Watermarking of 3D Images and Video

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

of

Doctor of Philosophy
in
Computer Science and Engineering

Submitted by
Shuvendu Rana

Under the guidance of
Dr. Arijit Sur

Department of Computer Science and Engineering
Indian Institute of Technology, Guwahati
August, 2017

Abstract

With the huge advancement in Internet technology as well as the availability of low cost 3D display devices, 3D image, and video transmission becomes popular in recent times. Since watermarking has been regarded as one of the potential digital right management (DRM) tools in the last decade; 3D image and video watermarking becomes an emerging research topic. In recent times, depth image based rendering (DIBR)-3D and depth based multi view 3D representation (MVD) are two popular 3D representations. In this dissertation, the entire thesis is primarily motivated to propose robust 3D image or video watermarking schemes for DIBR-3D and MVD representations against different hostile and non-hostile attacks.

In the first work, watermarking issues against 3D-HEVC compression and the view synthesis attack have been considered. It is observed from the literature that the most of the existing 3D video watermarking schemes not suit well against 3D-HEVC compression which is a state-of-art MVD based 3D video compression scheme. In the first phase of our first contributory chapter, a 3D video watermarking scheme has been proposed which can resist 3D-HEVC compression by employing independent view regions (Z-axis) based embedding. In the later part of this chapter, the proposed scheme has been extended to resist view synthesis attacks by exploiting the shift invariance property of the dual tree discrete wavelet transformation (DT-DWT) for embedding.

In 3D representation, depth is a crucial information and thus securing depth is an important requirement. In the first part of this chapter, a watermarking scheme is proposed to secure depth information of the

DIBR-3D image representation by scale invariant feature transform (SIFT) based embedding zone selection procedure. This work has been extended for video in the next part of this chapter by employing MCDCT-TF to reduce temporal noise in 3D-HEVC representation.

In recent literature, it is understood that view synthesis process in multi-view 3D representation can destroy the embedded watermark for most of the existing schemes. As a countermeasure, in the third chapter of this thesis, a view invariant image watermarking scheme has been proposed with efficient embedding zone selection by employing dependent view region identification and layer based depth partitioning algorithms.

In the fourth chapter of this dissertation, different depth based attacks such as depth modifications, depth blurring, changing baseline distance etc. have been considered along with view synthesis attack. In this chapter, a 3D image watermarking scheme has been proposed to resist against these attacks by exploiting the shift invariance and directional property of the dual tree complex wavelet transformation (DT-CWT) transformation.

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

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.

Shuvendu Rana

Copyright

Certificate

This is to certify that this thesis entitled **"Watermarking of 3D Images and Video"** being submitted by **Shuvendu Rana**, to Department of Computer Science and Engineering, Indian Institute of Technology, Guwahati, for partial fulfillment for the award of the degree of Doctor of Philosophy, is a bonafide work carried out by him under my supervision and guidance The thesis, in my opinion, is worthy of consideration for award of the degree of Doctor of Philosophy in accordance with the regulation of the institute. To the best of my knowledge, it has not been submitted elsewhere for the award of the degree.

.....

Dr. Arijit Sur
Associate Professor
Department of Computer Science and Engineering
IIT Guwahati

I would like to dedicate this thesis to my family...

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 am grateful to my guide **Dr. Arijit Sur** for allowing me to work in this research area. 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. G. Sajith, 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. Also I would like to thank Dr. Arnab Sarkar and Dr. Santosh Biswas, for their motivating support.

I would also like to acknowledge the acknowledge the services and support of the Staff of the Department of Computer Science and Engineering 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 Ashok sir, Mamata di, Shishendu, Sibaji, Satish, Brijesh, Sathisha B., Anirban, Niladri da, Mandar da, Senko da, Sangeet and many others for their unconditional help and support.

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.

Contents

1 Introduction **1**

1.1 3D Image and Video Representation 1

1.1.1 Depth 2

1.1.2 DIBR-3D Representation 3

1.1.3 Multi-view Representation 4

1.1.4 3D Video Encoder for MVD 5

View Synthesis 7

1.2 Digital Watermarking 8

1.2.1 Efficiency Parameters 8

1.3 Literature Survey 9

1.3.1 3D Image and Video Watermarking 9

1.3.1.1 3D Image Watermarking 9

1.3.1.2 3D Video Watermarking 12

1.3.2 3D Depth Watermarking 14

1.4 Motivation and Objectives 15

1.5 Contributions of the Thesis 16

1.5.1 Robust Watermarking for MVD Representation against 3D-HEVC Compression and View Synthesis Attack 16

1.5.1.1 Robust Watermarking for MVD Representation against 3D-HEVC Compression Attack 17

1.5.1.2 Watermarking for MVD Representation against 3D-HEVC Compression and View Synthesis Attack 17

1.5.2 Watermarking in Depth Information of 3D Image and Video Sequences against Depth Attack . . . 17
1.5.2.1 Watermarking in Image Depth against View Synthesis Attack . . . 18
1.5.2.2 Watermarking in Video Depth Sequences against Depth Modification and 3D-HEVC Compression Attack . . . 18
1.5.3 Depth-based View Invariant blind 3D Image Watermarking for Multi-view Representation . . . 18
1.5.4 View Invariant Watermarking Using DT-CWT for DIBR-3D Image Representation . . . 19
1.6 Thesis Organization . . . 19
1.7 Summary . . . 21

2 Research Background **23**
2.1 Motion Compensated DCT Temporal Filtering (MCDCT-TF) . . . 23
Inverse Motion Compensated DCT Temporal Filtering (IMCDCT-TF) . . . 25
2.2 Semantic Image Segmentation . . . 25
2.3 Scale Invariant Feature Transform . . . 26
2.4 Dual Tree Complex Wavelet Transform (DT-CWT) . . . 26
2.5 Attacks on 3D Image and Video . . . 28
2.6 Evaluation Parameter . . . 29
2.6.1 Visual Quality Parameter . . . 29
2.6.1.1 PSNR . . . 29
2.6.1.2 SSIM . . . 30
2.6.1.3 PSNRHVS . . . 30
2.6.1.4 MSSSIM . . . 31
2.6.1.5 VIFp . . . 31
2.6.1.6 Flicker Metric . . . 31
2.6.2 Robustness Parameter . . . 32
2.7 Experimental Dataset . . . 32
2.8 Summary . . . 32

3 Robust Watermarking for MVD Representation against 3D-HEVC Compression and View Synthesis Attack 35

3.1 Robust Watermarking for MVD Representation against 3D-HEVC Compression Attack . . . 37

3.1.1 Proposed Concept . . . 39

Z-axis Filtration . . . 39

3.1.2 Proposed Scheme . . . 41

3.1.2.1 Watermark Embedding . . . 41

3.1.2.2 Watermark Extraction . . . 43

3.1.3 Experimental Results . . . 48

3.1.3.1 Visual Quality Measurement . . . 48

3.1.3.2 Robustness Analysis . . . 49

3.1.4 Discussion . . . 53

3.2 Watermarking for MVD Representation against 3D-HEVC Compression and View Synthesis Attack . . . 53

3.2.1 Proposed Scheme . . . 55

3.2.1.1 Embedding Zone Selection . . . 56

3.2.1.2 Watermark Embedding . . . 57

3.2.1.3 Watermark Extraction . . . 60

3.2.2 Experiment Results . . . 60

3.2.2.1 Visual Quality . . . 62

3.2.2.2 Robustness . . . 62

3.2.3 Discussion . . . 64

3.3 Summary . . . 64

4 Watermarking in Depth Information of 3D Image and Video Sequences against Depth Attack 67

4.1 Watermarking in DIBR-3D based Image Depth against View Synthesis Attack . . . 69

4.1.1 Proposed Scheme . . . 69

4.1.1.1 Detection of Background and Foreground Objects 70

4.1.1.2 SIFT based coefficient selection . . . 71

4.1.1.3 Watermark Embedding . . . 72

4.1.1.4 Watermark Extraction . . . 73
4.1.2 Results . . . 75
4.1.2.1 Visual Quality . . . 76
4.1.2.2 Robustness . . . 78
4.1.2.3 Discussion . . . 79
4.2 Watermarking in Video Depth Sequences against Depth Modification and 3D-HEVC Compression Attack . . . 80
4.2.1 Proposed Scheme . . . 81
4.2.1.1 Motion Compensated Temporal Filtering (MCTF) 81
4.2.1.2 Detection of Background and Foreground Objects 82
4.2.1.3 SIFT based coefficient selection . . . 83
4.2.1.4 Watermark Embedding . . . 84
4.2.1.5 Watermark Extraction . . . 86
4.2.2 Results . . . 89
4.2.2.1 Visual Quality . . . 90
4.2.2.2 Robustness . . . 91
4.2.2.3 Discussion . . . 93
4.3 Summary . . . 94

5 Depth-based View Invariant Blind 3D Image Watermarking for Multi-view Representation 95
5.1 Embedding Zone Selection . . . 98
5.1.1 Dependent View Region Identification . . . 98
5.1.2 Layer Partitioning using Depth . . . 99
5.1.3 Block Partitioning . . . 101
5.1.4 Block Selection . . . 102
5.1.5 Visual Threshold Checking . . . 103
5.2 Watermark Embedding and Extraction Process . . . 104
5.2.1 Embedding of Watermark . . . 105
5.2.2 Extraction of Watermark . . . 105
5.2.2.1 Dependent View Region Identification for Synthesized Views . . . 107
5.2.2.2 Visual Threshold Checking . . . 109

5.2.2.3 Extraction of Watermark from the Selected Blocks 109
5.2.3 Robustness of the Proposed Scheme 111
5.3 Experimental Results 113
5.3.1 Visual Quality 114
5.3.2 Robustness 116
5.3.3 Discussion 120
5.4 Summary 124

6 View Invariant Watermarking Using DT-CWT for DIBR-3D Image Representation 127
6.1 Research Background 129
6.1.1 Rendering of DIBR-3D-Image 129
6.1.2 Dual Tree Complex Wavelet Transform (DT-CWT) 130
6.2 Proposed Scheme 133
6.2.1 Zone Selection 134
6.2.1.1 Zone Selection for Embedding 135
6.2.1.2 Zone Selection for Extraction 135
6.2.2 Watermark Embedding 136
6.2.3 Watermark Extraction 138
6.2.4 Robustness of the Proposed Scheme 141
6.3 Experimental Results 142
6.3.1 Visual Quality 142
6.3.2 Robustness 144
6.3.3 Discussion 147
6.4 Summary 150

7 Conclusion and future work 153
7.1 Summary of the Contributions 153
7.1.1 Robust Watermarking for MVD Representation against 3D-HEVC Compression and View Synthesis Attack 153
7.1.2 Watermarking in Depth Information of 3D Image and Video Sequences against Depth Attack 154

7.1.3 Depth-based View Invariant Blind 3D Image Watermarking for Multi-view Representation 155
7.1.4 View Invariant Watermarking Using DT-CWT for DIBR-3D Image Representation 156
7.2 Future Research Scope 156

References **159**

Appendix A: Summary of Publications **175**

List of Figures

1.1 Left and right view images with hidden pixels (say independent regions). . . . 2
1.2 Left and right video frames with hidden pixels. . . . 3
1.3 Dependent and independent regions of synthesized left view of Aloe image from center view in DIBR-3D representation 4
1.4 Basic encoding structure of 3D-HEVC encoder with inter component prediction 6
1.5 Rendering from a left camera position to a right camera position using depth maps [14] 7
1.6 Watermarking model 8

2.1 Motion Prediction of the Z-axis GOP 24
2.2 Semantic image segmentation 26
2.3 DTCWT Coefficients 27

3.1 Block diagram of 3D-HEVC codec using 2 view with inter component prediction 36
3.2 Left and right z axis extraction using depth value for Balloons video of camera view 2(left) and 4(right) 40
3.3 Motion Prediction of the Z-axis GOP 40
3.4 Connected regions of motion filtered Z-axis frame of Balloons video for GOP=8 41

LIST OF FIGURES

3.5 Selected blocks for embedding the watermark for each group of picture (GOP) after Z-axis filtration for balloons video (GOP=8) 42
3.6 Watermark embedding model (left eye video) 43
3.7 Watermark extraction model (left eye video) 45
3.8 PSNR comparison of the proposed scheme with Lee's scheme [65] for Balloons video 49
3.9 PSNR comparison of the proposed scheme with Lee's scheme [65] for Shark video 49
3.10 Flicker Metric comparison of the proposed scheme with Lee's scheme [65] for Balloons video (up to 100 frames) 50
3.11 Flicker Metric comparison of the proposed scheme with Lee's scheme [65] for Shark video (up to 100 frames) 50
3.12 BIR comparison of of the proposed scheme with Lee's scheme [65] for Balloons video 51
3.13 BIR comparison of the proposed scheme with Lee's scheme [65] for Shark video 51
3.14 Watermark embedding model 58
3.15 Watermark extraction model 61

4.1 Background of depth image detected using the main view image (kendo video frame) 70
4.2 Distribution of SIFT feature points in each group 71
4.3 Histogram of the depth block coefficients in first group for kendo video frame (a,b,c,d,e and f define the different blocks) 72
4.4 Watermark embedding model 74
4.5 Watermark extraction model 77
4.6 Foreground removal for balloons video frame number 17 82
4.7 Motion compensated connected pixels for balloons video frame number 17 (GOP 17-24) 83
4.8 Removal of foreground and unconnected pixels for balloons video frame number 17 (GOP 17-24) 83
4.9 Watermark embedding model 84
4.10 Watermark extraction model 88

5.1 Left and right dependent view region extraction using depth value for Aloe image of camera view 1(left view) and camera view 5(right view). 100

5.2 Dependent view layer partitioning using depth value for Aloe image of camera view 1(left view). 101

5.3 Watermark embedding model. 107

5.4 Rendering of left and right dependent view region to synthesized required view. 108

5.5 Watermark extraction model. 111

5.6 Watermarking zone shift. 112

5.7 (a) Original watermark, (b) (c) (d) Extracted watermark at JPEG compress level 50 from left view of Aloe image. 117

5.8 Hamming distance comparison of proposed scheme with existing schemes [55,62,76,101] for left view at Stereo-JPEG compression at quality 75. ($block_a = (I_w/8) \times (I_h/8)$, $block_b = (I_w/32) \times (I_h/32)$, $block_c$ = linearly embed) 118

5.9 Hamming distance comparison of proposed scheme with existing schemes [55,62,76,101] for right view at Stereo-JPEG compression at quality 75.($block_a = (I_w/8) \times (I_h/8)$, $block_b = (I_w/32) \times (I_h/32)$, $block_c$ = linearly embed and $block_d = 16 \times 256$) 119

5.10 Hamming distance comparison of proposed scheme with existing schemes [55,62,76,101] for synthesized center view at Stereo-JPEG compression at quality 75.($block_a = (I_w/8) \times (I_h/8)$, $block_b = (I_w/32) \times (I_h/32)$, $block_c$ = linearly embed and $block_d = 16 \times 256$) 120

5.11 Hamming distance comparison of proposed scheme with existing schemes [55, 62, 76, 101] against view synthesis attack at Stereo-JPEG compression at quality 75 (where 1 & 5 defines the left and the right views and others are the intermediate synthesized views). 121

5.12 Hamming distance comparison of proposed scheme with existing schemes [55,62,76,101] against Stereo-JPEG compression for right view. 122

5.13 Hamming distance comparison of proposed scheme with existing schemes [55, 62, 76, 101] against Gaussian noise attack at Stereo-JPEG compression at quality 75 for right view. 123
5.14 Hamming distance comparison of proposed scheme with existing schemes [55,62,76,101] against salt & pepper noise attack at Stereo-JPEG compression at quality 75 for right view. 124
5.15 Hamming distance comparison result after collusion attack on left image using the right image of the proposed scheme with Lin & Wu's scheme [55], Kim's scheme [76], Franco [62] and Rana's scheme [101] against Stereo-JPEG compression at quality 75 for right view ($block_a = (I_w/8) \times (I_h/8)$, $block_b = (I_w/32) \times (I_h/32)$, $block_c$ = linearly embed and $block_d = 16 \times 256$). 125

6.1 Dependent and independent regions of synthesized left view of Aloe image from center view 130
6.2 DTCWT Coefficients representation line features 132
6.3 Impact on original image due to change in DT-CWT wavelet coefficients [89] 133
6.4 Watermark embedding rule 137
6.5 Watermark embedding block diagram 139
6.6 Watermark extraction block diagram 141
6.7 PSNR comparison of the proposed scheme with existing schemes [55, 62, 76] for Middlebury Stereo 2006 Dataset [99] 143
6.8 SSIM comparison of the proposed scheme with existing schemes [55, 62, 76] for Middlebury Stereo 2006 Dataset [99] 143
6.9 VIFPcomparison of the proposed scheme with existing schemes [55, 62, 76] for Middlebury Stereo 2006 Dataset [99] 144
6.10 Hamming distance comparison of the proposed scheme with existing schemes [55, 62, 76] against view synthesis attack 146
6.11 Hamming distance comparison of the proposed scheme with existing schemes [55, 62, 76] against JPEG compression 147
6.12 Hamming distance comparison of the proposed scheme with existing schemes [55, 62, 76] against addition of Gaussian noise 148

6.13 Hamming distance comparison of the proposed scheme with existing schemes [55, 62, 76] against addition of salt & pepper noise . . 149
6.14 Hamming distance comparison of the proposed scheme with existing schemes [55, 62, 76] against baseline distance change attack . . 150
6.15 Hamming distance comparison of the proposed scheme with existing schemes [55, 62, 76] against blurring depth modification attack 151
6.16 Hamming distance comparison of the proposed scheme with existing schemes [55, 62, 76] against depth modification attack using DIBR technique 152

List of Algorithms

1 Watermark Embedding $(V_L, V_R, D_L, D_R, \alpha, W, MV)$ 44
2 Watermark Extraction $(W_{VL}, W_{VR}, D_L, D_R, MV)$ 47
3 Watermark Embedding $(V_L, V_R, D_L, D_R, \alpha, W)$ 59
4 Watermark Extraction (W_v) 61

5 Watermark Embedding (I, D, W) 74
6 Watermark Extraction (I, W_D) 76
7 Watermark Embedding (V_L, V_R, D_L, D_R, W) 85
8 Watermark Extraction (V', W_D) 87

9 Watermark Embedding $(I_L, I_R, D_L, D_R, \alpha, W)$. 106
10 Watermark Extraction (W_I, D_L, D_R). 110

11 Watermark Embedding (I, W). 138
12 Watermark Extraction (W_I). 140

List of Tables

2.1 Experimental Dataset. 33

3.1 Comparison of average PSNR, Flicker metric and BIR (of compressed video) of proposed scheme with existing scheme [65] . . . 48
3.2 Comparison of average hamming distance of extracted watermark from Balloons video 52
3.3 Average Hamming distance of extracted watermark for different *Quantization parameter* (QP) level of the proposed scheme 52
3.4 Average PSNR, SSIM, VIFp comparison result of the proposed scheme with Lee's scheme [65] and Rana's [100] 62
3.5 Hamming distance comparison result of the proposed scheme with Lee's scheme [65] and Rana's [100] for video sequences 63
3.6 Average hamming distance comparison result of left ans right view of the proposed scheme with Lee's scheme [65] and Rana's scheme [100] after collusion attack using DIBR technique 63

4.1 Average PSNR, SSIM comparison result of the proposed scheme with Guan's scheme [79] for depth 77
4.2 Average PSNRHVS, SSIM comparison result of the proposed scheme with Guan's scheme [79] for synthesized left and right view 78
4.3 Hamming distance comparison result of the proposed scheme with Guan's scheme [79] for image and video frame sequences against Compression and noise addition attack 78

4.4 Average hamming distance comparison result of the proposed scheme with Guan's scheme [79] at different baseline distance for image and video frame sequences 79
4.5 PSNR, SSIM comparison result of the proposed scheme with Guan's scheme [79] for depth 90
4.6 Average PSNR-HVS, SSIM comparison result of the proposed scheme with Guan's scheme [79] for left and right view 91
4.7 Hamming distance comparison result of the proposed scheme with Guan's scheme [79] against 3D-HEVC compression at different QP level 91
4.8 Hamming distance comparison result of the proposed scheme with Guan's scheme [79] at center view (view '0') and intermediate views in between center view and left & right views (view '±0.25') . . . 92
4.9 Hamming distance comparison result of the proposed scheme with Guan's scheme [79] at temporal scalable attack of frame dropping 93

5.1 Experimental set-up. 113
5.2 Visual quality measurement of the proposed scheme for different embedding block sizes. 115
5.3 Average PSNR and SSIM comparison result of the proposed scheme with Lin & Wu's scheme [55], Kim's scheme [76], Franco's scheme [62] and Rana's scheme [101]. 115

7.1 Summarization of the the proposed approaches. 157

List of Acronyms

3D-HEVC *3D-High efficient video compression*

AWGN additive white Gaussian noise

BIR bit increase rate

DCT discrete cosine transformation

DIBR depth image based rendering

DRM digital right management

DT-CWT dual tree complex wavelet transformation

DT-DWT dual tree discrete wavelet transformation

DWT discrete wavelet transformation

FFT fast fourier transform

GOP group of picture

HEVC *High efficient video compression*

HVS human visual system

IDCT inverse discrete cosine transformation

IMCDCT-TF DCT based inverse motion compensated temporal filtering (see §2)

JPEG *Joint photographic experts group*

MCDCT-TF DCT based motion compensated temporal filtering (see §2)

MCTF Motion compensated temporal filtering (see §2)

MSE mean square error

MSSSIM *Multi Scale Structural Similarity*

MVC *Multi-view video codec*

MVD *Multi-view video plus depth*

PSNR *Peak Signal-to-Noise Ratio*

PSNRHVS *Peak Signal-to-Noise Ratio taking into account Contrast Sensitivity Function*

QP *Quantization parameter*

RF Random Field

ROI region of interest

SIFT scale invariant feature transform

SIR stereo image representation

SSIM *Structural Similarity*

VIFp *Visual Information Fidelity, pixel domain*

List of Symbols

α	Embedding Strength/ Robustness threshold
δ	Grouping coefficients
γ	Grouping threshold
κ	Key
σ	SIFT scaling feature
τ	Threshold
ζ	Coefficient difference of DT-CWT (see §3.2.1)
θ	Angle in polar orientation
C	Original coefficients
C'	Watermarked coefficients
D	Depth
d_{isp}	Disparity
Ec	Embedding Capacity/ Payload
F	Frame
H	High pass coefficients of different transformation
H_{am}	Hamming distance
I	Image
I_h	Image height
I_w	Image width
L	Low pass coefficient of different transformation
m	Magnitude in polar orientation

mp_w	Watermark map
S_X	Shift of pixel X
T	Baseline Distance
V	Video
V_s	Synthesis video
V_d	Dependent view region
V_i	Independent view region
W	Original Watermark
W'	Extracted watermark
W_I	Watermarked image
W_V	Watermarked video
W_D	Watermarked depth
X_C	Center view pixels
X_L	Left view pixels
X_R	Right view pixels
Z**-axis**	Independent view region (refer to V_i)

Chapter 1

Introduction

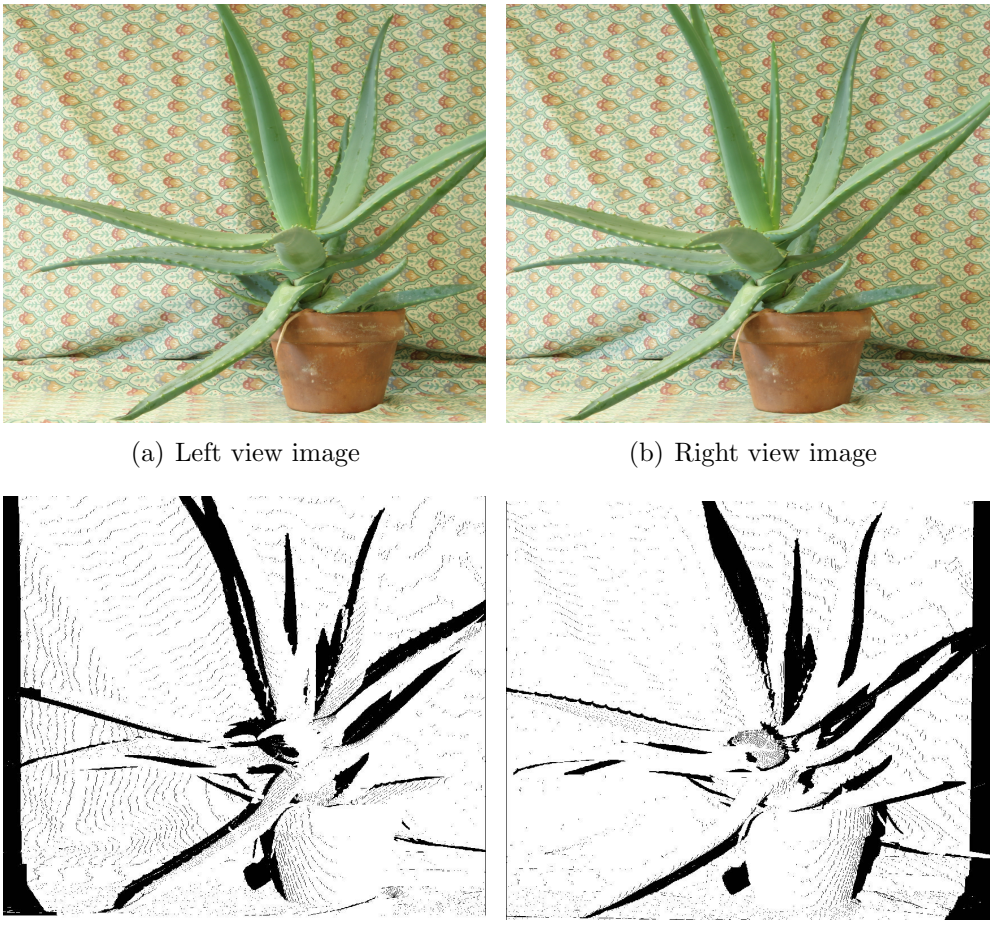
Recent improvements in media technology have led to growing interest in 3D media for its immersive experiences for viewers. The increased number of cinema screens capable of showing 3D movies, as well as the availability of low-cost 3D display devices, make 3D media transmission and distribution very popular. As a consequence, efficient content authentication or ownership authentication through watermarking for 3D images and video sequences is becoming an emerging research topic.

In scientific literature, depth image based rendering (DIBR)-3D image representation [1,2] has been introduced and it becomes popular due to its compression efficiency. In viewing 3D video, visual discomfort may occur [3] due to the unwanted horizontal and vertical parallax. Recently, another 3D representation, named depth based multi-view 3D [4, 5] using the DIBR [6] technique, becomes more popular due to its compression efficiency and better visual quality.

1.1 3D Image and Video Representation

3D is a combination of two views, one is the left eye view and other is the right eye view. The common view regions between left and the right view is called the dependent region of the 3D view or dependent view [7]. On the other hand, the uncommon region is called the independent view region, and it is used for

generating the 3D viewing experience to the respective left or right views. In literature, this independent view part is popularly known as the Z-axis. As shown in Fig. 1.1(a), 1.1(b) shows the left and the right views of the *Aloe* image and Fig. 1.1(c), 1.1(d) shows the dependent and the independent view of the left and the right view.



