2' = \frac{(Ct_1 + Ct_3)}{2} + \left|\frac{(Ct_1 + Ct_3)}{2}\right| * \alpha * W_i \tag{3.4}$$

where $W_i \in (0, 1)$ is the watermark bit and α is the robustness threshold (watermarking strength) (refer to Sec. 3.3.1). Ct_2' is the watermarked coefficient corresponding to Ct_2. The embedding location is saved in a location map ($Lmap$). An up-sampled version of the location map ($Lmap^U$) is used to locate the spatial coherent locations in the residual frame for a particular enhancement layer embedding. After embedding of watermark bits, the motion compensated inverse temporal filtering (IMCDCT-TF) is done to get base layer watermarked video sequence C'.

Enhancement Layer Embedding

Similar to the base layer, the residual frame sequence (R) is partitioned into non overlapping set of 3 residual frames in the temporal direction. For such a set of 3 residual frames (say R_1, R_2, R_3), R_1^2 and R_3^2 are predicted from R_1 and R_3 respectively using the motion vectors 1-D temporal DCT of R_1^2, R_2 and R_3^2 is calculated using Eqn. 2.3 to generate the low pass temporal filtered residual frame Rt. Then the Base layer location map ($Lmap$) is up-sampled to the size of residual layer to detect the watermark regions in the low pass temporal residual layer.

The watermarking region of the low pass residual frame (Rt) is again partitioned into a non-overlapping set of 3 consecutive coefficients (say $Rt(k)$, $Rt(k+1)$ and $Rt(k+2)$) and the watermark is embedded using the Eqn. 3.5.

$$Rt'(k+1) = \frac{(Rt(k) + Rt(k+2))}{2} + \left|\frac{(Rt(k) + Rt(k+2))}{2}\right| * \alpha * W_i \tag{3.5}$$

where W_i is the same watermark bit embedded in spatially coherent base layer (low pass DC frame) coefficients and α is the watermarking strength. Rt'_2 is watermarked coefficient corresponding to coefficient Rt_2 of the residual layer.

After embedding, the low pass residual frames are subjected to the IMCDCT-TF to get watermarked residual for the enhancement layer sequence as Rw. The base layer watermark video coefficient C' is up-sampled to the enhancement layer using Eqn. 3.1 to E' and added with Rw to get the watermarked video sequence. The overall watermark embedding procedure is narrated in Algorithm 3.1 and 3.2.

3.3.2 Extraction Scheme

Watermark extraction for spatial scalability is done for base layer as well as from residual layer separately. The enhancement layer watermark is derived by using the base and enhancement layer watermark.

Extraction of Base Layer Watermark

To extract the base layer watermark, the pre-processing steps are same as the embedding scheme as shown in Fig. 3.4. For the content authentication of base layer, first DC frame sequence ($\acute{C}$) is formed and MCDCT-TF is done on $\acute{C}$. Similar to the embedding scheme, the visual threshold is used to detect the embedding zone. Watermark is extracted from a set of three consecutive non-overlapping coefficient Ct'_1, Ct'_2, Ct'_3 from the detected watermarked block using Eqn. 3.6.

$$\left.\begin{array}{ll} W'_{bi} = 0 & if\ Ct'_2 \leqslant \frac{(Ct'_1+Ct'_3)}{2} \\ W'_{bi} = 1 & if\ Ct'_2 > \frac{(Ct'_1+Ct'_3)}{2} \end{array}\right\} \tag{3.6}$$

Similar to the embedding scheme, the watermark locations are saved in a location map ($Lmap$). A up-sampled version of the location map ($Lmap$) is used to locate the spatial coherent locations in the residual frame for a particular enhancement layer extraction. Extraction scheme is describe stepwise in Algorithm 3.3.

Algorithm 3.1: Embedding Algorithm (V, α, W)

Input: V:Raw Video, and α: Watermark Strength W: Watermark Bit stream

Output: V^w:Watermarked Video

1. /∗ Generate the DC frame (base layer) video C_i from the Raw video V ∗/

   for *each video frame* (V_i) *from raw video sequence* V **do**

   (a) Partition video frame (V_i) into non overlapping blocks of size $\frac{M}{P} \times \frac{N}{Q}$ as Fig. 3.1. $M \times N$ is the frame size and $P \times Q$ is fixed DC frame size.

   (b) Accumulate DC values of each blocks to obtain the DC frame (C_i) for the corresponding raw video frame (V_i) (refer to Fig. 3.1).

   (c) Up-sample DC frame (C_i) and subtract from the raw video frame (V_i) to get the residual video frame (R_i) using Eqn. 3.1, 3.2.

2. Partition the DC frame sequence (say C) corresponding to whole raw video sequence (V) into non overlapping set of k DC frames in temporal direction. In this algorithm, k is taken as 3 for experimental basis.

3. For such a set of 3 DC frames $\{C_1, C_2, C_3\}$, calculate the motion vectors (MV) and predict the C_2 from the C_1 and C_3 using MVs. Let C_2^1 and C_2^3 are predicted frames respectively.

4. Calculate coefficient wise temporal DCT of C_2^1, C_2 and C_2^3 using Eqn. 2.3 to generate the low pass temporal filtered DC frame Ct.

5. Partitioned the low pass DC frame (Ct) into a non-overlapping set of 3 consecutive coefficients

   for *each such set of 3 coefficients* **do**

   - **if** *the corresponding set satisfy the visual threshold (refer Eqn. 3.3)* **then**

     (a) Embed the watermark in the selected coefficient set using Eqn. 3.4.

     (b) Take watermark reference location from the base layer watermarking to location map $Lmap$.

6. Do the Inverse MCDCT-TF to get the watermarked DC frame sequence C'.

7. Up-sample the watermarked DC frame to a required enhancement layer.

8. /∗Call the residual layer embedding function to get the residual watermark frame Rw as described in Algorithm3.2 ∗/.

   $Rw = Residual\ Embedding(R, \alpha, W, Lmap, MV)$

9. Add the up-sampled watermark DC frame and watermark residual layer Rw to get the watermark video V^w.

Algorithm 3.2: Residual Embedding $(R, \alpha, W, Lmap, MV)$

Input: R:Residual Layer, α: Watermark Strength W: Watermark Bit stream, Lmap: Base layer location map and MV: Motion vector

Output: Rw:Watermarked Residual Layer

1. The Residual frame sequence (say R) is partitioned in to non overlapping set of k residual frames in temporal direction. in this algorithm, k is taken to 3 for experimental basis.
2. For such a set of 3 residual frames $\{R_1, R_2, R_3\}$ using the MV the $R_2^1$and R_2^3 are predicted from the $R_1$and R_3.
3. Calculate coefficient wise temporal DCT of R_2^1,$R_2$and R_2^3 using Eqn. 2.3 to generate the low pass temporal filtered residual frame Rt.
4. Take the Base layer location map (Lmap) and up-sample to size of residual layer($Lmap^u$) to detect the watermark regions in the low pass temporal residual layer.
5. The watermarking region of the low pass residual frame (Rt) is again partitioned into a non-overlapping set of 3 consecutive coefficients.

   for *each such set of 3 coefficients* **do**

   - Embed the watermark in the residual coefficient group using Eqn. 3.5.
6. Do the IMCDCT-TF on the watermark low pass residual layer to get the watermarked residual layer Rw.
7. **return** (Rw)

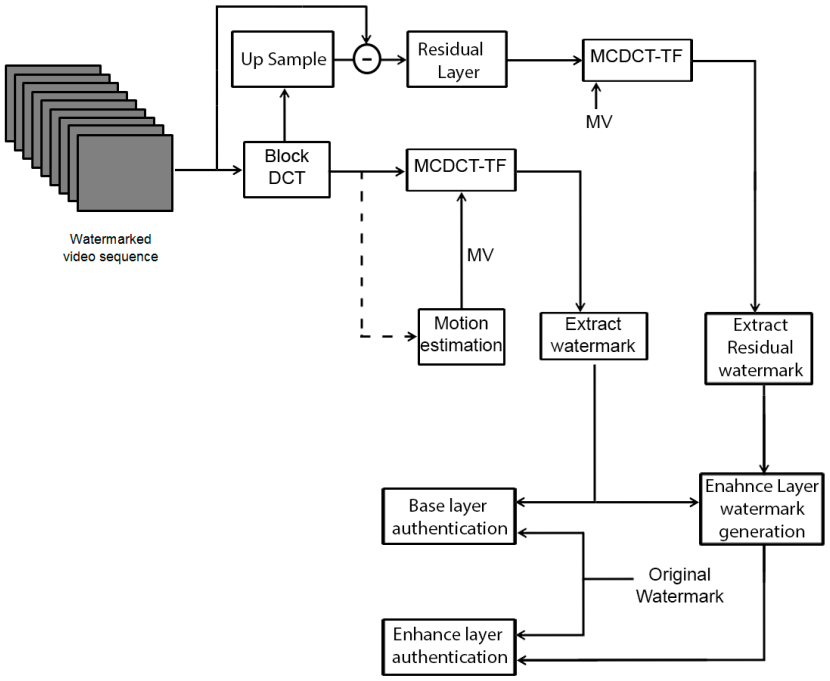


Figure 3.4: *Watermark Extraction Model*

Extraction of Enhancement Layer Watermark

Extraction of watermark from the residual layer is important for content authentication of the corresponding enhancement layer. The extraction scheme is same as the embedding scheme of the residual layer. Base layer location map (Lmap) is up-sampled to the size of residual layer ($Lmap^u$) to detect the watermark regions in the low pass temporal residual layer. The watermarked regions of the low pass watermarked residual frame (Rt') are again partitioned into a non-overlapping set of 3 consecutive coefficients (say $Rt'(k)$, $Rt'(k+1)$ and $Rt'(k+2)$ and extract the watermark using Eqn. 3.7.

$$Wt'_i = \sum_{k=0}^{\lfloor N/3 \rfloor} \frac{(Rt'(3k) + Rt'(3k+2))/2 - Rt'(3k+1)}{|(Rt'(3k) + Rt'(3k+2))/2 - Rt'(3k+1)|} \tag{3.7}$$

where N is number of residual coefficient in a block coherent to a single watermarked coefficient in base layer (low pass DC frame) as shown in Fig. 3.2. After extraction of each block, the binary watermark is generated using Eqn. 3.8. Stepwise residual layer extraction is presented in Algorithm 3.4.

$$\left.\begin{array}{ll} W'_{ri} = 0 & if\ Wt'_i \leqslant 0 \\ W'_{ri} = 1 & if\ Wt'_i > 0 \end{array}\right\} \tag{3.8}$$

The base layer and the residual layer watermark is combined to generate the enhancement layer watermark using Eqn. 3.9.

$$W' = W'_b \bigcup W'_r \tag{3.9}$$

3.3.3 Embedding Capacity

Embedding capacity of the proposed scheme depends on two factors, the visual quality threshold given in Eqn.3.3 and the number of connected coefficients obtained from temporal filtering. Let the number of connected coefficients per frame is η. So, at most $\eta/3$ watermarking bits can be embedded. The embedding capacity is further reduced by the visual quality threshold. Intuitively embedding capacity for GOP directly proportional to η and inversely proportional to V_{th}.

Algorithm 3.3: Extraction Algorithm (V^w, α)

Input: V^w:Watermarked Video, and α: Watermark Strength

Output: W'_b: Extracted base layer watermark, W'_e: Extracted enhancement layer watermark

1. /∗ Generate the DC frame (base layer) watermarked video C' from the watermarked video V^w as described in Sec. 3.2.1 ∗/

   for *each video frames* V_i^w *from watermarked video sequence* V^w **do**

   - The watermarked video frame V_i^w is partitioned to non overlapping blocks of size $\frac{M}{P} \times \frac{N}{Q}$ as on Fig. 3.1.
   - DC values of each blocks are accumulated to obtain the watermarked DC frame (C'_i) for the corresponding watermarked video frame V_i^w refer to Fig. 3.1.
   - Watermarked DC frame (C'_i) is up-sampled and subtracted from the watermarked video frame (V_i^w) to get the watermarked residual video frames (Rw_i) using Eqn. 3.1, 3.2.

2. The watermarked DC frame sequence (say C') corresponding to whole watermarked video sequence (V^w) is partitioned into non overlapping set of k DC frames in temporal direction. In this algorithm, k is taken as 3 for experimental purpose.
3. For such a set of 3 watermarked DC frames $\{C'_1, C'_2, C'_3\}$ the motion vectors (MV) is calculated and the $C'^1_2$and C'^3_2 are predicted from the $C'_1$and C'_3 using motion MV.
4. Calculate coefficient wise temporal DCT of C'^1_2,$C'_2$and C'^3_2 using Eqn. 2.3 to generate the low pass temporal filtered watermarked DC frame Ct'.
5. The low pass watermarked DC frame(Ct') is again partitioned into a non-overlapping set of 3 consecutive coefficients.

   for *each such set of 3 coefficients* **do**

   - **if** *the corresponding set satisfy the visual threshold (refer Eqn. 3.3)* **then**
     (a) Extract the watermark W'_b in the selected coefficient group using Eqn. 3.6.
     (b) Take watermark reference location from the base layer extraction to location map $Lmap$.

6. /∗ Call the residual layer extraction function to get the residual watermark W'_r as described in Algorithm 3.4.∗/

   $W'_r = Residual\ Extraction(Rw, \alpha, Lmap, MV)$

7. Generate the enhancement layer watermark W'_e using Eqn. 3.9 by combining the base and residual layer watermark.

Algorithm 3.4: Residual Extraction $(Rw, \alpha, Lmap, MV)$

Input: Rw: Watermarked Residual Layer, α: Watermark Strength, Lmap: Base layer location map and MV: Motion vector

Output: W'_r: Watermark extracted from the Residual Layer

1. The watermarked Residual frame sequence (say Rw) is partitioned into non overlapping set of k residual frames in temporal direction. In this algorithm, k is taken to 3 for experimental basis as on embedding scheme.
2. For such a set of 3 watermarked residual frames $\{Rw_1, Rw_2, Rw_3\}$ using the MV the Rw_2^1 and Rw_2^3 are predicted from the Rw_1 and Rw_3.
3. Calculate coefficient-wise temporal DCT of Rw_2^1, Rw_2 and Rw_2^3 using Eqn. 2.3 to generate the low pass temporal filtered watermarked residual frame Rt.
4. Take the Base layer location map (Lmap) and up-sample to size of watermarked residual layer to detect the watermarked regions in the low pass temporal watermarked residual layer.
5. The watermarking region of the low pass watermarked residual frame (Rwt) is again partitioned into a non-overlapping set of 3 consecutive coefficients.

   for *each such set of 3 coefficients* **do**

   - Extract the watermark from the residual layer using Eqn. 3.7.
6. Generate the residual binary watermark W'_r using Eqn. 3.8.
7. **return** (W'_r)

Table 3.1: *Experimental Setup*

Parameters for Watermarking	Values Taken
Encoder	H.264/SVC version
Reference Software	JSVM
Video Sequence Used	Bus, City, Crew, Coastguard, Mobile, Akiyo, Hall, Mother Daughter, Sunflower , Foreman, News
Video Resolution	4CIF, CIF, 720p
Watermark signal	32×32 and 64×64 binary image
Visual Quality Metrics	PSNR, flicker metric, VQM, SSIM
Robustness metric	Hamming distance

3.4 Experimental Results

The proposed method is tested on different High Definition (*Elephants dream, Sunflower, Pedestrian area*) and CIF (*Foreman, Akiyo, Bus*) video sequences. Videos with different motion characteristic are selected for the experimentation. *Pedestrian* area and *Bus* sequences are high motion video where *Sunflower* and *Akiyo* have relatively low motion . A 64×64 binary image is used as watermark signal. The fixed DC frame size is taken as 88×72. The watermark signal is embedded in the luma component using the proposed watermarking scheme. PSNR, SSIM [59] VQM [61] and flicker metric [60] are measured using the MSU VQM Tool [62] to evaluate the visual quality for the proposed scheme. To measure the robustness of the scheme, hamming distance is used. The robustness analysis is done at different resolutions extracted from H.264/SVC encoded bit stream. Whole experimental setup is tabulated in Table 3.1.

3.4.1 Visual Quality

The visual quality of the proposed scheme is compared with the related spatial scalable scheme proposed by Y. Wang and A. Pearmain [47] and quality scalable watermarking scheme proposed by Bhowmik et al. [39]. In Y. Wang and A.

Pearmain scheme [47], the watermark is embedded in the 2^{nd} frame of GOP of 3 frames. So for the comparison, PSNR, VQM, SSIM are calculated on the 2^{nd} frame of GOP of 3 consecutive frames. The flicker metric [60] is used to find out the blinking effect of the video, it has been measured on the full video sequence.

The comparative result of the proposed scheme with the Wang and A. Pearmain's scheme [47] and Bhowmik's scheme [39] with respect to PSNR, flicker metric, VQM and SSIM has been depicted in Fig. 3.5, 3.6, 3.7 and 3.8 respectively.

PSNR comparison for *Pedestrian area*, *Sunflower*, *Akiyo* and *Bus video* is shown in Fig. 3.5. The average PSNR for the proposed scheme is observed as close as 40dB for all the videos which may be regarded as acceptable quality. It can be also observed that the proposed scheme is much better than that of Wang's scheme [47] and average PSNR of all the frames is better than Bhowmik's scheme [39]. In Fig. 3.6, absolute difference of the flicker metric between original and watermarked video is shown. Because of motion coherent embedding, flicker difference is close to zero for the proposed scheme. In Bhwomik's scheme, despite use of MCTF, flickering is high because of shorter filter length.

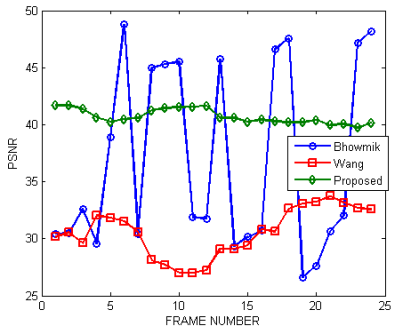
Fig. 3.7 depicts the VQM comparison of above mentioned three schemes. Lower value of VQM metric means better visual quality [61]. Fig. 3.7 proves that the visual quality of the proposed scheme is better than the existing schemes.

The comparison results for the *Pedestrian area* and *Sunflower video* shows that the proposed scheme produces less visual artifact than Wang's scheme [47] as well as Bhowmik's scheme [39] for both the very low and very high motion HD videos.

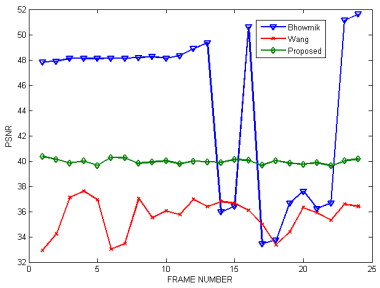
3.4.2 Robustness

Robustness of the proposed watermarking scheme is measured by means of Hamming distance between the original watermark and the extracted watermark after H.264/SVC content adaptation attack. The watermark is extracted from the different resolution layers after the watermarked video is compressed with the scalable encoder H.264/SVC with 5 different possible scaled versions.

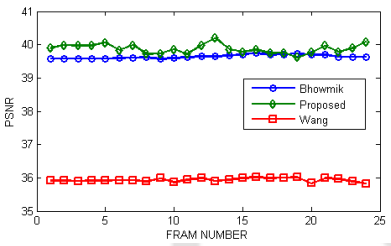
- Full resolution video.



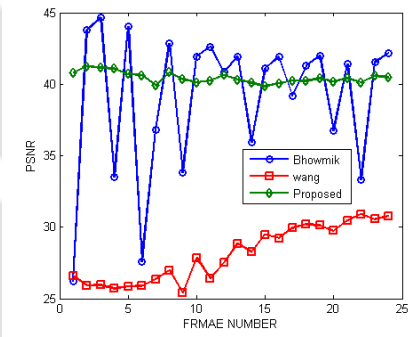
(a) *Pedestrian area video*



(b) *Sunflower video*



(c) *Akiyo video*

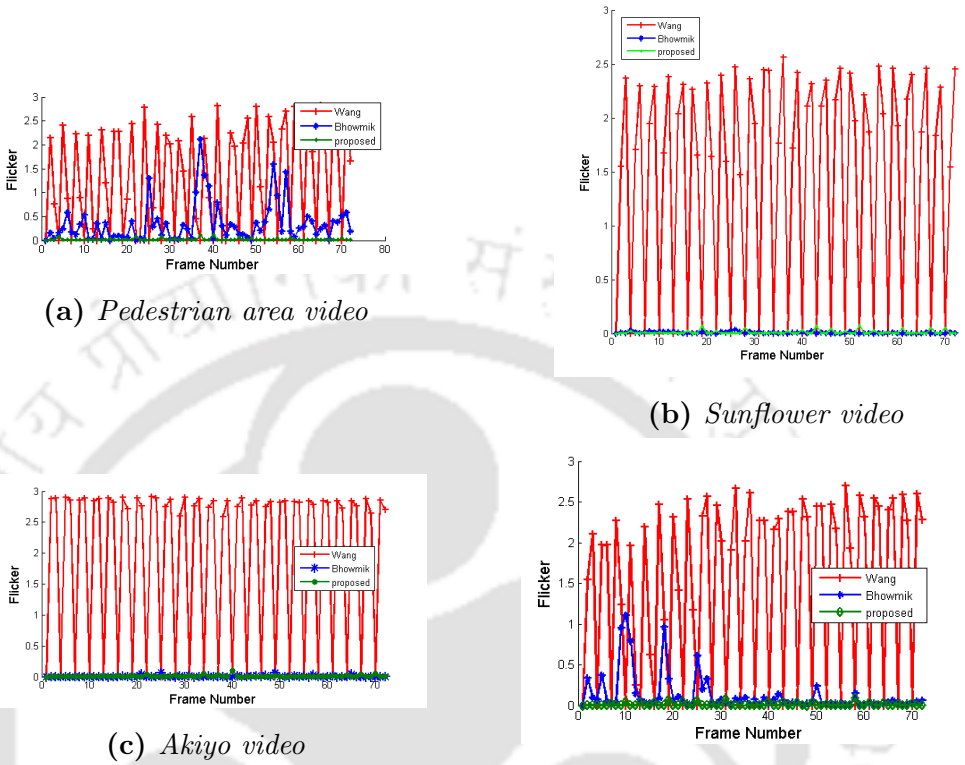


(d) *Bus video*

Figure 3.5: *PSNR comparison*

- Vertical $\frac{1}{2}$ resolution video
- Horizontal $\frac{1}{2}$ resolution video
- Vertical $\frac{1}{2}$ resolution & Horizontal $\frac{1}{2}$ resolution video
- Random (down sampled) resolution video

Video sequences with different resolutions at different bit rates (quality with respect to QP) are evaluated to analyze the robustness of the proposed scheme against resolution and quality adaptation attacks. Table 3.2 gives the the hamming distance of the watermark signal extracted from base layer and enhancement layers of watermarked *Bus* video at different bit rates. Tables 3.3, 3.4, 3.5 present similar results for *Akiyo* , *Pedestrian area* and *Sunflower* video.



(a) *Pedestrian area video*

(b) *Sunflower video*

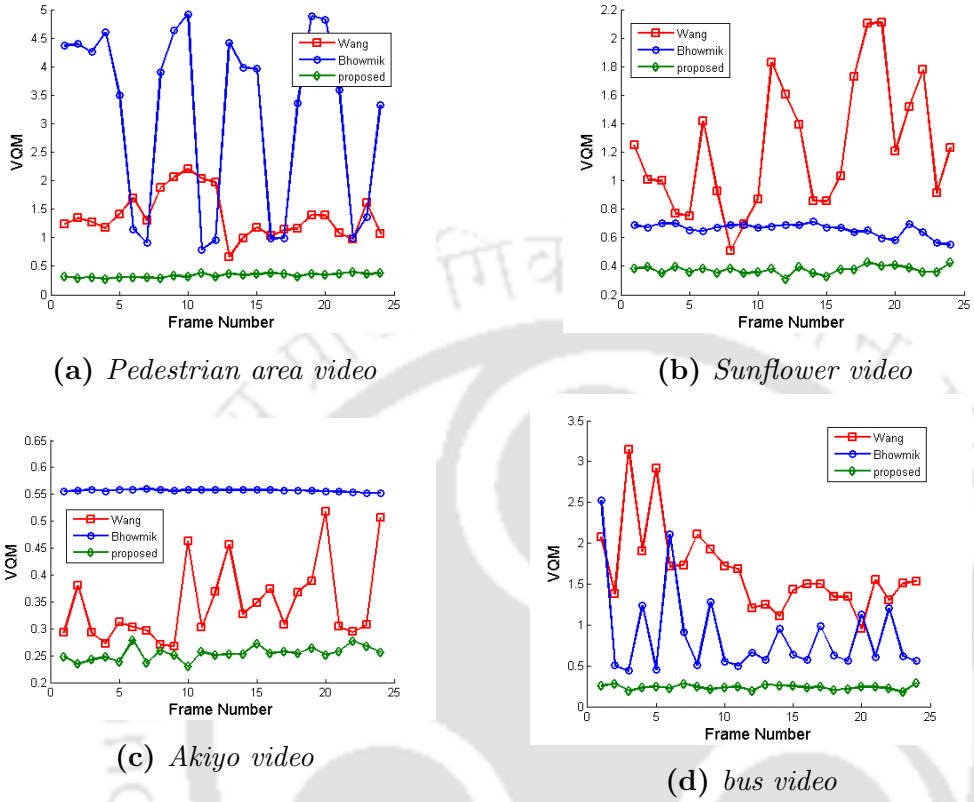
(c) *Akiyo video*

(d) *Bus video*

Figure 3.6: *Flicker Metric comparison*

From Tables 3.2 to 3.5, it is observed that the quality (hamming distance between extracted watermark and the original watermark) of the extracted watermark for enhancement layer is relatively higher than that of base layer. Moreover, it is also observed that robustness is increasing with higher levels of enhancement layer. This observation advocates the proposed claim of graceful improvement.