(a) Left view image (b) Right view image

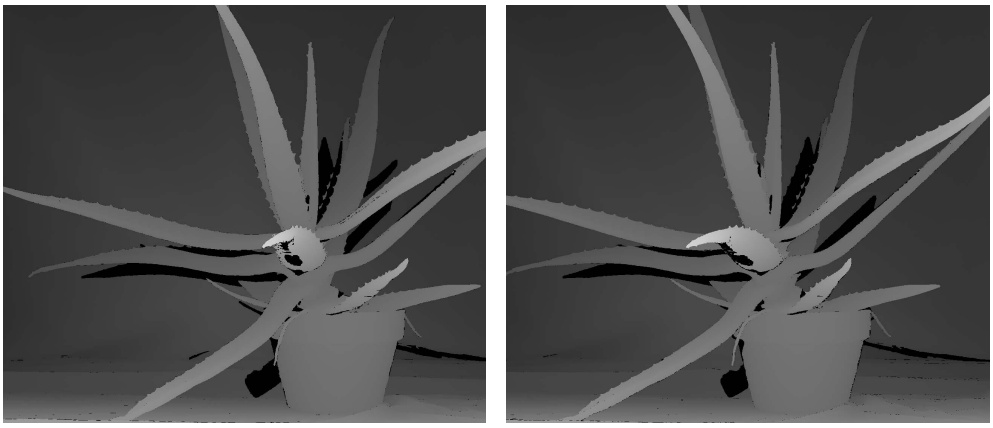
(c) Z-axis of left view image (d) Z-axis of view image

Figure 1.1: *Left and right view images with hidden pixels (say independent regions).*

1.1.1 Depth

In 3D image representation, depth is treated as an image containing the distance of the object with respect to the viewpoint. The depth represents a gray-scale

image, where the "0" represents the farthest point or the most distant point with respect to the view point in the scene and "255" represents the nearest point in the scene. The Fig. 1.2 shows the depth map of Fig. 1.1(a), 1.1(b) for *Aloe* image.



(a) Left video frame (b) Right video frame

Figure 1.2: *Left and right video frames with hidden pixels.*

1.1.2 DIBR-3D Representation

In this DIBR-3D [1,2,8] representation, only centre view and its depth map have been transmitted to the receiver end where left and right views are rendered with the help of centre view, its depth and corresponding disparity [1] according to the Eq. 1.1

$$\begin{aligned} X_L &= X_C + d_{isp}\, T_L\, \frac{D_{X_C}}{255} \\ X_R &= X_C - d_{isp}\, T_R\, \frac{D_{X_C}}{255} \end{aligned} \tag{1.1}$$

where d_{isp} is the disparity, D_{X_C} is the depth value of the center view for the location X_C and X_L & X_R are the corresponding shifted locations for the left and the right views respectively. T is the baseline distance between the reference view to the synthesis view. For the center view to the left and the right view, the baseline distance is taken as $T = 1/2$ for rendering from center view to left and right view.

The left and the right views are obtained by rendering the center view using the DIBR technique [8]. In this rendering process, dependent parts are fully obtained but the independent part is partially obtained from the center view (since the center view is the intermediate view of the left and the right view and independent region information is partially available). As a result, the unavailable regions creates holes in left and right views. In DIBR-3D views, the holes of the independent regions are filled up using the hole filling process [9], where the average values of the boundary pixels are used to fill the regions [1] as shown in Fig. 1.3. As the independent regions are less sensitive to human vision, the

(a) Center view image (b) Synthesized left view image

Figure 1.3: *Dependent and independent regions of synthesized left view of Aloe image from center view in DIBR-3D representation*

artefacts created using the hole filling technique may not degrade the quality with respect to the human visual perception. It is observed that, due to the hole filling technique, the independent regions gives a horizontal line pattern for the left and the right views of the DIBR-3D-image (refer to Fig. 1.3).

1.1.3 Multi-view Representation

Recently, the advance auto-stereoscopic display device has been introduce to produce immersion viewing experience without glass [10, 11]. The content im-

provement in the display device not only include the stereo view, it required a multiple number of views to make required display [12]. To make the surround position-invariant viewing experience, multi-view comes in the picture. The recent auto-stereoscopic display devices take more than 46 different views to provide 3D viewing experience. In case of video, these huge number of views which compressed using the *Multi-view video codec* (MVC), takes large storage space and huge bandwidth. The advance coding policy as *Multi-view video plus depth* (MVD) solves this problem [4,5,13] where few number of views (or boundary views) are transmitted with the depth. As the dependent view is common, the main view and the independent views of other views are used to represent the video stream to reduce the redundancy.

1.1.4 3D Video Encoder for MVD

Since for the 3D video encoding purpose, MVC extension of H.264/AVC is used to code a multiple number of views. But MVC is not appropriate for delivering 3D content for auto-stereoscopic displays.

A promising alternative is the transmission of 3D video in the MVD format [14]. In the MVD format, typically only a few views are actually coded, but each of them is associated with coded depth data, which represent the basic geometry of the captured video scene. Based on the transmitted video pictures and depth maps, additional views suitable for displaying 3D video content on auto-stereoscopic displays can be generated using DIBR techniques at the receiver side. *3D-High efficient video compression* (3D-HEVC) encoder comes to the picture to give support the MVD based encoder. Similar as for MVC, as shown Fig.1.4 all video pictures and depth maps that represent the video scene at the same time instant build an access unit and the access units of the input MVD signal are coded consecutively. Inside an access unit, the video picture of the so-called independent view is transmitted first directly followed by the associated depth map. Thereafter, the video pictures and depth maps of other

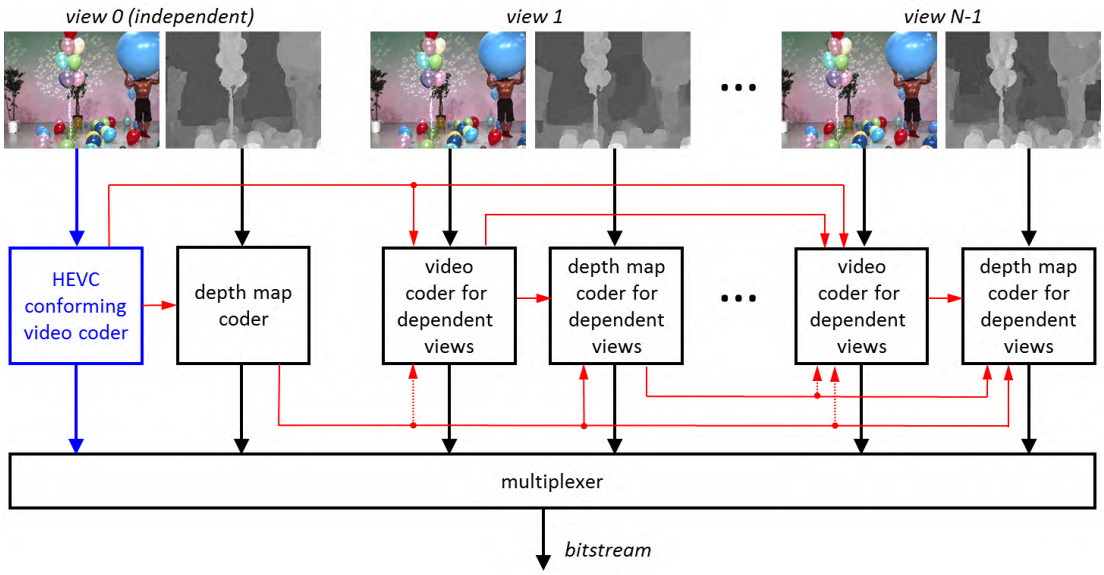


Figure 1.4: *Basic encoding structure of 3D-HEVC encoder with inter component prediction*

views are transmitted. A video picture is always directly followed by the associated depth map. In principle, each component signal is coded using an *High efficient video compression* (HEVC)-based coder. The corresponding bit-stream packets are multiplexed to form the 3D video bit-stream. The independent view is coded using a non-modified HEVC coder. The corresponding sub-bit-stream can be extracted from the 3D bit-stream, decoded with an HEVC decoder, and displayed on a conventional 2D display. The other components are coded using modified HEVC coders, which are extended by including additional coding tools and inter-component prediction techniques that employ already coded data inside the same access unit. For enabling an optional discarding of depth data from the bit-stream, e.g., for decoding a two-view video suitable for conventional stereo displays, the inter-component prediction can be configured in a way that video pictures can be decoded independently of the depth data.

View Synthesis

As an optional encoding technique, a mechanism is integrated by which regions independent views that can be rendered based on the transmitted independent view and the associated depth maps are identified. These regions are encoded by employing a modified cost measure, which mainly considers the required bit rates. After decoding, the renderable regions can be identified in the same way as in the encoder and replaced by rendered versions. The encoder identifies regions in the current picture that can be rendered from pictures of the same time instance in a reference view based on the reconstructed depth maps of the reference view. The rendering technique has been depicted in Fig. 1.5.

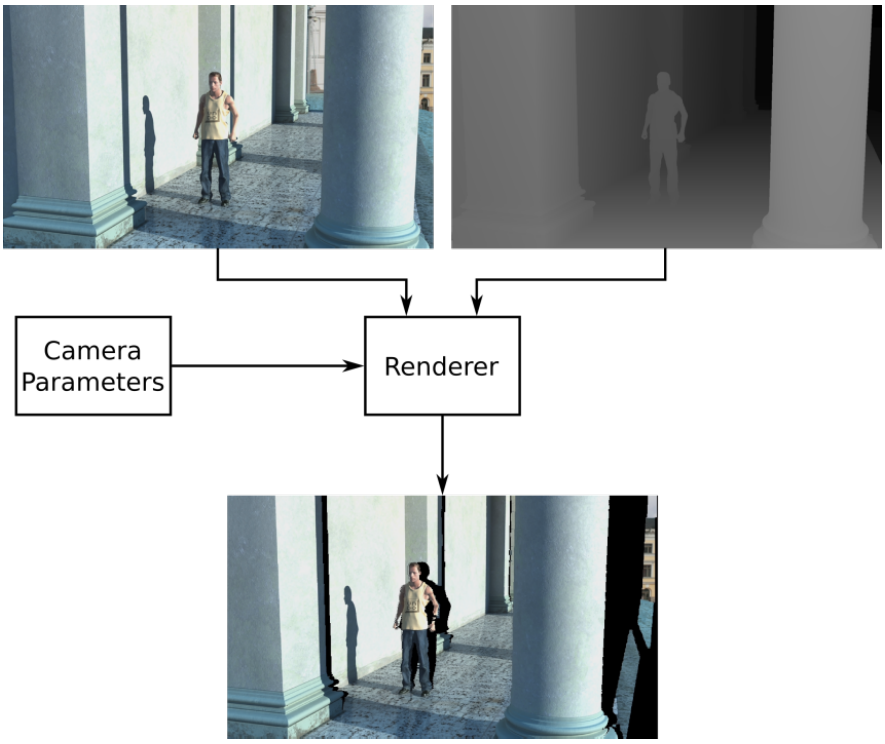


Figure 1.5: *Rendering from a left camera position to a right camera position using depth maps [14]*

1.2 Digital Watermarking

In digital image and video watermarking, a signature is inserted in the original content such a way that the signature can be extracted and the ownership of the image or the video sequence can be authenticated as shown in Fig. 1.6.

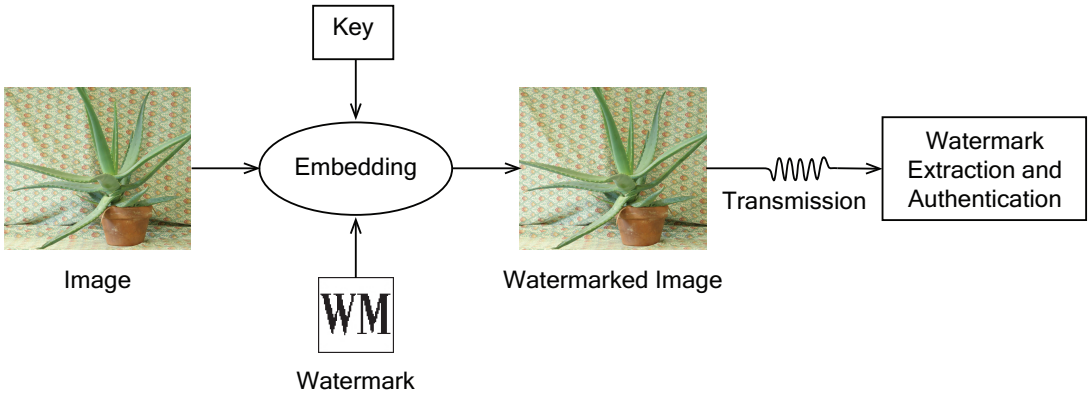


Figure 1.6: *Watermarking model*

The watermarking in image in video can be used for various security purposed like ownership authentication [15], video authentication [16], broadcast monitoring [17], trailer tracking [18] etc.

1.2.1 Efficiency Parameters

There are many parameters to check the efficiency of a watermarking scheme or some times used to categories them. Some time they conflicting to each other like payload may reduce the imperceptibility. Few important parameters described below [19–21].

1. **Robustness :** How efficiently watermark withstands against intentional and unintentional attacks. Depending on that a watermark it goes robust , fragile and semi-fragile.

2. **Imperceptibility :** The watermark should not noticeably degrade the video quality.

3. **Payload :** Measures the number of bits or the size of the watermark embedded.

4. **Blindness :** A watermarking scheme called blind if original video is not needed at extraction.

5. **Bit Increase Rate :** Often bit rate gets increased due to embedding. A good methodology does not increase bit increase rate (BIR) significantly.

1.3 Literature Survey

There are lots of robust blind watermarking schemes for 2D images [22–29] and 2D videos [30–40] to resist geometric modification attacks. But, implementation of 2D watermarking scheme over 3D image and video sequences might not give proper security due to the use of depth image based rendering (DIBR) [6] in different 3D representation.

1.3.1 3D Image and Video Watermarking

In recent literature, various digital watermarking schemes for stereo image representation (SIR) or 3D images [41–60] and and videos [57, 61–74] have been presented. Most of the schemes are implemented to secure the 3D content of the media by using the 3D characteristics. But those schemes are not very useful for DIBR [6] based rendering method. In other words, the main challenge is to embed the watermark in such a way that it resist the view synthesis attack and the 3D image and video compression methods.

1.3.1.1 3D Image Watermarking

In recent literature, Campisi [52] first introduced the concept of a 3D image watermarking technique using depth, where author argues that the watermarking scheme is semi-fragile to *Joint photographic experts group* (JPEG) and JPEG2000 compression. But, in the scheme, the author does not consider the dependent and

independent view regions of a 3D image. Being semi-fragile, the scheme could not resist the DIBR technique. Later, Halici and Alatan proposed a watermarking scheme [53] where the watermark is embedded with the middle view, which is synthesized from the left and right views by DIBR technique. The main problem in this scheme is that the watermark may get distorted if it is extracted from the left or right view at the receiver end due to the presence of the independent view region and the partial presence of the watermarked center view. So, the middle view has to be regenerated during watermark extraction. Thus, the scheme is not resistant to the synthesis view attack. Li et al. proposed a region of interest (ROI)-based watermarking scheme [54] and located the ROI using both depth and gray images. The watermark is embedded on both views of the selected locations. Those authors argued that the scheme is robust against JPEG compression attack. In Li's scheme [54], since the ROI locations contain portions of both dependent and independent view regions, the watermark may not survive the synthesis view attack. Moreover, the dependent view regions of both left and right views may contain different watermarks, which may be vulnerable to collusion attack [75]. Recently, Subramanyam et al. in [45] proposed a compressed domain image watermarking scheme. Later, Wang et al. proposed a discrete wavelet transformation (DWT) based watermarking scheme in [46] for image quality enhancement. Korus et al. in [47] proposed an image watermarking scheme for fast embedding. But none of these schemes [45–47] have considered the DIBR image representation. As a result, these schemes [45–47] may not resist the view synthesis attack for the DIBR-3D image. Recently, Lin and Wu [55] proposed a blind multiple watermarking scheme where the watermark is embedded in the left or right view image obtained by inverse rendering of the DIBR technique. In this scheme, since the embedding block size is 8×8 or 16×16, it may not withstand a synthesis view attack against high resolution (having high disparity) images because to resist a synthesis view attack, the block size should be greater than half of the disparity value between the left and right view images, which is in general around 200 pixels in the case of any natural image. In a similar direc-

tion, Jaipuria proposed a blind DWT-based watermarking scheme for DIBR-3D images [56] where the authors embed the watermark with the center view image obtained by a reverse rendering process as presented in [55]. Like the Lin and Wu's scheme [55] this scheme may not sustain a synthesis view attack because the block length for embedding is less than that of disparity. In another recent work, Kim et al. proposed a dual tree complex wavelet transformation (DT-CWT) base watermarking for 3D images [76]. In this scheme, 15° & -15° (1^{st}, 6^{th}) and 45° & -45° (2^{nd}, 5^{th}) coefficients of the 2D-DT-CWT of the center view image is used for embedding, and the authors have shown that embedding in these coefficients makes the scheme more robust because these coefficients are relatively more robust against the DIBR-based view synthesis process. This scheme may also suffer from a synthesis view attack for high-resolution images or images where the disparity is larger than the block size. Furthermore, this scheme may be vulnerable to collusion [75] attack because different watermark patterns are practically embedded within the visually similar left, right, or any other synthesized view. Later, Asikuzzaman et al. [59] extended the Kim et al. scheme [76] to video by incorporating the DT-CWT with the same embedding scheme and block size of 8×8 pixels. However, the scheme [59] might not improve the previous scheme [76] against high-resolution images and video sequences or images having disparity higher than the embedding block size because of the smaller block size. Yonggang proposed a robust image watermarking scheme [60] for the DIBR technique using 3D-discrete cosine transformation (DCT). However, this scheme may also not be robust against the view synthesis process of the DIBR scheme because the authors did not consider the 3D characteristics of the image at time of watermark embedding. Trick et al. [57] proposed a context-based 3D model watermarking technique that does not consider DIBR-based rendering at the time of embedding the watermark. As a result, the locations of the watermark may change (or may not be available) due to DIBR-based view generation, and the scheme may not be robust against a view synthesis attack. Arun and Paul proposed a 3D image watermarking scheme [58] for DIBR 3D images: In

this scheme, after the first level of DWT, the middle range frequencies are used to perform the second and third levels of DWT, respectively, and the 4×4 block DCT is used with the mid-range frequencies to make the embedded watermark imperceptible to human vision. It is observed that the embedding block size for the scheme [58] is 32×32 [4×4 for DCT and 8×8 for 3^{rd} level of DWT] for injecting a single watermark bit. This block size may not be sufficient to handle a disparity higher than 32 pixels (in a real scenario, the disparity is generally more than 200 pixels) and may be vulnerable to the view synthesis process.

1.3.1.2 3D Video Watermarking

In 3D video, possibly Louizis et al. [63] first proposed a 3D watermarking scheme where the bit sequences are inserted with in the luma component by altering the voxels of the host volume. Later, Tefas et al. [77] proposed an improve version of the previous scheme by embedding the additive watermark with the luma component of the views. But the motion compensation part of the video compression is not analysed in time of watermark insertion. As a result the schemes are fragile against video compression attack. Moreover in the watermarking schemes, 3D geometry part has not been considered which reduces the sustainability of the schemes [63, 77] in 3D representation. In 3D video compression, the dependent view regions does not occur in the other views to improve the efficiency. In the existing 3D watermarking schemes, watermark is inserted either in both the views [62, 65] or in the synthesized center view [74, 76]. In DIBR technique, it is observed that only one full view (left or right having both dependent and independent parts) along with independent part of other view(s) are communicated to the receiver side. So the watermark embedded in the other views (where only independent parts are encoded) may not be properly extracted. Moreover, in case of MVD [13, 14] based encoding, the view synthesis process using depth (say view synthesis attack) itself may destroy the watermark. In that case, a compromised receiver can generate a synthesized view from watermarked left and right views. Thus the existing 3D video watermarking scheme is not very suitable for handling

the DIBR [6] technique used in MVD [13, 14] based encoding.

Using the 3D characteristics, possibly first Ntalianis et al. [64] proposed a 3D video watermarking scheme. The authors used an object based watermarking scheme where the watermark embedded in each frame of the left and right eye video with a selected object from the video sequence to embed the 3D characteristics. But the watermark will not sustain due to motion compensation in time of compression because the real motion part in time of video compression is not categorised. Moreover, using of same objects in the left and the right view does not represent the dependent view regions. As a result, not using of the depth information to select object with common regions, the watermarking scheme [64] does not qualify for the 3D characteristics and remains unsustainable against 3D-HEVC compression attack. Later Wu et al. [66] proposed a 3D watermarking scheme where the uncommon regions of left and the right view video has been watermarked separately and the common regions of the 'I' frame has been watermarked separately. To insert the watermark, authors have used a block-wise DCT block matching to filter the common regions of the left and the right views. Also the watermark of the left and the right view is opposite of each other. Hence it seems to be a fragile watermarking scheme with respect to the view synthesis attack. Also 'B' and 'P' frames are not used dependent view region detection and watermark embedding. SO using reverse motion estimation the watermark can be easily removed. Moreover, DCT based block matching does not give the efficiency of the DIBR. So the scheme [66] does not sustain in the environment of MVD based encoder like 3D-HEVC. Chen and Huang proposed two different watermarking scheme in [67, 68] for 3D video where they used mean square error (MSE) to filter the common regions from the left and right eye views to insert the watermark. In natural 3D views, disparity may vary from 0 to 200 or more. So, using of MSE to select the dependent view regions goes much complex. Also there is a chance of getting dissimilar regions using MSE because the depth is not considered. As a result selected regions may overlapped in interviews, and the watermark bit can be loosen in time of embedding process. Furthermore, not us-

ing of DIBR based view selection method, the watermarking schemes [67,68] will not sustain for 3D-HEVC encoding. Lee et al. proposed a perceptual depth based 3D video watermarking scheme in [65] where the watermark is embedded with the hidden pixels and high motion Z-axis. Where the Z-axis is the uncommon regions extracted from the left and right eye video using DIBR [6]. Since the watermark is embedded by altering the Z axis only, the rendering process (DIBR) is not hampered due to embedding. But, since the watermark region is not synchronized with the continuous frames in the GOP, watermark bits may be destroyed at the time of motion compensation. Moreover, the scheme may cause bit rate increase due to embedding. Later Asikuzzaman et al. proposed a DT-CWT based 3D video watermarking scheme [73] by extending the Kim's scheme [76]. Where the authors embed the watermarking in the 3D level 2D-DT-CWT coefficients of the 2^{nd}, 5^{th} of the center view. The authors [73] claimed that the embedding of watermark in the chroma component will increase the robustness and improve the visual quality. Like the Kim's scheme [76], this scheme [73] may also suffer from a synthesis view attack for high-resolution images or images where the disparity is larger than the block size. Also using collusion attack [75] the scheme may remain un-sustainable if implementing is done on left and right views. Very recently an improvement has been done by the same authors in [74] where the same embedding policy is used for insertion of the watermark. Moreover, these schemes [73,74] are implemented in frame by frame manner and does not used the motion compensation part in the video sequence for insertion of the watermark. So, the schemes does not resist the motion based quantization part in the video compression. As a result, the schemes [73,74] may not sustain against 3D-HEVC compression attack as well as different 3D attacks.

1.3.2 3D Depth Watermarking

In recent trends, depth is an important part of 3D image and video sequences. Also for rendering purpose depth takes an important role to create 3D video

sequence in display devices. Capitalizing this characteristics of 3D viewing, unsecured depth parameter can be used to generate more distorted and illegitimate synthesized left and right views which may reduce the robustness of the 3D image and video watermarking schemes [41–57, 57–74]. So securing the depth of the 3D video content is an important requirement to ensure the media security.

To secure the depth of the 3D content, direct implementation of image and video watermarking schemes on depth part is not suitable because depth image is in general has low variance. So addition of watermark can be treated as a high frequency noise in the depth sequence and can be easily removed by smoothing attack [78]. Also most of the existing schemes [55, 65, 73, 74, 76] are not fully resilient to the DIBR process and collusion attack [75]. So the watermark can be easily destroyed by colluding two views. In depth based watermarking, Guan et al. proposed a blind multiple watermarking schemes for DIBR-3D videos in [79], where only the depth region is watermarked to secure the original image with depth. As depth is lesser important part than the original image, different smoothing attack can be used to destroy the watermarking scheme. Also, object based smoothing can be used to remove the watermark from the depth of the 3D image [78].

1.4 Motivation and Objectives

From the preceding discussion, it is obvious that most of the watermarking schemes are vulnerable to the 3D characteristics in image and video sequences. It is also observed in DIBR-based watermarking that the different watermark signals may be embedded in coherent locations of the left and right view images. This can be capitalized to mount a watermark estimation attack, popularly known as a collusion attack (type II) [75]. In this attack, the watermark signal can be removed by simply averaging the coherent image regions with different watermarks. So, it reveals that DIBR based 3D image/video watermarking schemes are vulnerable against collusion attacks. Additionally, it is also observed in the literature that

existing schemes are not performing well against different 3D domain attacks such as 3D video compression: 3D-HEVC, DIBR & multi-view image representation, view synthesis, depth blurring, baseline distance change, view collusion, depth collusion etc. and different image processing attacks. At the same time, maintaining acceptable visual quality of the embedded image/video is a challenging task [3]. Motivated by these issues, the main objective of this work is to enhance the robustness of the watermarking scheme for the 3D image and video while maintaining the decent visual quality of the watermarked sequences with respect to human visual system (HVS). The work has been carried out in the following ways:

1. Proposing robust watermarking for dependent and independent views of 3D image and video sequences against different 3D attacks.

2. Proposing robust watermarking for depth of the 3D image and video sequences to resist different depth based attacks [76, 80].

3. Maintaining the decent visual quality of the watermarked image and video sequences with respect to the HVS by using proper coefficient selection methods.

1.5 Contributions of the Thesis

1.5.1 Robust Watermarking for MVD Representation against 3D-HEVC Compression and View Synthesis Attack

In first chapter of this thesis, robust watermarking against 3D-HEVC compression and view synthesis attack is considered. There are two major contributions in this chapter, firstly a Z-axis based watermarking against 3D-HEVC compression and secondly, the scheme is extended to resist view synthesis attack.

1.5.1.1 Robust Watermarking for MVD Representation against 3D-HEVC Compression Attack

In the first phase of this work, a video watermarking scheme is proposed to resist the 3D-HEVC compression attack where the independent region (say Z-axis) is used for insertion of the watermark. In this scenario, the independent region does not blend in the time of 3D-HEVC compression (basically in the rendering part as described in previous) and the use of different watermark sequence for different views, improves the security. Using of DCT based motion compensated temporal filtering (MCDCT-TF) and DCT for embedding, robustness against 3D-HEVC has been enforced. The sustainability of the watermark signal is marked by outperforming the existing related schemes using an experimental set-up.

1.5.1.2 Watermarking for MVD Representation against 3D-HEVC Compression and View Synthesis Attack

In the next phase of this work, the proposed scheme is extended to resist view synthesis attack along with the 3D-HEVC video compression attack by employing DT-DWT based embedding. Center view is rendered from the left and right views of a 3D video frame to find DIBR [6] invariant coefficients. The temporal redundancy in the time of compression is reduced by using the modified MCDCT-TF over the center view. Shift invariant 2D DT-DWT is used for embedding purpose of reducing the spatial redundancy due to the view synthesis process using DIBR technique. A comprehensive set of experiments has been carried out to quantify the applicability over existing schemes with respect to compression of the 3D-HEVC video codec and synthesis view attack.

1.5.2 Watermarking in Depth Information of 3D Image and Video Sequences against Depth Attack

Depth is a crucial information in 3D representation and must be secured. In this chapter, two watermarking schemes are proposed. First is to secure depth information of the DIBR-3D image representation and second is the corresponding

extension for the MVD based 3D video sequence.

1.5.2.1 Watermarking in Image Depth against View Synthesis Attack

In the first phase, depth of the 3D image sequences is used to insert the watermark bit such that the embedding scheme can resist view synthesis attack for DIBR-3D representation. The scale invariant feature transform (SIFT) [81] is used to find the view invariant feature point locations from the main views and the watermark bits are inserted in the depth of the obtained locations to make the scheme invariant to the view synthesis process. In the time of location selection, foreground region (say more salient to the human vision) is removed to obtained satisfiable visual quality in the main view with respect to the HVS. A set of experiments is carried out to justify the robustness of the proposed scheme than the existing scheme.

1.5.2.2 Watermarking in Video Depth Sequences against Depth Modification and 3D-HEVC Compression Attack

In the second phase, an extended version of the first phase work is presented to secure the depth of the 3D video sequences in MVD representation. To sustain against 3D-HEVC compression attack, modified version of MCDCT-TF is used to find the temporal low-pass frames. In this embedding policy, watermark signal transmitted to the original views, that improves the sustainability of the watermark signal after different depth modification attacks. Using an experimental set-up the claimed robustness is justified by comparing with the related existing schemes.

1.5.3 Depth-based View Invariant blind 3D Image Watermarking for Multi-view Representation

In this work, a watermarking scheme for Multi-view Image Representation is presented to resist different 3D attacks on multi-view representation and image processing attacks. The horizontally shifted spatially coherent dependent view

regions of the left and right views are used for embedding purpose of resisting the view synthesis attack. Using of specific block selection method, reduce the spatial de-synchronization in the synthesis views as well as improve the robustness of the embedding policy. Watermark is inserted with the low-pass filtered dependent view region of 3D images using Block DCT to increase the strength of the watermark against different attacks. The applicability of the proposed scheme is experimentally established by a set of comparison with the existing schemes and with respect to view synthesis, Stereo JPEG compression and different noise addition attacks.

1.5.4 View Invariant Watermarking Using DT-CWT for DIBR-3D Image Representation

In the final phase of contribution, a watermarking scheme is proposed for DIBR-3D images to resist the different 3D attacks (like: depth modification, depth blurring, baseline distance change) using DT-CWT. 2D-DT-CWT coefficients of centre view are used for watermark embedding such that shift invariance and directional property of the DT-CWT can be exploited to make the scheme robust against view synthesis process. A specific coefficient selection scheme is presented to separate independent regions (or distorted regions due to hole filling) from the main views by checking the DT-CWT coefficient checking to improve the robustness. A comprehensive set of experiments has been carried out to justify the robustness of the proposed scheme over the related existing schemes with respect to the JPEG compression and synthesis view attack.

1.6 Thesis Organization

This Ph.D. dissertation is elaborated in seven chapters. In this first chapter, a brief introduction of 3D image and video representation and the encoding policy is explained with a brief literature of watermarking & their shortcomings, the motivation & objective of the thesis and the contribution of the thesis. The rest

of the organization is as follow:

- §2: In this chapter, a description is given on the background of the research which includes some preliminary concepts like MCDCT-TF, SIFT, DT-CWT etc., evaluation metrics, experimental data set etc. to use in following chapters.

- §3: In this chapter, a robust 3D video watermarking scheme is proposed to resist the rendering and video compression attack in time of 3D-HEVC encoding process. Then another DT-DWT based watermarking scheme is presented for MVD to resist the view synthesis as well as the 3D-HEVC compression attack.

- §4: In the first work of this chapter, depth of the 3D image is watermarked to make secure media transmission and to tesist the view synthesis attack. Then the scheme is extended for the 3D video sequences in the second phase to resist 3D-HEVC compression, view synthesis and different depth modification attacks.

- §5: In this chapter, a watermarking scheme is proposed to resist the different 3D image attacks in Multi-view Image Representation. Here the watermark is inserted with the DIBR invariant coefficients to resist view synthesis attack.

- §6: In this chapter, a watermarking scheme for DIBR-3D image representation is presented to resist the different 3D image attacks in DIBR-3D image representation. A novel watermark embedding scheme using the DT-CWT is presented the to resist the view synthesis attack.

- §7: A brief conclusion of the Ph.D. dissertation is presented in this chapter with the future scope of research directions.

1.7 Summary

In this chapter a brief introduction is presented over the 3D research domain to formulate the scope of research in this field. The representation of the 3D image and video sequences is explained with a brief literature of 3D image and video watermarking with the limitations of the existing schemes. Based on the limitations, motivation and the objective of the research work is formulated. Finally a brief description of contribution and the organization of the thesis is presented.

Chapter 2

Research Background

In this chapter, a brief overview of the mathematical preliminaries and theoretical foundation related to the research topic are presented. A brief discussion on MCDCT-TF, DT-CWT and SIFT is included along with the different evaluation parameters and the dataset used for experimental purpose.

2.1 Motion Compensated DCT Temporal Filtering (MCDCT-TF)

Motion compensated temporal filtering (MCTF) [82–84] is used to filter video sequences in temporal i.e. the motion direction. All the schemes are implemented using 2-tap Haar wavelet filter. Later, Atta et al. proposed the DCT based MCTF for temporal scalable video coding [85]. In [85], authors have proposed 9×1 MCDCT-TF by combination of 3×2 and 2×1 temporal filtration. Recently we have proposed an modified version of [85] using 3×1 in [86]. In [86], the motion compensation has been done to generate the middle frame from the boundary frames as proposed by Atta et al. [85] and 3×1 DCT has been carried out to filter the low pass temporal frame. But in this scenario, it is observed that the filtration can be done with a frames multiple of 3. Also if some de-synchronization is made, then the center view location may deviate and the same location might not be obtained. So a little modification is done on the motion compensation

model [86] as follow:

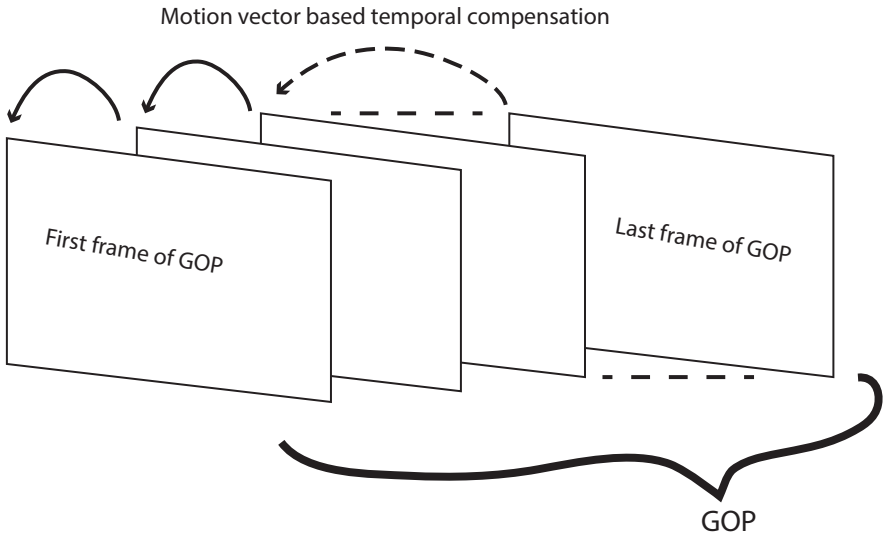


Figure 2.1: *Motion Prediction of the Z-axis GOP*

- To avoid the temporal adaptation, the main filterable frame is set to the first frame because the first frame is the I frame of the GOP and always present.
- Motion compensation is done progressively as shown in Fig. 2.1 and only the connected pixels are used [86]
- $F \times 1$ DCT filtration is carried out to extract the low pass temporal to filter frame from F number of frames as on Eq. 2.1

$$\begin{bmatrix} L(x,y) \\ H_1(x,y) \\ . \\ . \\ . \\ H_{F-1}(x,y) \end{bmatrix} = A_F \times \begin{bmatrix} I_1^1(x,y) \\ I_2^1(x,y) \\ . \\ . \\ . \\ I_F^1(x,y) \end{bmatrix} \tag{2.1}$$

where L is the low pass DC coefficient and H is the AC coefficients at location (x,y) for the motion compensated pixels $[I_1^1, I_2^1 ... I_F^1]$.

To generate the video sequence DCT based inverse motion compensated temporal filtering (IMCDCT-TF) is done over the temporal filtered video sequence.

Inverse Motion Compensated DCT Temporal Filtering (IMCDCT-TF)

IMCDCT-TF is the reverse process of MCDCT-TF. To obtain the motion compensated frame sequence $F \times 1$ inverse discrete cosine transformation (IDCT) is done on the filtered sequence as shown in Eq. 2.2

$$\begin{bmatrix} I_1'^1(x,y) \\ I_2'^1(x,y) \\ . \\ . \\ . \\ I_F'^1(x,y) \end{bmatrix} = A_F^T \times \begin{bmatrix} L(x,y) \\ H_1(x,y) \\ . \\ . \\ . \\ H_{F-1}(x,y) \end{bmatrix} \tag{2.2}$$

where $[I_1'^1, I_2'^1 ... I_F'^1]$ are the motion compensated frame sequence obtained after $F \times 1$ IDCT. To obtain the original frame sequence inverse motion compensation is done and the connected pixels are replaced in the original frame sequences as shown in Eq. 2.3

$$\begin{aligned} I_n(x,y) &= I_n'(x,y) \quad \textit{where connected pixels} \\ &= I_n(x,y) \quad \textit{where unconnected pixels} \end{aligned} \qquad \textit{for } n = 1 : F \tag{2.3}$$

where I_n' is the inverse motion compensated frame sequences obtained from IDCT and I_n' is the original frame sequences.

2.2 Semantic Image Segmentation

Semantic image segmentation is a procedure to label each pixel of an image. Usually an identifiable objects are more salient to human visual system. Using of semantic image segmentation, foreground objects are also filtered, which are more salient to human visual systems. In this scenario, combination of convolutional neural networks (CNNs) and conditional random fields (CRFs)-based probabilistic graphical modelling is used for semantic image segmentation [87]. The architecture of the segmentation process is depicted in Fig. 2.2. The FCN-8s

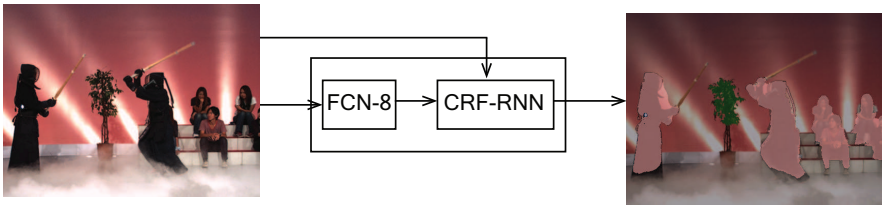


Figure 2.2: *Semantic image segmentation*

architecture [88] is used as the first part to unary potentials for the CRF. CRF-based probabilistic graphical modeling for structured prediction is processed with the pixel level prediction and semantic segmentation task is done by the end to end trainable deep network by combining a fully convolutional neural network with the CRF-RNN.

2.3 Scale Invariant Feature Transform

SIFT [81] is an algorithm to extract distinctive feature which enables the correct match for a key-point to be selected from a large database of other key point. It is observed that the SIFT feature is substantially robust against affine transformation, noise addition, and change in illumination. Due to its sustainability characteristics, SIFT feature can be used for matching key points in different views. Applying SIFT on an image gives feature point descriptor with the location (x, y), Scale feature (σ) and the orientation angle (θ)

The scale feature (σ) of the SIFT remain unchanged for the same size of the image. So, for same sift point the (σ) will remain constant for a same scale image.

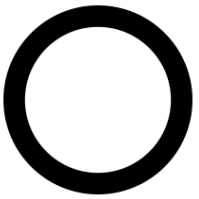
2.4 Dual Tree Complex Wavelet Transform (DT-CWT)

Kingsbury [89, 90] first presented the dual tree complex wavelet transformation (DT-CWT) to analyse it's shift invariant property. In dual tree discrete wavelet transformation (DT-DWT), two independent discrete wavelet transformations

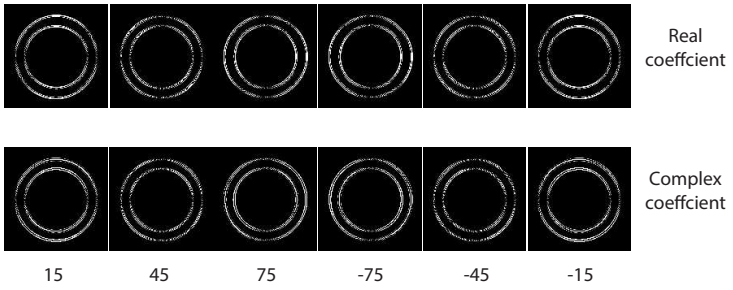
(DWTs) are exploited for each tree. Two real dual tree wavelet coefficient (where the second transform is done by picking the opposite samples of the first one) are combined to generate the complex wavelet coefficients as shown as

$$h_a + ih_b$$

where h_a and h_b are the real coefficients of the tree 'a' and tree 'b' as explained above for DT-CWT over$x$.



(a) Original Image



(b) coefficients of DT-CWT

Figure 2.3: *DTCWT Coefficients*

According to Kingsbury [89–91], it is observed that the DT-CWT is invariant to small shift or geometric distortion. Also, it is observed in the literature [76] that the *Peak Signal-to-Noise Ratio* (PSNR) between center view and left (or right) view, is relatively better when DT-CWT is used in comparison with DCT or fast fourier transform (FFT). Moreover, it is experimentally observed that the change of the magnitude of the DT-CWT coefficients due to the DIBR is very less. So the value of DT-CWT coefficients will remain constant than DCT of FFT coefficients against DIBR based view synthesis process. The six high

pass DT-CWT coefficients are represented as a combination of real and complex coefficients at orientation angle 15°, 45°, 75°, −75°, −45°, −15° (let H_1, H_2, H_3, H_4, H_5, H_6). Fig. 2.3(b) shows the of DT-CWT coefficients for the image 2.3(a).

2.5 Attacks on 3D Image and Video

For testing the strength of watermarking applied on the 3D image and video sequences, some attacks are carried out as follows:

- **View synthesis:** In this attack, DIBR technique [6] using the depth is applied to generate synthesis views to check the embedding strength. In this attack, pixels moves horizontally and the embedding position gets changed.
- **3D-HEVC video compression:** It is an attack incorporated by compressing the video at the different QP level to check the strength of the inserted watermark in the raw 3D video sequences using the MVD [13] based 3D-HEVC [14] compression technique.
- **Collusion attack:** It is used to collude the similar views with a different watermark to estimate the watermark [75]. For 3D, collusion attack is carried for interview sequences where using the DIBR [6] technique the dependent regions are synthesized to targeted views and colluded [75] to remove the embedded watermark.
- **Noise-addition attack:** It is mounted by adding a noise with the image sequences with a random change in brightness and the colour information [92]. In salt and pepper noise, white and black pixels of certain probability density are added randomly with image. In additive white Gaussian noise (AWGN), Gaussian noise of a certain variance is added with the images.
- **Baseline distance change:** It is an attack for DIBR-3D representation where the synchronization of the depth and the view is shifted to to a

certain baseline distance to reduce the embedding strength with very less degradation of the visual quality [76].

- **Depth blurring:** In this attack, the depth of the 3D image is blurred, so that the embedding locations in the synthesized left and right views got changed.

- **Depth modification:** In this attack, depth image is modified so that the location of the pixels in left and the right view got changes and the embedding location got changed.

2.6 Evaluation Parameter

Embedded image and video sequences are are evaluated by checking the visual degradation and the embedding strength by analysing different visual quality parameters and robustness parameter.

2.6.1 Visual Quality Parameter

In this work, PSNR, *Structural Similarity* (SSIM) [93], *Visual Information Fidelity, pixel domain* (VIFp) [94] and Flicker Metric [95, 96] are used to measure the image and the video quality to check the degradation of the quality after watermark embedding. *Peak Signal-to-Noise Ratio taking into account Contrast Sensitivity Function* (PSNRHVS) [97] and *Multi Scale Structural Similarity* (MSSSIM) [98] are used to measure the degradation in of video quality with respect to human perception, after embedding watermark in depth map.

2.6.1.1 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 Eq. 2.4

$$\begin{aligned} &PSNR = 10\log_{10}\frac{Max^2}{MSE} \\ &where \\ &MSE = \frac{\sum_{i=1}^{I_h}\sum_{j=1}^{I_w}(I(i,j)-W_I(i,j))^2}{I_h.I_w} \end{aligned} \tag{2.4}$$

where Max is maximum possible intensity (255), I_H and I_W are width and height of the frame/image respectively, I is the original image and W_I is modified image.

2.6.1.2 SSIM

SSIM index is an image quality metric. It is function of three components eg. luminance similarity ($l(I, W_I)$), contrast similarity ($c(I, W_I)$) and structural similarity ($s(I, W_I)$) as given in Eq. 2.5

$$SSIM(I, W_I) = [l(I, W_I)]^{\alpha}.[c(I, W_I)]^{\beta}.[s(I, W_I)]^{\gamma} \tag{2.5}$$

where α, β and γ are parameters used to adjust the relative importance of the three components [93].

2.6.1.3 PSNRHVS

In time of computing PSNR, HVS is taken in to the account for calculating PSNRHVS. The HVS version of PSNR is calculated using the following Eq. 2.6

$$PSNRHVS = 10\log_{10}\frac{Max^2}{MSE_{HVS}} \tag{2.6}$$

MSE_{HVS} is calculated taking account HVS as shown in following Eq. 2.7

$$MSE_{HVS} = \frac{\sum_{i=1}^{I_h-7}\sum_{j=1}^{I_w-7}\sum_{m=1}^{8}\sum_{n=1}^{8}(I(i,j)_{ij} - W_I(i,j)_{ij})^2}{(I_h-7).(I_w-7).64} \tag{2.7}$$

where C_{ij} denotes the DCT coefficients of 8×8 block [97].

2.6.1.4 MSSSIM

In MSSSIM, multi scale method is a convenient way to incorporate image details at different resolutions. MSSSIM index is calculated using the following Eq. 2.8

$$MSSSIM(I, W_I) = [l_M(I, W_I)]^{\alpha_M} . \prod_{i=1}^{M} [c_i(I, W_I)]^{\beta_i} . [s_i(I, W_I)]^{\gamma_i} \tag{2.8}$$

where $c_i(I, W_I)$ and $s_i(I, W_I)$ are the contrast and the structure comparison function at the i^{th} scale respectively, and $l_M(I, W_I)$ luminance comparison function at the M^{th} scale [98].

2.6.1.5 VIFp

VIFp is a procedure to quantifies the information fidelity for the entire image in a statistical HVS model. The VIFp is computed for a collection of $N \times M$ wavelet coefficients from each sub-band using the following Eq. 2.9

$$\text{VIFp} = \frac{\sum_{j \in \text{subbands}} I(\overrightarrow{C}^{N,j}, \overrightarrow{F}^{N,j} | s^{N,J})}{\sum_{j \in \text{subbands}} I(\overrightarrow{C}^{N,j}, \overrightarrow{E}^{N,J} | s^{N,j})} \tag{2.9}$$

where $I(\overrightarrow{C}^{N}, \overrightarrow{F}^{N} | s^{N})$ and $I(\overrightarrow{C}^{N}, \overrightarrow{E}^{N} | s^{N})$ represent the information that could ideally be extracted by the brain from a particular sub-band in the reference and the test images, respectively. The $\overrightarrow{C}^{N,j}$ represent N elements of the Random Field (RF) C_j that describes the coefficients from sub-band j [94].

2.6.1.6 Flicker Metric

Brightness Flicking Metric [95,96] is made to measure flicking quantity (difference between average brightness values) between neighbouring frames of the sequence (previous and current frame). To measure temporal degradation due to the watermark embedding Filicker difference of the original video and watermarked video is quantified.

2.6.2 Robustness Parameter

To measure the robustness of the watermarking scheme, normalized Hamming distance of the extracted watermark (W') and the original watermark (W) is calculated. The Hamming distance (H_{am}) between the W and W' is calculated using the following Eq. 2.10

$$H_{am} = \frac{1}{\|W\|} \sum (W \neq W') \tag{2.10}$$

where "$\|\ \|$" is denoted the count of the watermark bits.

2.7 Experimental Dataset

For experiment in image set, Middlebury Stereo 2006 Datasets of 21 images [99] (link: http://vision.middlebury.edu/stereo/data/scenes2006/) are used for watermark embedding. For video 10 high-resolution (1920×1080 and 1024×768) video sequences are used for watermarking [link:ftp.hhi.de,http://www.tanimoto.nuee.nagoya-u.ac.jp/~fukushima/mpegftv/,http://www.merl.com/pub/avetro/mvc-testseq/orig-yuv/ and http://www.merl.com/pub/tian/NICT-3D/]. In the experimentation, combination of 2 views are taken form the 5 views of the image and the video sequences to represent 3D. The dataset is enlisted in the following Table. 2.1.

2.8 Summary

In this chapter, few background concepts like MCDCT-TF, SIFT, DT-CWT etc. and few evaluation parameters are explained as the prerequisite of the different contributions of this dissertation. The concepts are used in the different phases of the proposed schemes in the later chapters. The evaluation parameters are used to justify the claims experimentally using the dataset summarised in this chapter.

Table 2.1: *Experimental Dataset.*

Image Sequence		Aloe, Baby1, Baby2, Baby3, Bowling1, Bowling2, Cloth1, Cloth2, Cloth3, Cloth4, Flowerpots, Lampshade1, Lampshade2, Midd1, Midd2, Monopoly, Plastic, Rock1, Rock2, Wood1, Wood2
Image Resolutions		1282×1110, 1240×1110, 1252×1110, 1270×1110, 1276×1110, 1300×1110, 1306×1110, 1312×1110, 1330×1110, 1366×1110, 1372×1110, 1396×1110
Video Sequence		Balloons, Dancer, Lovebird1 , Book_arrival, Kendo, Shark, Micro world, Champagne_tower, Pantomime, Newspaper
Video Resolutions		640×480, 1024×768, 1280×960, 1088×1920
Watermark Signal	Image & Video	64×64 binary image and random bit sequence
	Depth	random sequence of 100 bit

Chapter 3

Robust Watermarking for MVD Representation against 3D-HEVC Compression and View Synthesis Attack

Huge advancement in Internet technology and wide availability of the cheaper display devices made the 3D video more attractive due to its immersive experience. Due to the very nature of the digital media, it is easy to copy and redistribute the media content. This makes the video piracy and content tampering as a serious threat over efficient and secure video communication specially over Internet. Recently, depth image based rendering (DIBR) technique [6] has been introduced for generating this synthesis view and the corresponding display is popularly known as auto-stereoscopic display where the conventional *Multi-view video codec* (MVC) technology has been extended as *Multi-view video plus depth* (MVD) [14] with the advance *3D-High efficient video compression* (3D-HEVC) compression technology. As describe in §1, watermarking is being regarded as an efficient DRM (digital right management) tool for video transmission. Although, quite a few works have been reported on the video watermarking as described in literature §1.3, very less attention has been paid on 3D video watermarking until recently.

The 3D depth vision for the human perception is generated by merging two or more than two camera views (corresponding to the left eye and right eye views).

3. ROBUST WATERMARKING FOR MVD REPRESENTATION AGAINST 3D-HEVC COMPRESSION AND VIEW SYNTHESIS ATTACK

Hoffman et al. [7] have shown that the common pixels move horizontally with common left and right views respectively. This makes a common region, called dependent view and an uncommon region, called independent view, for the left and right views. In 3D-HEVC, the dependent view is not required to be encoded in the both the left and right frames [14]. As shown in Fig. 3.1, in time of the 3D video compression using MVD based 3D-HEVC encoder, the independent view zone of left and right frames does not collude [75] at the encoding [14] process. So securing independent region will secure the video sequence against 3D-HEVC

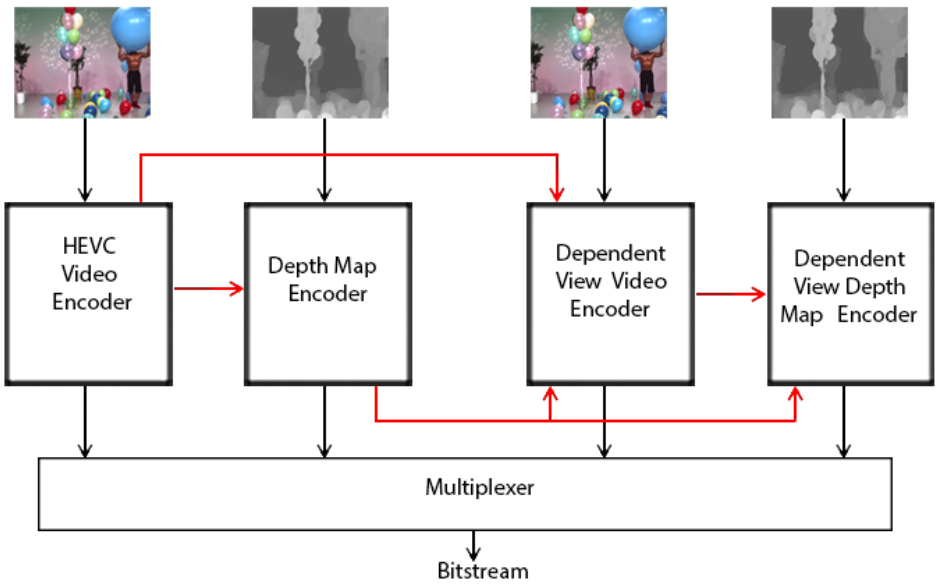


Figure 3.1: *Block diagram of 3D-HEVC codec using 2 view with inter component prediction*

compression attack. To encode the video efficiently, the depth and the camera parameters are used to determine the dependent and the independent view from the left and the right video frame views. At the time of viewing the stereo video, the independent view is used to generate the 3D view of objects. So this independent view part can be considered as the Z-axis. The independent view zone of left and right frames does not blend in the time of MVD based 3D-HEVC [14]

compression (basically in the rendering part as described in previous in §1.1.4). In the first phase of this chapter, a video watermarking scheme is proposed using the MVD characteristics, to resist the 3D-HEVC compression attack where the independent view region is used to insert the watermark.

Another challenge of 3D video watermarking is to secure not only both left and right views but these synthesized views also using DIBR based view synthesis process (as explained in §1.1.4). In DIBR technique, it is observed that only one full view (left or right having both dependent and independent parts) along with independent part of another view(s) are communicated to the receiver side. So the watermark embedded in the other views (where only independent parts are encoded) may not be properly extracted. Moreover, in the case of MVD based encoding, the view synthesis process using depth (say view synthesis attack) itself may destroy the watermark. In that case, a compromised receiver can generate a synthesized view from watermarked left and right views. To secure the main and the synthesis views of the video sequence, dual tree discrete wavelet transformation (DT-DWT) is introduced in the modified version of the proposed scheme. The shift invariant coefficients of the DT-CWT are used reducing the spatial redundancy due to the view synthesis process using the DIBR technique. These two schemes are described in details in the following sections.

3.1 Robust Watermarking for MVD Representation against 3D-HEVC Compression Attack

In this phase of work, the main challenge is to embed the watermark in such a way that the watermark could not be compromised in time of DIBR [6] technique used for 3D-HEVC compression. Various watermarking schemes [57, 63, 64, 66–74] are proposed to secure the 3D content. But most of them are not sustainable against DIBR technique (refer to §1.3.1)

In this scenario, Garcia and Dugelay [61] have proposed an object based wa-

termarking scheme for 3D video. Where the authors [61] used 3D object detection using 3D modelling and secure the texture of the 3D model. This scheme may suffer from spatial view de-synchronization error since the dependent and the independent view in successive frames are different. This may also suffer from temporal de-synchronization as the object motion is not considered. Later, Wu et al. in [66] proposed a 3D watermarking scheme where the uncommon region of left and the right view video has been watermarked separately and the common regions of the 'I' frame has been watermarked separately. In this scheme, they used block-wise discrete cosine transformation (DCT) to filter the left and right eye video and embedded the watermark in such a way that the embedding procedure for left and right eye video is opposite of each other. Here, the depth information is not used to predict the common region and 'B' and 'P' frames are not used to predict the common regions. So the predicted regions using DIBR for MVD based encoder does not match with the regions predicted in Wu's scheme [66]. So, the watermark can be easily compromised in time of 3D-HEVC encoding. Recently, Lee et al. proposed a perceptual depth based 3D video watermarking scheme [65] where the watermark is embedded with the hidden pixels (say the Z-axis) extracted using DIBR and high motion imperceptible regions. Since the watermark is embedded by altering the Z axis only, the rendering process DIBR is not hampered due to embedding. But, since the watermark region is not synchronized with the continuous frames in the group of picture (GOP), watermark bits may be destroyed at the time of motion compensation and quantization.

In the above discussion, it is observed that existing watermarking schemes are not very suitable to secure the 3D video sequence against MVD based 3D-HEVC compression. In DIBR representation, the independent region occurs either in the left or the right views. As a results, these regions remains untouched in time of 3D-HEVC compression. Moreover, in general the independent regions remain imperceptible to human visual system (HVS) [3, 7]. So, embedding in the independent region (Z-axis) may increase the robustness against 3D-HEVC compression and may be imperceptible with respect to the human vision.