Robustness of the proposed scheme has been compared with the Wang and A. Pearmain's scheme [47] and Bhowmik's scheme [39] for *Bus*, *Akiyo*, *Pedestrian area* and *Sunflower* videos. The watermarked raw videos are encoded with H.264/SVC video encoder and the extracted watermark is compared with the original watermark to measure the robustness of the scheme against scalable adaptation. Since, in the Wang et. al.'s scheme [47], the full resolution video is necessary for watermark extraction, the comparison with proposed scheme is done only for high resolution versions of the video sequences. In Fig. 3.9, the



(a) *Pedestrian area video*

(b) *Sunflower video*

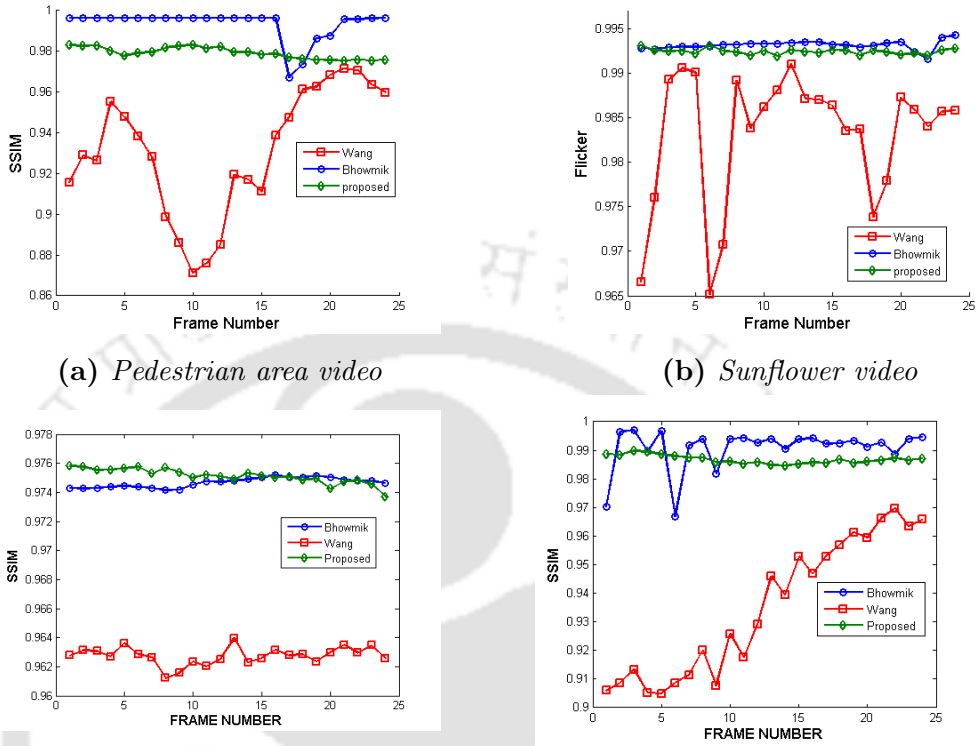
(c) *Akiyo video*

(d) *bus video*

Figure 3.7: *VQM comparison*

comparison of Hamming distance (as robustness metric) between proposed and existing schemes [39, 47] are presented. It is observed from the Fig. 3.9 that the proposed scheme for enhanced layer video as well as base layer video performs better than both schemes with respect to hamming distance.

From the results depicted from Fig. 3.9, it is observed that proposed scheme provides better performance for both base and enhancement layer video sequence than that of existing schemes [39, 47] . From Table 3.2 to 3.5, it is evident that both base layer and the enhanced layers are secured by the proposed scheme. Moreover, the graceful improvement has been achieved for successive enhancement layers of the on video sequences.



(a) *Pedestrian area video* **(b)** *Sunflower video*

(c) *Akiyo video* **(d)** *Bus video*

Figure 3.8: *SSIM comparison*

3.5 Conclusion

In this chapter, a DC frame based blind watermarking scheme has been proposed which can resist resolution scalability. A DCT based spatial filtering and a MCDCT based temporal filtering are used to find the suitable embedding zone for the embedding. Moreover, a robustness threshold and a visual quality threshold are used to enhance visual quality and robustness of the proposed scheme. A comprehensive set of experiments have been carried out to justify that the proposed scheme is performing better than the existing related works [39, 47] with respect to visual quality as well as robustness against resolution scalability. Proposed scheme ensures a graceful improvement of the extracted watermarking signal with successive enhancement layers. In this chapter, the watermarking issues are discussed only for the resolution and quality adaptation. In the sim-

Table 3.2: *Hamming distance of the extracted watermark from scalable CIF Bus video*

Hamming distance of Bus video Watermark				
Resolution	QP	Bit rate $kbit/s$	Base layer	Enhancement layer
352×288	26	2600	0.02356	0.01866
	28	2200	0.04121	0.03015
	30	1910	0.05453	0.03848
	32	1490	0.075921	0.05137
176×288	26	1238	0.03485	0.02626
	28	957	0.05914	0.04161
	30	827	0.07023	0.04824
	32	514	0.09756	0.06447
352×144	26	1352	0.03158	0.02453
	28	1138	0.06042	0.04168
	30	1023	0.07438	0.04978
	32	775	0.1048	0.06633
176×144	26	385	0.06584	0.04428
	28	294	0.1014	0.06394
	30	246	0.1252	0.07423
	32	156	0.1523	0.09213
272×192	26	691	0.0486	0.03536
	28	538	0.05996	0.04511
	30	445	0.06953	0.05487
	32	263	0.08659	0.07509

ilar line of thought, temporal adaptation should also be analyzed. In the next chapter, the temporal scalability issue has been considered.

Table 3.3: *Hamming distance of the extracted watermark from scalable CIF Akiyo video*

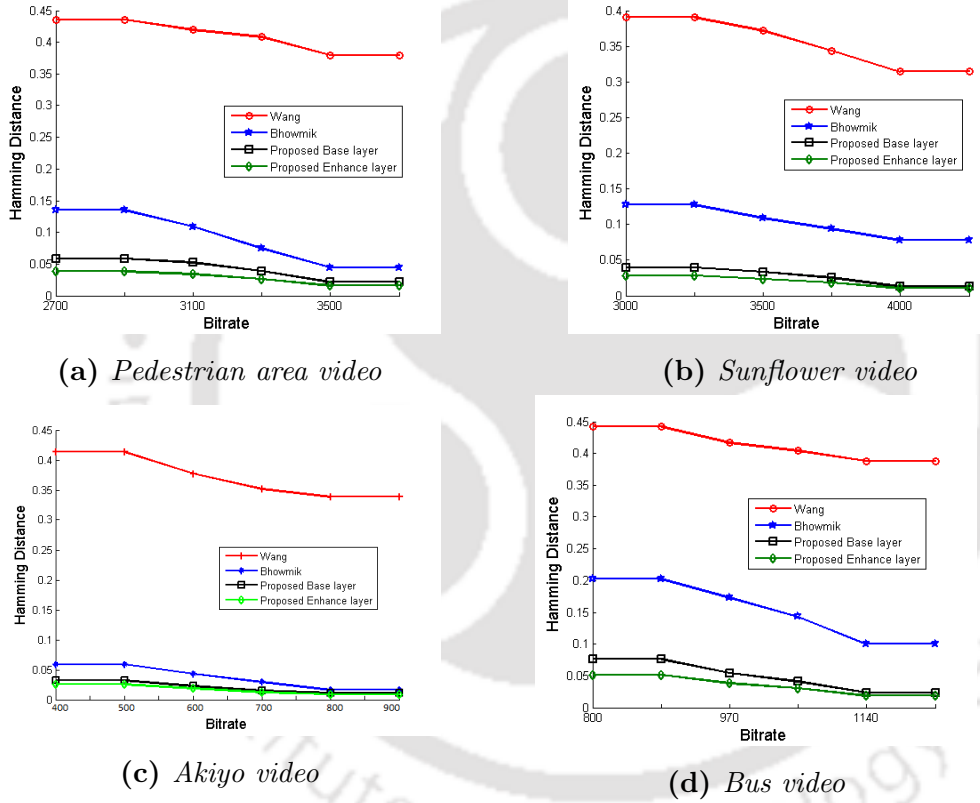
Hamming distance of Akiyo video Watermark				
Resolution	QP	Bit rate $kbit/s$	Base layer	Enhancement layer
352×288	26	752	0.01056	0.0091
	28	690	0.01489	0.01249
	30	566	0.02298	0.01874
	32	443	0.03156	0.02523
176×288	26	345	0.01659	0.01386
	28	238	0.02392	0.01958
	30	197	0.02792	0.02264
	32	84	0.03784	0.03009
352×144	26	433	0.01523	0.01286
	28	336	0.02352	0.01935
	30	295	0.02824	0.02301
	32	201	0.03679	0.02970
176×144	26	135	0.02153	0.01778
	28	90	0.03295	0.02637
	30	72	0.03825	0.03027
	32	28	0.05295	0.04118
272×192	26	222	0.01954	0.01622
	28	538	0.02652	0.02199
	30	445	0.02955	0.02437
	32	263	0.04895	0.03776

Table 3.4: *Hamming distance of the extracted watermark from scalable HD Pedestrian area video*

Hamming distance of Pedestrian area video Watermark				
Resolution	QP	Bit rate $kbit/s$	Base layer	Enhancement layer
1280×720	26	6621	0.02123	0.01583
	28	5624	0.0389	0.02627
	30	4800	0.05192	0.03404
	32	4214	0.05823	0.03837
640×720	26	3971	0.03156	0.02273
	28	957	0.05055	0.03463
	30	827	0.06486	0.04299
	32	514	0.07684	0.04993
1280×360	26	4166	0.03042	0.02201
	28	3313	0.04795	0.03134
	30	2782	0.05806	0.03667
	32	2120	0.07013	0.04291
640×360	26	2725	0.04296	0.02992
	28	1962	0.0564	0.03763
	30	246	0.06605	0.04273
	32	156	0.08265	0.05136
320×144	26	827	0.07958	0.04744
	28	600	0.127	0.07049
	30	480	0.15055	0.08214
	32	285	0.18754	0.10209

Table 3.5: *Hamming distance of the extracted watermark from scalable HD Sunflower video*

Hamming distance of Sunflower video Watermark				
Resolution	QP	Bit rate $kbit/s$	Base layer	Enhancement layer
1280×720	26	5622	0.013	0.01017
	28	4600	0.02534	0.01825
	30	3986	0.03306	0.02333
	32	3343	0.0397	0.02813
640×720	26	3008	0.0162	0.01256
	28	2200	0.02729	0.01977
	30	1500	0.03773	0.02639
	32	964	0.04675	0.03209
1280×360	26	3174	0.01454	0.01172
	28	2994	0.01751	0.01376
	30	2717	0.02206	0.01676
	32	1274	0.04256	0.02984
640×360	26	2063	0.02136	0.01647
	28	1390	0.0379	0.02651
	30	1109	0.04528	0.03104
	32	545	0.06185	0.03987
320×144	26	552	0.03846	0.02544
	28	362	0.0539	0.03584
	30	283	0.06085	0.04013
	32	125	0.07445	0.04859



(a) *Pedestrian area video*

(b) *Sunflower video*

(c) *Akiyo video*

(d) *Bus video*

Figure 3.9: *Robustness comparison*

Chapter 4

Robust Video watermarking against Temporal and Quality Scalability

4.1 Introduction

In this chapter, a scalable watermarking scheme is proposed where temporal and quality adaptations have been considered. The main issue of the temporal adaptation is to handle the reduction of the frame rate (frame dropping). The proposed work can also resist the temporal desynchronization attacks where random frame dropping or frame averaging is done intentionally. There are some non-hostile situations where the video frames can be dropped for example network congestion, buffer overload at end using devices etc. All of these cases essentially causes some sort of temporal desynchronization and can be handled by the proposed work. It is observed that the number of frames which have been dropped in case of scalable adaptation is very high in comparison with general frame dropping attacks. But the pattern of frame dropping in temporal adaptation is generally known a priory which is in general random in nature for the frame dropping attacks. Frame by frame watermarking (inserting watermark in each of the video frames) may be a naive solution to this problem but it has some serious disadvantages. Firstly, frame by frame watermarking is in general vulnerable against collusion attacks type I and II [17, 40] where simple frame averaging can be used to estimate the watermark. Moreover, simple frame by frame watermarking may

cause flickering artifacts [39, 17] due to presence of inter frame motion if proper motion compensation is not taken care off at the time of embedding.

There exists few watermarking schemes in the literature which have addressed the problem of temporal desynchronization. Chong et al. [67] proposed a RST invariant watermarking scheme which is also resilient to the frame dropping attack. In this scheme [67], authors have claimed that the average DC energy of a frame is RST invariant thus the DC energy histogram can be used to embed the watermark. Although this approach performs well against random frame dropping as well as temporal scalability, embedding capacity of the scheme is very low. In another scheme [68], authors have proposed a blind video watermarking method against frame dropping and frame averaging attack based on 3D-DWT transform. They showed that temporal high frequency coefficients are orthogonal to normally distributed watermark with zero mean and used high frequency band for embedding. It is experimentally observed that the scheme [68] fails if relatively large number of frames (more than 30%) have been dropped. In overall study , it has been observed that relatively less attention has been paid to the scalable watermarking to resist temporal adaptation until recently. Most of the existing schemes are not performing well against temporal scaling if frame rate adaptation (number of frames to be dropped) is relatively high.

In this chapter, a semi-blind watermarking scheme is proposed against temporal and quality scaling attacks where watermark is embedded in the motion compensated low pass frames. It is semi-blind because although original video is not required for the extraction but a location map describing watermark embedding locations is used. Due to embedding in low pass frames, watermark gets distributed among all the frames. Proposed scheme is tested over a large set of standard videos and the result shows that it performs well even at very high frame dropping rate. The scheme is described in subsequent sections.

4.2 Proposed Scheme

One of the main challenges in scalable video watermarking is to choose appropriate locations for embedding. In the proposed scheme, LL1 (Fig.[4.1]) subband of the each frame after 2-level of wavelet decomposition is chosen for embedding to

make it robust against quality scalability.

After spatial decomposition, motion compensated temporal decomposition is done on the LL1 subbands of all frames and low pass version of the LL1 sub-bands are used for embedding. Due to embedding in low pass frames of the LL1 sub-bands, watermark information gets spread over all the frames which can be extracted even after the frame dropping attack. Moreover, the embedded watermark in low pass frames spreads over the motion coherent locations due to motion compensated temporal filtering which reduces the flicker artifacts [39, 60]. Morover. embedding in the motion coherent regions helps to resist the collusion attacks [40]. In this work, DCT based motion compensated temporal filtering [56] is used. Temporal filtering is done on GOF of size 9 frames which can be generalized for any number of frames. At first, the temporal filtering is done on sub-GOF of 3 frames. Then 2^{nd} level of decomposition is done on 3 low pass frames generated from 3 consecutive sub-GOF's as shown in Fig. 4.2. During motion compensation among three frames, we get three different set of pixels having different motion characteristic. Those three categories are connected, unconnected and partially connected pixels (refer to chapter 2). These different pixels are needed to be handled differently. Only connected pixels are used for embedding. Unconnected and partially connected pixels are stored and used to get the watermarked video. Concept of MCDCT-TF is explained in details in Sec. 2.1.1 and how pixels are categorized into connected-unconnected sets is described in Sec. 2.1.3.

In the proposed scheme, watermark is extracted from each of the frames to make it robust against frame dropping. But during inverse motion compensation, watermark coefficients may change their locations in motion direction. To resist this desynchronization, in this work, embedding locations are saved during embedding in a map (say LocationMap). Generation of location map and whole embedding process is described in subsequent sub-sections.

4.2.1 Location Map

Because of temporal scalability, all embedded frames may not be available at the receiver side. So it may not be possible to get the same GOF structure during watermark extraction. But the additive watermark which is embedded in low

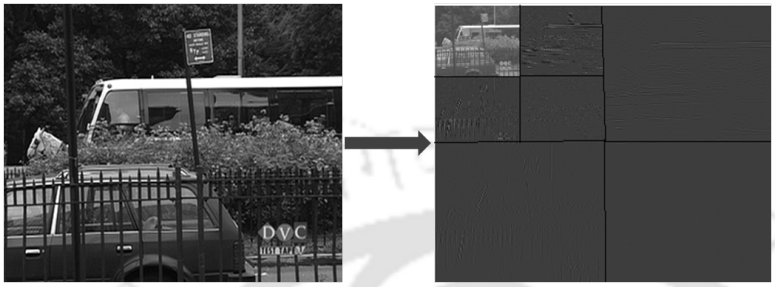


Figure 4.1: *Spatial Decomposition*

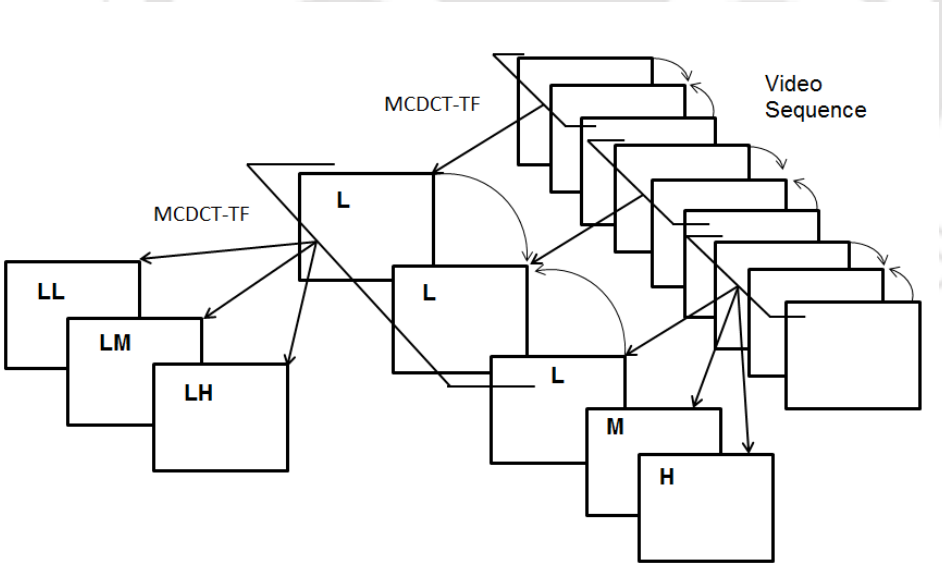


Figure 4.2: *DCT based Motion Compensated Temporal Filtering (MCDCT-TF)*

pass frames will be distributed in all the frames during the reconstruction of watermarked video. The non-zero coefficients obtained from 2^{nd} level temporal decomposition, satisfying a given selection criteria, are used for embedding. For every low pass frames, a location map is generated to store the embedding locations as shown in Fig 4.4. In Fig 4.4, I_{3t+1}, I_{3t+2} and I_{3t+3} are 3 consecutive original frames. Arrows on frame I_{3t+1} and I_{3t+3} are motion directions of 4×4 non-overlapping blocks. To predict the frame I_{3t+2} from I_{3t+1} and I_{3t+3}, motion compensation is done where I^1_{3t+2} and I^3_{3t+2} are the predicted frames. From the predicted frames, fully connected pixel locations are generated. All connected coefficients are passed through coefficient selection procedure. In the Fig.4.4, coefficients which are marked with non zero numbers in the location map are the embeddable coefficient. Zeros in gray area (in location map of Fig. 4.4) represents the coefficients which are rejected by the coefficient selection procedure. To capture the embedding locations in the upper layer frames, *Location maps* are also subjected to the (inverse) motion compensation using Eqn. 4.1-4.2. The procedure is shown in Fig 4.3.

$$Lmap^{l+1}_{3t+1}[m + H^{1->2}, n + V^{1->2}] = Lmap^{l}_{3t}[m, n] \tag{4.1}$$

$$Lmap^{l+1}_{3t+3}[m + H^{3->2}, n + V^{3->2}] = Lmap^{l}_{3t}[m, n] \tag{4.2}$$

where $Lmap^l_{3t}$ is the location map of $3t^{th}$ frame at l^{th} temporal layer. $(H^{1->2}, V^{1->2})$ $(H^{3->2}, V^{3->2})$ are MV's of I_{3t+2} with respect to I_{3t+1}, I_{3t+3} respectively. Size of the location map is directly proportional to size of the video frame. If frame size of the video is $M \times N$, then size of the location map for that frame will be $\frac{M}{2} \times \frac{N}{2}$.

4.2.2 Embedding Scheme

In this sub-section, proposed embedding scheme is described. For watermark embedding, a blind scheme has been employed. The watermark embedding rule is as follows:

$$\left.\begin{array}{ll} A < B & if\ wb = 0 \\ A > B & if\ wb = 1 \end{array}\right\} \tag{4.3}$$

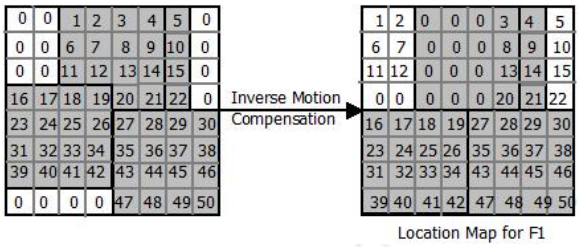


Figure 4.3: *Inverse motion compensation of the Location Map*

where wb is watermark bit and A, B are two embeddable coefficient. Embedding rule is implemented in Algorithm 4.1. Overall embedding scheme is depicted in Fig.[4.5]. A step by step embedding scheme is given in Algorithm 4.2.