In this chapter, a blind 3D video watermarking scheme has been proposed by embedding the watermark in the temporally connected Z-axis regions which makes the scheme robust against the 3D-HEVC compression. To increase the robustness of the scheme, the watermark embedding has been done by altering the low pass DC coefficient values of the 4×4 block of the Z-axis. The rest of the chapter is organised as follows. The proposed Z-axis region selection using DIBR technique and motion vector, is presented in §3.1.1. §3.1.2 describes the watermark embedding and extraction scheme with the selected pixels. Experiment results have been described and explained in §3.1.3.

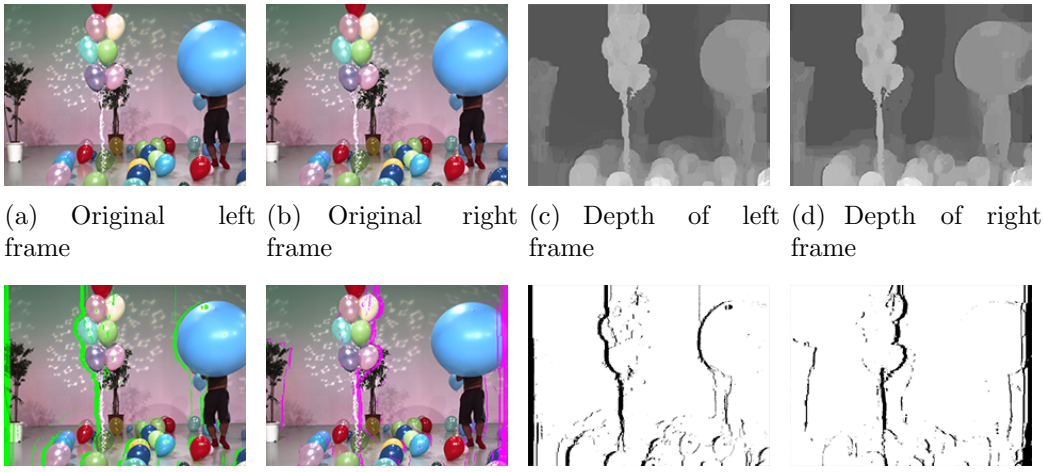
3.1.1 Proposed Concept

It has been observed that Z-axis based embedding helps to increase the robustness of the watermarking scheme for MVD based video sequence. This Z-axis can be extracted using DIBR technique. To increase the robustness of the watermarking scheme against video compression [86], motion compensated Z-axis regions are taken for watermark embedding in a given GOP of a 3D video sequence.

Z-axis Filtration

According to the scheme proposed by Gerhard et al. in [8] and the DIBR technique [6], stereo video is rendered using the depth information and the dependent and the independent (Z-axis) view parts which are detected in the 3D video sequences for both left and the right eye video. The independent part (Z-axis) is marked and filtered as depicted in Fig. 3.2. Here the motion vectors are used over a GOP to compensate the all frames to the last frame as shown in Fig. 3.3. The fully connected regions are filtered from the motion compensated GOP as described in [86]. The fully connected regions of Z-axis of the compensated frame for GOP(=8) is depicted in Fig. 3.4.

The part of the compensated frame of a GOP which contains the fully connected regions of Z-axis, is considered as low pass filtered Z-axis or Z^c. This filtered Z-axis components are used for watermark embedding.



(a) Original left frame (b) Original right frame (c) Depth of left frame (d) Depth of right frame

(e) Rendered left frame from right frame (f) Rendered right frame from left frame (g) *Z*-axis left frame (h) *Z*-axis right frame

Figure 3.2: *Left and right z axis extraction using depth value for Balloons video of camera view 2(left) and 4(right)*

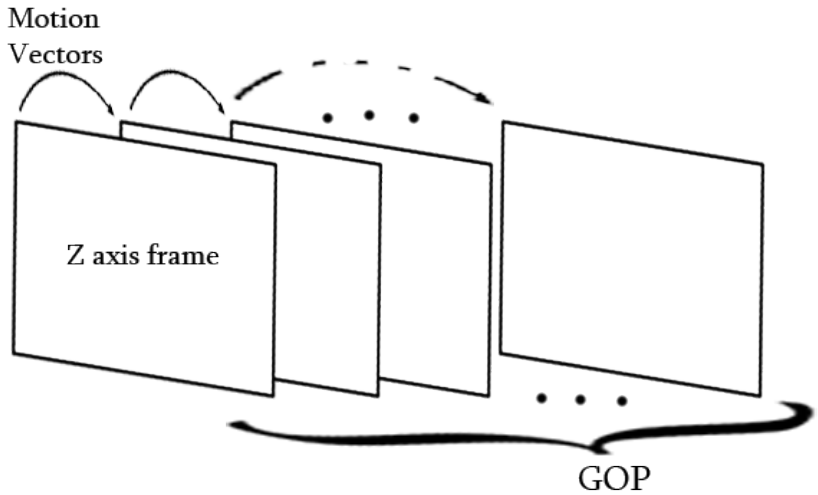
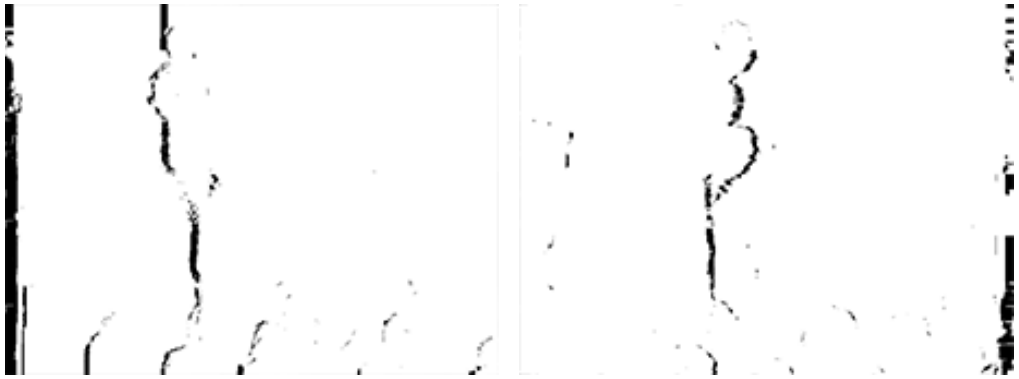


Figure 3.3: *Motion Prediction of the Z-axis GOP*



(a) Motion filtered z axis of left frame (b) Motion filtered z axis of right frame

Figure 3.4: *Connected regions of motion filtered Z-axis frame of Balloons video for GOP=8*

3.1.2 Proposed Scheme

3.1.2.1 Watermark Embedding

As discussed in the Section 3.1.1, the filtered Z-axis part is used for watermark embedding. The proposed scheme is partitioned into three subsections to embed the watermark in the selected regions of the Z-axis. In §3.1.2.1, the embedding capacity of the proposed method is described. How a location map is used to track the exact location of the watermark embedding is discussed in Section 3.1.2.1 and Section 3.1.2.1 narrates the watermark embedding scheme using the original video with the help of location map.

Embedding Capacity:

The connected filtered Z-axis part (say Z^c) [*connected pixels of the independent view part for a 3D video*] of a GOP is used for embedding (refer Section 3.1.1). Firstly the Z^c is partitioned into non overlapping 4×4 blocks. A repetitive sequence of binary bit-stream is used as a watermark. For embedding, a unique combination of two 4×4 blocks of the Z^c is considered to accommodate a single watermark bit. The number of watermark bits which are embedded in each GOP is called embedding capacity. Fig. 3.5 describes the embedding capacity (E_c) of the 3D Balloons video.

Figure 3.5: *Selected blocks for embedding the watermark for each GOP after Z-axis filtration for balloons video (GOP=8)*

Spreading of Watermark Bits to GOP:

As discussed in the last subsection, two 4×4 blocks are used to embed a single watermark bit. Selected pair of such 4×4 blocks are marked with a given watermark bit using a location map (L^{mp}) within the Z^c region in a GOP. Since Z^c is motion compensated, obtained location map is reverse motion compensated to get the embedded watermark, distributed through the GOP according to Eq. 3.1.

$$L^{mp}_{(i+1)} = L^{mp}_{i \leftarrow mv} \tag{3.1}$$

where mv denotes the motion vector between $i^{th} \rightarrow (i+1)^{th}$ frame. After motion compensation, a location map is generated for each frames, containing the watermark bit and the location of the two corresponding 4×4 blocks.

Embedding with the GOP of the Original Video:

A block-wise spatial DCT is done over the 4×4 blocks in the 3D video sequence using the location(L^{mp}). The DC coefficients of two blocks, C_1 and C_2 are used to embed the watermark according to location map (L^{mp}) using Eq. 3.2

$$\left.\begin{array}{ll} \left.\begin{array}{l} C'_{1_i} = \frac{C_1+C_2}{2}(1+\alpha) \\ C'_{2_i} = \frac{C_1+C_2}{2}(1-\alpha) \end{array}\right\} & if\ W_i = 1 \\ \left.\begin{array}{l} C'_{1_i} = \frac{C_1+C_2}{2}(1-\alpha) \\ C'_{2_i} = \frac{C_1+C_2}{2}(1+\alpha) \end{array}\right\} & if\ W_i = 0 \end{array}\right\} \tag{3.2}$$

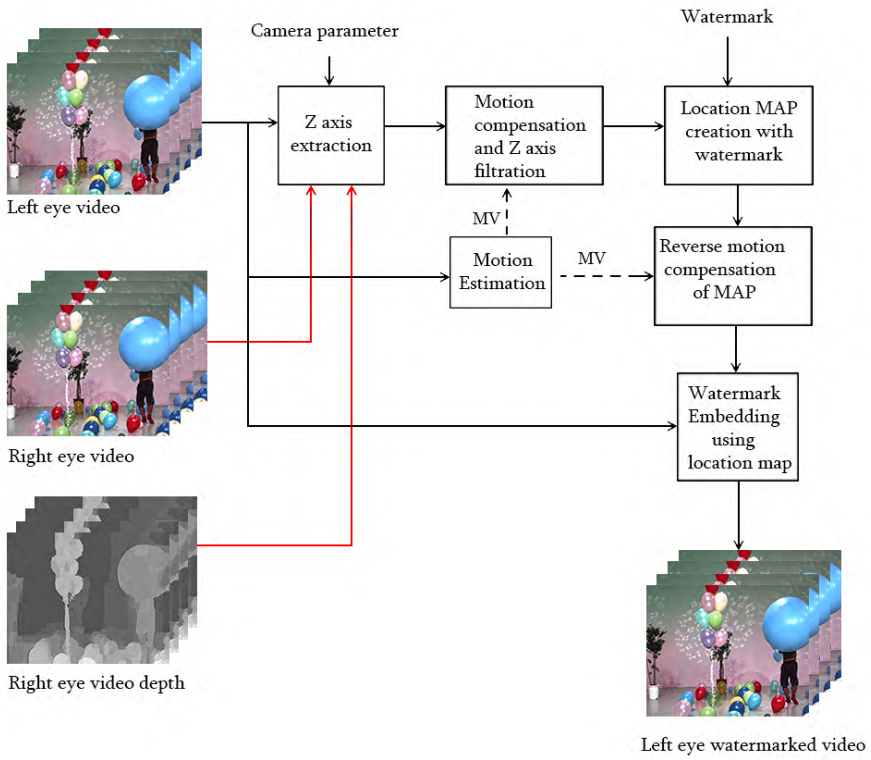


Figure 3.6: *Watermark embedding model (left eye video)*

where W_i is the watermarking bit according to the map on i^{th} location and α is the embedding strength. For the experiments, value of α is taken as 0.02.

After embedding, Inverse DCT is done over the 4×4 blocks to get the watermarked video sequence. The overall watermark embedding scheme is narrated in Algorithm 3 and Fig. 3.6 depicts a sample block diagram for the watermark embedding scheme for the left eye video. The similar approach is used for the right eye video.

3.1.2.2 Watermark Extraction

Extraction of the watermark from 3D video is a reverse process of the embedding scheme. To extract the watermark bit sequence from the embedded GOP, the extraction process has been partitioned in three different modules.

3. ROBUST WATERMARKING FOR MVD REPRESENTATION AGAINST 3D-HEVC COMPRESSION AND VIEW SYNTHESIS ATTACK

Algorithm 1: Watermark Embedding $(V_L, V_R, D_L, D_R, \alpha, W, MV)$

Input: V_L: Left eye video, V_R: Right eye video, D_L: Left eye video depth, D_R: Right eye video depth, α: Watermark strength, W: Watermark bit stream, and MV: Motion vector

Output: W_{VL}:Watermarked left eye video, W_{VR}:Watermarked right eye video

begin

1. The Z-axis for the left video is generated using the left video V_L, right video V_R and the depth of the right video D_R.
2. According to the GOP size, the connected component (Z^c) is extracted using the motion vector MV, as described in Section 3.1.2.1.
3. According to the proposed embedding policy, two 4×4 embeddable blocks are used to embed a single watermark bit and the embedding capacity (E_c) for each GOP is calculated.
4. Watermark bit stream is generated by using a binary random sequence.
5. The generated watermark sequence is partitioned according to the E_c.
6. **for** *each GOP in the video sequence* **do**
   (a) The location map (L^{mp}) is generated by marking the embedding locations with the corresponding watermark bit.
   (b) **for** *each frame in a GOP* **do**
      i. The Z^c is reverse motion compensated to generate the map for that frame with motion compensated locations.
      ii. The watermark is embedded in the DC coefficient of the selected 4×4 blocks using the location map obtained in Step 6(b)i for each frame of left eye video V_L using Eq. 3.2 to generate the watermarked left eye video W_{VL}.
7. Like the left eye video, repeat from step 1 to step 4 for right eye video to generate the watermarked right eye video W_{VR}.
8. **return** (W_{VL},W_{VR})

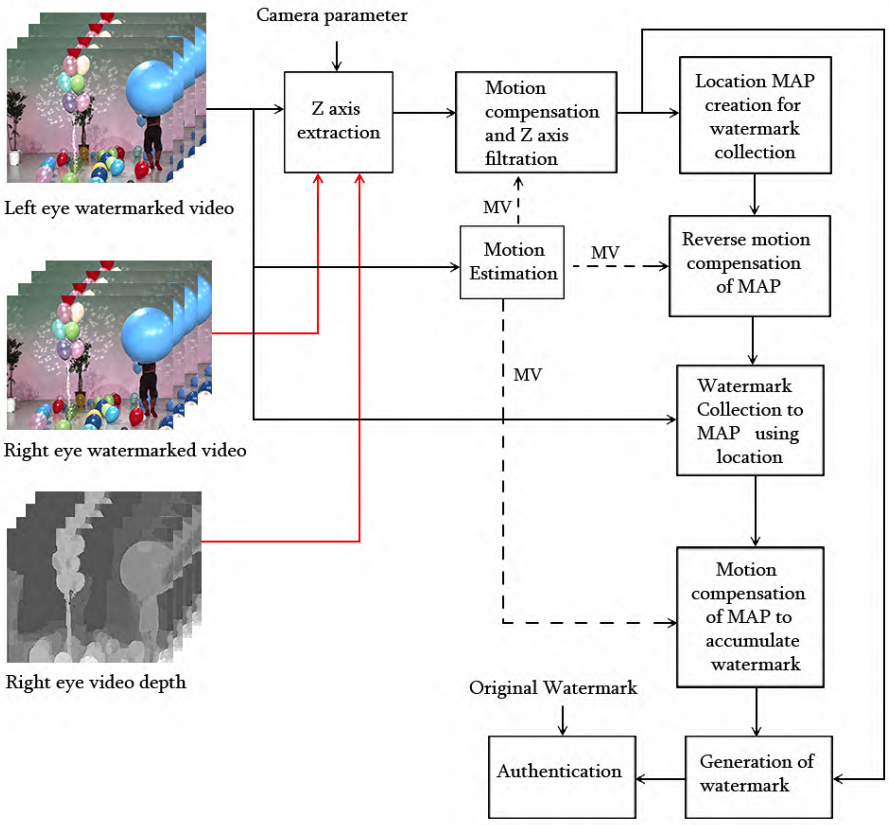


Figure 3.7: *Watermark extraction model (left eye video)*

Spreading of Extraction Map to GOP

In this proposed work, communication of the location map is not required as the location map is generated once again in the decoding side. Also, after attack or compression, location the independent view part (Z-axis) of the 3D video remains static (as the depth is remained untouched). So the generation of the location map is very similar to the encoding side. The reverse motion compensation is done according to the Eq. 3.3

$$mp_{w\,(i+1)} = mp_{w\,i \leftarrow mv} \tag{3.3}$$

where mv denotes the motion vector between $i^{th} \rightarrow (i+1)^{th}$ frame.

Extraction of the Watermark Map from Video

After getting the embedded 4×4 block, block wise DCT is done over the 4×4 blocks and the DC coefficients of two blocks, C'_1 and C'_2 are used according to the extraction map mp_w to extract the watermark bit using the Eq. 3.4

$$\left.\begin{array}{ll} mp_{w\,i} = 1 & if\ C'_{1_i} > C'_{2_i} \\ mp_{w\,i} = -1 & if\ C'_{1_i} < C'_{2_i} \\ mp_{w\,i} = 0 & if\ C'_{1_i} = C'_{2_i} \end{array}\right\} \tag{3.4}$$

where $mp_{w\,i}$ denotes the watermark bit in the i^{th} location.

Accumulate Watermark Map to Generate Watermark

The extracted watermarked map is motion compensated and according to the GOP and added with the previous map to obtain the final map using Eq. 3.5

$$mp_{w\,i} = mp_{w\,i} + mp_{w\,(i+1)\leftarrow mv} \tag{3.5}$$

where $mp_{w\,(i+1)\leftarrow mv}$ is the motion compensated map of the $(i+1)^{th}$ frame to i^{th} frame using the motion vector mv between $i^{th} \rightarrow (i+1)^{th}$.

After the scheme runs recursively over the GOP, the final watermark map Wm^f is generated. The final map is used to generate the part of the watermark (similar to the partitioning is done using the embedding capacity in §3.1.2.1) using Eq. 3.6

$$\left.\begin{array}{ll} W'_i = 1 & if\ Wm^f_i > 0 \\ W'_i = 0 & if\ Wm^f_i < 0 \end{array}\right\} \tag{3.6}$$

where W' denoted the extracted watermark. Using the embedding capacity, the watermark parts are joined and the final watermark W' is extracted.

For authentication of the extracted watermark from the left video W'_l and the right video W'_r, Hamming distance is used as described in §2.6.2 Eq. 2.10 for comparison with the original watermark W. Here H_L and H_R are the hamming distance of the extracted watermark from the left and the right view video.

The overall watermark extraction scheme is narrated in Algorithm. 4 and Fig. 3.7 depicts a sample block diagram for the watermark extraction scheme from the left eye video. The similar approach is used for the right eye video.

Algorithm 2: Watermark Extraction ($W_{VL}, W_{VR}, D_L, D_R, MV$)

Input: W_{VL}: Left eye watermarked video, W_{VR}: Right eye watermarked video, D_L: Left eye video depth, D_R: Right eye video depth, and MV: Motion vector

Output: W_l':Extracted left eye video Watermark, W_r':Extracted right eye video Watermark

begin

1. The Z-axis for the left watermarked video is generated using the left watermarked video W_{VL}, right watermarked video W_{VR} and the depth of the right video D_R.
2. According to the GOP size, the connected component (Z^c) is extracted using the motion vector MV, as described in Section 3.1.2.1.
3. According to the embedding policy that two 4×4 blocks are used to extract a single watermark bit, the embedding capacity (E_c) for each GOP is calculated.
4. **for** *each GOP in video sequence* **do**
   (a) Watermark location map (mp_w) is generated by marking the extraction location of the block to collect the corresponding watermark bit.
   (b) **for** *each frame in GOP* **do**
      i. The Z^c is reverse motion compensated to generate the extraction map for that frame with motion compensated location.
      ii. Extract the watermark using the motion compensated map using the Eq. 3.4.
   (c) Accumulate the extracted watermark mp_w by motion compensation for each GOP using Eq. 3.5.
   (d) Generate the watermark for left eye using Eq. 3.6
5. According to the embedding capacity (E_c), the watermark for the left eye video (W_l') is reconstructed by combining the bits extracted from the successive GOPs.
6. Like the left eye video, repeat from step 1 to step 5 for right eye video to extract the watermark from the right eye video W_r'.
7. **return** (W_l',W_r')

3.1.3 Experimental Results

For experimentation, the proposed scheme is tested over standard 3D video sequences (such as Balloons, Champagne_tower, Kendo, MicroWorld, Shark, etc.) as explained in §2.1 with different camera views. A 64×64 image is used to generate the watermark signal for embedding. The visual quality of the watermarked video is evaluated using the PSNR, and *Flicker metric* (as explained in §2.6). Here, the bit increase rate (BIR) implies the *increment of bit rate* due to watermark embedding. Hamming distance between the extracted watermark and the original watermark is used to measure the robustness of the proposed scheme. For experiment purpose, view 2 & 4 are taken as the left and the right view of the video dataset (refer to § 2.7) with a GOP of 8 frames for embedding purpose as in the 3D-HEVC encoder.

3.1.3.1 Visual Quality Measurement

To measure the watermarked video quality, visual quality and the BIR of the proposed scheme has been compared with the related existing scheme proposed by Lee et al. [65]. The Lee's scheme has been implemented with 2 views 3D scheme instead 3 views 3D scheme. For Balloons and Shark video, the comparison of PSNR and *Flicker Metric* are depicted in Fig. 3.8 to 3.11 respectively and the comparison of BIRs with respect to the 3D-HEVC encoder are depicted in Fig. 3.12 and 3.13.

The average PSNR, Flicker metric, and BIR comparison of the proposed scheme with existing Lee's scheme [65] for the video sequence (refer to § 2.7) are depicted in Table 3.1.

Table 3.1: *Comparison of average PSNR, Flicker metric and BIR (of compressed video) of proposed scheme with existing scheme [65]*

	PSNR	Flicker metric	BIR (%)			
			Overall	I frame	BFrame	P Frame
Proposed Scheme	49.326	0.0072	1.933	1.216	1.993	3.297
Lee's scheme	44.965	0.1288	5.903	2.100	6.225	9.750

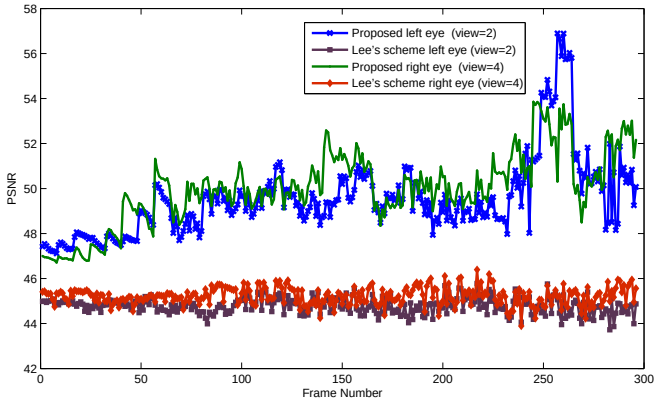


Figure 3.8: *PSNR comparison of the proposed scheme with Lee's scheme [65] for Balloons video*

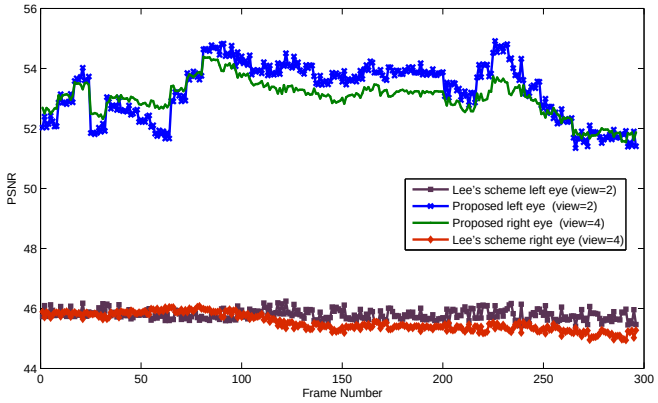


Figure 3.9: *PSNR comparison of the proposed scheme with Lee's scheme [65] for Shark video*

The results show that the proposed scheme gives better visual quality as well as less bit increase than Lee's scheme [65]for encoded video.

3.1.3.2 Robustness Analysis

To evaluate the robustness of the proposed scheme against 3D-HEVC compression, Hamming distance between original and extracted watermark is calculated. At the time of watermarking, the watermark bits are embedded in both left and

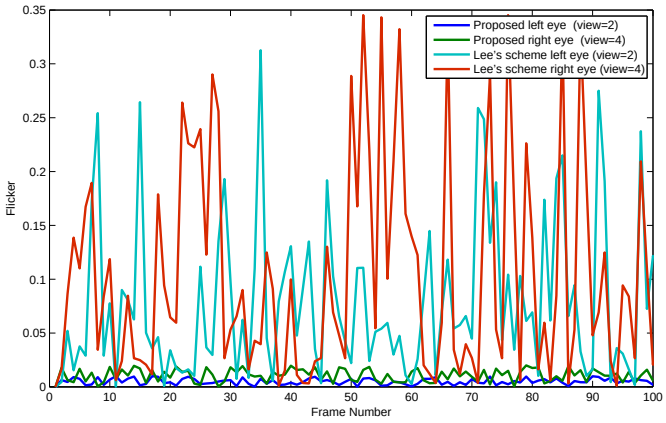


Figure 3.10: *Flicker Metric comparison of the proposed scheme with Lee's scheme [65] for Balloons video (up to 100 frames)*

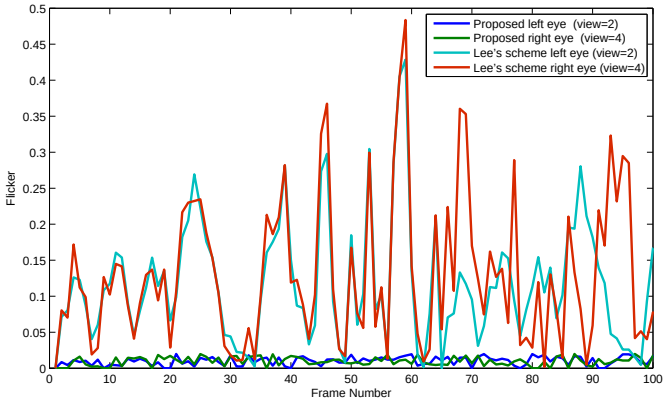


Figure 3.11: *Flicker Metric comparison of the proposed scheme with Lee's scheme [65] for Shark video (up to 100 frames)*

right views of the frame. So, for evaluating the robustness, watermark from both the views are extracted and compared with the original watermark. The comparison of the average hamming distance of extracted watermark from Balloons video & Shark video and average hamming distance of the extracted watermark from the video sequences are tabulated in Table 3.2 and the average hamming distance between extracted and original watermark for different QP values is tabulated in Table 3.3.

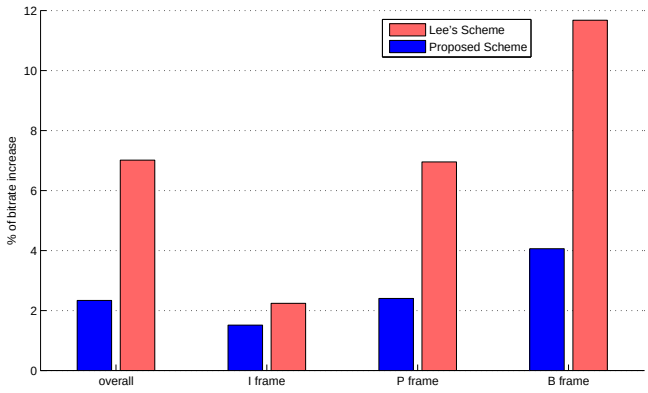


Figure 3.12: *BIR comparison of of the proposed scheme with Lee's scheme [65] for Balloons video*

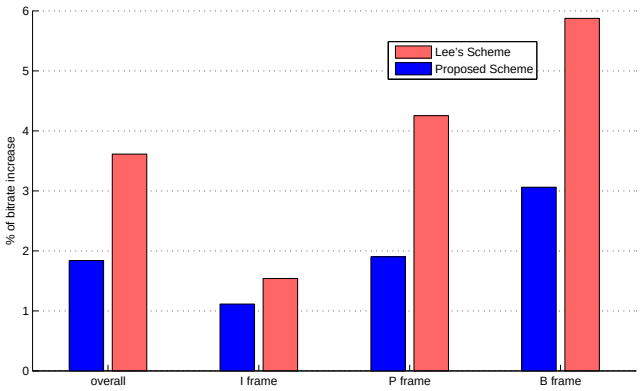


Figure 3.13: *BIR comparison of the proposed scheme with Lee's scheme [65] for Shark video*

From the comparison result presented in Tables 3.2, it can be observed that the proposed scheme shows relatively better robustness than Lee's scheme [65] for left and right views of video sequences against 3D-HEVC compression.