Algorithm 4.1: $Embed_{bit}(A, B, wb, \delta)$

Input: A *and* B: Two consecutive coefficients, and δ: Watermark Strength, wb: Watermark bit

Output: A' *and* B':Watermarked Coefficient

```
begin
    if wb==1 then
        if A − B < δ then
            A′ = A + (δ − (A − B))/2;
            B′ = B − (δ − (A − B))/2;
    else
        if B-A< δ then
            A′ = A − (δ − (A − B))/2;
            B′ = B + (δ − (A − B))/2;
```

4.2.3 Coefficient Selection

After watermark embedding in the low pass frames, IMCDCT-TF (Sec.2.1.2) is applied to get the watermarked video. Let $L(m, n)$ and $L(m, n+1)$ are two consecutive coefficients in low pass frames which are used for embedding and $WL(m, n)$ and $WL(m, n+1)$ are corresponding watermarked coefficient. Frames I^1_{3t+2}, I_{3t+2}

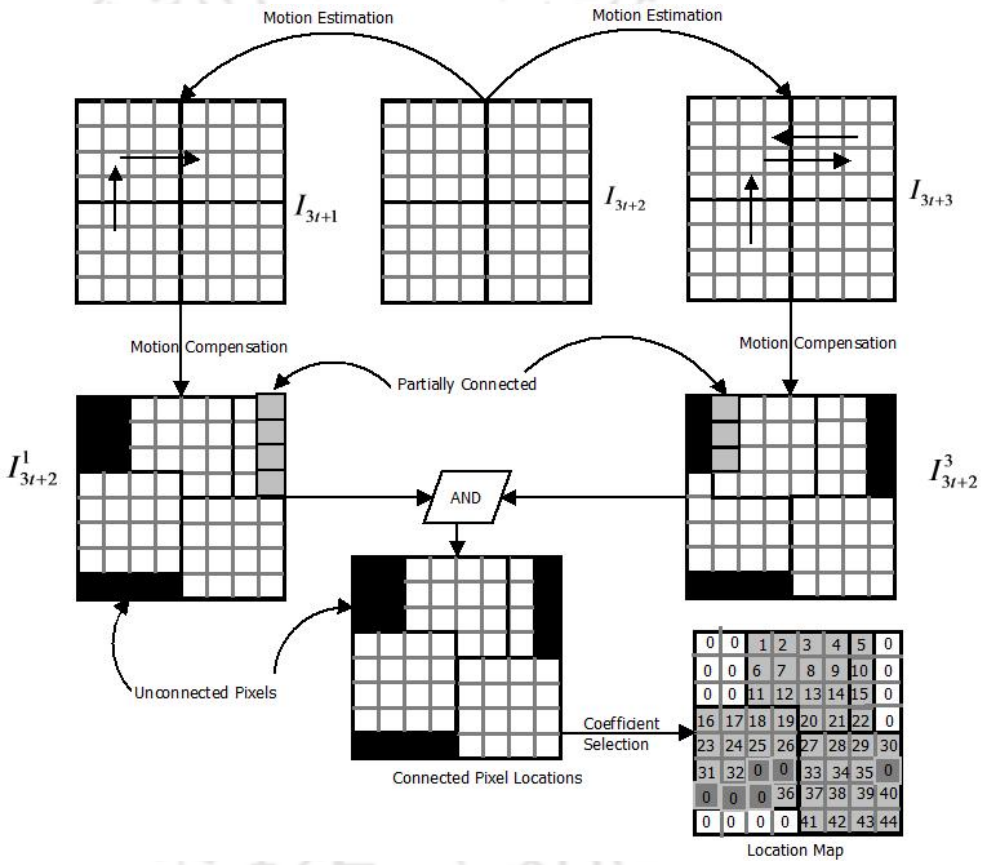


Figure 4.4: *Pixel categories and Location Map*

and I^3_{3t+2} in Fig. 4.4 are reconstructed using Eqn. 2.4. Let corresponding watermarked frames are WI^1_{3t+2}, WI_{3t+2} and WI^3_{3t+2} respectively. From Eqn. 2.4, value of $WI^1_{3t+2}(m,n)$ can be written as Eqn. 4.4.

$$WI^1_{3t+2}(m,n) = \frac{1}{\sqrt{3}}WL(m,n) + \frac{1}{\sqrt{2}}M(m,n) + \frac{1}{\sqrt{6}}H(m,n) \tag{4.4}$$

If the embedded watermark bit is 1 and locations are $WI^1_{3t+2}(m,n), WI^1_{3t+2}(m,n+1)$ for correct extraction

$$WI^1_{3t+2}(m,n) > WI^1_{3t+2}(m,n+1)$$

$$or,\ \frac{1}{\sqrt{3}}WL(m,n)+\frac{1}{\sqrt{2}}M(m,n)+\frac{1}{\sqrt{6}}H(m,n) > \frac{1}{\sqrt{3}}WL(m,n+1)+\frac{1}{\sqrt{2}}M(m,n+1)+\frac{1}{\sqrt{6}}H(m,n+1)$$

$$or,\ \frac{1}{\sqrt{3}}(WL(m,n)-WL(m,n+1)) > \frac{1}{\sqrt{2}}(M(m,n+1)-M(m,n))+\frac{1}{\sqrt{6}}(H(m,n+1)-H(m,n))$$

$$or,\ \sqrt{\frac{3}{2}}(M(m,n+1) - M(m,n)) + \frac{1}{\sqrt{2}}(H(m,n+1) - H(m,n)) < \delta \tag{4.5}$$

Similarly condition for WI^3_{3t+2} will be

$$\sqrt{\frac{3}{2}}(M(m,n) - M(m,n+1)) + \frac{1}{\sqrt{2}}(H(m,n+1) - H(m,n)) < \delta \tag{4.6}$$

and for WI_{3t+2}

$$\sqrt{2}(H(m,n) - H(m,n+1)) < \delta \tag{4.7}$$

So for blind extraction from every temporal layer, it is required to extract watermark from every frame. Now, let a watermark bit ($wb = 1$) is embedded at $L(m,n)$ and $L(m,n+1)$. To extract it from $WI^1_{3t+2}(m,n)$ and $WI^1_{3t+2}(m,n+1)$ (see Fig. 4.6), inequality given in the Eqn.4.5 must be satisfied.

Two coefficients of low pass frames are selected for embedding only if corresponding coefficients at middle frequency and high frequency frames satisfies inequalities given in Eqn(s).4.5, 4.6 and 4.7.

Algorithm 4.2: Embedding Algorithm (V, α, W)

Input V:Raw Video, and δ: Watermark Strength, W: Watermark Image

Output V^w:Watermarked Video, Lmap : Location Map

1. Divide the Raw video sequence (V) into group of (GOF) 9 non overlapping frames. Take one such GOF of 9 frames.

2. Each frame of a GOF is subjected to 2-level Haar Wavelet decomposition (Fig.4.1). Let LL_v is the low frequency subband sequence after 2-level of Haar Wavelet decomposition. $LL_v = \{LL1, LL2....LL9\}$

3. Divide the LL_v sequence into non-overlapping sub-group of 3 low frequency sub-band sequence.

4. Let such a sub-group of 3 low frequency subband named LL1, LL2 and LL3. Find motion vector from LL1 to LL2 ($MV_{1->2}$) and LL3 to LL2 ($MV_{3->2}$).

5. Determine the motion compensated version of LL1 as LL_{mc1} and LL3 as LL_{mc3}. Now calculate pixel by pixel temporal DCT of LL_{mc1}, LL2 and LL_{mc3}.

6. After 1D-DCT over LL_{mc1}, LL2 and LL_{mc3}, we get DCT coefficient frames as $L_{DCT(1)}$, $M_{DCT(1)}$ and $H_{DCT(1)}$ (low, medium and high frequency respectively)

$$[L_{DCT(1)}(i,j), M_{DCT(1)}(i,j), H_{DCT(1)}(i,j)] = 1D\ DCT\,[LL_{mc1}(i,j), LL2(i,j), LL_{mc3}(i,j)]$$

   for $i, j = 1$ to $\frac{M \times N}{4}$ where size of the frames are $M \times N$

7. Take $L_{DCT(1)}$, $L_{DCT(2)}$ and $L_{DCT(3)}$ from 3 sub-groups and apply step 4 to 6 to get LL_{DCT}, LM_{DCT} and LH_{DCT}.

8. Every two consecutive coefficient form LL_{DCT} is chosen for embedding if passes the selection criteria explained in Sec 4.2.3 and visual quality threshold discussed in Sec 4.2.4

9. Embed watermark in selected coefficients of LL_{DCT} sequence using the function $Embed_{bit}()$. Store the embedding location in Lmap

10. Do 2-level inverse temporal DCT and 2-level inverse spatial wavelet to get the watermarked video. Inverse motion compensation is done on Lmap and is stored for extraction.

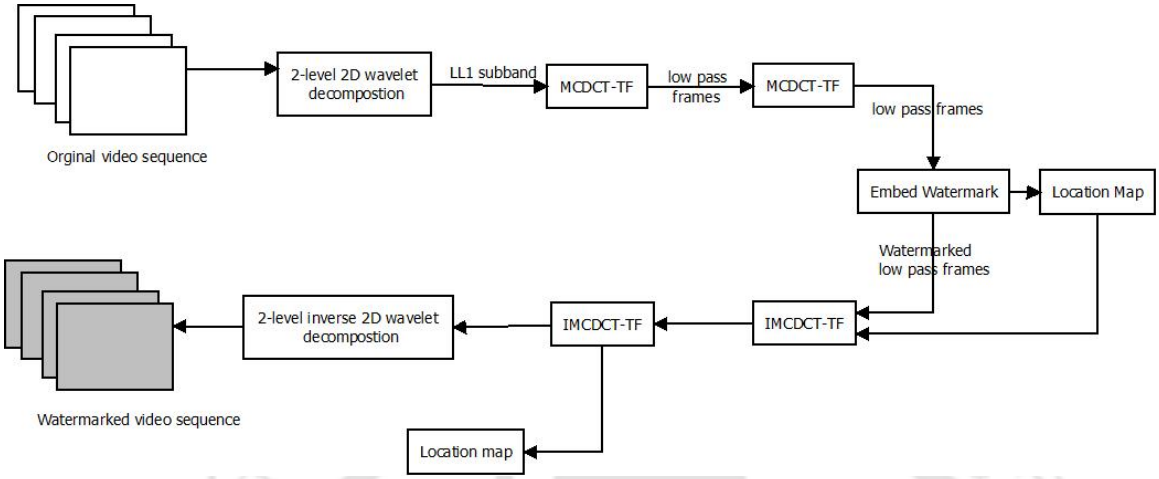


Figure 4.5: *Watermark Embedding Model*

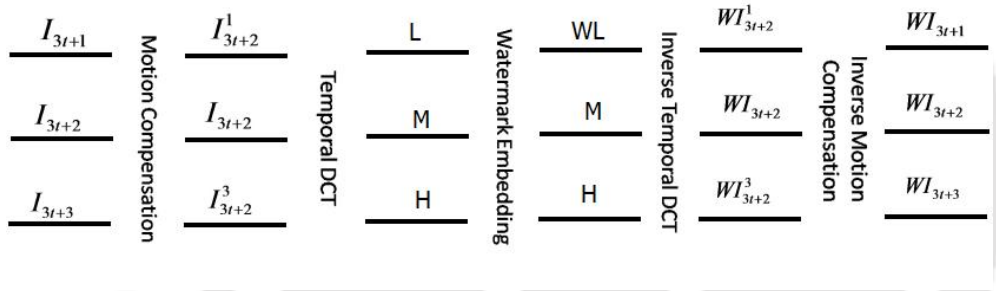


Figure 4.6: *Frames after every step*

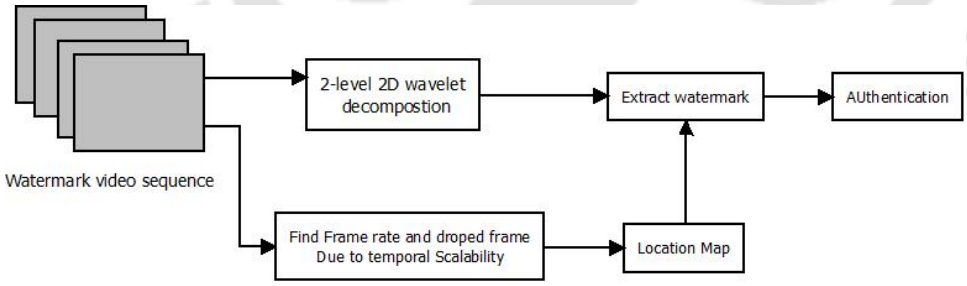


Figure 4.7: *Watermark Extraction Model*

4.2.4 Visual Quality Threshold

If difference between two consecutive coefficients (A,B) in Algorithm 4.1 is much less than δ (watermark strength) then distortion due to embedding will be high. To decrease this embedding distortion, further restriction is imposed on the se-

lection of coefficients. A coefficient pair is selected for embedding if it satisfies the visual quality threshold (V_{th}) in the Condition1. Value of V_{th} is dependent on the payload.

$$\left.\begin{array}{ll} A - B > V_{th} & when\ wb = 1 \\ B - A > V_{th} & when\ wb = 0 \end{array}\right\} \textbf{Condition1}$$

4.2.5 Extraction Scheme

The watermark bit is extracted by the Eqn. 4.8.

$$\left.\begin{array}{ll} \mathrm{W'_i} = 0 & if\ A' < B' \\ W'_i = 1 & if\ A' > B' \end{array}\right\} \tag{4.8}$$

where A' and B' are embedded coefficients. The extraction procedure is shown in the Fig.4.7. A step by step extraction procedure is given in Algorithm 4.3.

Algorithm 4.3: Extraction Algorithm (V^w)

1 **Data** V^w:Watermarked Video
2 **Result** W: Watermark bit stream

1. Each frame of the watermarked sequence is subjected to 2-level Haar Wavelet decomposition (refer to Fig.4.1). Let WLL_v are the low frequency subband sequence after 2-level of Haar Wavelet decomposition.
2. From the frame rate of the video, the number of frames dropped due to temporal scaling (if there is any) has been found.
3. For each LL subband, the corresponding Location map has been determined.
4. Using location map embedded coefficients are selected and watermark bit extracted using Eqn. 4.8.

4.3 Experimental Results

Proposed Scheme is evaluated on a set of standard video sequences with different motion characteristic and different size. In this paper, result of 4 video sequences

Table 4.1: *Experimental Setup*

Parameters for Watermarking	Values Taken
Encoder	H.264/SVC version
Reference Software	JSVM
Video Sequence Used	Bus, City, Crew, Coastguard, Mobile, Akiyo, Hall, Mother Daughter, Sunflower , Foreman, News
Video Resolution	4CIF, CIF
Watermark signal	32×32 and 64×64 binary image
Visual Quality Metrics	PSNR, flicker metric, VQM, SSIM
Robustness metric	Hamming distance

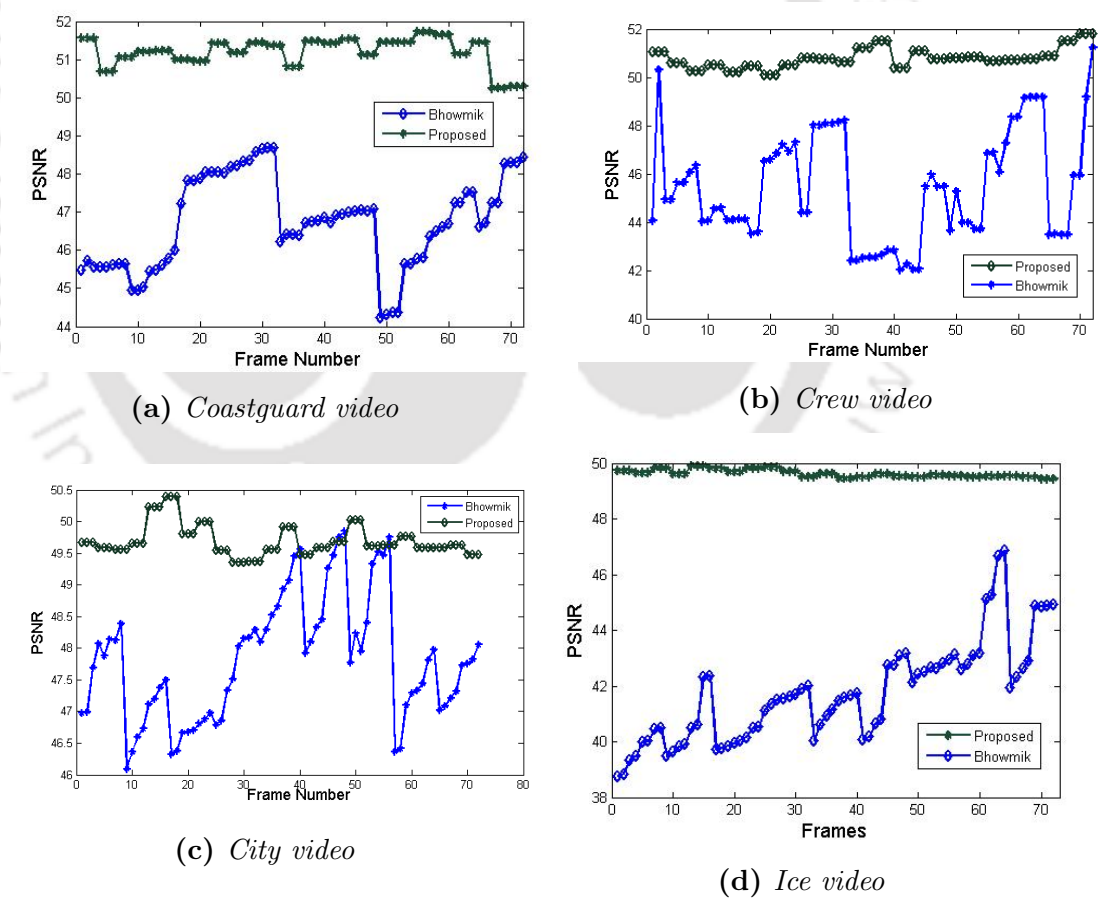
with different motion characteristics, e.g. *Crew* (with CIF size and high object motion), *Coastguard* (with CIF size having object motion as well as camera motion), *City* (with 4CIF size and only camera motion))are shown . A 32×32 and 64×64 binary logo is embedded in the luma component of the CIF and 4CIF sized video sequences respectively. All the watermarked video has been encoded using JSVM (Joint Scalable Video Model) reference software (H.264/SVC) and extracted in different quality (bit-rate) and temporal (frame rate) levels to evaluate the performance of the scheme. Experimental setup is tabulated in Table 4.1.

4.3.1 Visual Quality

In this sub-section, visual quality of the proposed scheme is compared with scheme proposed by Bhowmik et al. [39]. PSNR between cover video and watermarked video is compared with Bhowmik et al.'s scheme [39] in Fig. 4.8 for 4 video sequences. It is observed from the Fig.(s) that the visual quality (with respect to PSNR) of the proposed scheme (green line) is better than that of Bhowmik's scheme [39] for all the video sequences. Intuitively use of visual quality threshold (V_{th}) in the proposed scheme helped to improve the PSNR value of the watermarked video. Although V_{th} has trade off with the payload, it chooses the best

coefficients for embedding for a given payload.

Flicker metric calculates inter frame distortion by taking three consecutive frames into consideration. In the proposed scheme, watermark is embedded in motion coherent location upto 9 frames (filter length) as a result absolute flicker difference of original video and watermarked video is close to zero. Absolute flicker difference is compared in Fig. 4.9. Fig. 4.10 depicts the SSIM comparison of proposed and Bhowmik's scheme. We can see that the mean SSIM is better or close to Bhowmik's scheme. Comparison of VQM with Bhowmik's scheme is given in Fig. 4.11. Lower VQM means better video quality. So, video quality with respect to VQM is better than existing work [39] for all the video sequences.



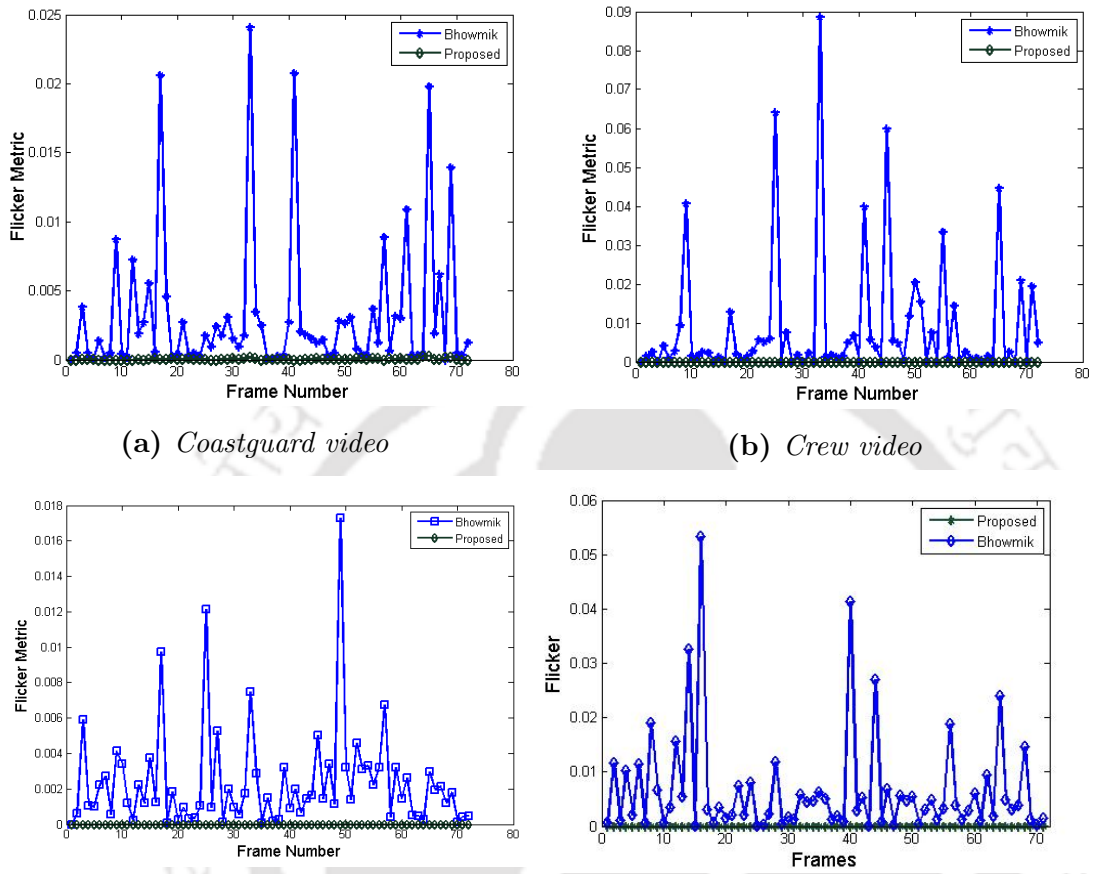
(a) *Coastguard video*

(b) *Crew video*

(c) *City video*

(d) *Ice video*

Figure 4.8: *PSNR comparison*



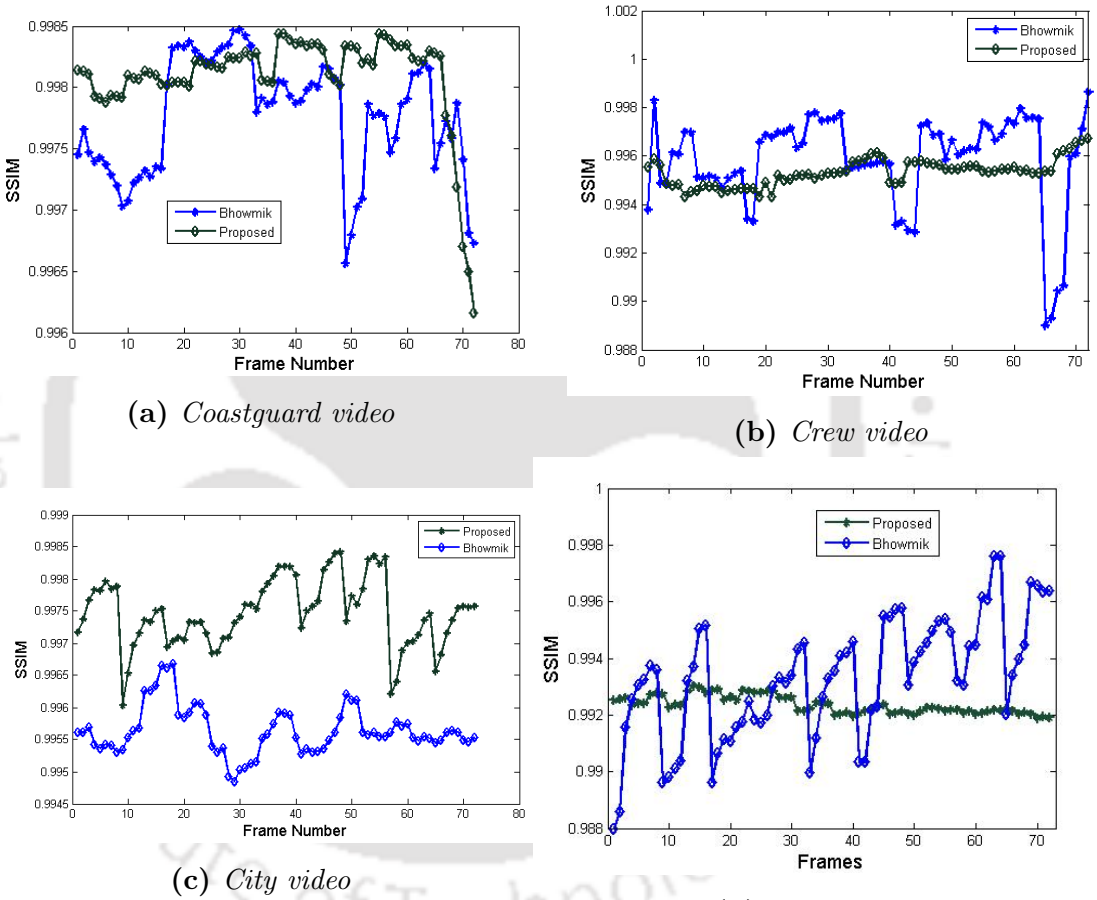
(a) *Coastguard video* **(b)** *Crew video*