From Tables 3.3, it is observed that robustness of the proposed scheme (hamming distance between original and extracted watermark) performs decently for different QP vales of 3D-HEVC compression.

Table 3.2: *Comparison of average hamming distance of extracted watermark from Balloons video*

Video	View	Lee's Scheme	Proposed Scheme
Balloons video	Left video (view=2)	0.3846	0.1253
	Right video (view=4)	0.3696	0.1024
Shark video	Left video (view=2)	0.4236	0.0826
	Right video (view=4)	0.3983	0.0787
Average	Left video (view=2)	0.3902	0.0981
	Right video (view=4)	0.3752	0.0920

Table 3.3: *Average Hamming distance of extracted watermark for different QP level of the proposed scheme*

Video	QP for encoding	Hamming Distance	
		Left video	Right video
Balloons video	24	0.0424	0.0384
	28	0.0623	0.0549
	32	0.0988	0.0820
	35	0.1253	0.1024
	40	0.1976	0.1653
Shark video	24	0.0312	0.0277
	28	0.0488	0.0428
	32	0.0686	0.0651
	35	0.0826	0.0787
	40	0.1319	0.1238
Average	24	0.0373	0.0342
	28	0.0543	0.0494
	32	0.0758	0.0746
	35	0.0981	0.0920
	40	0.1513	0.1471

3.1.4 Discussion

In this work, watermark is embedded in the Z-axis (independent regions) of the left and the right view. As the independent regions of the left and the right views remained independent in time of MVD based compression (say 3D-HEVC encoder), the embedding scheme shows decent robustness against the rendering precess occur in time of video compression. Using the block DCT coefficients of the temporal filtered Z-axis as embedding coefficients improves the sustainability against the motion estimation and against re-compression with the different compression quality. Proper coefficient selection method improves the robustness as well as the visual quality of the proposed scheme over the existing literature.

It has been observed that the independent regions of the left and the right views are partially present in the intermediate synthesis views [6,9,14], so it may not sustain against the synthesis view attack. To improve the robustness against view synthesis process, another watermarking scheme is proposed in the next section.

3.2 Watermarking for MVD Representation against 3D-HEVC Compression and View Synthesis Attack

The 3D depth vision for the human perception is generated by merging two or more than two camera views (corresponding to the left eye and right eye views). Hoffman et al. [7] have shown that the common pixels move horizontally with common left and right views respectively. This makes a common region, called dependent view and an uncommon region, called independent view, for the left and right views. To achieve higher compression efficiency, few intermediate views are synthesized in the DIBR technique. One of the main challenge of 3D video watermarking is to secure not only both left and right views but these synthesized views also. In the existing 3D watermarking schemes, watermark is

inserted either in both the views [65, 100] or in the synthesized center view [76]. In DIBR technique, it is observed that only one full view (left or right having both dependent and independent parts) along with independent part of other view(s) are communicated to the receiver side. So the watermark embedded in the other views (where only independent parts are encoded) may not be properly extracted. Moreover, in case of MVD based encoding, the view synthesis process using depth (say view synthesis attack) itself may destroy the watermark. In that case, a compromised receiver can generate a synthesized view from watermarked left and right views. Thus the existing 3D watermarking scheme is not very suitable for handling the DIBR technique used in MVD based encoding.

In this phase of work, the main challenge is to insert the watermark in such locations of the video sequences, so that the watermark can be extracted from the embedded views as well as the other synthesis views. In the previous section (§3.1), the independent region based watermarking scheme using Z-axis is proposed where the watermark is embedded on the independent part of the left and right view for a video frame to sustain against 3D-HEVC encoding. The main motivation for using Z-axis based scheme is that the independent view is mutually exclusive and the watermark region could not be colluded. But due to the partial presence of the independent view in the synthesized view, the watermarking scheme could not resist the synthesis view attack using the DIBR technique. Moreover, the independent regions of the left and the right views occur either in the left view or the right view [14]. So tampering of the independent regions does not make visual degradation with respect to HVS [3]. So, the watermark can be easily be destroyed by tampering the independent region of the 3D video sequences.

There are few relevant observations that can be made from the above discussion, firstly, most of the existing schemes as discussed in the literature (refer to §1.3)are not resilient to DIBR based encoding and view synthesis process. Secondly, although independent view based embedding is not visually perceptible due to parallax display [3], the embedding can easily be destroyed by modifying

this region. Thirdly, with the availability of the depth information at the receiver, it is possible to regenerate the other views from a single view which may be a serious threat to the view based watermarking. Finally, since there is a strong similarity between different views of a frame, generation of watermarks for different views should be homogeneous to resist the collusion attack. Taking these facts into consideration, in this work, a 3D blind video watermarking scheme is proposed which embeds the watermark in the synthesized center view to resist the synthesis view attack. The challenge of the proposed scheme is to protect the left and right views which are generated from the center view by a reverse synthesis process and to handle the temporal de-synchronization due to motion in the successive frames. To achieve this, motion compensated temporal filtering is applied on center views of the successive video frames and DT-DWT [89] coefficients of the low pass center view are modified for watermark embedding. In this scheme, block-wise approximate shift invariant DT-DWT [89] is used to make the watermark robust against the DIBR technique. A comprehensive set of experiments is carried out to justify the applicability of the proposed scheme over the existing literature. The rest of the work is organized as follows. In §3.2.1, proposed watermarking scheme is presented. The experimental results are presented in the §3.2.2.

3.2.1 Proposed Scheme

In the proposed scheme, watermark is embedded in the synthesized center view which is obtained by rendering the left and right views using DIBR. After embedding, the watermarked center view is inverse rendered to get watermarked version of the left and right views which are being communicated to the receiver side. So the primary goal of this work is to embed the watermark in center view in such a way that the watermark propagates to left and right views during inverse rendering process. An efficient block selection strategy has been employed to meet this requirement. Moreover, DT-DWT coefficients of a selected block

are used for embedding to make the scheme invariant to the inverse rendering process. The details of block and coefficient selection processes are discussed in the successive subsections. It is observed in that most of the existing 3D watermarking schemes [57,63,64,66–74] did not consider the motion along the temporal axis which may produce temporal flickering due to embedding. To reduce this temporal artifacts, the motion compensated temporal filtering has been employed on successive center views in a GOP and a low pass version is chosen for embedding. In the proposed scheme, since the dependent portion of both left and right views are watermarked with homogeneous watermark (generated from same center view), the type II collusion attack can also be resisted. In this section, first, an embedding zone selection process of the proposed scheme is given. Watermark embedding and extraction process are presented in subsequent subsections.

3.2.1.1 Embedding Zone Selection

In the proposed scheme, a GOP of a cover 3D video sequence is taken for watermarking. Firstly, left and right views for each frame of a 3D video are identified and a center view corresponding to each frame has been generated. Then all the center views for the GOP are subjected to motion compensated temporal filtering as discussed in [86] to get the low pass version of such views. In this work, using DCT based motion compensated temporal filtering (MCDCT-TF) (refer to §2.1) where $k \times 1$ (where k is the number of frames in a GOP) temporal DCT is used to extract 1 low pass frame from GOP of k frames.

Block Partitioning:

The low pass center view obtained from the MCDCT-TF is partitioned into non overlapping blocks of size $M \times N$. It is observed that the disparity between left and right views is approximately around the 200 pixels for the natural 3D video frames. It implies that the disparity between center view and any of the left or right view will be around 100 pixels (half of the disparity between left and right). To assure the retaining of the watermarkable coefficients in the reverse rendering process, the block width of the center view for embedding should be more than

100 pixels (the disparity between center view and any of the left or right view). Keeping this fact in mind, the block size of the center view for embedding is taken as 16×256. The blocks of the low pass version of the center view are subjected to 2D dual tree discrete wavelet transformation (DT-DWT) which is discussed in the next subsection.

2D Dual Tree Discrete Wavelet Transform (DT-DWT):

As it is observed in the literature that the dual tree complex wavelet transformation (DT-CWT) or its real counter part (DT-DWT) are approximately shift-invariant [89] and can be used to handle the disparity occurred due to rendering the left or right views from the center view [76]. It is also observed that the PSNR between of center view with left or right view for H_1, H_2, H_5, H_6 coefficients are higher than that of H_3, H_4 coefficients where H_1, H_2, H_3, H_4, H_5, H_6 are representing the coefficients having orientation angle $15°$, $45°$, $75°$, $-75°$, $-45°$, $-15°$ respectively for 2D-DT-DWT transformation. It is observed in the experimentation that the absolute difference between H_1 and H_2 and that of H_5 and H_6 are generally very less (approximately 5% of each other). Additionally, absolute difference between $|\frac{1}{n}\sum_n (H_1 - H_6)|$ and $|\frac{1}{n}\sum_n (H_2 - H_5)|$ is negligibly small (almost close to zero).

3.2.1.2 Watermark Embedding

To embed the binary watermark sequence (W), the 3^{rd} level 2D DT-DWT coefficients (say H_1^3, H_2^3, H_5^3, H_6^3) are used. Let $H_1^3 - H_2^3$ and $H_6^3 - H_5^3$ are taken as ζ_1 and ζ_2. In this embedding process, if watermark bit is "0", then $\sum_n |\zeta_1|$ should be greater than $\sum_n |\zeta_2|$ by atleast τ value and if the watermark bit is "1", then $\sum_n |\zeta_1|$ should be less than $\sum_n |\zeta_2|$ by atleast τ value where τ is a threshold and is defined as $(\sum_n |\zeta_1| + \sum_n |\zeta_2|) * \alpha$. In case, if watermark bit is "0" but the corresponding embedding rule $(\sum_n |\zeta_1| - \sum_n |\zeta_2| \geq \tau)$ does not hold, the coefficient H_2^3, H_5^3 are modified to meet the requirement. Similar measure is taken for embedding bit "1". The watermark embedding process is described by the Eq. 3.7

$$
\left.
\begin{array}{l}
if\ W_i == 0 \\
\quad if\ \sum\limits_n |\zeta_1| - \sum\limits_n |\zeta_2| \geq \tau\ then \qquad no\ change; \\
\quad else \\
\qquad H^3_{2_n} = H^3_{2_n} - (H^3_{1_n} - H^3_{2_n}).(\alpha + \kappa) \\
\qquad H^3_{5_n} = H^3_{5_n} + (H^3_{6_n} - H^3_{5_n}).(\alpha + \kappa) \\
else \\
\quad if\ \sum\limits_n |\zeta_2| - \sum\limits_n |\zeta_1| \geq \tau\ then \qquad no\ change; \\
\quad else \\
\qquad H^3_{2_n} = H^3_{2_n} + (H^3_{1_n} - H^3_{2_n}).(\alpha + \kappa) \\
\qquad H^3_{5_n} = H^3_{5_n} - (H^3_{6_n} - H^3_{5_n}).(\alpha + \kappa)
\end{array}
\right\} \tag{3.7}
$$

where α is the robustness threshold, and κ is defined as $\frac{\sum\limits_n |\zeta_1| - \sum\limits_n |\zeta_2|}{\sum\limits_n |\zeta_1| + \sum\limits_n |\zeta_2|}$. From experimental observations, the value of α is taken as 0.02.

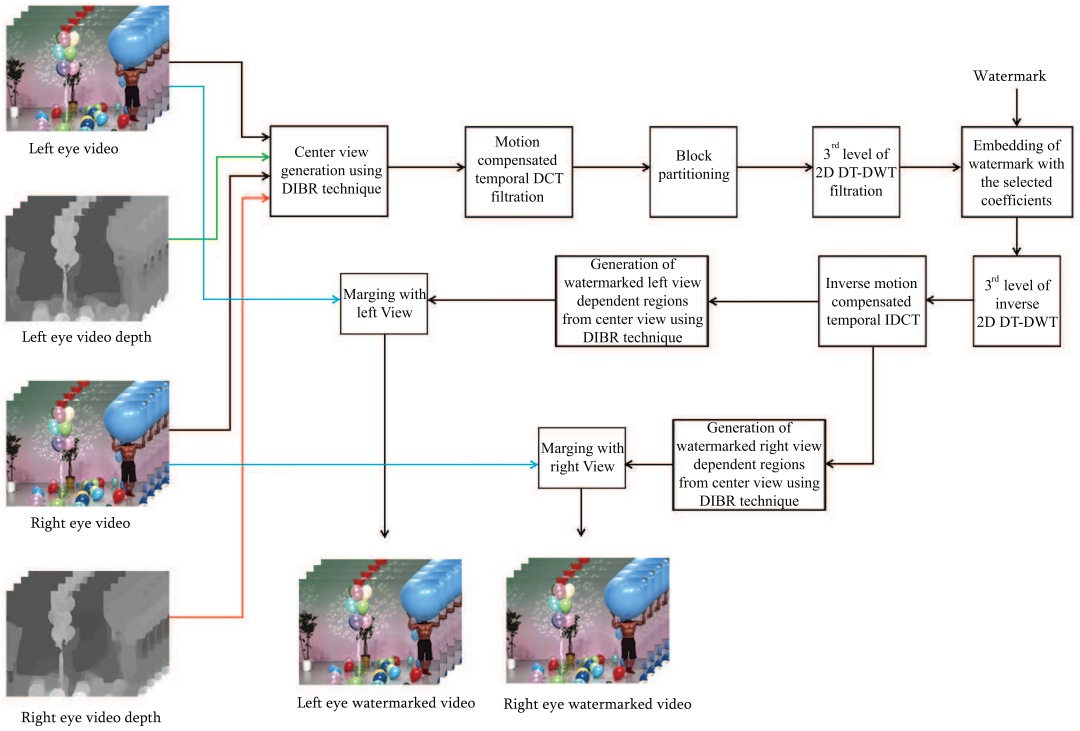


Figure 3.14: *Watermark embedding model*

The watermarked coefficients are subjected to 3^{rd} level of 2D inverse dual tree DWT (IDT-DWT) and subsequently to the inverse MCDCT-TF to generate the

Algorithm 3: Watermark Embedding $(V_L, V_R, D_L, D_R, \alpha, W)$

Input: V_L: Left eye video, V_R: Right eye video, D_L: Left eye video depth, D_R: Right eye video depth, α: Watermark strength, and W: Watermark bit stream

Output: W_{VL}:Watermarked left eye video, W_{VR}:Watermarked right eye video

begin

1. The center view video is synthesized using the left video V_L, left eye video depth D_L, right video V_R and the right eye video depth D_R.
2. According to the GOP size, the center view video frames are motion compensated to a single frame using the motion vector.
3. A temporal DCT is done over the motion compensated frame and low-pass frame is generated.
4. Partition the low-pass temporal filtered frame in to non overlapping $M \times N$ blocks as described in Sec. 3.2.1.1.
5. **for** *each Block* **do**
   (a) Apply $3^r d$ level of 2D DT-DWT and select the coefficients of 15°, 45°, 75°, −75°, −45°, −15° orientation angle (H_1^3, H_2^3, H_5^3, H_6^3).
   (b) Embed the watermark with H_1^3, H_2^3, H_5^3, H_6^3 using the Eq. 3.7
   (c) Apply $3^r d$ level of inverse 2D DT-DWT to generate the temporal low-pass filtered watermarked frame
6. Apply inverse temporal DCT and inverse motion compensation to generate the watermarked center view.
7. Using DIBR technique, the left and right views are inverse rendered from the watermarked center view and they are used to replace the dependent part to generate the watermarked left and right video W_{VL}, W_{VR}.
8. **return** (W_{VL}, W_{VR})

watermarked center view for the video frame. These watermarked center view is inverse rendered to generate the dependent portion of the left and right view and merged to finally generate the watermarked views. The overall embedding process is narrated in Alg. 3 and the overall block diagram is depicted in Fig. 3.14 .

3.2.1.3 Watermark Extraction

As a view invariant watermarking, the watermark can be extracted from any views for a frame (left, right or any other synthesized views) in the proposed scheme. The watermark extraction process is almost reverse as of the embedding process. The 3^{rd} level 2D DT-DWT is carried out over the motion compensated temporal filtered blocks of the video frames (any view). Let $H_1'^3$, $H_2'^3$, $H_5'^3$, $H_6'^3$ are the filtered high frequency watermarked coefficients of orientation angle 15°, 45°, −45°, −15° and let $H_1'^3 - H_2'^3$ and $H_6'^3 - H_5'^3$ are taken as ζ_1' and ζ_2'. The watermark is extracted using ζ_1' and ζ_2' as described in Eq. 3.8.

$$\left.\begin{array}{l} if \ \sum\limits_n |\zeta_1'| > \sum\limits_n |\zeta_2'| \ then \\ \quad W_i' = 0 \\ else \\ \quad W_i' = 1 \end{array}\right\} \tag{3.8}$$

where W' is the extracted watermark. For authentication of the extracted watermark from the 3D video W', Hamming distance is used as described in §2.6.2 Eq. 2.10 to compare with the original watermark W. Here H is the hamming distance of the extracted watermark from the video sequence. The overall watermark extraction process is narrated in Algo. 4 and the overall block diagram is depicted in Fig. 3.15.

3.2.2 Experiment Results

For the experimental purpose, the proposed scheme is tested over a set of standard 3D video sequences (as esplained in §2.7) with distinctive camera views. The experiment results of the proposed scheme is compared with the Lee's scheme [65]

Algorithm 4: Watermark Extraction (W_v)

Input: W_v: Watermarked video
Output: W':Extracted Watermark
begin

1. According to the GOP size, the video frames of watermarked video (V_W) are motion compensated to a single frame using the motion vector.
2. A temporal DCT is done over the motion compensated frame and low-pass frame is extracted.
3. Partition the low-pass temporal filtered frame in to non overlapping $M \times N$ blocks as described in Sec. 3.2.1.1.
4. **for** *each Block* **do**
   (a) Apply $3^r d$ level of 2D DT-DWT and select the coefficients of 15°, 45°, 75°, −75°, −45°, −15° orientation angle (H_1^3, H_2^3, H_5^3, H_6^3).
   (b) Extract the watermark W' by comparing H_1^3, H_2^3, H_5^3, H_6^3 using the Eq. 3.8
5. **return** (W')

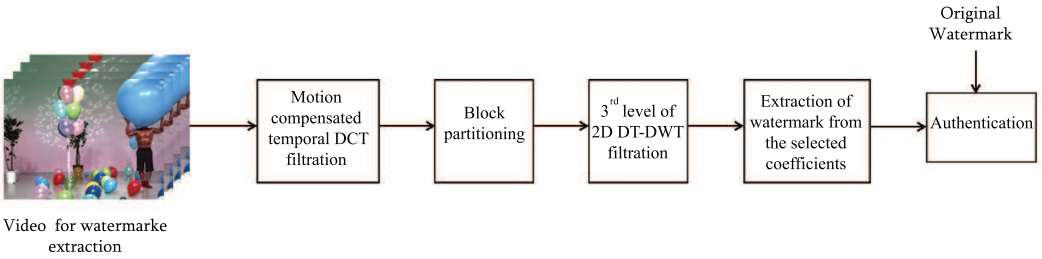


Figure 3.15: *Watermark extraction model*

and previous scheme (§3.1) say Rana's scheme [100] as these two schemes are most recent watermarking scheme for DIBR based 3D video sequence.

3.2.2.1 Visual Quality

To evaluate the visual quality of the proposed scheme, PSNR, SSIM and VIFp are measured. The comparison results for proposed scheme with Lee's scheme [65] and Rana's scheme [100] are tabulated in Table 4.5 for these visual quality measures. From Table 3.4, it is observed that the proposed scheme gives almost comparable result for the PSNR, SSIM and VIFp against the Lee's scheme [65]. It shows a bit inferior results against Rana's scheme [100] as only independent part of the left or right view is alter in the Rana's scheme [100].

Table 3.4: *Average PSNR, SSIM, VIFp comparison result of the proposed scheme with Lee's scheme [65] and Rana's [100]*

		Average PSNR	Average SSIM	Average VIFp
Proposed scheme	Balloons	45.12	0.9980	0.9896
	Shark	45.22	0.9984	0.9899
	Average	44.87	0.9907	0.9846
Lee's scheme	Balloons	45.35	0.9978	0.9882
	Shark	44.24	0.9973	0.9861
	Average	44.97	0.9884	0.9812
Rana's scheme	Balloons	49.55	0.9985	0.9912
	Shark	52.14	0.9982	0.9904
	Average	49.33	0.9981	0.9894

3.2.2.2 Robustness

The robustness of the proposed scheme is evaluated against 3D-HEVC compression by measuring the hamming distance between original and extracted watermark. To show the robustness against synthesis view attack, watermark is extracted from intermediate synthesis view generated by left and right view. For experimental purpose, camera view 2 and 4 is taken as left and the right view. Camera view 2 and 4 are used to generate the synthesis camera view 2.5, 3, 3.5 video sequence using DIBR technique. Table 3.5 tabulates the comparison result of the proposed scheme with Lee's scheme [65] and Rana's scheme [100] for video

sequences of camera view 2, 2.5, 3, 3.5, 4.

Table 3.5: *Hamming distance comparison result of the proposed scheme with Lee's scheme [65] and Rana's [100] for video sequences*

Video	View number	Proposed scheme	Lee's scheme	Rana's scheme
Balloons video	2	0.1811	0.3846	0.1253
	4	0.1781	0.3696	0.1024
	3	0.0926	0.6612	0.6011
	2.5	0.1304	0.4805	0.5126
	3.5	0.1253	0.4712	0.5211
Shark Video	2	0.1755	0.4236	0.0825
	4	0.1719	0.3983	0.0787
	3	0.0892	0.5824	0.5686
	2.5	0.1223	0.5102	0.5482
	3.5	0.1205	0.4996	0.4562
Average	2	0.1783	0.3902	0.0981
	4	0.1744	0.3752	0.0920
	3	0.0911	0.5940	0.5263
	2.5	0.1283	0.5207	0.5374
	3.5	0.1237	0.4836	0.4829

It is observed from the Table 3.5 that proposed scheme outperforms other two schemes, Lee's scheme [65] and Rana's scheme [100], for synthesized views (i.e. for 3, 2.5 and 3.5). It gives a bit inferior results than the Rana's scheme [100] when extracting from original embedded views (i.e. 2 and 4 views) but outperforms the Lee's scheme [65].

Table 3.6: *Average hamming distance comparison result of left ans right view of the proposed scheme with Lee's scheme [65] and Rana's scheme [100] after collusion attack using DIBR technique*

Video	Proposed scheme	Lee's scheme	Rana's scheme
Balloons	0.1961	0.4216	0.4141
Shark	0.1902	0.4656	0.3988
Average	0.1947	0.4387	0.4030

To test the robustness of the proposed scheme against collusion attack, the

left (and the right) view video is rendered to right (and left) view video and colluded with the right (and the left) eye video. The extracted watermark from the colluded video is depicted in Table 3.6. It is observed that the proposed scheme performs better against collusion attack than Lee's scheme [65] and Rana's scheme [100] because the watermark in the dependent view will create co-located regions for the left and right view.

3.2.3 Discussion

In this work, watermark is embedded in the motion compensated center view part of the left and the right view video. As the dependent regions of the left and the right views are available in the center view and also common for both the views, the robustness of the embedding scheme is increased against MVD based compression (say 3D-HEVC encoder). Using of the shift invariant coefficients of the DT-DWT and selection of proper block size improves the sustainability against the synthesis view attack.

3.3 Summary

In this chapter, two 3D video watermarking schemes have been proposed against MVD based compression. In the first work, a Z-axis based 3D video watermarking scheme has been proposed which can resist the 3D-HEVC compression (and quantization) attack. To increase the robustness of the scheme, the block DCT coefficients of the temporal filtered Z-axis are used for embedding the watermark. A set of experiments has been carried out for different video sequences to justify the applicability of the proposed scheme over the recent existing schemes. For experimental purpose, all the parameters have been set according to the 3D-HEVC compression.

In the second phase of this chapter, an improvement of the prior scheme has been done by proposing a shift invariant DT-DWT based watermarking scheme to resist the synthesis view and DIBR based rendering process. Motion com-

pensated temporal filtering is used over the synthesized center view to make the scheme robust against 3D-HEVC compression attack. To improve the robustness 3^{rd} level 2D DT-DWT has been used for watermark embedding and specific coefficients have been selected to make the scheme DIBR invariant and collusion attack invariant. The experiment results demonstrate the applicability of the proposed scheme over the existing methods. In the next chapter, the issues for securing the depth of the image and the image and video sequences has been considered.

3. ROBUST WATERMARKING FOR MVD REPRESENTATION AGAINST 3D-HEVC COMPRESSION AND VIEW SYNTHESIS ATTACK

Chapter 4

Watermarking in Depth Information of 3D Image and Video Sequences against Depth Attack

The advancement in network technology and the wide availability of the cheaper display devices make the 3D media more attractive due to its immersive experience. Due to the very nature of the digital media, it is easy to copy and redistribute the media content. Watermarking is being regarded as an efficient DRM (digital right management) tool for media transmission. Although, quite a few works have been reported on the image or video watermarking, very less attention has been paid on 3D image video watermarking until recently as discussed in §1.3.

In the 3D vision, the depth impression is created for human viewing by combining the left and the right views, where the common pixels in both the views are horizontally shifted from the left to right respectively [7]. Recently, the depth based 3D image and video representation [1, 2, 13, 14] becomes popular due to its compression efficiency and is popularly known as DIBR-3D and *Multi-view video plus depth* (MVD)(refer to §1.1). In this DIBR-3D representation, depth map is used to generate the 3D view (say left view and the right view) at the user end from the center view while the same is generated from the dependent and

the independent view information for MVD representation using the depth image based rendering (DIBR) technique. In this kind of 3D virtualization, a visual discomfort, called vertical and horizontal parallax [3] can be occurred in time of viewing the 3D images and video. By analysing this characteristics of 3D viewing, it is observed that small distortion in the viewing plane may not be sensitive to the human visual system (HVS). In this scenario, unsecured depth parameter can be used to generate more distorted and illegitimate synthesized left and right views which may make the recent 3D image and video watermarking schemes vulnerable against different attacks [41–57, 57–74].

For securing the depth of 3D video sequence, Guan et al. proposed a blind multiple watermarking scheme for DIBR-3D representation [79], where only the depth region is watermarked to secure the original image with depth using quantized index modulation embedding policy. However as, depth is relatively less important part than the original image, different smoothing attack can be used to destroy the watermarking scheme. Also, object based smoothing can be used to remove the watermark from the depth of a 3D image [78]. Moreover, the object based embedding location does not support the DIBR based view synthesis process. So the robustness of the scheme [79] may reduce with the increasing of the baseline distance.

Since in MVD based video representation, separate watermark having embedded for the left and the right video, the embedded watermark can be estimated by collusion like attacks (as discussed in §1.3). Moreover, the image based watermarking scheme is not sustainable against 3D-HEVC compression due to not considering the motion component in time of embedding. So, an improved version is proposed for MVD based video sequence to resist 3D video compression as well as view synthesis attack. These two works are described in details in the following sections.

4.1 Watermarking in DIBR-3D based Image Depth against View Synthesis Attack

In this module of the work, the main challenge is to secure the depth of the 3D image in the DIBR-3D representation. It is observed by analysing the depth of the 3D image (refer to §1.1) that the depth map can be viewed as an image having relatively less high frequency components. So, large depth modification may create object discontinuity in the view plane after creation of 3D using DIBR technique which results visual discomfort [3]. To reduce the visual degradation with respect to the human eye, in the proposed scheme (in §4.1.1), foreground objects (generally inviting higher attention of human eyes) are excluded from the embedding zone. Additionally, the scale invariant feature transform (SIFT) locations are used for watermarking to make the scheme view invariant. A comprehensive set of experiments are carried out (in the §4.1.2) to justify the applicability of the proposed scheme over the existing literature.