(c) *City video* **(d)** *Ice video*

Figure 4.9: *Flicker comparison*

4.3.2 Robustness Comparison

Robustness of the proposed scheme is measures by Hamming distance as given in Eqn. 2.19. Robustness at different temporal layers of different videos are shown in Fig. 4.12. Here weighted average of extracted watermarks from different layers is taken. Then hamming distance from the original frame is calculated. Base layer frames are given more weight because bit-rate of base layer frames are more (less compression) than other layer frames. Figure shows progressive improvement of robustness over frame rate. Robustness of the proposed scheme is compared with scheme proposed by Bhowmik et al. [39]. Fig.4.13 shows the robustness comparison for the same set of video sequences. It is observed from the Fig.4.13



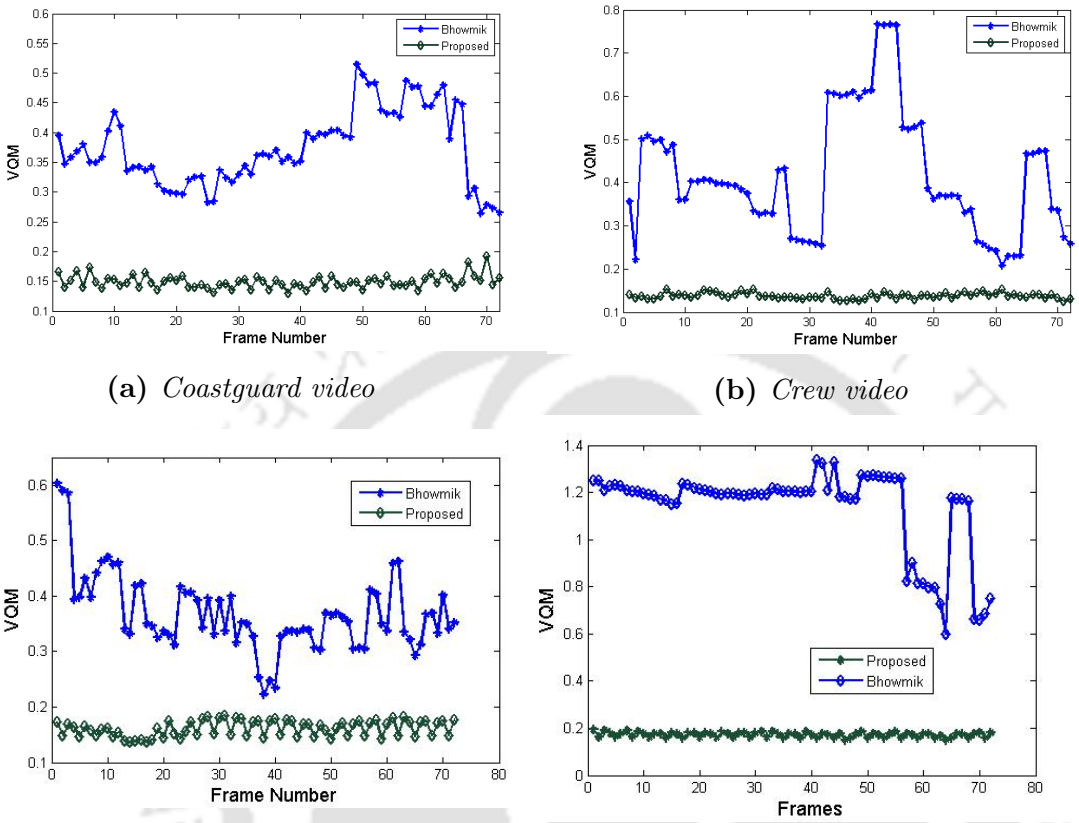
(a) *Coastguard video*

(b) *Crew video*

(c) *City video*

(d) *Ice video*

Figure 4.10: *SSIM comparison*



(a) *Coastguard video* **(b)** *Crew video*

(c) *City video* **(d)** *Ice video*

Figure 4.11: *VQM comparison*

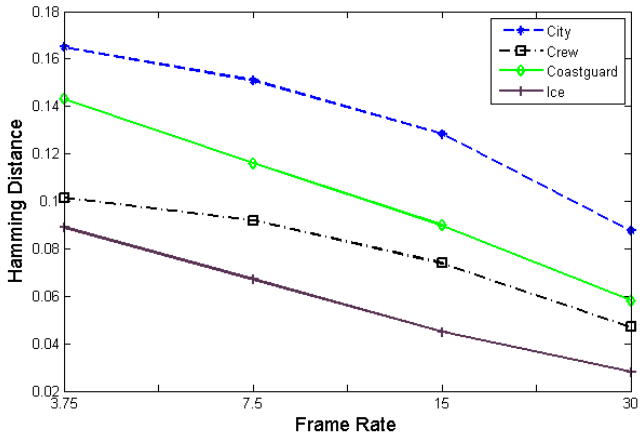
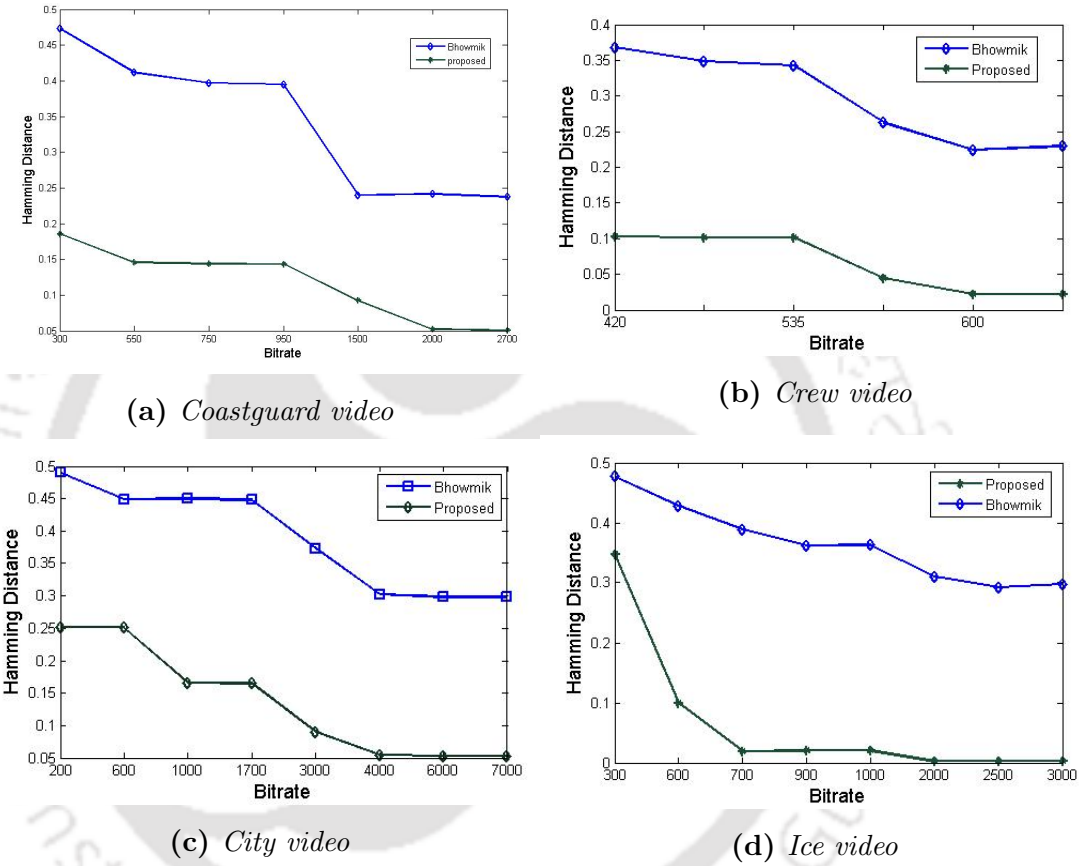


Figure 4.12: *Robustness at different Temporal layer*

that the robustness (Hamming Distance) of the proposed scheme is better than that of Bhowmik's scheme [39] and it shows the required graceful improvement in robustness as bitrate gets increased.



(a) *Coastguard video*

(b) *Crew video*

(c) *City video*

(d) *Ice video*

Figure 4.13: *Robustness comparison*

4.3.3 Explanation

In this watermarking scheme, watermark is embedded in the low pass spatio-temporal frame of the video. Here the watermark is embedded only in the connected pixels of the spatio-temporal low pass video frames, so the watermark will sustain after quality scaling (bitrate scaling), temporal scaling or frame dropping. As the watermark is embedded in temporal low pass frames, watermark information are distributed in all the frames, so even after the frame dropping due to

temporal scaling, watermark information can be extracted from the remaining frames.

4.4 Conclusion

In this chapter, a MCDCT-TF based watermarking scheme has been proposed. MCDCT-TF uses longer length filter than the existing schemes which exploits more correlation in consecutive frame's. Proposed MCDCT-TF based watermarking scheme shows better robustness and less embedding distortion than existing MCTF based watermarking scheme. The robustness is analyzed against compression in H.264/SVC coding. In this chapter, the temporal scalability and the quality scalability have been considered. In the previous chapter and this chapter, robust watermarking schemes are proposed for three different scalability like resolution, temporal and quality scalability. It has been observed that the proposed schemes are doing well if resolution or temporal scaling are relatively low. But there is still scope for improvement when amount of scaling is bit high. In the next chapter, a SIFT based approach is devised to handle the large resolution scaling.

Chapter 5

SIFT based Robust Image Watermarking against Resolution Scalability

5.1 Introduction

In chapter 3, a watermarking scheme against resolution scaling is proposed where up-sampled base layer watermark is embedded in the enhancement layers. But experimental results shows that the performance of the proposed scheme is not up to the mark when the scaling factor is relatively high. In this chapter, a scale invariant image watermarking scheme based on Scalable Invariant Feature Transform (SIFT) [30] features is proposed. The proposed image watermarking scheme can be easily extended to video by taking the motion information into consideration to avoid the flickering artifact. The SIFT [30] algorithm (refer to chapter 2) extracts the distinctive features of local image patches and is proved to be invariant to image scaling and rotation. When it comes to scale invariance, Morel et al. [69] shown that SIFT is the best feature extraction methods and outperforms all other image feature extraction methods. SIFT descriptors are robust against noise, changes in illumination and viewpoint etc. These local invariant features are highly distinctive and are matched with a high probability against large image distortions. SIFT features have been used in many applications like multi view matching [31, 32], object recognition [33], object classification [34, 35], robotics [36] etc. It is also being used for robust image watermarking against ge-

ometric attacks [24, 38, 37]. Miyaki et al. [37] proposed a RST invariant object based watermarking scheme where SIFT features are used for the object matching. In the detection scheme of [37], the object regions are first detected by feature matching. The transformation parameters are then calculated to detect the hidden message. Though the method produces quite promising results but it is a type of informed watermarking as the register file has to be shared between the sender and receiver. Kim et al. [24] inserted watermark into the circular patches generated by the SIFT. The detection ratio of the method varies from 60% to 90% depending upon the intensity of the attack. Under strong distortions due to attenuation and cropping, the additive watermarking method may fail to survive for several images. Jing et al. [38] used SIFT points to form a convex hull, which are then optimally triangulated. The watermark is then embedded into the circles centered around the centroid of each triangle. This scheme also fails to large resolution scaling.

All these scheme used SIFT feature to select a embedding zone and to synchronize watermark location during extraction. In this work, a novel watermarking scheme is proposed where SIFT feature descriptor itself is used as watermark signal. In the proposed scheme, intensity of an image patch is altered in such a way that it generates some new feature points. Descriptor of this new feature points are stored as watermark signal. The patch is chosen in such way that it creates less perceptual distortion.

5.2 Proposed Scheme

In this scheme, the invariance property of the SIFT features to the rotation, scaling and translation is exploited. The original image is modified in a context coherent way such that the perceptual meaning of the image is not changed. The said perturbation generates new set of SIFT features. The new SIFT descriptors, (the 128 bit dimensional vectors associated with this new SIFT features) act as the watermark message in the image. These new features are extracted and registered in the database. During the watermark detection, the SIFT features are extracted from the image in consideration. Then SIFT matching is done between the registered SIFT features in the database and the features which are

extracted from the attacked image. A high degree of matching (greater than a prescribed threshold), denotes the authenticity of the image. Since SIFT features are invariant to geometric distortions, it is more likely that these features would produce a match with high probability in case of image scaling. The proposed watermarking scheme is thus robust to the resolution and quality adaptation.

5.2.1 Watermark Zone Selection

The first step of the proposed scheme is to find an image region such that alteration to the region will generate least perceptual error. To do this, contrast of the image is increased (refer Step 2 of the Algorithm 5.1), so that some background objects which are interleaved in the background are detected in the next step. In Step 4, for removing small objects from binary image, an area opening is performed which removes all connected components having less number of pixels. Now connected components are decided by 4-connected neighborhood operator [70]. The areas and pixel position of all the objects are obtained.

It is observed that in the obtained areas A, the top 5% of the objects with large areas capture almost the whole of the image, leaving a collection of small objects which are insignificant to the human visual system (HVS). The distribution of the objects is such that there are maximum number of objects with less area and very few objects with appreciable area. Choosing the area generically, for applying the patch from all the objects wouldn't give good result as the distribution of the object areas varies from image to image. Unique areas of objects U (Step 6) are hence considered because their distribution will be nearly same for all the images. The selection of an object which is neither be too large to be noticeable for the human eye nor too small to give fragile SIFT points is decided by β. The value of β is empirically found to be 1.5. The details are given in Sec. 5.2.2. Now the object with area $U(\lfloor \frac{n}{\beta} \rfloor)$ is found. In very rare cases, the value $\lfloor \frac{n}{\beta} \rfloor$ may be zero where it is considered as $U(1)$. This gives the pixels where the image has to be altered.

5. SIFT BASED ROBUST IMAGE WATERMARKING AGAINST RESOLUTION SCALABILITY

Algorithm 5.1: Watermark Zone selection

Input: Image I, quality parameter β.

Output: Pixel Locations in I

1. Convert I to gray scale image.
2. Increase the contrast of I.
3. Convert the gray scale image I to a binary image using threshold computed using Otsu's method [71].
4. Remove small objects from binary image.
5. Get a list of connected components along with their pixel positions and areas in the modified binary image. Let each object in the list be denoted by $\{A(i), P(i)\}$ where $A(i)$ is the area and $P(i)$ contains pixel locations of the i^{th} object.
6. Get a list of unique areas in A sorted in the ascending order. Let the list be U and length of the list be n.
7. Let $p = \lfloor \frac{n}{\beta} \rfloor$. Obtain a connected object with the area $U(p)$ i.e. find i such that $A(i) = U(p)$.
8. Get the pixel locations of the object i which is $P(i)$

Figure 5.1: *Lena Binary Image*

5.2.2 Derivation of Quality Parameter β

Lower β gives the object having larger area in Algorithm 5.1. This results a watermark patch which is more perceptible to the HVS and is more robust. Similarly higher β gives less robustness and good visual quality. Hence, there is a trade-off between these two factors. The proposed algorithm uses an empirical value of 1.5 for quality parameter β. The value of β directly affects the visual quality and the robustness of the watermark.

In the experimental setup, the value of β is varied form 1.1 to 2.0 in 0.1 increments. The Watson metric and robustness for each scale is calculated for every alpha. The optimal value for β is the one which has maximum robustness along with minimum perceptual error. For total robustness of an image at a particular β, the arithmetic mean of robustness at each scale is considered as $\frac{1}{|s|}\sum_{s} R_s$ where R_s is the robustness at that scale s. In this experimental study, the scaling factor is taken from 0.3 to 1.2 and robustness R_s is calculated as given in equation 5.2. Now, arithmetic mean over all the images is considered. In this way, the robustness at each β is obtained. Since the perceptual error is very less and the robustness is high relative to it, both are separately normalized.

The difference between Robustness (R) and Perceptual Error (PE) is considered to determine the value of β. The plot of the graph can be seen in Fig. 5.2. It can be observed that after $\beta = 1.5$ there is very little increase in the slope of the curve. Moreover as the value of β goes above 1.5, the robustness will be very less which makes the watermarking itself ineffective. This is similar to the parametric

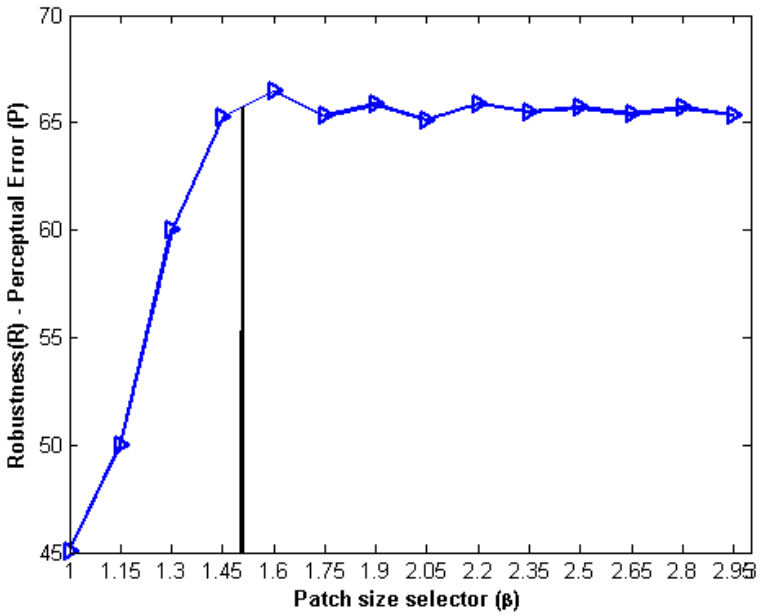


Figure 5.2: *Plot for finding best possible* β

estimation done by Lowe [30], where it is mentioned that one must settle for a solution that trades off efficiency with completeness. So even if there is a slight increase in the curve at the end, the value of β is chosen as 1.5.

5.2.3 Watermark Embedding

The entire watermark embedding process is summarized in Algorithm 5.2. Algorithm 5.1 is applied on the given image to get the pixel locations which are to be modified. In this work, the pixels obtained are changed using Eqn. 5.1 so that the new SIFT features obtained are strong and do not match with the original image. Now the SIFT features D^w which are not in the original image but are present in modified image are extracted. These set of features act as watermark message and are registered in the database. Each members of the set D^w will be a 128 bit dimensional vector representing the SIFT descriptors which are exclusively present only in watermarked image.

Algorithm 5.2: Watermark Embedding

Input: Image I

Output: Watermark Descriptors D^w

1. Extract the SIFT features of the original image I. Let the set of SIFT descriptors obtained for the original image be D.
2. Apply Algorithm 5.1 to get the pixel locations P in the image I where the patch has to be applied.
3. Modify the original image I to I' by modifying the intensities at the pixel locations non-linearly. Intensity of pixels at location P is modified as
$$I'(P) = mod(I(P)^2, 256). \tag{5.1}$$
4. Extract the SIFT features of the modified image I'. Let the set of SIFT descriptors obtained from the modified image be D'.
5. Take the difference of the two sets. $D^w = D' \setminus D$.

5.2.4 Watermark Extraction & Authentication

In the proposed scheme, a blind watermark extraction method has been employed. First SIFT features are extracted from the attacked image. Then the feature matching is done between the extracted features and registered features. For authentication, matching percentage must be greater than some predefined threshold. Step wise description is given in Algorithm 5.3.

Algorithm 5.3: Watermark Extraction & Authentication

Input: Attacked Image I'

1. Extract the SIFT features of the Image to be checked for authenticity I'. Let the set of SIFT descriptors obtained from the modified image be D'.
2. Get the registered feature descriptors D^w stored in the database.
3. Apply SIFT matching to find the features in D^w which match with D'. If the degree of matching is high (greater than a prescribed threshold), then the image is matched and is authenticated.

5.2.5 Experimental Results

In this section, a comprehensive set of experimentations have been carried out to justify the efficiency of the proposed scheme over the existing literature.

Experimental Setup

Data Set : Proposed scheme is tested on a huge dataset of approximately 82000 images. The images are collected from various computer vision standard databases such as Complex Scene Saliency Dataset (CSSD) and Extended Complex Scene Saliency Dataset (ECSSD) [64]. Caltech 256 dataset [65] and The LabelMe-12-50k dataset [66]. So Images used for experimentations are of different characteristics and of various categories. The collection also includes 24 standard images such as *lena, baboon, airplane* etc.

Image	baboon	barbara	boat	girl	lenna	mountain	serrano	tulips	zelda
GPE ($\times 10^{-4}$)	1.55	7.89	2.82	2.54	2.89	16.7	2.72	3.14	14.2

Table 5.1: *GPE for Standard Images*

Evaluation Parameters :

Robustness : The percentage of matching is taken as the **robustness** of the descriptor i.e.

$$Robustness\ R = \frac{m}{|D^w|} \tag{5.2}$$

where D^w is the feature point descriptors in the database and m is the number of points matched in D^w when compared with D'.

Perceptual Quality: For assessing the perceptual quality of the proposed scheme, Watson metric [63] is used where the global perceptual error (GPE) of the watermarked image with respect to original image is computed. In addition RARE2012 [58] is used to evaluate the mean visual saliency of the patch in the original image to compare the efficiency of the embedding location selection process.

Visual Quality

The embedding patch is selected in such a way that the modified image will be very close to the original image in terms of its visual perception. Table 5.1 gives the statistical metrics of the GPE for 10 standard images.

Robustness against Resolution Scaling

Table 5.2 gives the median value of robustness of the watermark for all images in the database. The results for the standard images are shown in Table 5.3 for the blind watermarking scheme. The watermark is observed to be highly robust against resolution scaling. This is due to the selection of highly stable SIFT points.

Scale	1.2	0.9	0.7	0.5	0.3
Robustness	89.91	86.72	71.87	69.36	50.59

Table 5.2: *Average Robustness for all the images in the dataset*

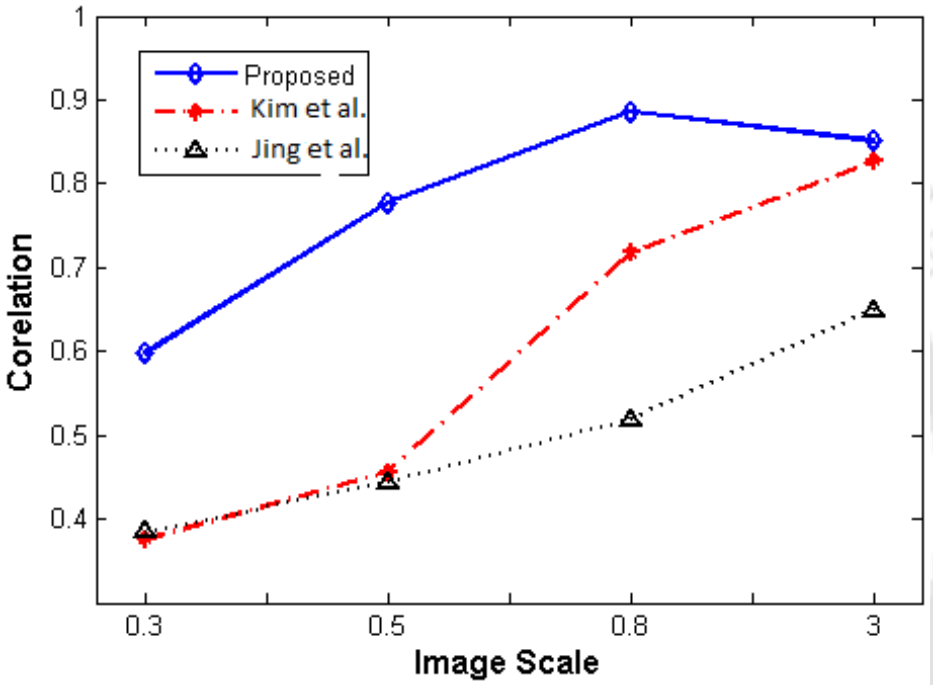


Figure 5.3: *Robustness Comparison with scheme [24] (red) and [38](black)*

A comparison of the proposed scheme with [24] and [38] with respect to the robustness against the resolution scaling is depicted in Fig. 5.3. It is observed that the proposed scheme outperforms the existing schemes for the standard images.

Image	1.2	0.9	0.5	0.3
baboon	91.67	83.33	76.67	53.33
barbara	85.71	82.22	80.00	60.00
boat	83.33	81.13	69.81	47.17
girl	90.00	86.67	86.00	62.22
lena	88.89	85.29	73.52	52.94
mountain	89.66	85.21	70.69	55.17
serrano	96.88	90.63	90.63	68.75
tulips	85.71	85.716	81.63	59.18
zelda	83.33	86.36	86.36	81.81

Table 5.3: *Robustness for Standard images when scaled*

5.3 Improvement over the proposed scheme

Although proposed scheme in the previous sub-section performs quite well against resolution scalability, it can be further improved. In this subsection, the improvements over the previous scheme have been discussed. In this improved version, a visual saliency based zone selection method has been employed to get better visual quality of the watermarked video. In addition, the stability (or robustness) of the SIFT features is considered in the improved version. The new SIFT features generated due to patch insertion are sorted with respect to their stability and 10 most robust (stable) features are stored as the watermark.