4.1.1 Proposed Scheme

In this scheme, the watermark is embedded in the depth map (image) of the center view image. Eventually, the depth map is required in the receiver side to render the left and the right view. So, change in depth map may change the object position in the Z-axis (say in the depth plane) which creates visual discomfort [3]. Thus, one of the main challenges of this scheme is to insert the watermark in such a way that the modified depth parameter does not make any visual discomfort with respect to the human visual system (HVS). It is important to find such embeddable locations which are less sensitive to HVS. Also, it is observed that foreground object locations are in general more sensitive to human vision [7,87]. So avoiding of the foreground object regions for embedding locations may achieve the desire visual quality. A foreground object detection method is explained in the next section. Using the less sensitive regions, the scale invariant feature transform (SIFT) is used to find a suitable embedding coefficients for

view invariant watermarking. The watermark embedding & extraction schemes are described in the subsequent subsections.

4.1.1.1 Detection of Background and Foreground Objects

In the HVS, the foreground object is, in general, more sensitive than the background object. To detect the foreground object (or to extract the background scene), Semantic Image Segmentation (as explained in §2.2) is used to extract the foreground [87]. Fig. 4.1(b) shows the foreground object segmentation for Fig. 4.1(a) and Fig. 4.1(c) shows the background and the foreground region separation for the given image. Fig. 4.1(e) represents the background regions after removal of the foreground objects from the center view depth image Fig. 4.1(d).

(a) Original Image (b) Foreground object detection (c) Background and foreground detection (d) Depth image

(e) Depth without foreground

Figure 4.1: *Background of depth image detected using the main view image (kendo video frame)*

It is observed in the Fig. 4.1(e) that there are some locations near to the camera (the floor) are not detected as foreground because those regions are not sensitive to the human vision (or say less salient regions). So embedding of watermark with those regions in the depth image will not effect the visual quality of the image subsequently.

4.1.1.2 SIFT based coefficient selection

To make the scheme invariant to the view synthesis process, embedding zone selection is carried out based on the scale invariant feature transform (SIFT) feature points. In this scheme, SIFT locations of the original image is used for embedding the watermark in the depth image. After view synthesis process, some embedding locations may be missing due to the hole filling process [9]. So a coefficient partitioning is implemented on the basis of the scale feature (say σ) of the SIFT. Here the SIFT coefficients are grouped with separation gap (say δ) from 2 to 10 of the scaling features (as the strong features are available in those locations [81]). In constant separation gap among the scaling features, the distribution of SIFT feature points in each groups goes decreasing with the increasing of the σ. So the δ is calculated using an relative polynomial function to make a better distribution of SIFT coefficients in each group. The separation (δ) is calculated using the following Eq. 4.1.

$$\delta_i = \sigma_i^2 \gamma \tag{4.1}$$

where σ_i is the starting value of the i^{th} group and γ is the grouping threshold. So the value of σ_{i+1} is $(\sigma_i + \delta_i)$. The distribution of SIFT points in group of scaling feature value are depicted in Fig. 4.2.

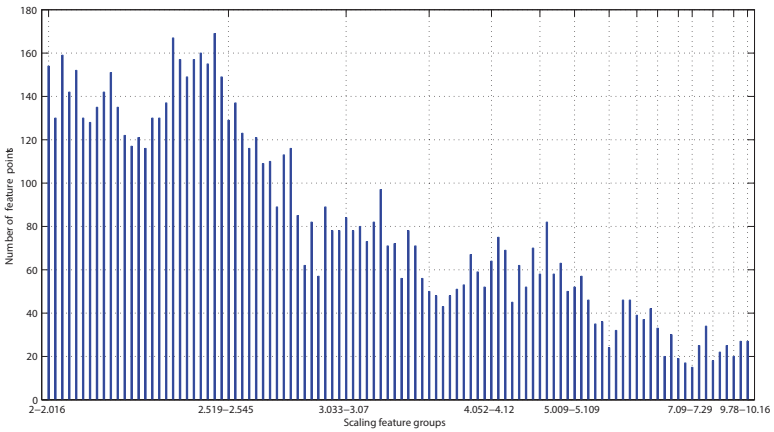


Figure 4.2: *Distribution of SIFT feature points in each group*

Here each group locations are used for insertion of a single watermark bit. As the scaling feature points are invariant to DIBR, the locations of the feature points in synthesis view will identify the same at pixel locations as the original center image. Embedding of watermark with the depth in locations of the feature points will make the scheme invariant to DIBR view synthesis process.

4.1.1.3 Watermark Embedding

The selected portion of the depth image after analysing the SIFT features and foreground subtraction of the original image, is partition in to $n \times n$ blocks (here we set n=8). For experimental purpose, the value of n is taken as 8 to increase the robustness by using more coefficients for a single watermark bit insertion and each block starting position indicates the SIFT point location. A binary watermark bit sequence (W) is used for watermarking purpose where a single bit is embedded with the such 8×8 blocks of each group. For embedding of the watermark, depth values are altered and each group is used to insert a single watermark bit. It is observed that, Depth is a very smooth region with less variation with the near bound pixels. As a result, the variation of coefficients in each 8×8 block are very less as shown in Fig. 4.3.

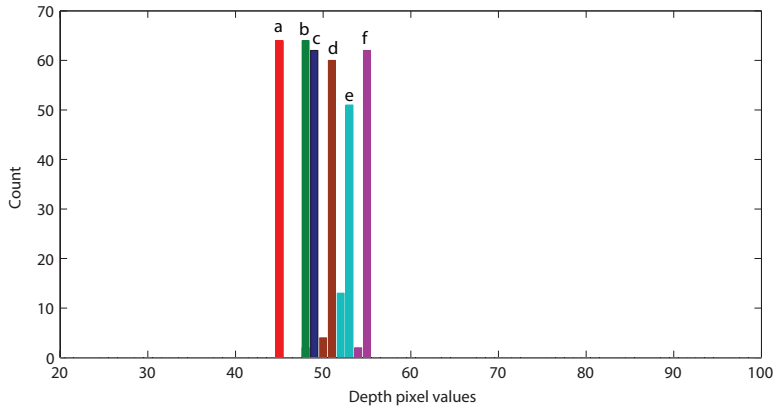


Figure 4.3: *Histogram of the depth block coefficients in first group for kendo video frame (a,b,c,d,e and f define the different blocks)*

Also, depth is used to generate the left and the right views of the 3D image

using the Eq 4.2

$$[I_{L_n}, I_{R_n}] = I_{C_n} + d_{isp} T \frac{D_{I_{C_n}}}{255} \tag{4.2}$$

where d_{isp} is the disparity, D_{I_C} is the depth value of the center view image I_C for the location n and I_L & I_R are the corresponding shifted locations for the left and the right view respectively. T is the baseline distance between the reference view to the synthesis view. To synthesis the left and the right view from the center view the value of T is taken to ± 0.5 (where $+.5$ for left view and $-.5$ for the right view).

So, change of single bit in depth, makes very less degradation in the original views by using Eq. 4.2. It is experimentally observed that making perturbation of a single bit change of each coefficients in the depth image pixels, does not make any perceptible changes in the 3D vision with respect to HVS. In this watermarking scheme, all coefficient values of the depth blocks are changed to even or odd by using the embedding rule as depicted in Eq. 4.3

$$\left.\begin{array}{l} if\ W == 0 \\ \quad mod(C_i, 2) = 0; \\ else \\ \quad mod(C_i, 2) = 1; \end{array}\right\} \tag{4.3}$$

where W is the watermark bit and C defines all the coefficients of i^{th} block. The bean values are altered for embedding the watermark as depicted in Eq. 4.4

$$C'_{j_i} = C_{j_i} - mod((C_{j_i} + W_k), 2) \tag{4.4}$$

where C_{j_i} and C'_{j_i} defines the j^{th} coefficient of the i^{th} block of the group k.

The overall embedding process is narrated in Algorithm. 5 and the block diagram is depicted in Fig. 4.4.

4.1.1.4 Watermark Extraction

The watermark extraction scheme is the reverse procedure of the embedding scheme. The foreground object removal and the block partitioning schemes are same as the embedding scheme. Using of SIFT based coefficient selection (refer

Algorithm 5: Watermark Embedding (I, D, W)

Input: I: Center view Image, D: Center view depth image, W: Watermark

Output: W_D: Watermarked Depth image

begin

1. Semantic Image Segmentation [87] is used on the center view image I to extract the foreground.
2. The extracted foreground is removed from the depth image D and generate background depth image D_b (as described in Section 4.1.1.1).
3. SIFT is used on the center view image I to find embedding location.
4. Grouping is done using the SIFT points using Eq. 2.4.
5. Partitioned the background depth D_b to $n \times n$ blocks on the location of feature points where value of n is 8.
6. Using embedding rule (Eq. 4.3) insert the watermark (W) with the selected coefficients of depth image using Eq. 4.4.
7. **return** (W_D)

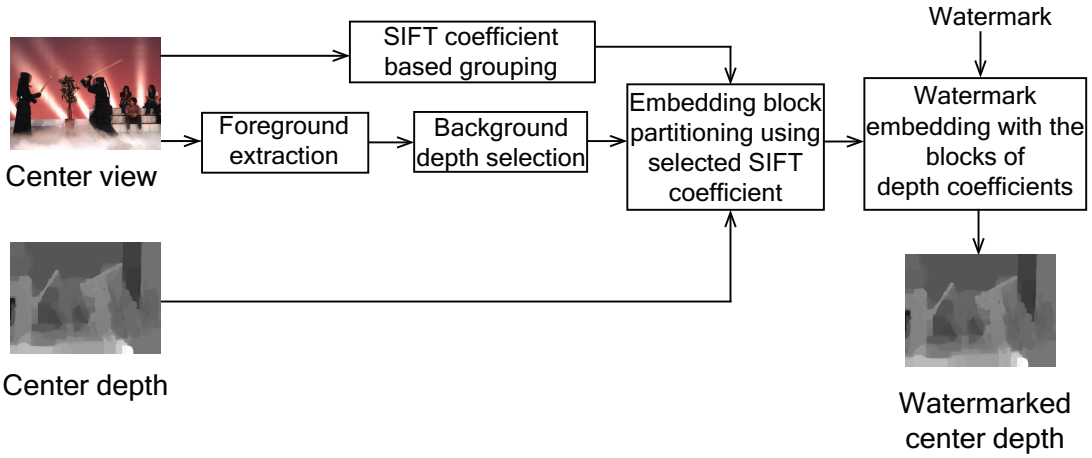


Figure 4.4: *Watermark embedding model*

to §4.1.1.2), identical coefficients are selected for watermark extraction as in the embedding scheme. According to the embedding policy (refer to Eq. 4.3), the watermark is extracted by comparing the majority of the all groups. From each block of a group, watermark is extracted using the Eq. 4.5

$$\begin{aligned} &if \;\; \|mod(C'_i, 2) == 0\| > \|mod(C'_i, 2) == 1\|; \\ &\quad W'_{i_k} = 0 \\ &else \\ &\quad W'_{i_k} = 1 \end{aligned} \tag{4.5}$$

where W'_i is the extracted watermark by comparing the maximum number coefficients C' are even or odd from i^{th} block of the group k and $\|mod(C'_i, 2) == x\|$ defines the coefficient count having $mod(C', 2) = x | x \in \{0, 1\}$. For each group, the final watermark is calculated by combining the watermark from each block using the Eq. 4.6

$$\begin{aligned} &if \;\; \|W'_i == 0\| > \|W'_i == 1\|; \\ &\quad W'_k = 0 \\ &else \\ &\quad W'_k = 1 \end{aligned} \tag{4.6}$$

where W'_k is the final extracted watermark by combining the obtained watermark W'_i from the group k and $\|W'_i == x\|$ defines the coefficient count having $W' = x | x \in \{0, 1\}$. For authentication of the extracted watermark W, Hamming distance is used as described in §2.6.2 Eq. 2.10 for comparison with the original watermark W.

The overall extraction process is narrated in Algorithm 6 and block diagram is depicted in Fig. 4.5.

4.1.2 Results

In this proposed watermarking scheme, the watermark is embedded with the depth image of the 3D image dataset of the Middlebury Stereo 2006 Datasets of 21 images [99] and the frames of 3D video sequences (such as Balloons, Champagne_tower, Kendo, MicroWorld, Shark, etc.) as explained in §2.1 with different camera views. As the scheme is view invariant, the watermark can be extracted

Algorithm 6: Watermark Extraction (I, W_D)

Input: I: Image, W_D: Watermarked Depth image
Output: W':Extracted Watermark
begin

1. Semantic Image Segmentation is used on the center view image I to extract the foreground.
2. The extracted foreground is removed from the watermarked depth W_D and generate watermarked background depth W_{Db} (as described in Section 4.1.1.1).
3. SIFT is used on the center view image I to find embedding location.
4. Grouping is done using the SIFT points using Eq. 2.4.
5. Partitioned the selected watermarked background depth W_{Db} to $n \times n$ blocks on the location of feature points where value of n is 8.
6. Using embedding rule (Eq. 4.3) extract the group wise watermark (W'_k) with the selected coefficients using Eq. 4.5.
7. Count the group wise watermark (W'_k) to generate the final extracted watermark (W') using Eq. 4.6.
8. **return** (W')

from the center view depth map as well as the synthesized depth map of the other views.

4.1.2.1 Visual Quality

To evaluate the visual quality of the proposed scheme depth map image as well as synthesized 3D are used. *Peak Signal-to-Noise Ratio* (PSNR), *Structural Similarity* (SSIM) are measured to evaluate the depth image (refer to §2.6) and the corresponding comparison results for proposed scheme with related depth watermarking scheme proposed by Guan et al. [79] are tabulated in Table 4.1 for these

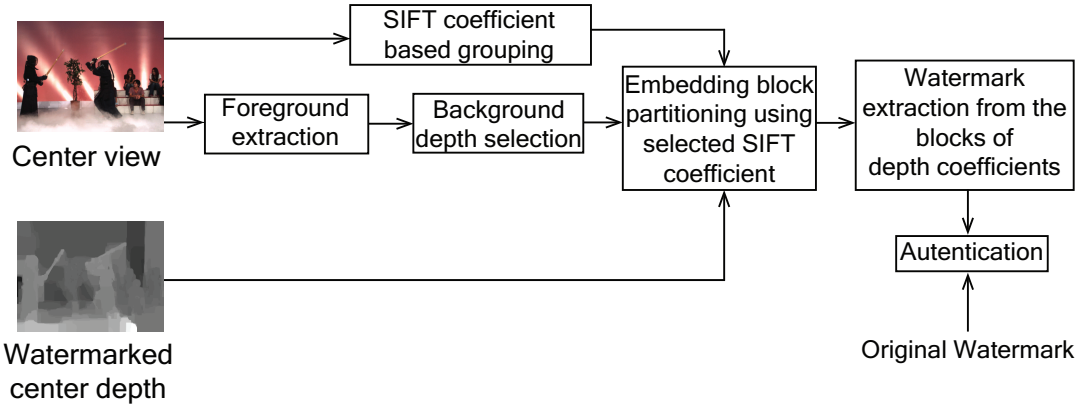


Figure 4.5: *Watermark extraction model*

visual quality metrics. To compare the visual quality of the 3D image, center

Table 4.1: *Average PSNR, SSIM comparison result of the proposed scheme with Guan's scheme [79] for depth*

		Average PSNR	Average SSIM
Proposed scheme	Image set	51.23	0.9984
	Video frame Set	50.46	0.9980
Guan's scheme	Image set	42.48	0.9897
	Video frame set	41.23	0.9843

view is synthesized to left and the right view using the watermarked depth and original depth for proposed scheme and Guan's scheme [79]. *Peak Signal-to-Noise Ratio taking into account Contrast Sensitivity Function* (PSNRHVS) and *Multi Scale Structural Similarity* (MSSSIM)(refer to §2.6) metrics are used to check the visual degradation due to watermark embedding with respect to the human vision of the proposed scheme and the existing scheme [79] as tabulated in Table 4.2.

From Table 4.1 and 4.2, it is observed that the proposed scheme produces better visual quality than Guan's [65] with respect to depth map as well as synthesized 3D image. Due to embedding of watermark in the imperceptible regions with respect to the human vision, the proposed scheme gives better result for PSNRHVS ans MSSSIM comparison.

Table 4.2: *Average PSNRHVS, SSIM comparison result of the proposed scheme with Guan's scheme [79] for synthesized left and right view*

		Average PSNRHVS	Average MSSSIM
Proposed scheme	Image set	52.13	0.9996
	Video frame Set	51.96	0.9994
Guan's scheme	Image set	41.35	0.9848
	Video frame set	42.74	0.9923

4.1.2.2 Robustness

The robustness of the proposed scheme is evaluated against the JPEG compression of the depth image. Since depth image has very less variance, it can sustain even a low quality JPEG compression. To justify the robustness against compression attack, watermark is extracted from JPEG quality =50. Also, different noise addition attacks are implemented over the proposed and existing scheme [79] to evaluate the robustness as shown in Table 4.3.

Table 4.3: *Hamming distance comparison result of the proposed scheme with Guan's scheme [79] for image and video frame sequences against Compression and noise addition attack*

	Sequences	JPEG compression quality = 50	Gaussuan Noise variance= 100	Salt & Papper density= 0.1
Proposed scheme	Image	0.011	0.0253	0.0331
	Video	0.017	0.0297	0.0486
Guan's scheme	Image	0.1719	0.0953	0.159
	Video	0.1223	0.1021	0.148

It is observed from the Table 4.3 that the proposed scheme outperforms Guan's scheme [79] for JPEG compression (quality= 50) and image processing attacks (say noise addition attacks with Gaussian noise of variance = 100 and salt & pepper noise of density = 0.1).

To test the robustness of the proposed scheme against synthesized depth, the depth rendered with different baseline distance rather than the the center view

(where baseline distance is 0 refer to §1.1.2). As the left and the right view refers with baseline distance +0.5 and -0.5 respectively, an intermediate baseline can be represented as +0.25 and -0.25. The comparison result of the extracted watermark is tabulated in Table 4.4.

Table 4.4: *Average hamming distance comparison result of the proposed scheme with Guan's scheme [79] at different baseline distance for image and video frame sequences*

	Baseline distance	Proposed scheme	Guan's scheme
Image sequence	±0.25	0.0487	0.2341
	±0.5	0.1485	0.4843
Video frame sequence	±0.25	0.0481	0.2913
	±0.5	0.1501	0.5196

It is observed from the Table 4.4 that the proposed scheme outperforms the existing scheme [79] for different synthesized depth at different baseline.

4.1.2.3 Discussion

In this work, foreground objects are excluded for watermark embedding as they are more sensitive to the HVS. Embedding of watermark in the less sensitive background locations makes the watermarking scheme imperceptible to the human visual system. It is experimentally observed that change of single bit in depth does not degrade the 3D visual perception as shown in the MSSSIM and the PSNRHVS of the constructed 3D. Since SIFT has been used to find the embedding locations, the embedding locations remains spatial shift invariant due to view synthesis process. In the embedding scheme, a single watermark bit is embedded with a group of block sequences, which improve the robustness against noise addition attack as depicted in Table 4.3.

The proposed scheme (in §4.1) can not be directly used for video sequence as it may create some temporal noise. Moreover, the embedding policy is not suitable for MVD [13] representation. So in next section (refer to 4.2), the same has been extended by incorporation the motion compensation and the MVD based

3D video characteristics in the time of embedding.

4.2 Watermarking in Video Depth Sequences against Depth Modification and 3D-HEVC Compression Attack

In this phase of work, the main challenge is to secure the depth of the 3D video sequence in MVD representation against 3D-HEVC compression attack as well as different depth based attack as explained in §2.5. In the previous section (§4.1), a 3D image depth watermarking scheme is presented for the the DIBR-3D representation. The main motivation was to make that scheme robust against view synthesis attack and noise addition attacks on the depth map. To extend the scheme for depth coefficients of the MVD based 3D-HEVC video compression., temporal noise needs to be reduced by categorising the motion component. Moreover, the similar watermark should be embedded in the common regions (say dependent view) of the left and the right views to improve the robustness against collusion attack, and depth modification attack (by synthesizing the depth). To achieve this, the center view (generated from the left and the right view) is used to cover the dependent region of the left and the right view. Considering the video compression environment, motion compensation temporal filtration is carried out in the proposed scheme (in §4.2.1) to reduce temporal noise. To reduce the visual degradation with respect to the human eye, foreground objects (generally inviting the higher attention of human eyes) are excluded from the embedding zone. Additionally, the SIFT (scale invariant feature transform) feature points are used for watermarking to make the proposed scheme to sustain in different views. A comprehensive set of experiments are carried out (in the §4.2.2) to justify the applicability of the proposed scheme over the existing literature.

4.2.1 Proposed Scheme

The main aim of this work is to add the watermark with depth map of the video sequence such a way that the watermark can be extracted from the embedded view as well as different synthesis views. It is observed that an alteration in the depth map of the video sequence changes the position of the object in the view plane which creates visual discomfort [3]. For insertion of watermark, the left and the right views along with the depths are synthesized to the center view and center depth using the DIBR [6] technique for synchronization of the watermarking sequence in the left and the right view depth. Like the previous scheme (refer to §4.1), foreground objects are removed to achieve the visual quality with respect to HVS. To make the scheme robust against video compression attack, watermark is inserted in the motion compensated regions as explained in the next subsection. Also, a foreground object detection method (like the previous scheme as explained in §4.1) is used to identify less sensitive regions to HVS. A SIFT location based embedding and extraction schemes are proposed to sustain against view synthesis attack as described in the subsequent subsections.

4.2.1.1 Motion Compensated Temporal Filtering (MCTF)

One of the main goals of this work is to make the embedding process robust against video compression as well as frame dropping attack. To achieve this, robust watermark embedding location from the center view video are filtered using motion compensated temporal filtering as described in the motion compensation part of §2.1 [100]. After MCTF, only the fully connected regions [86] in the center view video sequences of a GOP are extracted as shown in Fig. 2.1 (refer to §2.1). After motion compensation, the connected locations are mapped to the first frame of the GOP. In this scheme, the size of the GOP is taken as 8 to make the first frame of the GOP as the 'I' frame in the 3D-HEVC compression process. These locations are used to do the further filtration to make the watermarking scheme invariant to HVS system.

4.2.1.2 Detection of Background and Foreground Objects

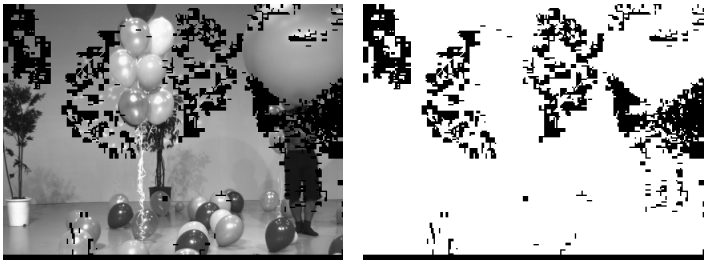
Foreground objects in any video sequence are more sensitive to HVS than the back ground objects. To detect the foreground object, (like the previous scheme §4.1) Semantic Image Segmentation (as explained in §2.2) is used to extract the foreground [87]. Here, Fig. 4.6(b) shows the foreground selected object using the said segmentation method for Fig. 4.6(a) and Fig. 4.6(c) shows foreground removes map for the video frame.

(a) Original frame (b) Foreground object detection (c) Removal of foreground map

Figure 4.6: *Foreground removal for balloons video frame number 17*

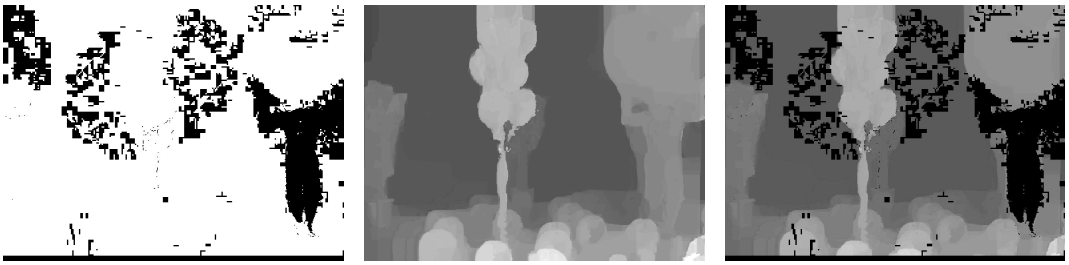
It is observed in the Fig. 4.6(b) that there are some locations which are nearer to the camera are not detected as foreground objects because those regions are not sensitive to the human vision (or say less salient regions like the balloons) [87]. So embedding with those regions in the depth map will not effect the visual quality of the video sequence subsequently.

A pictorial representation for embedding zone selection has been depicted in Fig. 4.7 and Fig. 4.8. After removing unconnected pixels, the embedding locations are shown in Fig. 4.7(a) and corresponding map is generated for the connected pixel as shown in Fig. 4.7(b). Fig. 4.8(a) shows the combination of connected and the background map. Originally this map is used to find the watermark embedding location in the depth as shown in Fig 4.8(b) & 4.8(c) (where the black locations are removed from the embeddable location).



(a) Motion compensated Y chanel (b) Unconnected pixel removal map

Figure 4.7: *Motion compensated connected pixels for balloons video frame number 17 (GOP 17-24)*



(a) Foreground and unconnected pixels removal map (b) Depth (c) Watermarkable depth region

Figure 4.8: *Removal of foreground and unconnected pixels for balloons video frame number 17 (GOP 17-24)*

4.2.1.3 SIFT based coefficient selection

Like the previous work (§4.1), SIFT feature points of the original frame are used to locate the watermarking locations and corresponding depth values of these positions are altered for the embedding to make the proposed scheme invariant to the view synthesis process . The similar coefficient partitioning is implemented on the basis of the scale feature (say σ) of the SIFT as explained in the §4.1.1.2 with the separation (δ) as depicted in the following Eq. 4.7 (similar to §4.1.1.2 Eq. 4.1).

$$\delta_i = \sigma_i^2 \gamma \tag{4.7}$$

where σ_i is the starting value of the i^{th} group and γ is the grouping threshold.

So the value of σ_{i+1} is $(\sigma_i + \delta_i)$. To get better result the value of γ is 0.004 and σ_i is ranged between 2 and 10 for experimental purpose (similar to §4.1.1.2).

4.2.1.4 Watermark Embedding

For watermark embedding, stereo 3D video, and the depth is rendered to the center view using DIBR technique. MCTF is done to each GOP of the center view video frame (as described in Sec. 4.2.1.1) and only the connected regions are detected. Foreground object removal is done on the original center view first frame of each GOP (as described in Sec. 4.2.1.2). To get watermark embedding zone, foreground removal carried out on the motion compensated zone map as depicted in Fig. 4.8(a). To detect the view invariant coefficients SIFT feature points from the first frame of the GOP is taken and a grouped to embed the watermark (like §4.1).