5.3.1 Strength of Individual SIFT Feature

SIFT is extensively used in computer vision, object matching and image retrieval. The SIFT algorithm generates a large number of feature points. Matching strength or robustness of all feature points are not same. In [72], a method is described to measure the stability of a feature point. In the scheme [72], first a large set of SIFT descriptor is clustered into 4096 clusters, then each cluster is assigned a quality metric. Quality metric is measured by the false match rate where less false match rate signifies more stability. To get the false match rate

of a cluster, features from a large image set is calculated and the mapping between clusters and features have been identified. Then, each image from the image set is passed through a set of transformation (scaling rotation etc.) and the false match is calculated for each image and for each transformation. The ratio of total number of false match from a cluster and total number of feature in that cluster is determined as false match rate of that cluster. In the proposed scheme, newly generated features due to change of intensity of a patch are stored as the watermark. Features which belongs to cluster with less false match rate are chosen from the new feature set.

5.3.2 Modified Watermark Zone Selection

Step by step procedure of modified zone selection is described in Algorithm 5.4. The image zone is determined using visual attention model named RARE2012 [58]. RARE2012 [58], which assigns saliency based on multi-scale spatial rarity is described in Sec.2.3. First saliency map of the input image is calculated. Then the saliency map is converted into binary values where lowest 15% saliency values are changed to white(1). Then all the connected components of the binary saliency map are determined.

Algorithm 5.4: Watermark Zone selection

Input: Image I.
Output: Pixel Locations in I

1. $R = \text{RARE2012}(I)$
2. Convert saliency map into binary by changing lowest 15% values to one and rest zero.
3. Get a list of connected components in this binary map.
4. Return the all connected components and their locations

5.3.3 Watermark Embedding

The modified watermark embedding process is summarized in Algorithm 5.5. Algorithm 5.4 is applied on the input image to find the candidate pixel locations for embedding. The object with the lowest saliency values in the map is chosen for embedding. In this scheme, the pixel's intensity perturbed by δ. The value of δ is varied from -255 to 255 and is chosen such that it generates a large number of new features with relatively less visual degradation. If the number of features thus generated is less than some threshold (F^{th}) then next object is chosen. After required number of new features are generated, they are sent to selection procedure described in Sec. 5.3.1. In this scheme, 10 features are stored as watermark having lowest false match rate. Let D and D' be set of descriptors generated from the original image and watermarked image respectively. In the previous scheme, new feature set is generated by taking set difference of D and D'. Due to change in intensity of the patch, lot of original descriptor will change to a new descriptor with very less difference/distance between them. Those features will be stored as watermark. But when matching with the original descriptor set, those feature will match with high probability because in SIFT matching algorithm (as described in Sec. 2.2), two matched features does not require to have zero distance (exact match). As a result in the previous scheme, there is a high percentage of matching with the stored watermark and the original feature set (false positive). The matching ratio of watermark descriptor (D^w) with the original descriptor (D) in previous scheme is plotted in Fig. 5.4.

In this modified embedding scheme, new feature set is calculated (in step 7 of algorithm 5.5) by taking set difference of D' and D^m, where D^m is sub set of D' which are matched with the original descriptor set D.

5.3.4 Watermark Extraction

Watermark extraction is same as the previous scheme as described in Sec. 5.2.4.

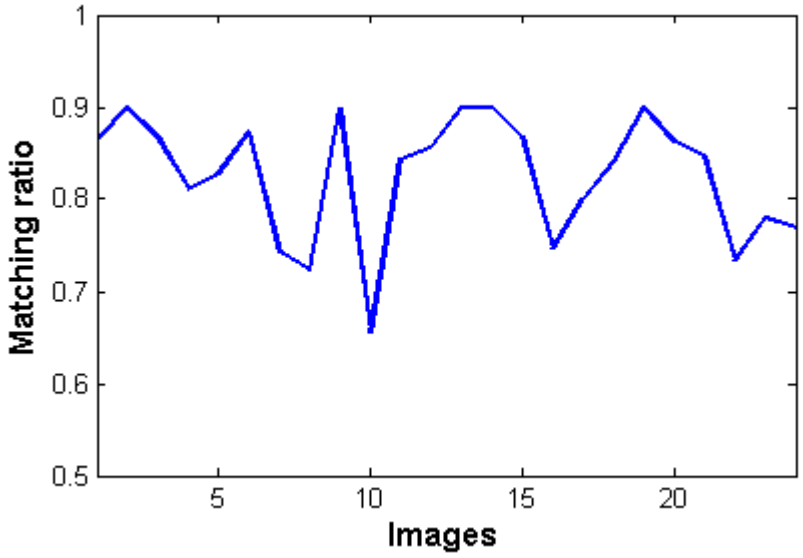


Figure 5.4: *Matching ratio of watermark descriptor with original descriptor in previous scheme*

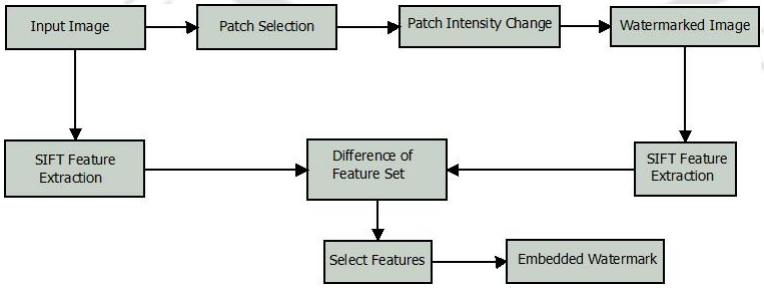


Figure 5.5: *Watermark Embedding Scheme*

Algorithm 5.5: Watermark Embedding

Input: Image I

Output: Watermark Descriptors D

1. Extract the SIFT features of the original image I. Let the set of SIFT descriptors obtained for the original image be D.
2. Apply Algorithm 5.4 to get the connected components P with lowest saliency.
3. Select smallest component say P_i
4. Modify the original image I to I' by modifying the intensities at the pixel locations P_i i.e. $I'(P_i) = I(P_i) + \delta$ where δ is constant.
5. If number of new features is less than F^{th} then select next small connected component and go to step 4
6. Extract the SIFT features of the modified image I'. Let the set of SIFT descriptors obtained from the modified image be D'.
7. Find SIFT matching between D and D'. Let D^m is the set of feature in D' matched with D. Then new feature set $D = D' \setminus D^m$.
8. Find out quality of each new SIFT feature as described in Sec. 5.3.1.
9. Sort descriptor set D in ascending order according to their quality.
10. Store the first 10 descriptor from sorted D as watermark.

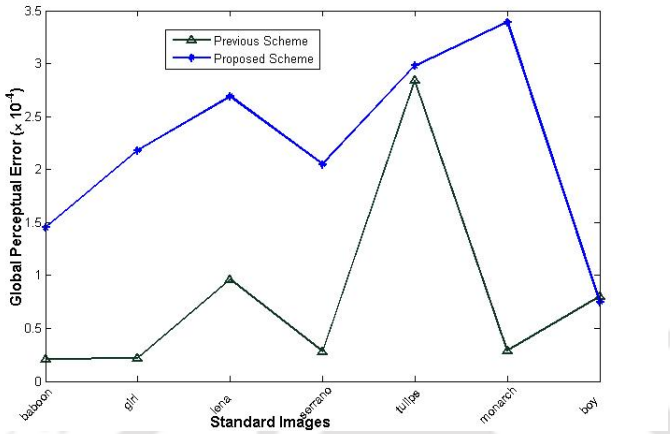


Figure 5.6: *Visual Degradation Comparison between proposed scheme and existing schemes for Lena*

5.3.5 Experimental Result

Modified scheme is evaluated with the same experimental setup as previous scheme (5.2.5).

Comparison with previous scheme

In a previous work [73], a SIFT based image watermarking scheme is proposed where watermark was embedded in an object of a chosen size. All the newly generated features are stored in the register. The proposed algorithm results less visual degradation than our previous scheme [73] as the changes are made in the least salient image regions in this work. Comparison between two schemes against GPE for 7 standard images is shown in Fig. 5.6. There is also a significant improvement in robustness over the previous scheme. This improved result may be caused due to the selection of the stable SIFT features for watermarking. Comparison of robustness of *Boy* and *Serano* image is shown in Fig(s). 5.7 and 5.8 respectively.

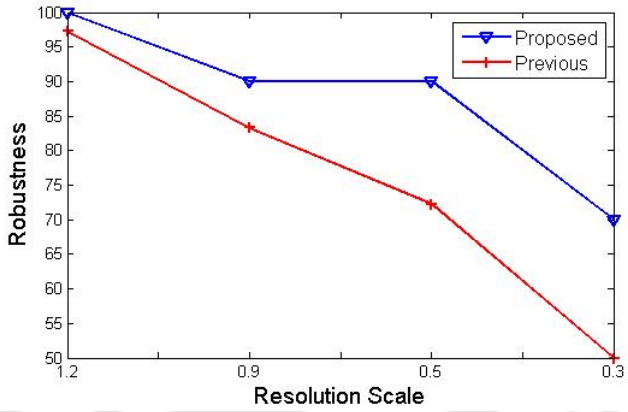


Figure 5.7: *Robustness Comparison between proposed scheme and existing schemes for Boy*

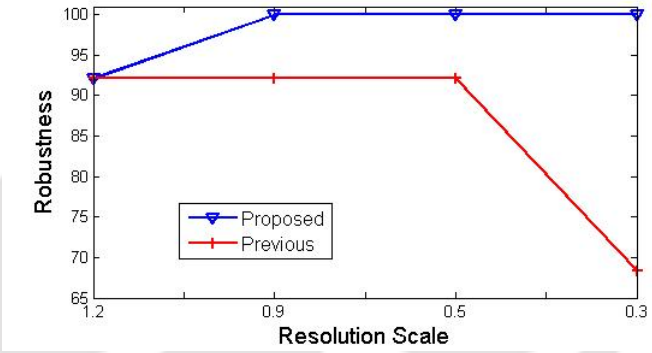


Figure 5.8: *Robustness Comparison between proposed scheme and existing schemes for Serano*

Visual Quality

Changes made to the images in the proposed scheme are very less perceptible to the HVS because the changes are made in the least salient region of the image. The watermarked images and the SIFT features generated for *baboon, barbara* and *lena* are shown in Fig. 5.9. It can be observed from the Fig. 7 that there are no perceptible visual artifacts in the watermarked images due to embedding. The GPE [63] for 10 standard images along with the saliency values using RARE2012 [58] for the corresponding image locations are tabulated in Table 5.4. As least salient regions are chosen for embedding, saliency values for selected image loca-

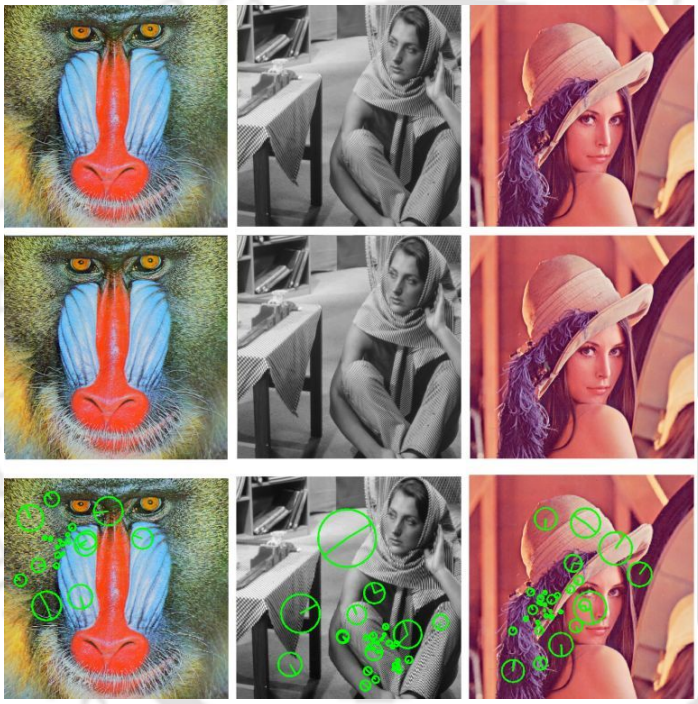


Figure 5.9: *Embedding of watermark for baboon, barbara and lena. Top row shows the original images. Middle row shows the watermarked images with patch. Bottom row shows the newly generated SIFT points due to insertion of patch.*

tion are zero for the proposed scheme.

Image	**GPE** ($\times 10^{-4}$)	**Saliency Values**
baboon	0.21	0.0
girl	0.22	0.0
lena	0.96	0.0
serrano	0.28	0.0
tulips	2.84	0.0
monarch	0.29	0.0
boy	0.80	0.0

Table 5.4: *GPE and Saliency for Standard Images*

Robustness against Resolution Scaling

Table 5.5 tabulates the median value of robustness of the watermark for sample images from the databases. The results for the standard images are shown in Table 5.6. From the Table, the proposed watermarking scheme is observed to be highly robust against resolution scaling. Intuitively, this is due to the selection of highly stable SIFT points.

Relation of change in candidate pixel (P) intensity with feature stability and quality of the image

In this section, how stability of the new feature points and quality of the watermarked image varies with the change in intensity of the candidate pixels (P) is tested. The intensity of the patch (P) is gradually varied and the changes in the

Scale	1.2	0.9	0.7	0.5	0.3
Robustness	76.34	76.39	71.02	67.08	52.73

Table 5.5: *Median Robustness for all the images*

Image	1.2	0.9	0.5	0.3
baboon	50.0	83.33	83.33	83.33
girl	85.71	71.42	71.42	42.85
lena	100.0	83.33	83.33	83.33
monarch	100.0	100.0	50.0	25.00
boy	70.0	70.0	70.0	60.00
serrano	60.00	90.00	60.00	30.00
tulips	76.47	64.71	58.82	23.53

Table 5.6: *Robustness for Standard images when scaled*

stability of new feature points are observed. It can be seen from the plots of the Fig. 5.10 that the stability of the SIFT features increase with increasing change in patch intensity. Since the images which are used for experimentation are 8 bit color images, plot stabilizes when intensity changed to 0 or 255 (intensity can not be decreased/increased). The point at which this occurs clearly depends on the initial image intensity. It has been observed that the value of the maximum stability depends on the image. This is because the number of newly generated SIFT points and their stability varies depending on the context of the patch.

Similar experiments are done to observe the change in image quality (perceptual error) due to change in intensity. The intensity of the patch selected is gradually varied and the changes in the perceptual error, when compared with original image is observed. Plots of the results of the experiments are shown in Fig. 5.11. It is seen that the perceptual error magnitude is very less. Overall the images, it has been observed that the perceptual error never crosses 3.235×10^{-3}. This is due to the significantly small size of the patch when compared to the image size. Similar to the stability plot, perceptual error also stabilizes at certain point when changed intensity reaches to maximum (255) or minimum (0) possible value.

These observations are reconfirmed by observing the change in the stability with the perceptual error. The resulting graphs are shown in Fig. 5.12. As expected, with the increase in the stability, greater perceptual error results. Hence, there is a trade-off between the two factors of stability and visual perceptual

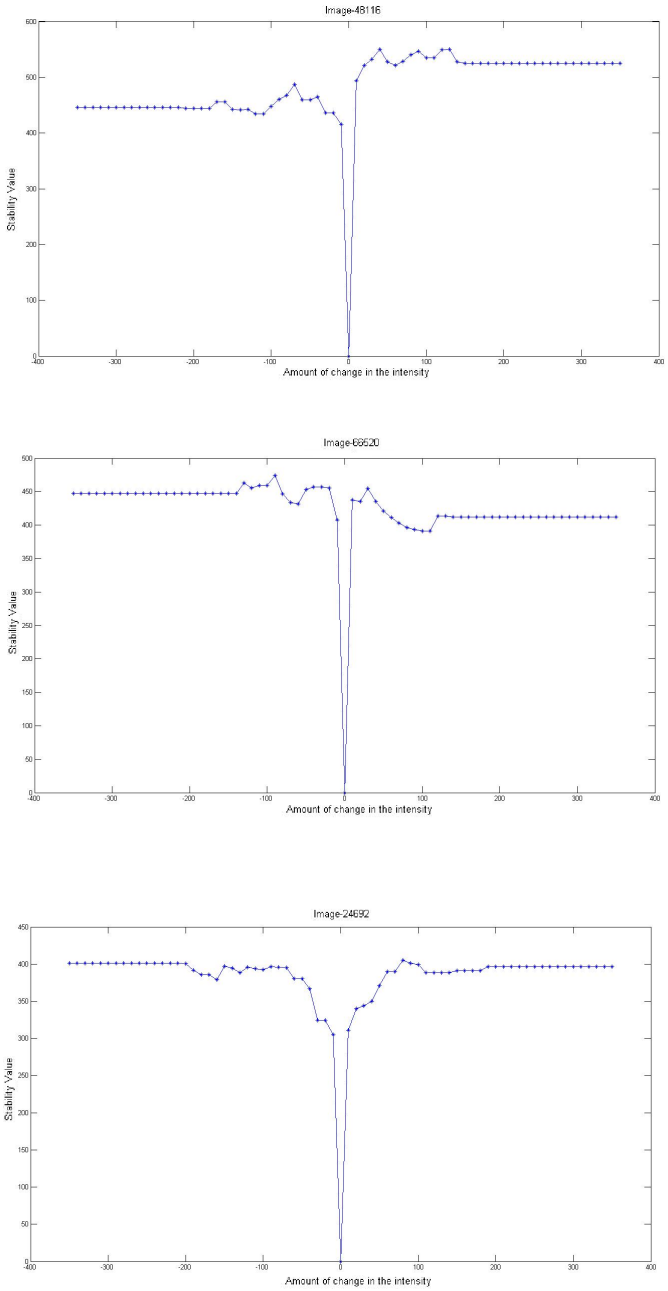


Figure 5.10: *Plot depicting variation of intensity with stability*

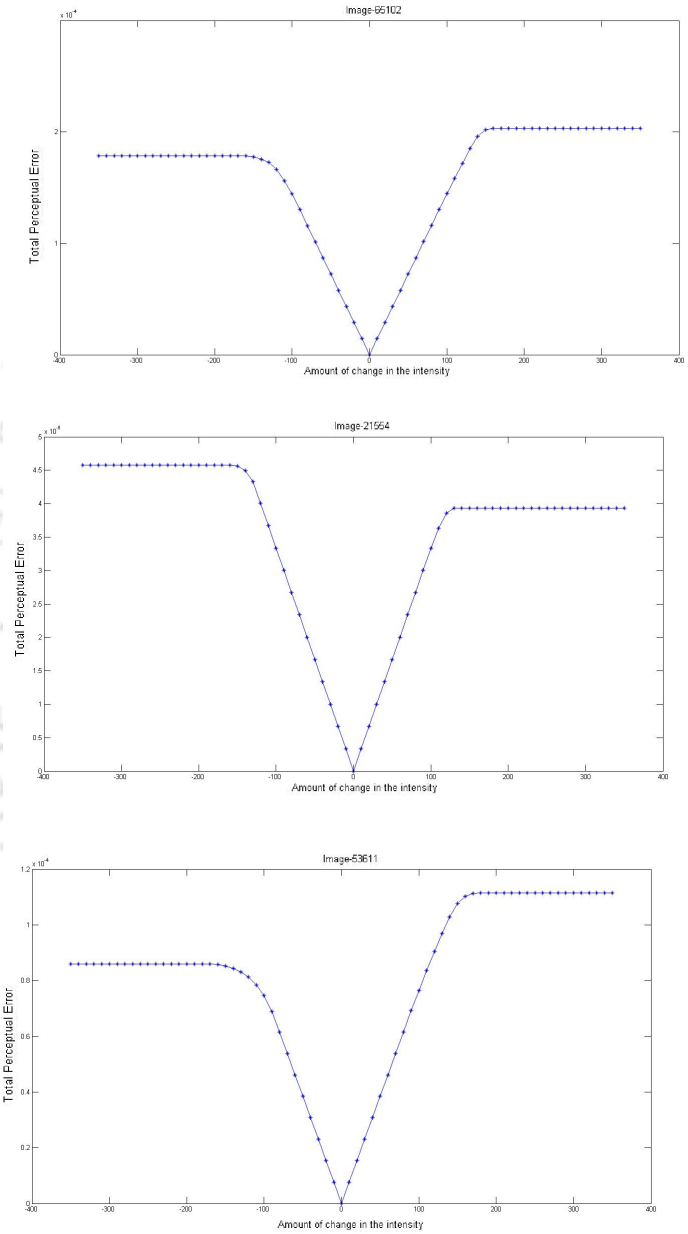


Figure 5.11: *Plot depicting variation of intensity with perceptual error*

error. Another trade-off between Robustness and Visual Quality is explained in the next section.

Robustness and Visual Quality Trade-off

In this section, we observe the trade off, by modifying Step 4 of Algorithm. 5.1 used for zone selection. Throughout the scheme, the watermark is obtained by complementing the intensity values of the original patch. Since the user has only the watermarked image and not the original image, it is difficult for the user to predict the location of the patch. In spite of that, a patch can be inserted by just increasing the pixel intensity by some value. In our experimentation, the value is taken as 20 based on experimental evidence.

It is observed that the amount of alteration in the pixel value within the selected patch due to injection of watermark has a direct consequence to the perceptual quality and robustness of the watermarking scheme. Increasing the intensity value by small amount instead of complementing the image clearly leads to decrease in the perceptual error rate. Since SIFT extracts distinctive features of local image patches, increasing the intensity by a small amount would generate less points when compared to complementing the same pixel values. Hence, there will be a decrease in robustness. This has been experimentally verified and the comparative graph of the visual quality metric and robustness for the two cases is shown in Fig. 5.14 and Fig. 5.13.

5.4 Conclusion

In this chapter, a content-based image watermarking scheme is proposed where an image is modified by inserting a context coherent patch. Newly generated SIFT descriptor are stored as the watermark. The proposed scheme can be used on each video frame to develop a video watermarking scheme resilient to resolution scaling. The scheme belongs to the class of second generation watermarking schemes which uses SIFT descriptor points. Experimental results showed that proposed watermarking scheme is highly robust to the scaling attacks. In the second part of this chapter, the proposed scheme is improved by incorporating

saliency based embedding zone selection and selecting stable SIFT features for watermarking. Also, various results regarding the stability of the feature points are described. In the next chapter, the SIFT features are used to develop a video watermarking scheme against the temporal scalability.