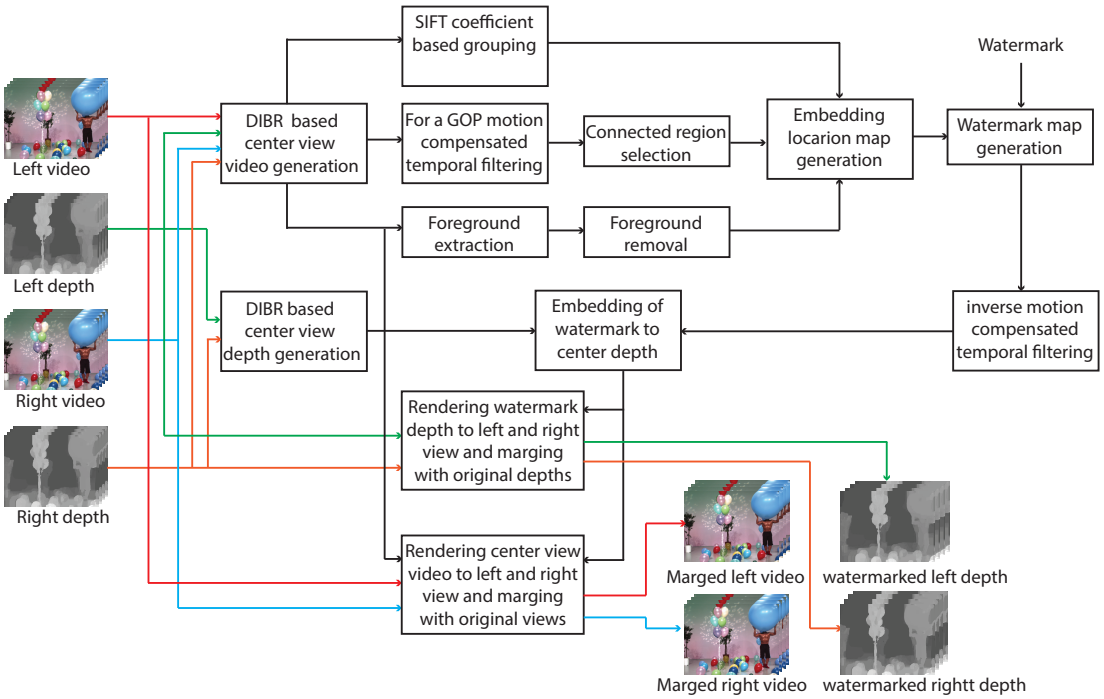


Figure 4.9: *Watermark embedding model*

To embed the watermark in the GOP, the watermark bit is mapped with the

Algorithm 7: Watermark Embedding (V_L, V_R, D_L, D_R, W)

Input: V_L: Left view video, V_R: Right view video, D_L: Left view depth, D_R: Right view depth, W: Watermark

Output: W_{DL}: Watermarked left view depth, W_{DR}: Watermarked right view depth, V'_L: Merged left view video, V'_R: Merged right view video

begin

1. Render left view and the right view video and the depth (V_L, V_R, D_L, D_R) to the center view video and depth (V, D).
2. Split the center view video sequence in GOP
3. **for** *each GOP* **do**
   (a) Motion compensated temporal filtering is done over the GOP of video sequence to compensate to the first frame.
   (b) Only the full connected location map is created (mp_c).
   (c) Semantic Image Segmentation is used on the original first frame to extract foreground.
   (d) The foreground is removed from the mp_c to generate motion compensated foreground subtracted map mp_{cs} (refer to §4.2.1.2).
   (e) SIFT is used on the first frame of GOP to make grouping using Eq. 4.7.
   (f) **for** *each SIFT group* **do**
      i. Partition the map mp_{cs} to $n \times n$ blocks where $n = 8$.
      ii. map the watermark bit sequence (W) with the map mp_{cs} to generate watermark map mp_{wc}.
   (g) Do inverse motion compensation the map mp_{wc} to distribute over the GOP and generate mp_w
   (h) Using embedding rule (Eq. 4.8) insert the watermark (W) obtained from the map mp_w with the selected coefficients using Eq. 4.9 to obtained watermarked center view Depth.
4. Render the center view and the watermarked depth to original left and the right view and merged with the original ones to get Watermarked left view depth (W_{DL}), Watermarked right view depth (W_{DR}), Merged left view video (V'_L), Merged right view video (V'_R).
5. **return** $(W_{DL}, W_{DR}, V'_L, V'_R)$

8×8 of shift group at the motion compensated foreground subtraction map. The map is inverse motion compensated for distributing the embedding location all over the GOP. A frame by frame watermarking is done using the watermark bit from the map. In this scheme, all coefficient values of the depth blocks are changed to even or odd by using the embedding rule in the previous section (refer to §4.1.1.3) as depicted in Eq. 4.8

$$\left.\begin{array}{l} if\ W == 0 \\ \quad mod(C_i, 2) = 0; \\ else \\ \quad mod(C_i, 2) = 1; \end{array}\right\} \tag{4.8}$$

where W is the watermark bit and C defines all the coefficients of depth in i^{th} inverse motion compensated map block. The bin values are altered for embedding the watermark as depicted in Eq. 4.9

$$C'_{j_i} = C_{j_i} - mod((C_{j_i} + W_k), 2) \tag{4.9}$$

where C_{j_i} and C'_{j_i} defines the j^{th} coefficient of the i^{th} block of the group k.

After embedding the watermark, the watermarked center depth and the video frames are rendered to the original left and the right view using the DIBR technique and replace the dependent portion of the left and the right depth and the original video.

The overall embedding process is narrated in Algorithm. 7 and the block diagram is depicted in Fig. 4.9.

4.2.1.5 Watermark Extraction

The watermark extraction scheme is the reverse procedure of the embedding scheme. It is said that the watermark can be extracted from the main view, as well as the synthesis views or the each independent views. So rendering to the center view is not required for watermark extraction process. Like the embedding scheme MCTF and foreground object removal is done on a GOP of the given view. Like the embedding scheme SIFT feature based grouping is done to get embedding

Algorithm 8: Watermark Extraction (V', W_D)

Input: V':Video W_D:Watermarked depth
Output: W':Extracted Watermark
begin

1. Split the video sequence (V') in GOP
2. **for** *each GOP* **do**
   (a) Motion compensated temporal filtering is done over the GOP of video sequence to compensate to the first frame.
   (b) Only the full connected location map is created (mp_c) (refer to Sec. 4.2.1.1).
   (c) Semantic Image Segmentation is used on the original first frame to extract foreground.
   (d) The extracted foreground is removed from the mp_c and motion compensated foreground subtracted map mp_{cs} (refer to Sec. 4.2.1.2).
   (e) SIFT is used on the original first frame to make grouping using Eq. 2.4.
   (f) **for** *each SIFT group* **do**
      i. Partition the map mp_{cs} to $n \times n$ blocks for watermark collection.
   (g) Do inverse motion compensation the collection map to distribute over the GOP and generate mp'_w
   (h) Using embedding rule (Eq. 4.8) extract the group wise watermark (W'_k) with the selected coefficients using Eq. 4.10 & 4.11 from each depth and srote in the collection map mp'_w.
   (i) Use motion compensation on the map mp'_w to combine the watermark bit.
   (j) Generate the final extracted watermark W' using Eq. 4.12.
3. **return** (W')

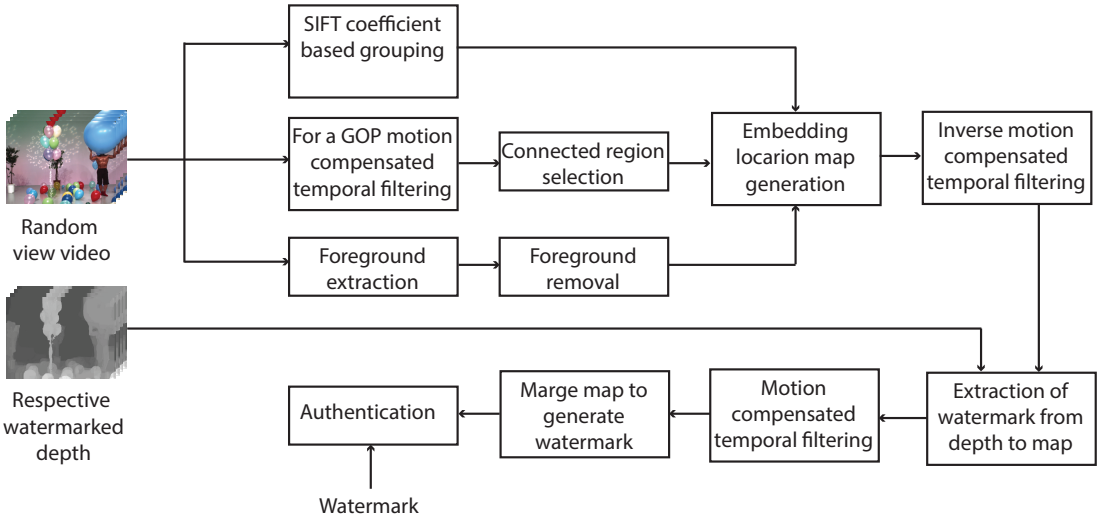


Figure 4.10: *Watermark extraction model*

location map. Now, map in each group is reversed motion compensated (as the watermark is embed in the normal video sequences) to distribute over the GOP. According to the embedding policy (refer to Eq. 4.8), the watermark is extracted by comparing the majority of all groups using the depth pixels. For each depth of the video frame, watermark map is extracted from each block of a group using the Eq. 4.10

$$\begin{aligned} &if \quad \|mod(C'_i, 2) == 0\| > \|mod(C'_i, 2) == 1\|; \\ &\qquad mp'_{wi_k} = 0 \\ &else \\ &\qquad mp'_{wi_k} = 1 \end{aligned} \tag{4.10}$$

where mp'_{wi} is the extracted watermark map by comparing the maximum number coefficients C' are even or odd from i^{th} block of the group k and $\|mod(C'_i, 2) == x\|$ defines the coefficient count having $mod(C', 2) = x | x \in \{0, 1\}$. For each group, the final watermark is calculated by combining the watermark from each block using the Eq. 4.11

$$\begin{aligned} &if \quad \|mp'_{wi} == 0\| > \|mp'_{wi} == 1\|; \\ &\qquad mp'_{wk} = 0 \\ &else \\ &\qquad mp'_{wk} = 1 \end{aligned} \tag{4.11}$$

where mp'_{wk} is the final extracted frame by frame watermark map by combining the obtained watermark mp'_{wi} from the group k and $\|mp'_{wi} == x\|$ defines the coefficient count having $mp'_w = x|x \in \{0,1\}$. Again motion compensation is carried out over the mp'_w to compensate all the bits to the embedding frame ('I' frame)to generate the map mp'_{wc}. and combine the bits to generate the final watermark bit using Eq. 4.12

$$\begin{aligned} &if \ \ \|mp'_{wcf} == 0\| > \|mp'_{wcf} == 1\|; \\ &\quad W'_k = 0 \\ &else \\ &\quad W'_k = 1 \end{aligned} \tag{4.12}$$

where W'_k is the extracted watermark by combining the compensated map mp'_{wc} of frame f from the group k and $\|mp'_{wcf} == x\|$ defines the coefficient count having $mp'_{wc} = x|x \in \{0,1\}$. For authentication of the extracted watermark W, Hamming distance is used as described in §2.6.2 Eq. 2.10 for comparison with the original watermark W.

The overall extraction process is narrated in Algorithm 8 and block diagram is depicted in Fig. 4.10.

4.2.2 Results

In this proposed watermarking scheme, the watermark is embedded with the depth of the 3D video dataset (such as Balloons, Champagne_tower, Kendo, MicroWorld, Shark, etc.) as explained in §2.1 with different camera views. As the scheme is view invariant, the watermark can be extracted from the embedded center view depth map, main views as well as the synthesized depth map of the other views. The recent depth based 3D video watermarking scheme proposed by Guan et al. [79] is used for comparison. Here for embedding and extraction purpose the GOP is taken 8 which is similar to 3D-HEVC encoding GOP.

4.2.2.1 Visual Quality

To evaluate the visual quality the proposed scheme is compared with the Guan's scheme [79] for PSNR, SSIM of the depth video using vqm tool [link: http://mmspg.epfl.ch/vqmt]. The comparison results are tabulated in Table 4.5 for these visual quality measures.

Table 4.5: *PSNR, SSIM comparison result of the proposed scheme with Guan's scheme [79] for depth*

Video sequence	Proposed scheme		Guan's scheme	
	PSNR	SSIM	PSNR	SSIM
Balloons	49.95	0.9964	42.23	0.9747
Shark	50.41	0.9958	44.16	0.9836
Champagne tower	52.44	0.9975	43.17	0.9804
Kendo	48.94	0.9941	41.32	0.9771
MicroWorld	49.87	0.9961	42.95	0.9829
Average	50.11	0.9951	42.86	0.9798

In the time of embedding perturbed center view depth is used to render the center view video to the left and the right views, a distortion may occur due to the original pixel shift. So, the visual quality degradation of the original views (say left and the right views) need to compare with the watermarked views (synthesis using watermarked depth and merged with the originals) in the perspective of the human visual system. PSNRHVS and MSSSIM metrics are used to check the visual degradation due to watermark embedding with respect to the human vision of the proposed scheme and the existing scheme [79] as tabulated in Table 4.6.

From Table 4.5 and 4.6, it is observed that the proposed scheme gives a better result than the Guan's [79] with respect to depth map as well as image visualization. Due to the embedding of the watermark in the imperceptible regions with respect to the human vision, the proposed scheme gives a better result for PSNRHVS and MSSSIM comparison.

Table 4.6: *Average PSNR-HVS, SSIM comparison result of the proposed scheme with Guan's scheme [79] for left and right view*

Video sequence	Proposed scheme		Guan's scheme	
	PSNRHVS	MSSSIM	PSNRHVS	MSSSIM
Balloons	52.19	0.9993	42.82	0.9894
Shark	51.43	0.9996	43.41	0.9471
Champagne tower	50.48	0.9947	41.44	0.9887
Kendo	51.70	0.9984	42.17	0.9913
MicroWorld	53.85	0.9941	40.91	0.9896
Average	51.62	0.9991	43.48	0.9810

4.2.2.2 Robustness

The robustness of the proposed scheme is evaluated against the 3D-HEVC compression attack. As depth map having a lot of smooth regions, the loss does not occur at higher QP level also using of group-based selection neutralize the error rate of the proposed scheme. In this scenario for Guan's scheme [79], the center view is rendered to the left and the right view and compressed with the 3D-HEVC encoder. The comparison of the proposed with existing scheme [79] for different QP level is tabulated in Table 4.7.

Table 4.7: *Hamming distance comparison result of the proposed scheme with Guan's scheme [79] against 3D-HEVC compression at different QP level*

Video sequence	Proposed scheme			Guan's scheme		
	QP=24	QP=28	QP=32	QP=24	QP=28	QP=32
Balloons	0.1642	0.1792	0.192	0.4143	0.4714	0.5127
Shark	0.1594	0.1822	0.2091	0.4417	0.4808	0.5418
Champagne tower	0.1510	0.1788	0.2121	0.4429	0.4847	0.5441
Kendo	0.1380	0.1547	0.1810	0.3973	0.4585	0.4982
MicroWorld	0.1741	0.1906	0.2171	0.4017	0.4441	0.5277
Average	0.1523	0.1718	0.1931	0.4187	0.4632	0.5152

It is observed from the Table 4.7 that the proposed scheme outperforms Guan's scheme [79] for different QP level of video compression attack.

To test the robustness of the proposed scheme against view synthesis , depth of the video sequence is synthesized to the different position between left and the right view. Here the left and the right views are considered as view 0.5 and view -0.5 with respect to the center view baseline distance. The center view is considered as view 0 and intermediate views in between of center view and main views are considered and ± 0.25. The comparison result of the extracted watermark for different views of at QP=24 (for 3D-HEVC compression) are tabulated in Table 4.8.

Table 4.8: *Hamming distance comparison result of the proposed scheme with Guan's scheme [79] at center view (view '0') and intermediate views in between center view and left & right views (view '± 0.25')*

Video sequence	Proposed scheme		Guan's scheme	
	View '0'	View '± 0.25'	View '0'	View '± 0.25'
Balloons	0.0426	0.0812	0.2730	0.4834
Shark	0.0310	0.0774	0.2822	0.4472
Champagne tower	0.0397	0.0716	0.2232	0.5289
Kendo	0.0481	0.0902	0.2875	0.4931
MicroWorld	0.0518	0.0941	0.2513	0.5398
Average	0.0436	0.0861	0.2614	0.5006

It is observed from the Table 4.8 that the proposed scheme outperforms the existing scheme [79] for the different synthesized view of depth. As in both the schemes, the watermark is embedded in the center view, the hamming distance decreases near to the center view.

As in this scheme, the watermark is embedded using the motion compensated temporal filtering. So the embedding policy can resist the temporal scalability (say frame dropping). To test the robustness of the temporal scalability, 25% & 50% frames in the GOP are dropped. In this scenario, the GOP is taken 6 and 4 for extraction of the watermark at frame dropping rate of 25% and 50% (as the embedding time GOP was taken as 8). The comparison result for temporal scalability attack using different frame dropping rate at QP=24 for 3D-HEVC

compression are tabulated in Table 4.9.

Table 4.9: *Hamming distance comparison result of the proposed scheme with Guan's scheme [79] at temporal scalable attack of frame dropping*

Video sequence	Proposed scheme		Guan's scheme	
	25% frame drop	50% frame drop	25% frame drop	50% frame drop
Balloons	0.1712	0.1756	0.4232	0.4495
Shark	0.1741	0.1894	0.4547	0.4681
Champagne tower	0.1637	0.1744	0.4511	0.4603
Kendo	0.1464	0.1521	0.4093	0.4288
MicroWorld	0.17911	0.1832	0.4101	0.4301
Average	0.1671	0.1763	0.4246	0.4472

It is observed from Table4.9 that the proposed scheme sustain against the different frame dropping ratio (say temporal scalable attack) and outperforms the existing scheme.

4.2.2.3 Discussion

In this work, Watermark is inserted in the motion compensated part of the visually less sensitive regions of the 3D video sequence. As center view covers the most of the portion of the left and the right views, the major portion of the embedding locations is available on the left and the right views. It is observed that the foreground objects are more visually sensitive to HVS. Hence, using the motion compensated background improves the robustness as well as the imperceptibility of the scheme. To quantify the visual degradation in the original left and the right views, MSSSIM and the PSNRHVS are used as depicted in §4.2.2. Like the previous scheme (refer to §4.1), using of SIFT locations for embedding purpose will increase the robustness of the proposed scheme against view synthesis attack. Also, using the motion compensated regions for embedding and the GOP watermark distribution technique makes the scheme robust against frame dropping attack.

4.3 Summary

In this chapter, two 3D depth watermarking schemes are presented for the images and video in the DIBR-3D and MVD representation respectively. In the first phase of work, depth of the DIBR-3D images is watermarked such a way that the embedded locations do not make any visual artefacts with respect to HVS. Removing of front object locations from the embedding locations, not only improve the visual quality but the robustness of the scheme. A novel SIFT location based watermark embedding policy is used to improve robustness against view synthesis attack. A comprehensive set of experiments has been carried out to justify the applicability of the proposed scheme over the existing literature against the different attacks.

In the second phase of work, the depth image watermarking scheme in DIBR-3D representation is extended for the 3D video sequences in MVD representation. Here the watermark is inserted in the synthesized center view depth to cover the dependent regions of the left and the right depths. The foreground removal is done using the center view to make the embedding scheme imperceptible to HVS and SIFT location based watermark embedding policy is used on the embeddable locations of the center view depth to improve robustness against view synthesis attack. Moreover, using of the connected pixels after motion compensated temporal filtering, improve the sustainability against 3D-HEVC compression as well as frame dropping attack. A set of experiments has been carried out to demonstrate the applicability of the proposed scheme over the existing literature against the different attacks. In the next chapter, dependent region based watermarking scheme is proposed to improve the robustness against view synthesis attack for depth based multi-view image representation.

Chapter 5

Depth-based View Invariant Blind 3D Image Watermarking for Multi-view Representation

The recent improvements in media technology led to a growing interest in 3D media for its immersive experiences to the viewers. The increased number of cinema screens capable of showing 3D movies as well as the availability of low cost 3D display devices make the 3D media transmission and distribution more popular. As a consequence, content authentication or ownership authentication through watermarking for 3D image and video sequence is becoming an emerging research topic. The 3D depth vision for the human perception is generated by combining the left eye and right eye views where the common pixels move horizontally with common left and the right view respectively [7] as explained in §1.1. As a result, for each of the view, there will be a common region (dependent view region), and uncommon region (independent view region) in the left and the right views. Due to advancement of display devices, auto-stereoscopic display comes for advance viewing experience [10, 11]. To support this another 3D image representation, named depth based multi view 3D image [4, 5] using the DIBR [6] technique, becomes more popular due to its compression efficiency and better visual quality (refer to §1.1.3). In this depth based multi view 3D technique, one

of the boundary views (either left or right view), the independent view region of the other views and the depth image of both the views are communicated to the receiver [5]. At the receiver end, the main views and the intermediate synthesized views are generated using the DIBR technique. The efficient watermarking scheme for authenticating multi view 3D image requires to secure both virtual left and right views from illegal distribution of the 3D content. In addition, it may also require to secure each single view, including the original boundary views (left view and right view) such that they should not also be illegally distributed. Due to certain inherent features, like pixel disparity and changes in the depth image etc., the direct extension of the conventional watermarking schemes for the 2D and the stereo images [41–60] are not very useful for DIBR based encoding as discussed in the literature (refer to §1.3). In other words, for DIBR based multi-view 3D encoding, the main challenge is to embed the watermark in such a way that the watermark should resist the view synthesis process (or synthesis view attack).

In recent literature as explained in §1.3.1, it is observed that most of the existing schemes are not sustainable against the multiview representation. Recently, Lin and Wu [55], proposed a blind multiple watermarking scheme where the watermark is embedded in the left or right view image obtained by inverse rendering of the DIBR technique. But the scheme may not withstand synthesis view attack for the high resolution (having high disparity) images. In another recent work, Kim et al. proposed a dual tree complex wavelet transformation (DT-CWT) base watermarking for 3D images [76] by altering the 15° & -15° (1^{st}, 6^{th}) and 45° & -45° (2^{nd}, 5^{th}) coefficients. This scheme may also suffer from synthesis view attack for the high resolution images or the images where the disparity is larger than the block size. Furthermore, this scheme may be vulnerable against collusion [75] attack as different watermark patterns are practically embedded within the visually similar left, right or any other synthesized views. Moreover, in the previous scheme as proposed in §3.2, the watermarking in MVD representation against 3D-HEVC compression and View Synthesis Attack [101], the

watermark is embedded by altering the 2D-DT-DWT coefficients of the center view of the video frame to make the scheme robust against view synthesis attack. In this method, the embedding block width is taken larger than the disparity to resist the view synthesis attack. However, this makes the embedding block size to be dependent on the corresponding disparity value which may not be good design criteria. Also, it is noticed that the block which is taken 256 to make it larger than the disparity. For normal scenario, the scheme gives better result but for very high resolution image, where the disparity is greater than the 256, the scheme may not sustain against the view synthesis attack.

From the above discussion, it is observed that most of the existing watermarking schemes are vulnerable against the DIBR technique [6]. It is also observed in DIBR based watermark embedding that the different watermark signal may be embedded in coherent locations of left and right view images. This can be capitalized to mount a watermark estimation attack popularly known as collusion attack (type II) [75]. In this attack, the watermark signal can be removed by simply averaging the coherent image regions having different watermarks. In [7], it is observed that the independent regions are less sensitive to the human visual system (HVS). Since, tampering of the independent view does not make any substantial degradation of visual quality with respect to the HVS [2,81], the embedded watermark can be destroyed by damaging the independent view without any severe visual artefacts. So, one of the main challenges for the dependent view region based 3D watermarking scheme is to make the scheme robust against the view synthesis process. In this work, a blind 3D image watermarking scheme is proposed where the identical watermark is embedded in the co-located dependent view regions of the left and right view images such that the proposed scheme can resist synthesis view attack as well as collusion attack. The embedding block size is not dependent on the disparity which is the one of the main drawback of the previous scheme [101]. To make the scheme robust against compression as well as other image processing attacks, DC coefficients of two blocks have been altered to embed the watermark. The rest of the work is organized as follows.

The proposed filtration of dependent and independent view regions using DIBR technique and the layer wise block partitioning scheme have been presented in §5.1 and the watermark embedding and extraction algorithm has been presented in §5.2. Finally experiment results are presented and explained in §5.3

5.1 Embedding Zone Selection

In the depth based multi view 3D image representation, pixels may move horizontally from left to right view (as well as intermediate synthesis views) in time of rendering using the DIBR technique [6]. The common region between the left and right view is called dependent view region which is required to render the main part of left or right view. On the other hand, the uncommon region is called independent view region which is responsible to generate the 3D viewing experience to the respective left or right views. Intuitively, the dependent view region is more suitable for watermarking as it is common to both the views. In this work, identical watermark signal is embedded in the dependent view regions of the left or right views, so identifying the dependent view regions in left or right view is an important task which is described in the next subsection.

5.1.1 Dependent View Region Identification

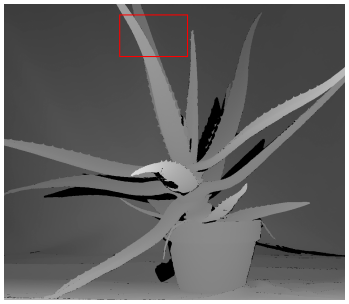
In DIBR based view rendering [6], the fast 1-D view synthesis [5, 14] process is used to synthesized the dependent view region of right view using the left view image and corresponding depth information. In this work, identical watermark is embedded in the dependent view regions of the both left and right views. To generate the dependent view regions of the left view, the depth of the right image (as shown in Fig. 5.1(b)) has been rendered to the left view using disparity information. Similarly, the depth of the left image (as shown in Fig. 5.1(a)) has been rendered to the right view to generate the dependent view regions of the right view. This left and right dependent view generation methods are depicted in Fig. 5.1(c) and Fig. 5.1(d) respectively by horizontally shifting the pixels (as

shown in Fig. 5.1(e) the generation of the left depth synthesis from the right depth). If a single depth view is given (left or right view), the dependent view region of the other view (right or left view) is obtained by rendering the given view by above-mentioned process (refer to Fig. 5.1(b), 5.1(d) for not available of right view depth). Then the dependent and independent view regions (the remaining parts of the generated view) are marked and filtered as depicted in Fig. 5.1. Extracted dependent view regions, marked as white color as shown in Fig. 5.1(f), 5.1(g), will participate in the watermarking purpose.