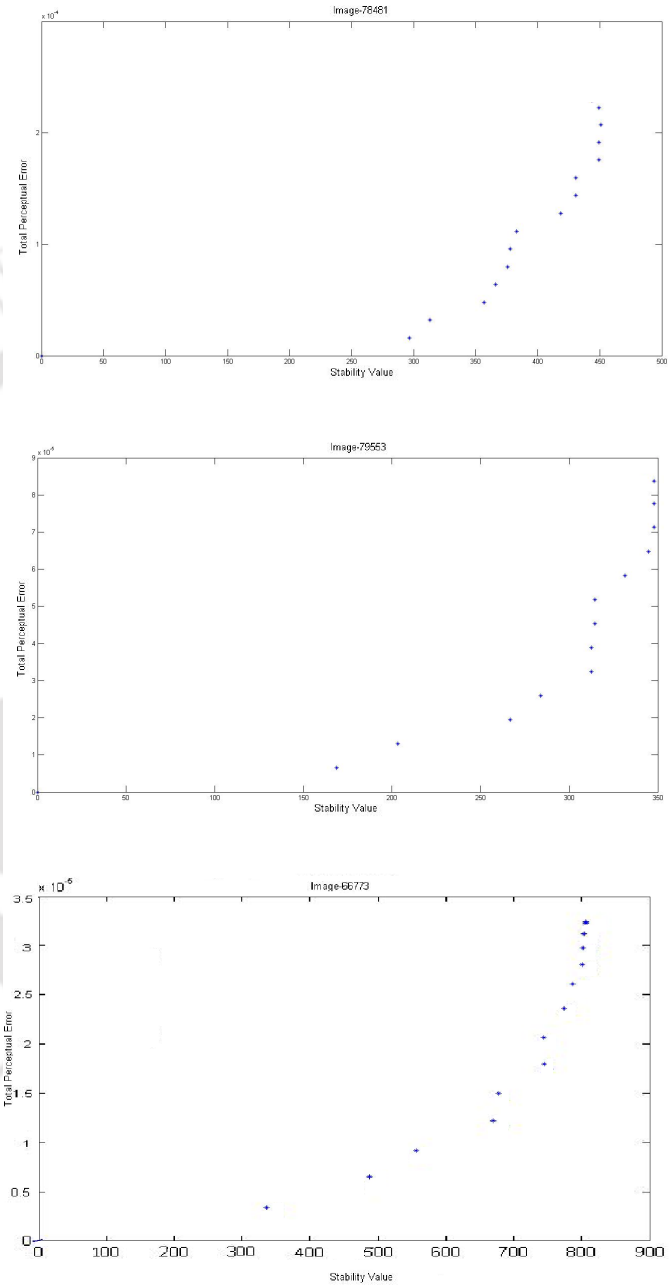


Figure 5.12: *Plot depicting variation of stability with perceptual error*

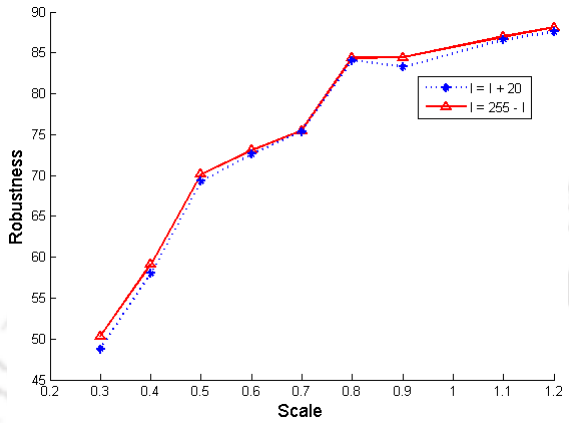


Figure 5.13: *Variation of robustness with change in intensity of the patch by 20*

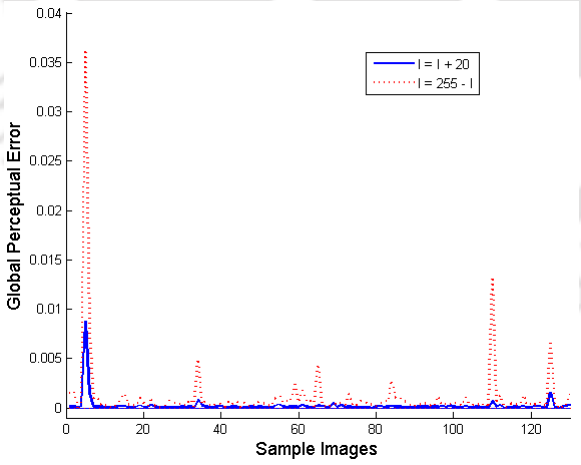


Figure 5.14: *Variation of GPE with change in intensity of the patch by 20*

Chapter 6

Robust Video Watermarking against Temporal Scalability

As it has been observed in chapter 3, temporal de-synchronization is a potential threat to the watermarking system and it can be employed by simply dropping some random frames. It was also observed that temporal scaling or frame rate scaling is a common de-synchronization attack. In chapter 3, a semi-blind watermarking algorithm has been proposed to resist the attack. In that scheme, a location map is used to find the watermark locations. Size of the location map is directly proportional to the size of the video. Storing or communicating the location map (of size close to the size of original video frame) for every video frame is an extra overhead and may not be always feasible. In this chapter, two blind video watermarking schemes are proposed which can resist temporal de-synchronization attacks (eg. frame dropping) without the help of any location map. In the first method, a SIFT (Scale Invariant Feature Transform) based watermarking scheme is proposed to resist temporal scaling where SIFT features of side views of the video are used for watermarking. In the second scheme, different watermarking signals are embedded in different hierarchical layers of the H.264/SVC encoding to ensure the graceful improvement in successive enhancement layers. These two schemes are described in details in the following sub-sections.

(a) *Side Plane*

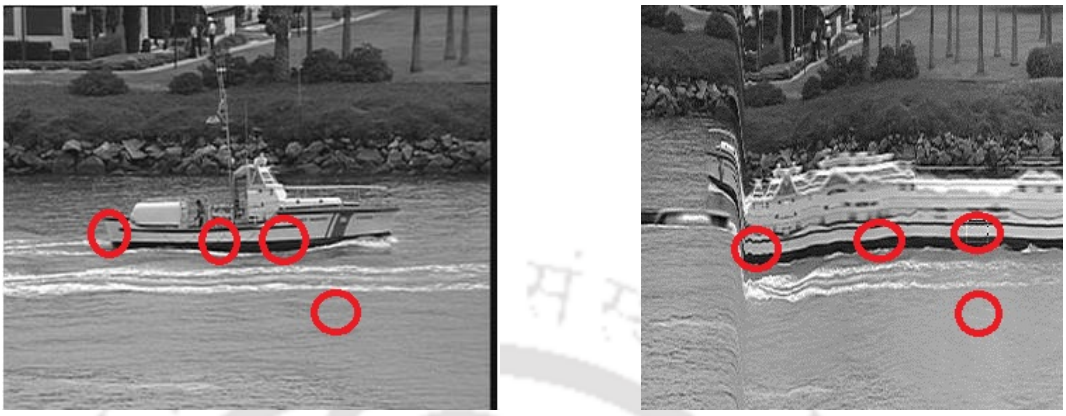
(b) *Watermarked region and corresponding Side Planes*

Figure 6.1: *Side Plane and Embedding zone*

6.1 SIFT based Video Watermarking Resilient to Temporal Scalability

In this work, a watermarking scheme has been proposed which is robust against the temporal adaptation for scalable video coding. The SIFT has been used in the proposed scheme to make it robust against frame rate adaptation due to temporal scaling. The frame dropping attack can also be resisted by the proposed scheme. In this method, the video sequence (of a given no. of frames) is modeled as a 3D signal. It essentially forms a three dimensional cuboid having width equivalent to no. of frames in the given sequence. If the dimension of the video frame is $m \times n$ and k no. of such frames are considered then a cuboid of dimension $m \times n \times k$ has been formed where it's height is m, width is n and it has a depth of k pixels. Now, if a side face of this cuboid is imagined as an image (having dimension $m \times k$), without loss of generality, it can be said that there exists n such images. So the new form of cuboid is having its height as m, width as k and depth as n. The scenario has been depicted in Fig. 6.1.

Intuitively, the side face image of above defined cuboid depicts the motion characteristics of the given video sequence. In the proposed scheme, a watermark patch (let the size of the patch is $u \times v$) is inserted in one of the side face images. The patch is embedded in such a way that it generates strong SIFT features. These newly generated SIFT features themselves work as a watermark. The frame rate adaptation or frame dropping attacks may be considered as width scaling of such images. Now, if such width scaling is known a-priory, correspond-



(a) *Cover Video Frame* **(b)** *Side View*

Figure 6.2: *Embedding Zones*

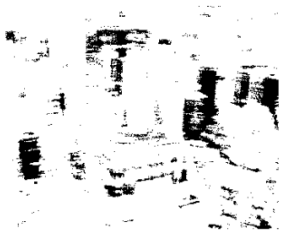
ing height scaling can also be possible to get a resized side view image. Since the SIFT features are invariant to scaling, these features can be extracted from any image resolution i.e. from the video after frame rate adaptation or frame dropping attacks. The different steps of proposed scheme such as watermarking zone selection, watermark embedding and extraction are narrated in successive subsections.

6.1.1 Watermarking Zone Selection

In the previous subsection, it is said that a patch is inserted in the side face image such that it generates strong SIFT features. In this work, although the SIFT feature has been calculated over the side face images of the video cuboid, the watermarking zones have been selected for embedding with respect to the spatial and temporal characteristics of the normal video frames (in the rest of the paper, this actual video orientation is called main view). Intuitively, the video zones with low motion and high texture are suitable for embedding as it helps to reduce the flickering artifacts and usually masks the spatial additive noise due to watermarking. The idea of the embedding zone selection of the proposed scheme has been depicted in Fig. 6.2, where the suitable embedding zones on the main view are located in Fig. 6.2a and the corresponding side view image regions are located in Fig. 6.2b.

(a) *Original Frame*

(b) *Motion Map*



(c) *Binary Motion Map after Thresholding*

Figure 6.3: *Motion Map*

Block Selection in the Main View

In the proposed scheme, the watermarking zone selection with respect to the main view is done firstly. Low motion zones in the main view are selected for embedding as embedding in high motion zones may cause flickering artifacts [60]. To find the low motion zones for a video sequence, a motion map has been generated. Difference of consecutive frames is directly proportion to the motion of the video. That is if there is high motion then difference will be high, unless there is scene change. Using this characteristics, motion map is devised as the summation of absolute difference of consecutive frames in a video sequence (having n frames) as shown in Eqn. 6.1.

$$MMap(i,j) = \sum_{k=1}^{n} |frame_{k+1}(i,j) - frame_k(i,j)| \tag{6.1}$$

where (i, j) is the pixel (spatial) index of the frame and k is the frame (temporal) index of a video sequence having n no. of frames.

The main view of the *city* video and its corresponding motion map are depicted in Fig. 6.3a and Fig. 6.3b respectively. A motion threshold has been employed to select the 10% of the frame pixels having lowest motion according to the motion map. A binary motion map after thresholding is depicted in Fig. 6.3c where black pixels are representing low motion pixels in the map. After getting the motion map, the frame (main view) is divided into non overlapping blocks of size $w \times h$ and a block DCT is done to each block. The energy of the block

is measured by taking the sum of the squared AC coefficients of the transformed block as given in Eqn. 6.2.

$$E_{BL} = \left(\sum_{x=0}^{x=h-1} \sum_{y=0}^{y=w-1} [C(x,y)]^2 \right) - [C(0,0)]^2 \tag{6.2}$$

where $C(0,0)$ is the DC coefficient, $C(x,y)$ is DCT coefficient at location (x,y), h and w is height and width of the block. An energy threshold has been used to select the higher energy blocks for embedding. The blocks are sorted in non increasing order separately with respect to the energy of the blocks and the higher number of black pixels in the motion map. Now blocks having highest energy (priority one) with maximum black pixels in the motion map (priority two) has been used for embedding. The number of blocks thus selected for embedding may be decided according to the desired payload which depends on the applications.

Block Selection in the Side View

Once the blocks with respect to the main view image are selected, the embedding suitability of these blocks with respect to side view image has been analyzed. The blocks which are selected with respect to the main view image can easily be identified in the side view image. The selected block locations with respect main view image and their corresponding locations in the side view image are depicted in Fig. 6.2a and Fig. 6.2b respectively. Corresponding blocks in side view image will be at different depth. In Fig. 6.2b, it is shown in single side view image.

In side view image, a subset of the selected blocks have been chosen such a way that it generates strong SIFT features. In this work, the strength of the SIFT features are quantified by the number of new SIFT features which are generated when a patch is inserted in a selected block. In other words, blocks / regions which generates more new SIFT features due to patch insertion are more suitable for watermarking. An experiment has been done where number of newly generated SIFT features due to patch insertion are compared between low frequency (smooth background) area and high frequency area (busy areas). The results are plotted in the Fig. 6.4. It can be observed from the Fig. 6.4 that the number of newly generated SIFT features are usually more for the low frequency (smoother) areas than that of high frequency (busy) areas.

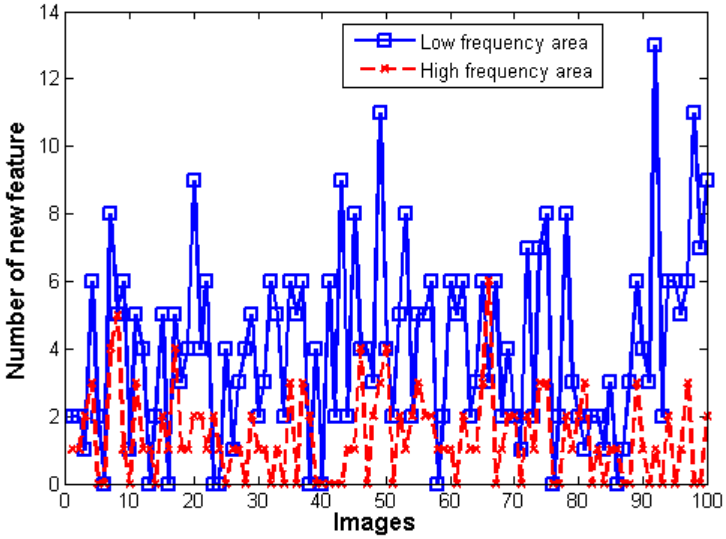


Figure 6.4: *New SIFT features for Smooth Area and Busy Area*

According to the above observation, blocks resided in the relatively smother areas are chosen for embedding from the previously selected blocks. In this work, the blocks residing in relatively smoother area in the side view image are determined by measuring the average energy for the surrounding area of that block. To calculate the average energy of the surrounding area of a block, a bigger block (of size $U \times V$) is imagined covering the block of size $u \times v$. Thus, there are $U/u \times V/v$ no.s of blocks of size $u \times v$ within a single bigger block. To decide the relative smoothness among the bigger blocks, the mean intensity values of each smaller block pixels have been calculated and $U/u \times V/v$ block DCT has been employed on these mean intensity values. The smoothness of a bigger block is quantified by the squared sum of the AC coefficients obtained from the block DCT. The bigger blocks (of size $U \times V$) are sorted in non-decreasing order of squared sum of AC coefficients (higher the squared sum of AC coefficients, lower the smoothness) and initial blocks are taken for embedding where the number of blocks per frame depends on the prescribed payloads. The selection procedure for the blocks located in relatively smoother regions are depicted in the Fig. 6.5. In Fig. 6.5, it is shown that a block (in green) has been selected for embedding as it resides in relatively smoother region within a side view. Over all zone selection

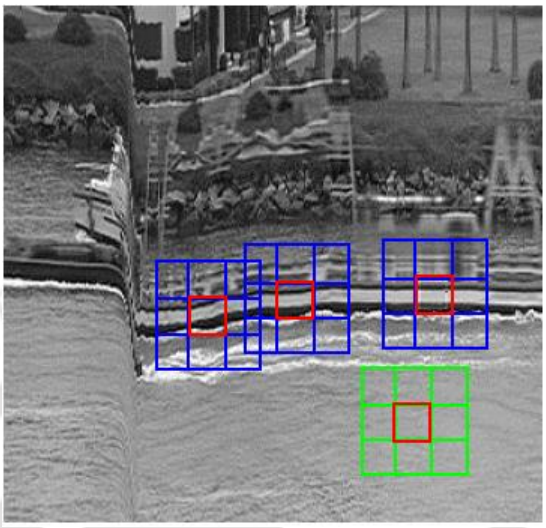


Figure 6.5: *Selection of blocks belonging to relatively smoother region*

algorithm is shown in Fig. 6.6.

6.1.2 Watermark Embedding

In the proposed scheme, watermark embedding is done using a patch insertion (altering the pixel intensities of the selected region) in the selected regions as described in the previous subsection (refer to Sec. 6.1.1). Essentially, the new SIFT features generated due to patch insertion are treated as the watermark signal.

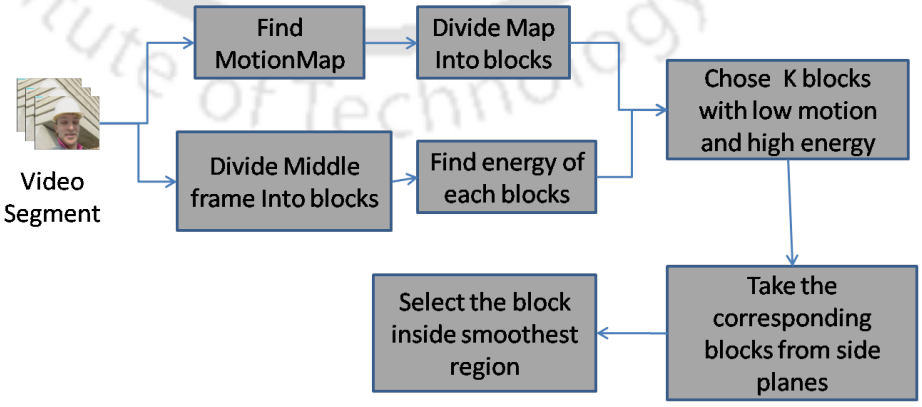
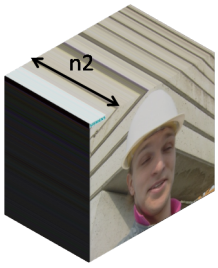
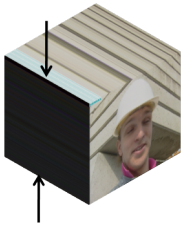


Figure 6.6: *Zone Selection Procedure*



(a) *Temporally adapted video*



(b) *The required resizing for watermark extraction*

Figure 6.7: *Temporally adapted video and corresponding resizing for extraction*

In this watermark embedding process, pixels intensities of the single side plane of the selected cuboid region are altered to generate new SIFT features. The image plane is chosen randomly from multiple image planes (depth of the cuboid). In our experiments, and such randomly chosen image plane is used for patch insertion as described in [73]. Whole watermarking scheme is described in the Algorithm 6.1.

6.1.3 Watermark Detection & Authentication

In the proposed scheme, one of the main goals is to detect watermark from the temporally adapted video sequence. In general, amount of temporal scaling (the number of frames are actually dropped) is known a-priory to the watermark extraction process. In this work, the height of the side view is scaled in accordance with the temporal scaling (width scaling) so that 2D SIFT can be used for watermark extraction. In case of random frame dropping attack (intentional or unintentional), number of dropped frames is very less. In that situation manual height scaling is not required. So if the frame dropped ratio is greater than a threshold (D_{th}) then only height scaling of side frames are required. This resizing process has been depicted in Fig. 6.7. The overall watermark detection and authentication method is summarized in Algorithm 6.2.

Algorithm 6.1: Watermark Embedding

Input: Video V

Output: Watermark Descriptors D^w, Watermarked frame index y

1. Watermark embedding zones have been selected as discussed in Sec. 6.1.1.
2. From the side view, there exists multiple image planes (depth of the cuboid).
3. One side image plane (let say SVI_k) has been randomly chosen for the watermarking.
4. Extract the SIFT features of the side view image (SVI_k). Let the set of SIFT descriptors obtained for this chosen side view image (SVI_k) is D.
5. Change the pixel intensities of the region for the side image plane (SVI_k) using Eqn. 6.3. Let the modified side image plane be SVI'_k

$$C' = min(C + \delta, 255) \tag{6.3}$$

where C is the original pixel value of the selected region and δ is change in intensity

6. Extract the SIFT features of the modified side view image (SVI'_k). Let the set of SIFT descriptors obtained for the modified side view image is D'.
7. Find set difference of two set $D^W = D' \backslash D$. D^W is newly generated SIFT features.
8. Store D^W in database as watermark.

Algorithm 6.2: Watermark Extraction & Authentication

Input: Watermarked Video V^W, Number of frames in original video $n1$, side plane index where watermark is embedded y, watermark descriptor D^W

1. Let number of frames in the watermarked video V^W is n_2
2. **if** $n_2/n_1 > D_{th}$ **then**
   Resize side planes to keep its height-width ratio same(Fig.6.7b)
   end
3. Extract SIFT descriptors from y^{th} side plane of V^W. Let (D') is.
4. Apply SIFT matching on D' and D^W.
5. If there is high percentage of matching then the video is authenticated.

6.1.4 Experimental Results

In this section, a comprehensive set of experimentations have been carried out to justify the efficiency of the proposed scheme. Proposed scheme is applied on many standard videos of CIF resolution. Few of theme are presented here. All the videos are encoded with H.264/SVC (Scalable Video Coding) using JSVM [74] reference software after watermarking.

Evaluation Parameters :

Robustness : The percentage of matching is taken as the **robustness** of the descriptor i.e.

$$Robustness\ R = \frac{m}{|D|} \tag{6.4}$$

where D is the feature point descriptors in the database and m is the number of points matched in D when compared with D' (descriptors generated from attacked video).

Perceptual Quality: For assessing the perceptual quality of the proposed scheme, PSNR, SSIM and Flicker metric are used. MSU-VQMT [62] tool is used to calculate these metrics.

Video \ Frame Drop	15/300	30/300	50/300
city	87	84	80
crew	91	85	80
coast guard	91	86	84
mobile	92	85	83

Table 6.1: *Robustness against random frame dropping*

Video \ Frame Drop	25%	50%	75%
city	75	71	45
crew	70	56	50
coast guard	81	80	52
mobile	82	80	56

Table 6.2: *Robustness against temporal scaling*

Robustness against Frame Drop Attack

The robustness of the proposed scheme [as defined in Eqn. 6.4] has been evaluated against frame dropping and temporal adaptation attacks respectively in Table 6.1 and Table 6.2. Table 6.1 shows robustness of the scheme when 15, 30 and 50 (out of 300) frames are dropped randomly. In Table 6.2 robustness against temporal scaling attack is tabulated where dyadic/non-dyadic scaling is done using H.264-SVC encoder. From the tables, it is evident that robustness is more than 70% even after 50% frames are dropped. Table 6.3 shows the robustness of the proposed scheme against frame averaging where every frame is replaced by the average of previous and next frame. Robustness against frame dropping attack and temporal scaling of the proposed scheme is compared with Chong's scheme [67] in Fig. 6.8. Figure shows that robustness of the proposed scheme is better than Chong's scheme [67].

video	Robustness
city	88
crew	75
coast guard	80
mobile	82

Table 6.3: *Robustness against frame averaging*

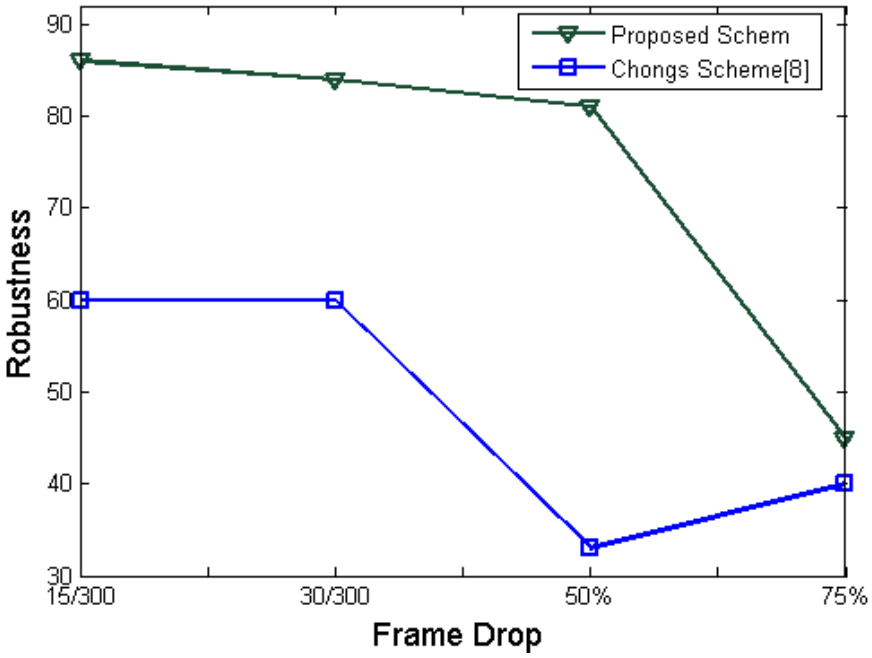


Figure 6.8: *Robustness comparison with Chong's scheme [67] for City video*

Perceptual Quality of the Watermarked Video

In this section, PSNR, Structural Similarity (SSIM) and flicker metric [60] are measured using MSU video quality measurement tool [62] to quantify the visual distortion due to watermark embedding.

In Fig. 6.9, comparison of PSNR of the watermarked *city* video of proposed scheme and Chong's scheme [67] is presented. In this figure, only watermarked frames are compared. Although number of frames altered in Chong's scheme is higher than the proposed scheme, only same number of frames are compared. PSNR of proposed scheme for all frames are constant because number of pixel altered and value of δ is same for all the frames. In Fig. 6.10 and Fig. 6.11, SSIM and flicker metric of the watermarked frames are compared with Chong's scheme [67] for *City* video. As it can be seen from the figures that the proposed scheme gives better result in terms of flickering artifacts.

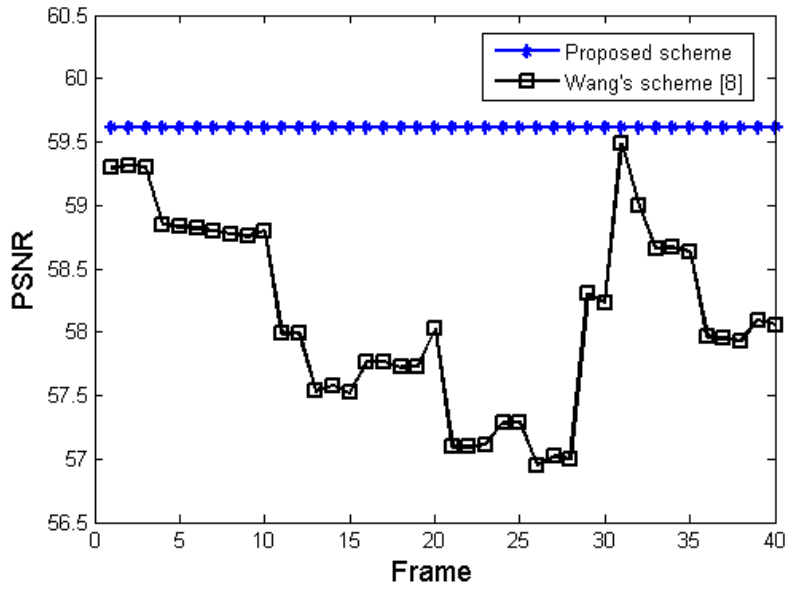


Figure 6.9: *PSNR comparison of the watermarked frames*

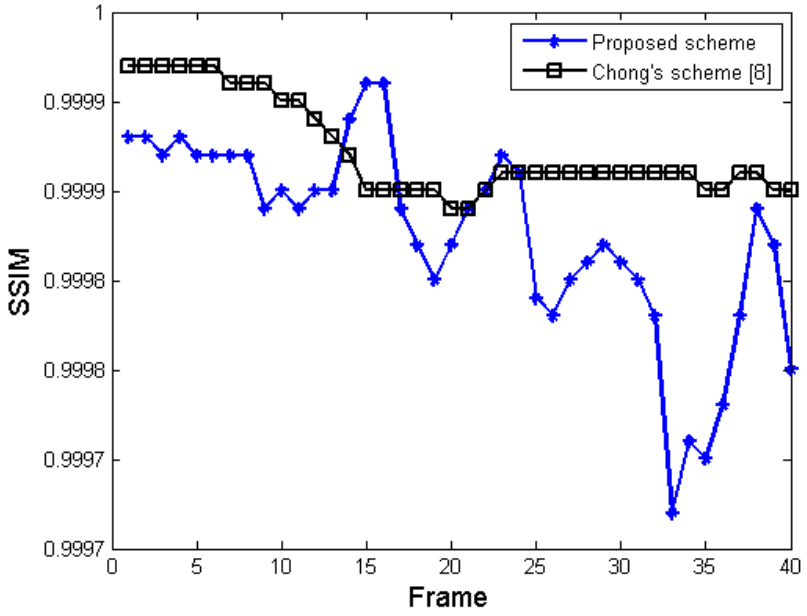


Figure 6.10: *SSIM comparison of the watermarked frames*

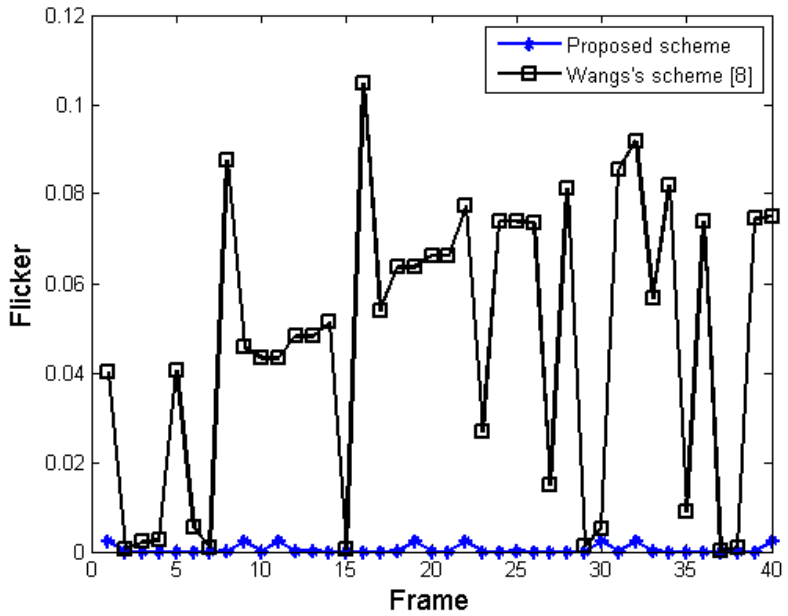


Figure 6.11: *Flicker comparison of the watermarked frames*

6.2 Robust video watermarking against Temporal Scalability

In the previous chapters, it has been shown that how temporal adaptation attacks can be resisted using a location map based temporal desynchronization scheme and a SIFT based approach. In this sub-section another very simple approach has been employed against temporal adaptation to meet another important requirement for scalable video watermarking called graceful improvement. In another words graceful improvement means that the robustness of the watermark increases with the addition of successive enhancement layers for scalable encoding. In the proposed scheme, each temporal layers of the scalable video has been separately embedded with a different watermark signals which are generated by DCT domain decomposition of a single watermark image. A zigzag sequence of block wise DCT coefficients of the watermark image is partitioned into non overlapping sets and each set is embedded separately into different temporal layers. The base layer is embedded with the first set of DCT coefficient (which includes DC coefficient of each block) and successive layers are embedded with successive non-overlapping coefficient sets. The coefficients of each set is chosen in such a fashion that uniform energy distribution across all temporal layers can be maintained. Experimental results show that the proposed scheme is robust against temporal scalability and robustness of the watermark increases with the addition of successive enhancement layers to achieve the graceful improvement.

6.2.1 Proposed scheme

In this section, proposed watermarking scheme has been illustrated in four subsections. Firstly, the process of watermark signal generation is described, followed by embedding and extraction algorithms are presented and finally, how the proposed scheme has achieved graceful improvement, is justified. H.264/SVC supports dyadic and non-dyadic temporal scalability and the nature of the scalability as well as number of temporal layers are defined by some input parameters (e.g size of GOP, input frame rate, output frame rate etc.). In the proposed scheme, first, all frames are divided into groups, each containing frames of a specific layer

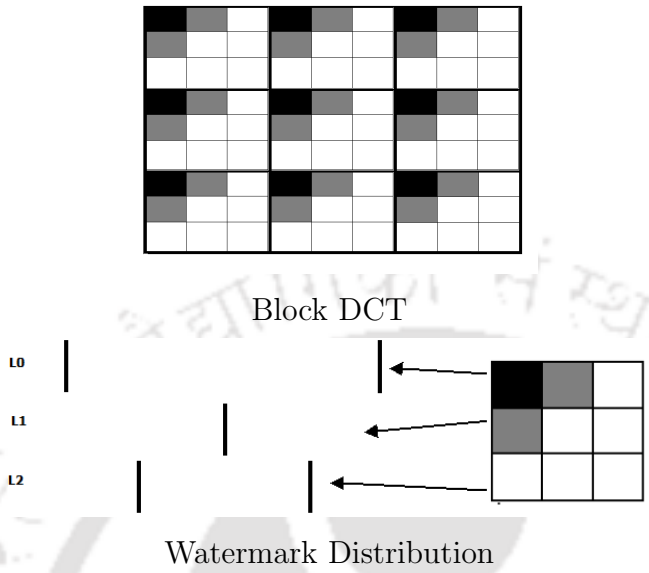


Figure 6.12: *Watermark Generation*

as shown in Fig. 6.12. In the Fig. 6.12, L0, L1 and L2 are group of frames. Frames are categorized using the input parameters of H.264/SVC. Then each set of frame is embedded with different watermark as discussed in Sec. 6.2.2.

6.2.2 Watermark Generation

In the proposed scheme, gray scale image is used as watermark. The watermark image is divided into non overlapping blocks of the size $k \times k$. Each blocks then are subjected to $k \times k$ block DCT transform. Coefficients of each block are divided into N non overlapping sets where N is the number of temporal layers. Each set consists of DCT coefficients in zigzag order as shown in Fig.6.12. In Fig.6.12, watermark generation and distribution of the watermark is shown for 3 dyadic temporal layers. Black colored coefficient from each block is embedded in all frames of layer L0. Watermark for next layer (L1) frames consist of gray coefficients of each block and so on. Number of coefficients in each set is selected in such a way that the energy of the watermark gets distributed uniformly across the every set of coefficients.

6.2.3 Watermark Embedding

In the proposed scheme, watermark is embedded in the approximation sub band obtained from 2-level of wavelet decomposition of each frame to make it robust against compressions. After the watermark generation, first set of coefficients which includes DC coefficient is embedded in the each frame of the base layer frame set (L0 in Fig. 6.12). Successive set of coefficients are embedded in the successive enhancement layers as shown in Fig. 6.12. The overall embedding scheme is depicted in Fig. 6.14. Watermark is embedded using Eqn.6.5

$$C_1' = C_1 + \alpha w \tag{6.5}$$

where C_1 is wavelet coefficient after 2-level wavelet decomposition and α is the watermark strength. In Fig. 6.13, plot of absolute values of DCT coefficients of 16×16 block of a gray scale natural image in zigzag order is shown. It is observed from the Fig. 6.13 that the scale of values decreases toward the end of the right bottom corner. So the use of same α value in Eqn.6.5 for every layer may not be useful in this case. So in the proposed scheme, value of α is incremented in every layer depending on the energy distribution of the successive coefficient sets.

6.2.4 Watermark Extraction

Same as embedding, watermarked video is also divided into sets of frames during extraction. Each set contains frames from different temporal layers. From different set of frames, watermark information are extracted separately. If only the base layer is available at the user end then only the first set of watermark coefficients can be extracted. If one or more enhancement layers are available then successive set of coefficients can be extracted. After extraction, coefficients are arranged in a block structure. If all coefficients are not available then blocks are filled with zeros and inverse-DCT is done on the blocks to extract the watermark. The whole extraction scheme is depicted in Fig.6.15.

6.2.5 Graceful Improvement

As mentioned in [20], graceful improvement is one of the important characteristics of scalable watermarking. For temporal scalability, with addition of higher

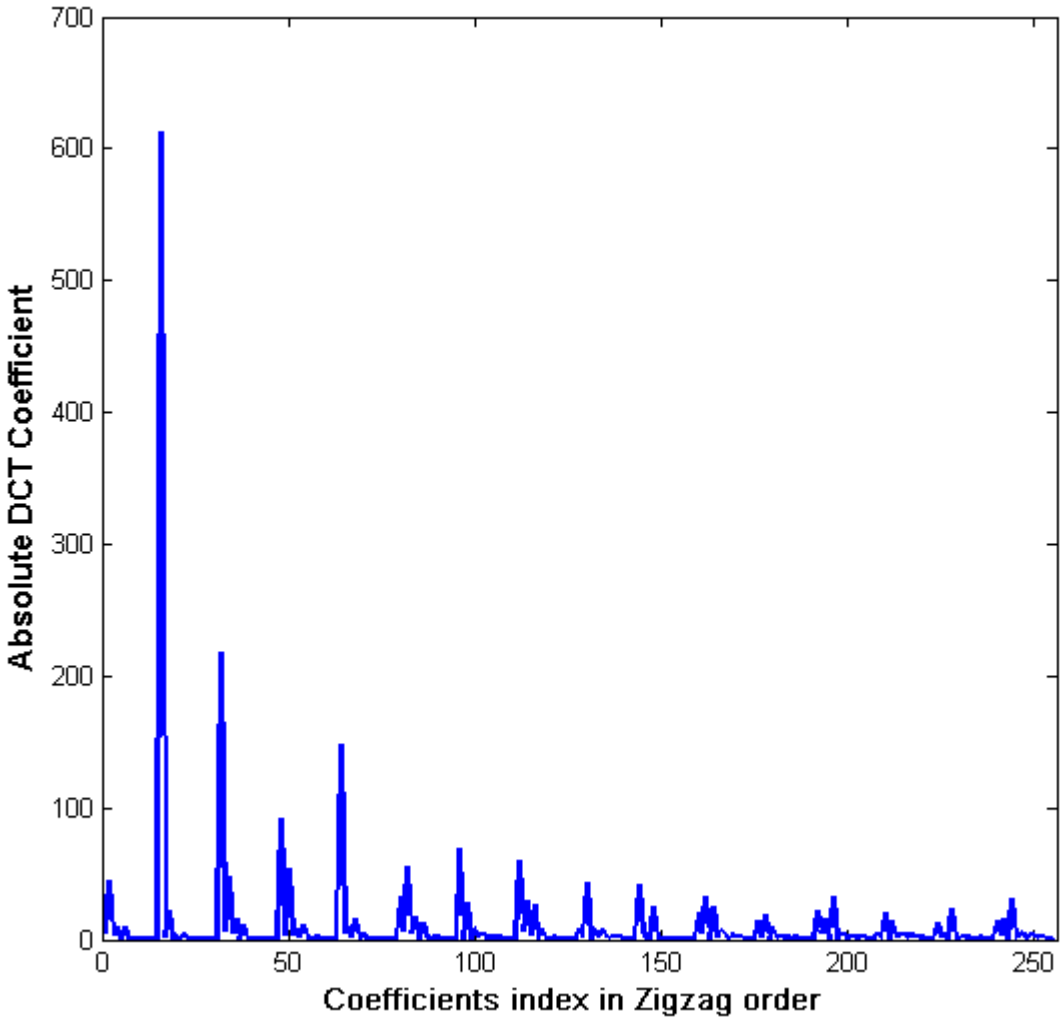


Figure 6.13: *DCT coefficients in zigzag scan*

temporal layers, robustness of the extracted watermark should increase. Fig. 6.16 depicts the mechanism employed to achieve graceful improvement within the proposed scheme. If only the base temporal layer of the video is available at user end, then only a crude approximation of the watermark image is obtained from the first set of extracted watermark coefficients. Now with the addition of the higher temporal layer, higher frequency coefficients of the watermark image can be extracted, thus improving the correlation of the extracted image with the original

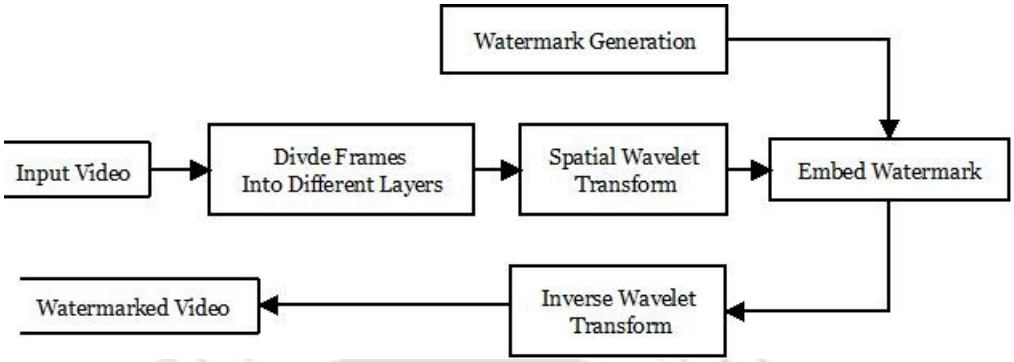


Figure 6.14: *Watermark Embedding*

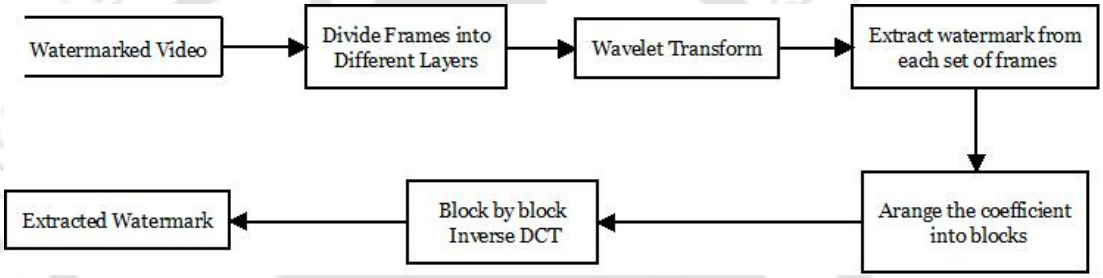


Figure 6.15: *Watermark Extraction*

watermark.

6.2.6 Experimental Result

The proposed scheme has been tested on different video sequences with different motion characteristics (e.g. *Akiyo* with low motion and *Bus* with high motion) and different size (e.g *City* and *Crew* with 4CIF resolution and *Bus* and *Akiyo* with CIF resolution). *Joint Scalable Video Model* (JSVM) [74] reference software version 9.19.12 is used for temporal scalable video encoding. Robustness of the proposed scheme is calculated by finding the correlation of the extracted watermark image and original watermark image.

In Fig.6.17, robustness of the proposed scheme is compared with the Bhowmik's [39] scheme at different frame rate for different video sequences. The graph shows that with increasing frame rate, robustness of the proposed scheme is increasing and it is consistently better than Bhowmik's[39] scheme for all the video sequences. During extraction in Bhowmik's approach, all the dropped frames

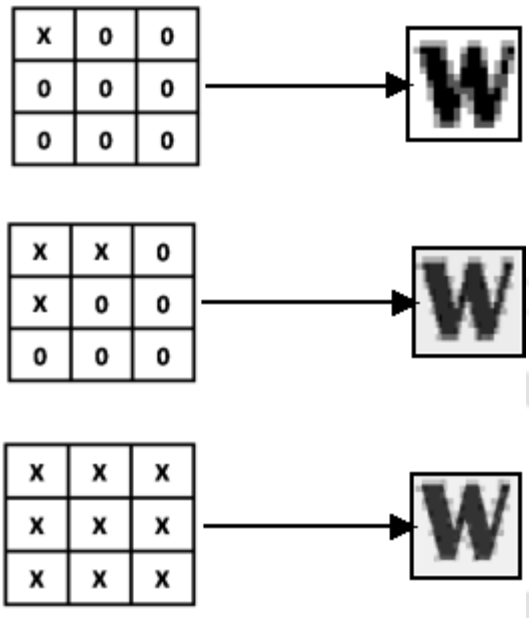


Figure 6.16: *Graceful Improvement*

are replaced with original frames to maintain the GOP structure. Visual quality of the watermarked video for the proposed scheme performs better than the Bhowmik's scheme. PSNR comparison of the corresponding watermarked video sequences is shown in Fig.6.18. Payload of each layer frame is different which generates the pattern in the PSNR plot.

6.3 Conclusion

In this chapter, two watermarking schemes resilient to temporal adaptation attack are described. In the first scheme, instead of embedding any watermark information, watermark is generated from the video itself. In that scheme, intensity of few patches from side faces of the video are changed. Due to this change, some new features are generated, which are used as watermark. Experimental result shows that watermark sustains even after 75% frame dropping. In the second scheme, each temporal layer frames are watermarked with different watermark, which is generated by block DCT of a single watermark image. Proposed scheme achieves graceful improvement over the robustness with addition of higher layer

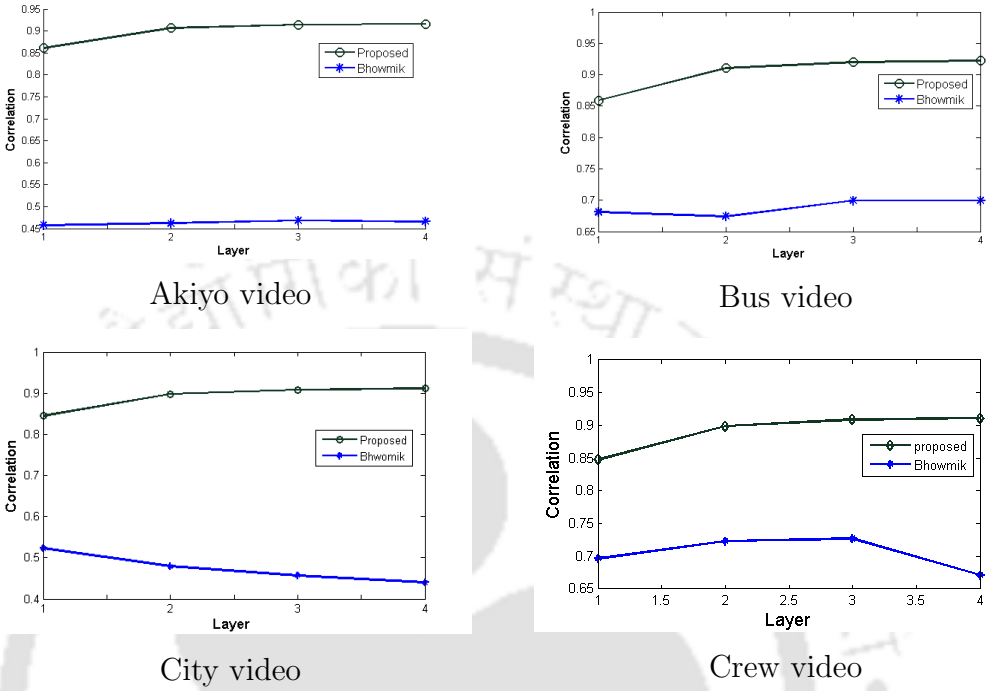


Figure 6.17: *Robustness comparison*

frames.

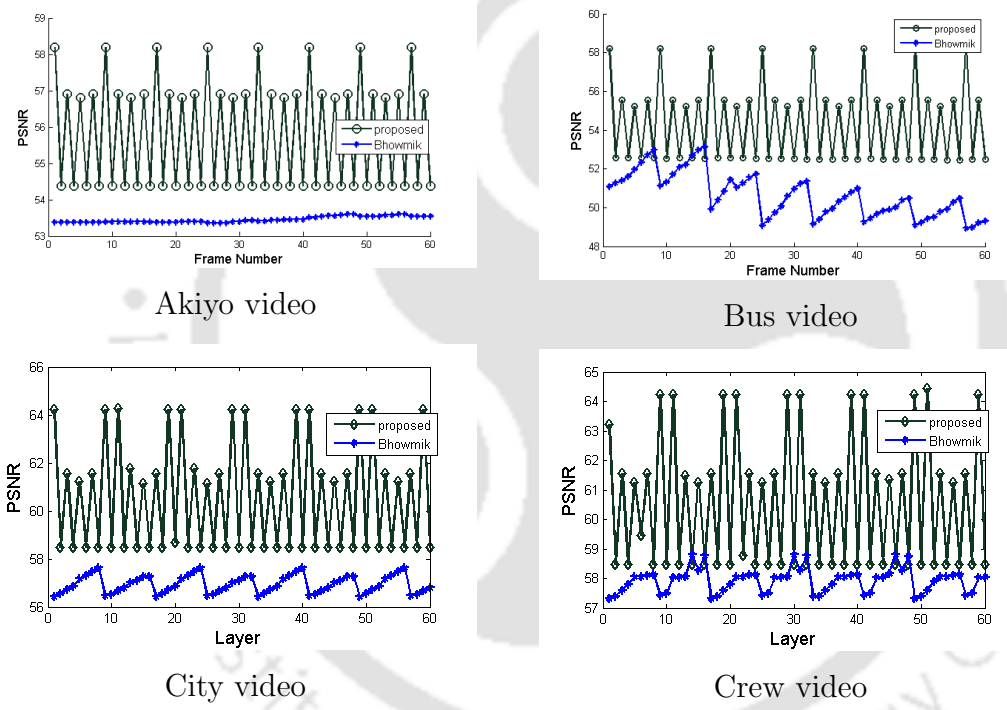


Figure 6.18: *PSNR comparison*

Chapter 7

Conclusion and Future Works

With the recent popularity of the scalable video coding, secure scalable video transmission become an important requirement. In this dissertation, the entire work is primarily motivated to propose the robust watermarking solutions for different scalable adaptations like resolution, temporal and and quality scalability. A brief summary of contributions made toward these is provided below.

7.1 Watermarking against Resolution and Quality Scalability

It has been observed in the literature that most of the existing schemes against resolution and quality adaptation fail to meet two basic requirements of the scalable watermarking, firstly the watermark should be extracted from each of the scalable layers and secondly reliability of the extracted watermark should be increased with increase of the video quality layers i.e. achieving graceful improvement.

In the first work of this dissertation, a uncompressed domain blind video watermarking scheme is proposed against resolution and quality scalability where enhancement layers are embedded with up-sampled base layer watermark. The spatial synchronization between successive layers are maintained using a location map to achieve the graceful improvement. For base layer, watermark is embedded in a DC frame which is generated by accumulating DC values of non-overlapping

blocks for every frame in the input video sequence. DC frame sequence is up-sampled and subtracted from the original video sequence to generate residual frame sequence. Then DCT based temporal filtering is applied on DC frame sequence as well as residual frame sequence. The watermark is embedded in low pass DC frames and the up sampled watermark is embedded in the low pass residual frames. It is experimentally shown that the proposed scheme performs well against resolution and quality adaptation and outperforms existing related schemes.

7.2 Watermarking against temporal and quality scalability

There exist few schemes [67, 68] in the literature which are resilient to the random frame dropping where number of dropped frames are very less. But these schemes fails against temporal adaptation where number dropped frames are more. In the second work of this thesis, a scalable video watermarking scheme has been proposed, which is robust against quality and temporal scalability. In the proposed scheme, wavelet based spatial filtering and DCT based temporal filtering are used for selecting watermark embedding zone. Temporal filtering is used on Group of Picture (GOP) to exploit the correlation among frames and the watermark is embedded in the low pass frames. To extract the watermark, a location map is required which essentially stores locations of embedded watermark in each frame so watermark can be extracted after the temporal adaptation.

7.3 Image watermarking based on SIFT against resolution scaling

Although, the proposed scheme against the resolution scalability outperforms recent existing schemes, its performance may be improved especially when the resolution scaling is relatively large. In the third work of the dissertation, a novel SIFT based image watermarking scheme is proposed which is robust to the resolution scaling, which can be easily extended to video by taking the temporal

dimension (motion) into consideration. In this work, a context coherent image patch has been inserted in the image such a way that it generates new and stable SIFT features. These newly generated SIFT feature descriptors are themselves used as the watermark. Since the SIFT features are invariant to scaling, these features can be extracted from any image resolution with high probability. Experiment on large image data set have been carried out to prove the efficiency of the scheme over the existing literature against high degree of resolution scaling.

7.4 Watermarking against Temporal Scalability

In the fourth chapter of this thesis, a watermarking scheme is proposed for temporal scalability which outperforms the existing methods. But it requires location map for the extraction of the watermark. In the final phase of the work, two blind watermarking schemes are proposed against temporal scalability which requires no extra information for the watermark extraction. In the first work, SIFT features are used to handle the temporal scalability. In this work, a patch of a side plane of the video is modified to generate a set of new SIFT feature, which then stored in database as watermark. Modification is done in a low motion area of a randomly selected frame set to avoid flickering artifacts. Effectiveness of the scheme is experimentally justified against the temporal adaptation and frame dropping attacks.

In the second work, frames of each temporal layer has been embedded with a different watermark which is generated by block DCT decomposition of a single watermark image. A zigzag sequence of block DCT coefficients of the watermark image is partitioned into non overlapping sets such that energy are distributed uniformly among the sets. Each set of coefficients are then embedded separately into different temporal layers to achieve graceful improvement. The base layer is embedded with the first set of DCT coefficient (which includes DC coefficient of each block) and successive layers are embedded with successive non-overlapping coefficient sets. Experimental result shows good robustness and graceful improvement over the temporal scalability.

7.5 Future Research Scope

The present study of this dissertation is mainly restricted for the uncompressed domain watermarking. Since uncompressed domain schemes are a bit slow because of decoding and further re-encoding, the equivalent study in compressed domain may be an interesting future scope. In addition, combining different scalability is always an difficult task and may also be another important future scope of this work. Finally, extension of these schemes for HD, beyond HD and 3D video sequence may also another good topic for further research.

References

[1] P. Meerwald and A. Uhl, "Toward robust watermarking of scalable video," in *SPIE, Security, Forensics, Steganography, and Watermarking of Multimedia Contents X*, vol. 6819, Jan 2008. [Pg.1]

[2] T. Stutz and A. Uhl, "A survey of h.264 avc/svc encryption," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 22, no. 3, pp. 325 –339, march 2012. [Pg.2]

[3] K. Mokhtarian and M. Hefeeda, "Authentication of scalable video streams with low communication overhead," *Multimedia, IEEE Transactions on*, vol. 12, no. 7, pp. 730 –742, nov. 2010. [Pg.2]

[4] National Institute of Standards and Technology, "Advanced encryption standard (AES)," *FIPS-197*, Nov 2001. [Pg.2]

[5] P. K. Atrey, W.-Q. Yan, E.-C. Chang, and M. S. Kankanhalli, "A hierarchical signature scheme for robust video authentication using secret sharing," in *Proceedings of the 10th International Multimedia Modelling Conference*, ser. MMM '04. Washington, DC, USA: IEEE Computer Society, 2004, pp. 330–. [Online]. Available: http://dl.acm.org/citation.cfm?id=968883.969463 [Pg.2]

[6] I. Cox, J. Kilian, F. Leighton, and T. Shamoon, "Secure spread spectrum watermarking for multimedia," *IEEE Transactions on Image Processing*, vol. 6, no. 12, pp. 1673 –1687, dec 1997. [Pg.2], [Pg.6]

REFERENCES

[7] F. Hartung and M. Kutter, "Multimedia watermarking techniques," *Proceedings of the IEEE*, vol. 87, no. 7, pp. 1079–1107, Jul 1999. [Pg.2]

[8] M. Swanson, M. Kobayashi, and A. Tewfik, "Multimedia data-embedding and watermarking technologies," *Proceedings of the IEEE*, vol. 86, no. 6, pp. 1064–1087, Jun 1998. [Pg.2]

[9] G. Langelaar, I. Setyawan, and R. Lagendijk, "Watermarking digital image and video data. a state-of-the-art overview," *Signal Processing Magazine, IEEE*, vol. 17, no. 5, pp. 20–46, Sep 2000. [Pg.2], [Pg.5]

[10] Y. Tew and K. Wong, "An overview of information hiding in h.264/avc compressed video," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 24, no. 2, pp. 305–319, Feb 2014. [Pg.2]

[11] S. P. MAITY and M. K. KUNDU, "Performance improvement in spread spectrum image watermarking using wavelets," *International Journal of Wavelets, Multiresolution and Information Processing*, vol. 09, no. 01, pp. 1–33, 2011. [Online]. Available: http://www.worldscientific.com/doi/abs/10.1142/S0219691311003931 [Pg.2]

[12] R. B. Wolfgang and E. J. Delp, "Fragile watermarking using the vw2d watermark," in *Proc. SPIE/IS&T Inter. Conf. Security and Watermarking of multimedia Contents*, 1999, pp. 204–213. [Pg.5]

[13] J. Haitsma and T. Kalker, "A watermarking scheme for digital cinema," in *Image Processing, 2001. Proceedings. 2001 International Conference on*, vol. 2, Oct 2001, pp. 487–489 vol.2. [Pg.5]

[14] S. Emmanuel and M. S. Kankanhalli, "Mask-based interactive watermarking protocol for video," pp. 247–258, 2001. [Online]. Available: http://dx.doi.org/10.1117/12.448209 [Pg.6]

[15] R. Anderson and F. A. Petitcolas, "On the limits of steganography," *Selected Areas in Communications, IEEE Journal on*, vol. 16, no. 4, pp. 474–481, May 1998. [Pg.6]

[16] P. Singh and R. Chadha, "A survey of digital watermarking techniques, applications and attacks," *International Journal of Engineering and Innovative Technology (IJEIT)*, vol. 2, no. 9, 2013. [Pg.6]

[17] G. Doerr and J. Dugelay, "Security pitfalls of frame-by-frame approaches to video watermarking," *Signal Processing, IEEE Transactions on*, vol. 52, no. 10, pp. 2955–2964, Oct 2004. [Pg.6], [Pg.8], [Pg.59], [Pg.60]

[18] G. Dorr and J.-L. Dugelay, "A guide tour of video watermarking," *Signal Processing: Image Communication*, vol. 18, no. 4, pp. 263 – 282, 2003, special Issue on Technologies for Image Security. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0923596502001443 [Pg.6]

[19] A. Piper, R. Safavi-Naini, and A. Mertins, "Coefficient selection methods for scalable spread spectrum watermarking," in *Digital Watermarking*, ser. Lecture Notes in Computer Science, T. Kalker, I. Cox, and Y. Ro, Eds. Springer Berlin Heidelberg, 2004, vol. 2939, pp. 235–246. [Online]. Available: http://dx.doi.org/10.1007/978-3-540-24624-4_18 [Pg.6], [Pg.13]

[20] ——, "Resolution and quality scalable spread spectrum image watermarking," in *Proceedings of the 7th workshop on Multimedia and security*, ser. MM&Sec '05. New York, NY, USA: ACM, 2005, pp. 79–90. [Online]. Available: http://doi.acm.org/10.1145/1073170.1073186 [Pg.6], [Pg.8], [Pg.35], [Pg.119]

[21] J. Seo and H. Park, "Data protection of multimedia contents using scalable digital watermarking," in *Proceedings of the Fourth Annual ACIS International Conference on Computer and Information Science*, ser. ICIS '05. Washington, DC, USA: IEEE Computer Society, 2005, pp. 376–380. [Online]. Available: http://dx.doi.org/10.1109/ICIS.2005.42 [Pg.6]

[22] P. Bas, J.-M. Chassery, and B. Macq, "Geometrically invariant watermarking using feature points," *Image Processing, IEEE Transactions on*, vol. 11, no. 9, pp. 1014–1028, 2002. [Pg.6], [Pg.7]

[23] H.-Y. Lee, C.-h. Lee, H.-K. Lee, and J. Nam, "Feature-based image watermarking method using scale-invariant keypoints," in *Advances in Multimedia Information Processing-PCM 2005*. Springer, 2005, pp. 312–324. [Pg.6]

[24] H.-Y. Lee, H. Kim, and H.-K. Lee, "Robust image watermarking using local invariant features," *Optical Engineering*, vol. 45, no. 3, pp. 037 002–037 002–11, 2006. [Online]. Available: +http://dx.doi.org/10.1117/1.2181887 [Pg.xx], [Pg.6], [Pg.7], [Pg.78], [Pg.86]

[25] J. J. O'Ruanaidh and T. Pun, "Rotation, scale and translation invariant digital image watermarking," in *Image Processing, 1997. Proceedings., International Conference on*, vol. 1. IEEE, 1997, pp. 536–539. [Pg.6]

[26] J. Radon, "1.1 über die bestimmung von funktionen durch ihre integralwerte längs gewisser mannigfaltigkeiten," *Classic papers in modern diagnostic radiology*, p. 5, 2005. [Pg.6]

[27] S. Pereira and T. Pun, "Robust template matching for affine resistant image watermarks," *Image Processing, IEEE Transactions on*, vol. 9, no. 6, pp. 1123–1129, 2000. [Pg.6]

[28] G. Sharma and D. J. Coumou, "Watermark synchronization: Perspectives and a new paradigm," in *Information Sciences and Systems, 2006 40th Annual Conference on*. IEEE, 2006, pp. 1182–1187. [Pg.7]

[29] M. Kutter, S. K. Bhattacharjee, and T. Ebrahimi, "Towards second generation watermarking schemes," in *Image Processing(ICIP). Proceedings. 1999 International Conference on*, vol. 1. IEEE, 1999, pp. 320–323. [Pg.7]

[30] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vision*, vol. 60, no. 2, pp. 91–110, Nov. 2004. [Online]. Available: http://dx.doi.org/10.1023/B:VISI.0000029664.99615.94 [Pg.7], [Pg.24], [Pg.25], [Pg.77], [Pg.82]

[31] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. New York, NY, USA: Cambridge University Press, 2003. [Pg.7], [Pg.77]

[32] S. E. Chen and L. Williams, "View interpolation for image synthesis," in *Proceedings of the 20th annual conference on Computer graphics and interactive techniques*, ser. SIGGRAPH '93. New York, NY, USA: ACM, 1993, pp. 279–288. [Online]. Available: http://doi.acm.org/10.1145/166117.166153 [Pg.7], [Pg.77]

[33] D. Lowe, "Object recognition from local scale-invariant features," in *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on*, vol. 2, 1999, pp. 1150–1157 vol.2. [Pg.7], [Pg.77]

[34] A. Bosch, A. Zisserman, and X. Muoz, "Scene classification via plsa," in *Computer Vision ? ECCV 2006*, ser. Lecture Notes in Computer Science, A. Leonardis, H. Bischof, and A. Pinz, Eds. Springer Berlin Heidelberg, 2006, vol. 3954, pp. 517–530. [Online]. Available: http://dx.doi.org/10.1007/11744085_40 [Pg.7], [Pg.77]

[35] J. Mutch and D. Lowe, "Object class recognition and localization using sparse features with limited receptive fields," *International Journal of Computer Vision*, vol. 80, no. 1, pp. 45–57, 2008. [Online]. Available: http://dx.doi.org/10.1007/s11263-007-0118-0 [Pg.7], [Pg.77]

[36] P. Saeedi, P. Lawrence, and D. Lowe, "Vision-based 3-d trajectory tracking for unknown environments," *Robotics, IEEE Transactions on*, vol. 22, no. 1, pp. 119–136, 2006. [Pg.7], [Pg.77]

[37] V.-Q. Pham, T. Miyaki, T. Yamasaki, and K. Aizawa, "Geometrically invariant object-based watermarking using sift feature," in *Image Processing, 2007. ICIP 2007. IEEE International Conference on*, vol. 5, 2007, pp. V – 473–V – 476. [Pg.7], [Pg.78]

[38] L. Jing, L. Gang, and Z. Jiulong, "Robust image watermarking based on sift feature and optimal triangulation," in *Information Technology and Applications, 2009. IFITA'09. International Forum on*, vol. 3. IEEE, 2009, pp. 337–340. [Pg.xx], [Pg.7], [Pg.78], [Pg.86]

[39] D. Bhowmik and C. Abhayaratne, "Video watermarking using motion compensated 2d+t+2d filtering," in *Proceedings of the 12th ACM workshop on Multimedia and security*. New York, NY, USA: ACM, 2010, pp. 127–136. [Online]. Available: http://doi.acm.org/10.1145/1854229.1854254 [Pg.xix], [Pg.8], [Pg.10], [Pg.11], [Pg.12], [Pg.20], [Pg.23], [Pg.48], [Pg.49], [Pg.51], [Pg.52], [Pg.53], [Pg.60], [Pg.61], [Pg.70], [Pg.71], [Pg.72], [Pg.75], [Pg.121]

[40] P. Vinod and P. Bora, "Motion-compensated inter-frame collusion attack on video watermarking and a countermeasure," *Information Security, IEE Proceedings*, vol. 153, no. 2, pp. 61 – 73, june 2006. [Pg.8], [Pg.9], [Pg.10], [Pg.20], [Pg.59], [Pg.61]

[41] H.-S. Jung, Y.-Y. Lee, and S. U. Lee, "Rst-resilient video watermarking using scene-based feature extraction," *EURASIP J. Appl. Signal Process.*, vol. 2004, pp. 2113–2131, Jan. 2004. [Online]. Available: http://dx.doi.org/10.1155/S1110865704405046 [Pg.8], [Pg.9], [Pg.10]

[42] W. Lu, R. Safavi-Naini, T. Uehara, and W. Li, "A scalable and oblivious digital watermarking for images," in *Signal Processing, 2004. Proceedings. ICSP '04. 2004 7th International Conference on*, vol. 3, aug.-4 sept. 2004, pp. 2338 – 2341 vol.3. [Pg.8]

[43] C.-C. Wang, Y.-C. Lin, S.-C. Yi, and P.-Y. Chen, "Digital authentication and verification in mpeg-4 fine-granular scalability video using bit-plane watermarking." in *IPCV*, H. R. Arabnia, Ed. CSREA Press, 2006, pp. 16–21. [Online]. Available: http://dblp.uni-trier.de/db/conf/ipcv/ipcv2006-1.html#WangLYC06 [Pg.8], [Pg.9], [Pg.33]

[44] P. Meerwald and A. Uhl, "Robust watermarking of h.264-encoded video: Extension to svc," in *Proceedings of the 2010 Sixth International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, ser. IIH-MSP '10. Washington, DC, USA: IEEE Computer Society, 2010, pp. 82–85. [Online]. Available: http://dx.doi.org/10.1109/IIHMSP.2010.28 [Pg.xix], [Pg.8], [Pg.12], [Pg.13], [Pg.33], [Pg.36]

[45] A. Alattar, E. Lin, and M. Celik, "Digital watermarking of low bit-rate advanced simple profile mpeg-4 compressed video," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 13, no. 8, pp. 787–800, Aug 2003. [Pg.9]

[46] F.-C. Chang, H.-C. Huang, and H.-M. Hang, "Layered access control schemes on watermarked scalable media," in *Circuits and Systems, 2005. ISCAS 2005. IEEE International Symposium on*, may 2005, pp. 4983 – 4986 Vol. 5. [Pg.9]

[47] Y. Wang and A. Pearmain, "Blind mpeg-2 video watermarking in dct domain robust against scaling," *Vision, Image and Signal Processing, IEE Proceedings* -, vol. 153, no. 5, pp. 581 –588, oct. 2006. [Pg.9], [Pg.35], [Pg.48], [Pg.49], [Pg.51], [Pg.52], [Pg.53]

[48] L. Yan and Z. Jiying, "Rst invariant video watermarking based on 1d dft and radon transform," in *Visual Information Engineering, 2008. VIE 2008. 5th International Conference on*, 29 2008-aug. 1 2008, pp. 443 –448. [Pg.10]

[49] Z. Huai-yu, L. Ying, and W. Cheng-ke, "A blind spatial-temporal algorithm based on 3d wavelet for video watermarking," in *Multimedia and Expo, 2004. ICME '04. 2004 IEEE International Conference on*, vol. 3, 2004, pp. 1727–1730 Vol.3. [Pg.10]

[50] A. Essaouabi and E. Ibnelhaj, "A 3d wavelet-based method for digital video watermarking," in *Networked Digital Technologies, 2009. NDT '09. First International Conference on*, 2009, pp. 429–434. [Pg.10]

[51] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the scalable video coding extension of the h.264/avc standard," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 17, no. 9, pp. 1103 –1120, sept. 2007. [Pg.10]

[52] M. Flierl and B. Girod, "Video coding with motion-compensated lifted wavelet transforms," *Signal Processing: Image Communication*, vol. 19, no. 7, pp. 561 – 575, 2004. [Online]. Available: http:

//www.sciencedirect.com/science/article/pii/S0923596504000372 [Pg.10], [Pg.19]

[53] S.-J. Choi and J. Woods, "Motion-compensated 3-d subband coding of video," *Image Processing, IEEE Transactions on*, vol. 8, no. 2, pp. 155 – 167, feb 1999. [Pg.10], [Pg.19]

[54] F. Verdicchio, Y. Andreopoulos, T. Clerckx, J. Barbarien, A. Munteanu, J. Cornelis, and P. Schelkens, "Scalable video coding based on motion-compensated temporal filtering: complexity and functionality analysis." in *ICIP*, 2004, pp. 2845–2848. [Online]. Available: http://dblp.uni-trier.de/db/conf/icip/icip2004-5.html#VerdicchioACBMCS04 [Pg.xix], [Pg.10], [Pg.19], [Pg.20]

[55] M. Noorkami and R. M. Mersereau, "A framework for robust watermarking of h.264-encoded video with controllable detection performance," *Information Forensics and Security, IEEE Transactions on*, vol. 2, no. 1, pp. 14 –23, march 2007. [Pg.12]

[56] R. Atta and M. Ghanbari, "Spatio-temporal scalability-based motion-compensated 3-d subband/dct video coding," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 16, no. 1, pp. 43 – 55, jan. 2006. [Pg.20], [Pg.23], [Pg.61]

[57] J.-R. Ohm, "Advanced packet-video coding based on layered vq and sbc techniques," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 3, no. 3, pp. 208 –221, jun 1993. [Pg.23]

[58] N. Riche, M. Mancas, M. Duvinage, M. Mibulumukini, B. Gosselin, and T. Dutoit, "Rare2012: a multi-scale rarity-based saliency detection with its comparative statistical analysis," *Signal Processing: Image Communication*, vol. 28, no. 6, pp. 642–658, 2013. [Pg.xix], [Pg.26], [Pg.27], [Pg.85], [Pg.88], [Pg.93]

[59] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *Image Processing, IEEE Transactions on*, vol. 13, no. 4, pp. 600 –612, april 2004. [Pg.28], [Pg.48]

[60] F. X, G. W, L. Y, and Z. D, "Flicking reduction in all intra frame coding," JVT-E070, Tech. Rep, October 2002. [Pg.28], [Pg.39], [Pg.48], [Pg.49], [Pg.61], [Pg.106], [Pg.115]

[61] F. Xiao, "Dct-based video quality evaluation," Final Project for EE392J, Final Project for EE392J, December 2000. [Pg.28], [Pg.29], [Pg.48], [Pg.49]

[62] D.Vatolin, M.Smirnov, A.Ratushnyak, and V.Yoockin, "Msu video quality measurement tool," MSU Graphics and Media Lab, Tool, 2001-2008. [Online]. Available: http://www.compression.ru/video/ [Pg.28], [Pg.29], [Pg.48], [Pg.112], [Pg.115]

[63] A. B. Watson, "Dct quantization matrices visually optimized for individual images," in *IS&T/SPIE's Symposium on Electronic Imaging: Science and Technology*. International Society for Optics and Photonics, 1993, pp. 202–216. [Pg.29], [Pg.30], [Pg.85], [Pg.93]

[64] Q. Yan, L. Xu, J. Shi, and J. Jia, "Hierarchical saliency detection," in *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*. IEEE, 2013, pp. 1155–1162. [Pg.31], [Pg.84]

[65] G. Griffin, A. Holub, and P. Perona, "Caltech-256 object category dataset," 2007. [Pg.31], [Pg.84]

[66] R. Uetz and S. Behnke, "Large-scale object recognition with cuda-accelerated hierarchical neural networks," in *Intelligent Computing and Intelligent Systems, 2009. ICIS 2009. IEEE International Conference on*, vol. 1. IEEE, 2009, pp. 536–541. [Pg.31], [Pg.84]

[67] C. Chen, J. Ni, and J. Huang, "Temporal statistic based video watermarking scheme robust against geometric attacks and frame dropping," in *Digital Watermarking*, ser. Lecture Notes in Computer Science, A. Ho, Y. Shi, H. Kim, and M. Barni, Eds. Springer Berlin Heidelberg, 2009, vol. 5703, pp. 81–95. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-03688-0_10 [Pg.xxi], [Pg.60], [Pg.113], [Pg.114], [Pg.115], [Pg.126]

[68] C. Wang, C. Zhang, and P. Hao, "A blind video watermark detection method based on 3d-dwt transform," in *Image Processing (ICIP), 2010 17th IEEE International Conference on*, Sept 2010, pp. 3693–3696. [Pg.60], [Pg.126]

[69] J. Morel and G. Yu, "Is sift scale invariant?" *Inverse Problems and Imaging*, vol. 5, no. 1, pp. 115–136, 2011. [Pg.77]

[70] R. Haralick, "Some neighborhood operators," in *Real-Time Parallel Computing*. Springer, 1981, pp. 11–35. [Pg.79]

[71] N. Otsu, "A threshold selection method from gray-level histograms," *Automatica*, vol. 11, no. 285-296, pp. 23–27, 1975. [Pg.80]

[72] H. Su, W.-H. Chuang, W. Lu, and M. Wu, "Evaluating the quality of individual sift features," in *Image Processing (ICIP), 2012 19th IEEE International Conference on*, Sept 2012, pp. 2377–2380. [Pg.87]

[73] P. Bollimpalli, N. Sahu, and A. Sur, "Sift based robust image watermarking resistant to resolution scaling," in *Image Processing (ICIP), 2014 21st IEEE International Conference on*, Sept 2014. [Pg.92], [Pg.110]

[74] J. Reichel, H. Schwarz, and M. Wien, "Joint scalable video model 11 (jsvm 11)," Jul. 2007. [Pg.112], [Pg.121]

List of Publication

Journal Publication :

1. **Nilkanta Sahu**, Shuvendu Rana, Arijit Sur, "MCDCT-TF based video watermarking resilient to temporal and quality scaling", Multimedia Tools and Application, doi :10.1007/s11042-015-2949-y.

2. Shuvendu Rana, **Nilkanta Sahu**, and Arijit Sur, "Robust watermarking for resolution and quality scalable video sequence", Multimedia Tools and Applications, Springer US, 2015, 74, 7773-7802.

3. Arijit Sur, Sista Venkat Madhav Krishna, **Nilkanta Sahu** and Shuvendu Rana, "Detection of Motion Vector Based Video Steganography", Multimedia Tools and Applications, Springer, pp. 1-16, 2014

Conference Publication :

1. Priyatham Bollimpalli, **Nilkanta Sahu** and Arijit Sur, "SIFT Based Robust Image Watermarking Resistant To Resolution Scaling,", IEEE International Conference on Image Processing (ICIP), pp. 5507-5511, Paris, France, October, 2014

2. **Nilkanta Sahu**, Vivek Tiwari and Arijit Sur "Robust Video Watermarking Resilient to Temporal Scalability", National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG 2015), Patna, India.

Book Chapter :

1. **Nilkanta Sahu**, Arijit Sur, "Scalable Video Watermarking : A Survey", In R. Pal (Ed.), Innovative Research in Attention Modeling and Computer Vision Applications (pp. 365-387). Hershey, PA: Information Science Reference. doi:10.4018/978-1-4666-8723-3.ch015