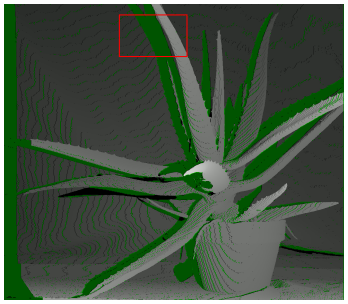
5.1.2 Layer Partitioning using Depth

As described in §5.1.1, the dependent view regions of left and right images are used for embedding. In this work, each dependent view region has been watermarked separately with the identical watermark. One of the main challenges of this work is to embed the watermark in such a way that the embedding locations should remain intact with respect to different views. In other words, the embedding scheme should secure both the (left and the right) views as well as the intermediate synthesis views (generated using DIBR technique). In depth based multi view 3D image representation, the common pixels move to the left or the right direction with respect to the disparity vector which is related to the depth parameter. The depth image has in general 256 gray values which implies that there are 256 types of different ways to (depth wise) correlate the left and right views and also for any other intermediate synthesis views. So, in this work, the dependent view region is partitioned in 256 number of layers according to the depth and the identical watermark has been embedded in same depth layer for both left and right view images. This embedding policy helps watermark to be remained consistent with respect to the disparities of pixels in time of the view generation. In the layer partitioning of the dependent view region, regions which do not come under a specific layer are considered as the independent view region for that layer. The above-mentioned depth partitioning process is depicted in

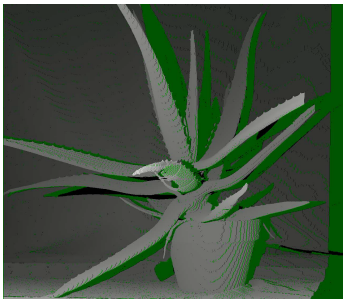
(a) Depth of left image



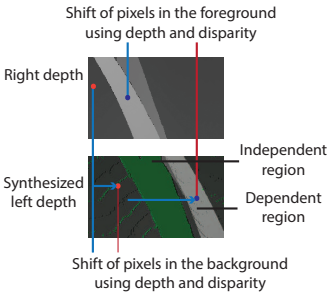
(b) Depth of right image



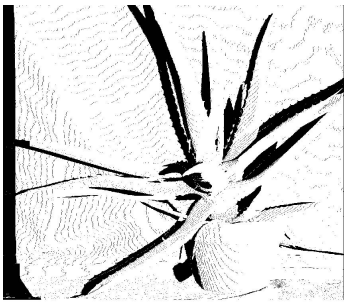
(c) Rendered left depth image from right depth image using DIBR



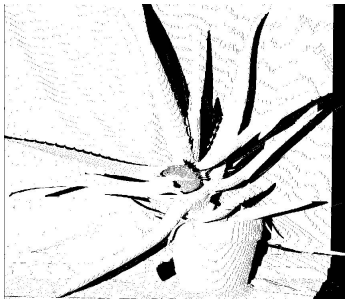
(d) Rendered right depth image from left depth image using DIBR



(e) Shift of pixels from right view to left view using disparity



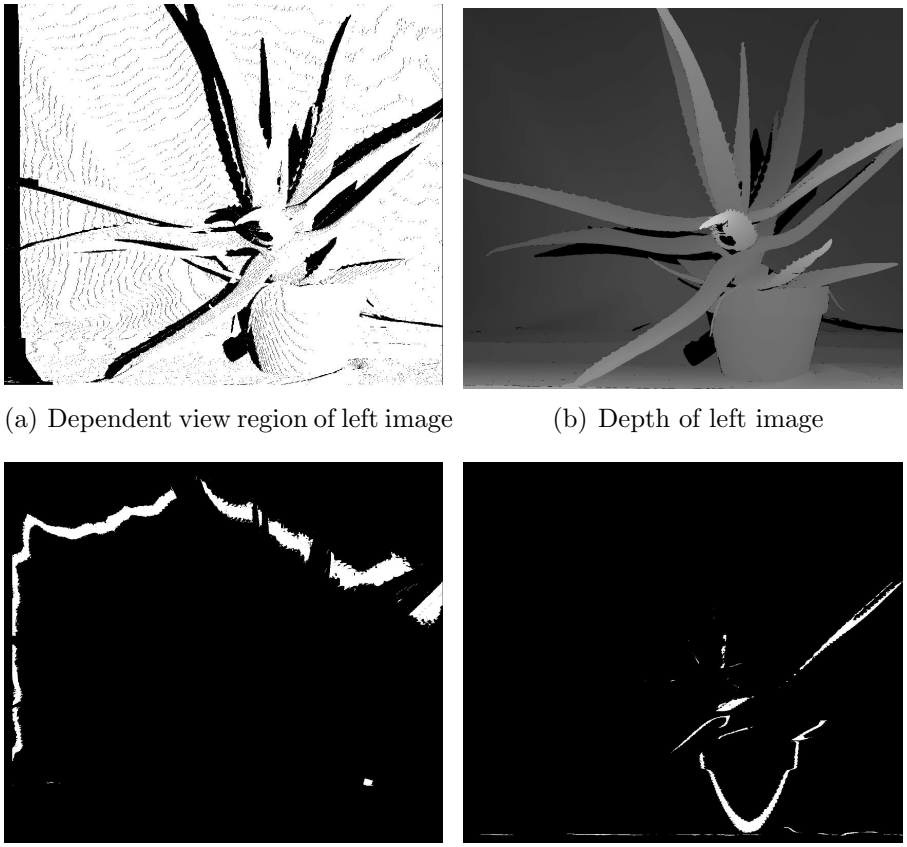
(f) Dependent and independent view regions of left image



(g) Dependent and independent view regions of right image

Figure 5.1: *Left and right dependent view region extraction using depth value for Aloe image of camera view 1(left view) and camera view 5(right view).*

Fig 5.2.



(a) Dependent view region of left image

(b) Depth of left image

(c) Layer filtered dependent view region for depth=50

(d) Layer filtered dependent view region for depth=110

Figure 5.2: *Dependent view layer partitioning using depth value for Aloe image of camera view 1(left view).*

5.1.3 Block Partitioning

As discussed in the previous section, the left or the right view consists of 256 different depth layers. The watermark is embedded in each depth layers of the left or the right view. Here, it is important to note that the horizontal pixel shift between dependent view regions of the left and right view may be different for different depth layers. In this work, for a given depth, the dependent view region for that depth layer of the left view (and the right view) is partitioned into non overlapping $n \times n$ blocks and the same watermark bit is embedded in such blocks in both the left and the right view image. In other words, block partitioning

process ensures that no block pixels are from independent view regions of the left or right view and all block pixels are from the same depth layer as given in Condition 1.

$$\left.\begin{array}{l} \textit{Each pixel of a } n \times n \textit{ block} \\ \textit{is in the dependent view.} \\ \\ \textit{Each pixel of a } n \times n \textit{ block} \\ \textit{has the same depth layer.} \end{array}\right\} \textbf{Condition 1}$$

A block is used for embedding if the Condition 1 is satisfied for that block. In this work, the horizontally shifted coherent blocks in left and right views at given depth layer are embedded with the similar watermark such that collusion attack can be resisted. Moreover, at the time of extraction, having knowledge of the depth and the disparity, watermark can be extracted from left, right or any other synthesized views.

5.1.4 Block Selection

A pair of blocks is selected from the set of blocks obtained in the previous section (§5.1.3) to embed a single watermark bit. Before that blocks are grouped according to the number of row and depth layer. In this grouping, blocks are chosen such a way that all the blocks in each group are placed in a single row (say co-linear) and must belong to the same dependent view region at a given depth layer. In other words, the dependent view regions consist of block rows and two of such blocks from a group are chosen to embed a bit. Moreover, the width of such block rows are ensured to be larger than the disparity between left and right view. No blocks are used for embedding from a block row having width less than the disparity. The detailed grouping criteria has been explained

in Condition 2.

Each group of $n \times n$ blocks should be aligned in the same horizontal line.

Each group of $n \times n$ blocks should remain in the same depth layer.

Width of the depth layer in horizontal line $\not<$ disparity

} **Condition 2**

This condition has been imposed to ensure that the dependent view regions for both left and right views should remain intact in any other synthesis views to make the scheme view invariant.

A security key (κ) is used to select each pair of blocks from each group to make the scheme more secure against naive randomization based attacks by increasing the cryptographic search space for the block selection. How the security key (κ) is used to increase the cryptographic security is described below:

Let η be the number of blocks in each group. Then the cryptographic length of the function will be a combination of number of groups and the expected number of coefficient in each group as shown in Eq. 5.1

$$\kappa =^{\eta} C_2 \frac{I_h}{n} 256 \tag{5.1}$$

where $\frac{I_h}{n}256$ represents the number of groups available for block selection and I_h represents the image height and n represents the block size. In this scheme, the same selection method is implemented for the left and the right view image. In this scheme each pair of block belongs to a single depth. For a specific depth, the horizontal shift in the left and the right view is fixed. As a result, relative position of the coherent block remains same against DIBR based view synthesis process.

5.1.5 Visual Threshold Checking

Due to embedding of watermark, visual artefacts may occur [29] and may degrade the visual quality. To maintain the visual quality of the watermarked image,

another round of checking (say visual threshold checking) is done over the selected block pairs like the [29]. For this checking, block wise DCT has been done for both the blocks for a given block pair. Let C_1 and C_2 are the DC coefficients of the first and second block respectively. In this threshold checking process, a block pair is selected for embedding if it satisfies the condition given in Eq. 5.2, otherwise their DC coefficients are altered according to the Eq. 5.2 to mark the pair unsuitable for embedding even after any possible attacks. This marking scheme helps to identify the embedded block pair at the time of extraction. The experimental results confirm that the noise addition due to this alteration is negligible with respect to the human visual system. The visual threshold is defined using the Eq. 5.2

$$\left.\begin{array}{ll} if & |C_{1_i} - C_{2_i}| \leq \tau \\ & Accept\ for\ watermarking \\ else\ if & C_{1_i} > C_{2_i} \\ & C_{1_i} = C_{1_i} + (C_{1_i} + C_{2_i})\alpha \\ & C_{2_i} = C_{2_i} - (C_{1_i} + C_{2_i})\alpha \\ else & \\ & C_{1_i} = C_{1_i} - (C_{1_i} + C_{2_i})\alpha \\ & C_{2_i} = C_{2_i} + (C_{1_i} + C_{2_i})\alpha \end{array}\right\} \tag{5.2}$$

where α is the embedding strength and τ is the threshold. For experiment purpose, value of α has been taken 0.03 and the threshold τ is taken as $(C_{1_i} + C_{2_i})\alpha$.

5.2 Watermark Embedding and Extraction Process

In this section, the watermark embedding and extraction algorithms are presented. As discussed in §5.1.1, the watermark is embedded in the dependent view regions for the left and the right views of the 3D images and the DC coefficients of the selected block pairs are used for embedding.

5.2.1 Embedding of Watermark

In this work, a binary image of given dimension (according to the prescribed payload) is used as the watermark sequence. A simple embedding rule has been employed for the proposed scheme as described in the Eq. 5.3

$$\left.\begin{array}{l} if \quad W_i = 0 \quad then \; C_{1_i} \geq C_{2_i} \\ if \quad W_i = 1 \quad then \; C_{1_i} < C_{2_i} \end{array}\right\} \tag{5.3}$$

where W_i is the i^{th} watermark bit to be embedded and (C_1 and C_2) are the DC coefficients of the selected block pair. Now, according to the above embedding rule, watermark is embedded in a block pair using the following Eq. 5.4

$$\left.\begin{array}{l} \left.\begin{array}{l} if \quad C_{1_i} - C_{2_i} \geq \frac{C_{1_i}+C_{2_i}}{2}\alpha \\ \quad\; no\ change \\ else \\ \quad\; C'_{2_i} = C_{1_i} - \frac{C_{1_i}+C_{2_i}}{2}\alpha \end{array}\right\} \quad where\ W_i = 0 \\ \left.\begin{array}{l} if \quad C_{2_i} - C_{1_i} \geq \frac{C_{1_i}+C_{2_i}}{2}\alpha \\ \quad\; no\ change \\ else \\ \quad\; C'_{1_i} = C_{2_i} - \frac{C_{1_i}+C_{2_i}}{2}\alpha \end{array}\right\} \quad where\ W_i = 1 \end{array}\right\} \tag{5.4}$$

where W_i is the i^{th} watermarking bit and α is the embedding strength. For experiment purpose, value of α has been taken 0.03. After embedding the watermark by altering the DC coefficient of the blocks, block wise inverse DCT has been done to get back the watermarked left view image. Both left and right view have been watermarked using the same procedure. The overall watermarking scheme is described in Algorithm 9 and the block diagram of the embedding procedure is depicted in Fig. 5.3.

5.2.2 Extraction of Watermark

Extraction scheme is mostly the reverse process of the embedding scheme. As the proposed watermarking scheme is blind in nature, the cover image is not required for watermark extraction. Only the left and the right depth are necessary

Algorithm 9: Watermark Embedding $(I_L, I_R, D_L, D_R, \alpha, W)$.

Input: I_L: Left view image, I_R: Right view image, D_L: Left view image depth, D_R: Right view image depth, α: Watermark strength and W: Watermark bit stream.

Output: W_{IL}:Watermarked left view image, W_{IR}:Watermarked right view image.

begin

1. Using DIBR technique and the right image depth D_R the left dependent view region V_{dl} for the left image I_L is generated as described in Fig.5.1.
2. Like Step 1 the right dependent view region V_{dr} for right I_R is generated as described in Fig.5.1.
3. Using depth D_L and D_R the dependent view region V_{dl} and V_{dr} is partitioned in 256 number of layers as described in §5.1.2, Fig. 5.2.
4. Use block partitioning as described in §5.1.3
   for *depth value = 0 : 255* **do**
   (a) **if** *$n \times n$ unique block from the same dependent view region* **then**
   if *All coefficients of $n \times n$ blocks belongs to same depth level* **then**
   Partition the block.
5. **if** *pair of two $n \times n$ blocks aligned to same horizontal line & same depth layer* **then**
   (a) **if** *Width of the depth layer in horizontal line $\not<$ disparity* **then**
   Select the block for embedding.
6. Do block wise DCT on the pair of selected blocks of original image and let C_1 and C_2 are the DC coefficients.
7. **if** *C_1 and C_2 satisfied the visual threshold as depicted in Eq. 5.2* **then**
   Embed the watermark with the C_1 and C_2 using the Eq. 5.4.
8. Do inverse DCT over the block to get the watermarked image.
9. Repeat Step 1 to Step 8 over left and right view image to generate the watermarked left view image W_{IL} and right view image W_{IR}.
10. **return** (W_{IL},W_{IR})

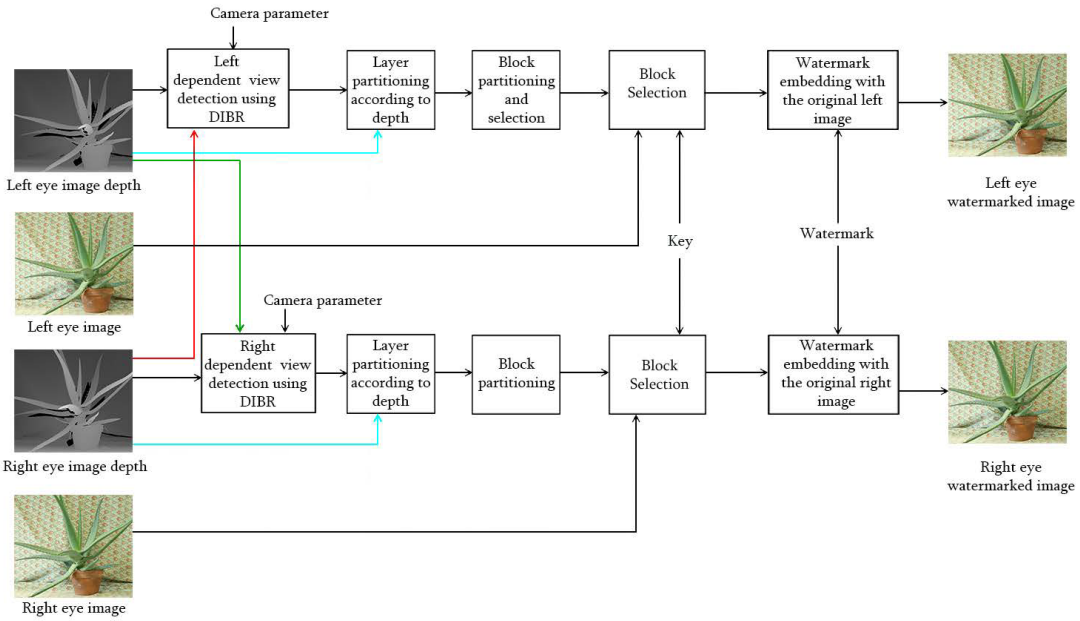
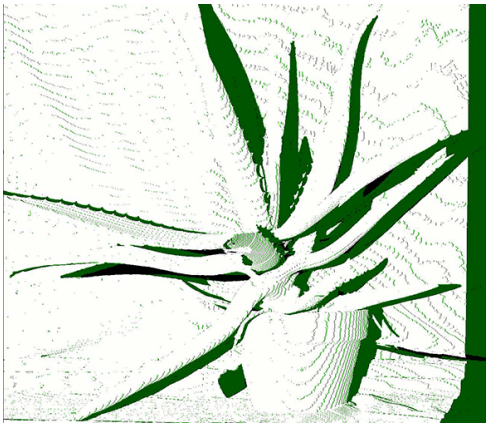


Figure 5.3: *Watermark embedding model.*

to detect the dependent view regions at the time of extraction. Similar to the embedding process, dependent view region identification (refer to §5.1.2) has been employed to find the dependent view regions for left and right views. The obtained left or right dependent view region can be used to identify the dependent view region for any other synthesized view images using the corresponding disparity information as described in the next subsection.

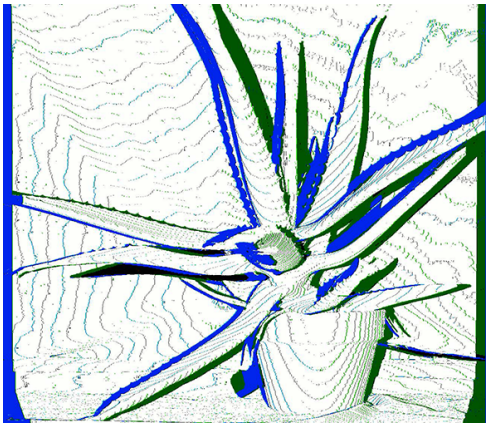
5.2.2.1 Dependent View Region Identification for Synthesized Views

In multi-view 3D image representation, the dependent view region for right (left) view can be generated from the left (right) view depth information using corresponding disparity between left and right views [6]. In this work, this concept is extended to find the dependent view regions of any other middle level synthesized views having the knowledge of the corresponding disparity with the main view. The procedure is depicted in Fig. 5.4. In Fig. 5.4(a), the green color portion is the independent view region and the white color portion is the dependent view region which is obtained from the left view depth and the disparity value.



(a) Dependent view region of right view generated from left view

(b) View to detect dependent view region



(c) Rendering of dependent view region from right view to geven view

(d) Dependent view region from left and right view to the given view

Figure 5.4: *Rendering of left and right dependent view region to synthesized required view.*

The given watermarked view is shown in Fig. 5.4(b). Now, using the baseline distance (or disparity) and the depth of the right view, the dependent portion of the right view can be rendered to the given watermarked view (as shown in Fig. 5.4(c)). As shown in the Fig. 5.4(c), blue and green color region is the independent view region obtained by the rendering process with respect to the left and right view. Thus, the dependent region of the given intermediate synthesized view (Fig. 5.4(d)) with respect to the original left and the right view can be ob-

tained for the watermark extraction. After getting the dependent view regions of the given view, block partitioning (refer to §5.1.3) and block selection (refer to §5.1.4) and followed by a visual threshold checking have been carried out to find the embedded watermark locations. The visual threshold checking method is narrated in the next subsection.

5.2.2.2 Visual Threshold Checking

The visual threshold checking is done to find the embedded watermarked block pair at the time of extraction. It has been observed from the Eq. 5.2 that unsuitable coefficients are marked such a way that the absolute difference between C_1 and C_2 becomes greater than $3(C_{1_i} + C_{2_i})\alpha$. So, at the time of extraction, the block pairs having $|C'_{1_i} - C'_{2_i}| < 2(C'_{1_i} + C'_{2_i})\alpha$ are selected as the valid candidates for watermark extraction as presented in the Eq. 5.5

$$\begin{array}{l} if \quad |C'_{1_i} - C'_{2_i}| < 2(C'_{1_i} + C'_{2_i})\alpha \\ Accept\ for\ watermark\ extraction \end{array} \tag{5.5}$$

where α is the embedding strength and (C'_1 and C'_2) are the watermarked DC coefficients of the selected block pair.

5.2.2.3 Extraction of Watermark from the Selected Blocks

The selected DC coefficients C'_1 and C'_2 are compared for extraction of the watermark according to the Eq. 5.6

$$\left.\begin{array}{ll} W'_i = 0 & if\ C'_{1_i} \geq C'_{2_i} \\ W'_i = 1 & if\ C'_{1_i} < C'_{2_i} \end{array}\right\} \tag{5.6}$$

where W'_i is the extracted i^{th} watermark bit. After extraction of the watermark bit sequence, Hamming distance is used to compare the extracted watermark bits (W') with the original watermark bits (W) as described in §2.6.2 Eq. 2.10. The step by step extraction scheme is narrated in Algorithm 10 and a block diagram of the overall extraction process is depicted in the Fig. 5.5.

Algorithm 10: Watermark Extraction (W_I, D_L, D_R).

Input: W_I: Image for watermark extraction D_L: Left view image depth, D_R: Right view image depth.
Output: W': Extracted watermark.
begin

1. Using DIBR and the right image depth D_R the left dependent view region V_{dl} for the left image depth D_L is generated.
2. Like Step 1 the right dependent view region V_{dr} for right depth D_R is generated.
3. The dependent view regions are filter and marked.
4. Using depth D_L and D_R the dependent view region V_{dl} and V_{dr} is separated in 256 number of layers as described in §5.1.2, Fig. 5.2.
5. **for** *depth value = 0 : 255* **do**
   (a) Select $n \times n$ unique block from the dependent view region of the layer.
   (b) **if** *All coefficient of* $n \times n$ *belongs to dependent view region* **then**
      if *All coefficients of* $n \times n$ *blocks belongs to same depth level* **then**
         Partition the block.
6. **if** *pair of two* $n \times n$ *blocks aligned to same horizontal line and same depth layer* **then**
   if *Width of the depth layer in horizontal line* $\not<$ *disparity* **then**
      Select the block for extraction.
7. Do block wise DCT on the pair of selected blocks of watermarked image IW and let C'_1 and C'_2 are the DC coefficients.
8. **if** C'_1 *and* C'_2 *satisfied the extraction visual threshold as depicted in Eq. 5.5* **then**
   Extract the watermark W' by comparing C'_1 and C'_2 by using the Eq. 5.6.
9. **return** (W')

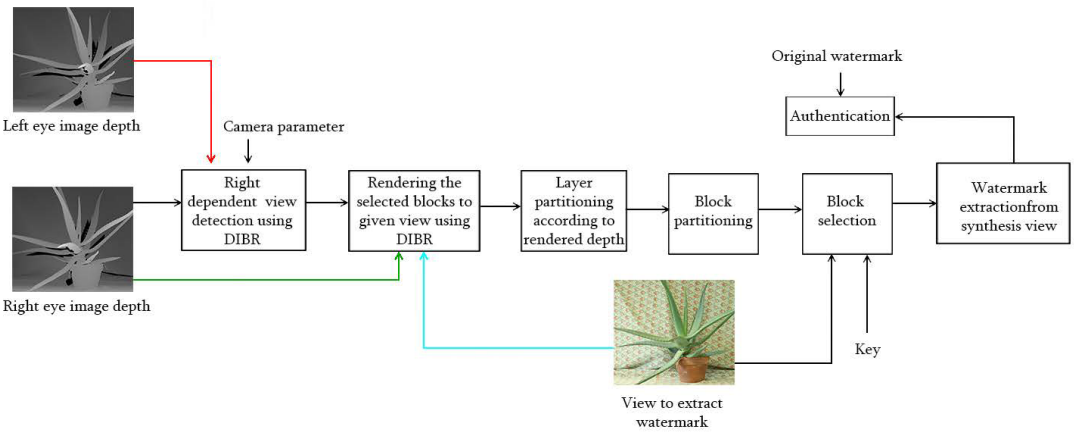


Figure 5.5: *Watermark extraction model.*

5.2.3 Robustness of the Proposed Scheme

In the proposed scheme, one of the major contributions is the watermarking zone selection process to make the scheme robust against view synthesis process. Firstly, it has been shown that the dependent view region of any synthesized view can be obtained by a trivial extension of DIBR [6] technique. It is also observed that the horizontal shift of the dependent view region between left and right view image for a particular depth layer depends on the depth parameter [1] as described in the following Eq. 5.7

$$\begin{aligned} S_{X_L} &= -d_{isp} T_L \frac{D_{X_L}}{255} \\ S_{X_R} &= d_{isp} T_R \frac{D_{X_R}}{255} \end{aligned} \tag{5.7}$$

where d_{isp} is the disparity, D_{X_L} & D_{X_R} are the depth value of the left and right view, respectively for location X and S_{X_L} & S_{X_R} are the corresponding shifted locations. T is the baseline distance between the reference view and the synthesis view. For the left and the right main view, the base line distance is taken as $T = 1$.

Since the horizontal shift is different for different depth layers, if the watermark is embedded without considering the depth information for a particular dependent view region, it may introduce some spatial de-synchronization noise at the time of extraction. Now, if the horizontal shift of the dependent view region

between left and right view image for a particular depth layer is known a priory, the spatial de-synchronization error at the time of extraction can be reduced. To exploit this fact, the layer wise separate embedding has been employed for the proposed watermarking scheme. In this process, at a given depth layer, similar watermark has been embedded in horizontally shifted coherent dependent zones for a different view so that above-mentioned de-synchronization noise can be substantially reduced. Moreover, it may also help to resist the inter view collusion attacks.

In this work, the block partitioning has been done in row-wise for a dependent view region. If the width of the row is less than the disparity between left and right view, this portion of the dependent view region (although appears both in left and right views) may not be available in the intermediate synthesis view as illustrated in the Fig. 5.6.

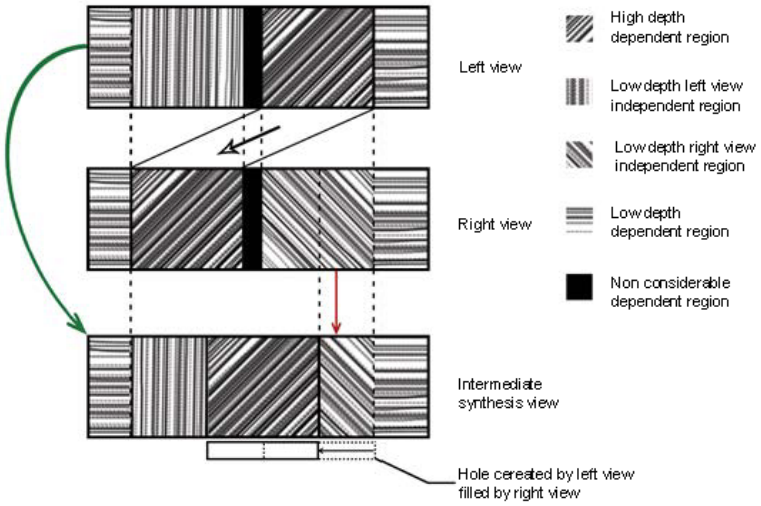


Figure 5.6: *Watermarking zone shift.*

In the Fig. 5.6, the black marked dependent view region is one of such regions. Embedding watermark in such region makes the scheme fragile against intermediate view synthesis process. To handle this situation, in the proposed scheme, the block pairs are selected from the dependent view regions rows having width

greater than the disparity value (refer to 5.1.4).

5.3 Experimental Results

In this proposed watermarking scheme, the watermark is embedded with the coherent locations of the left and the right views of the Middlebury Stereo 2006 Datasets of 21 images [99] (refer to §2.7 Table 2.1). As the scheme is view invariant, the watermark can be extracted from the main views (i.e. the left and the right view) as well as the intermediate synthesized views. The full experimental setup for the proposed scheme is tabulated in the Table 5.1, where α is the embedding strength and the disparity value represents the shift of pixels for DIBR technique.

Table 5.1: *Experimental set-up.*

Image name	Aloe	Baby1	Baby2	Baby3	Bowling1	Bowling2	Cloth1	Cloth2	Cloth3	Cloth4	Flowerpots	Lampshade1	Lampshade2	Midd1	Midd2	Monopoly	Plastic	Rocks1	Rocks2	Wood1	Wood2
Resolution	1282×1110	1240×1110	1240×1110	1312×1110	1252×1110	1330×1110	1252×1110	1300×1110	1252×1110	1300×1110	1312×1110	1300×1110	1300×1110	1396×1110	1366×1110	1330×1110	1270×1110	1276×1110	1276×1110	1372×1110	1306×1110
Disparity	270	300	300	250	290	240	290	260	290	260	251	260	260	196	214	237	280	274	274	210	254
Camera view	1& 5																				
α **value**	0.03																				
Block for embedding	4×4, 8×8, 16×16																				

With this experimental setup, three different experiments have been carried out with three different embedding block sizes (4×4, 8×8, 16×16). The proposed scheme has been compared with different recent schemes such as Lin and Wu's scheme [55], Kim's scheme [76], Kim's scheme* [76], Franco's scheme [62] and multi-view based based Rana's scheme [101] to justify its applicability over the existing literature. The Kim's scheme* is a modified version of the Kim's scheme [76] using the similar embedding payload as the proposed scheme where

the embedding sub-block size in Kim's scheme* [76] has been changed to $(I_w/32)\times(I_h/32)$ to attain the desired payload for comparison purpose, where I_w and I_h represent the image width and height respectively. To compare under similar fidelity as the proposed scheme, the DIBR-3D image watermarking schemes (Lin and Wu's scheme [55], Kim's scheme [76] and Kim's scheme* [76]) have been implemented on depth based multi view 3D image [5]. Here, the center view is generated by rendering the left and the right view and the watermark is embedded using the existing schemes [55, 76]. Using the DIBR technique, the dependent view regions are reverse rendered and replaced with the left and the right view to compare with the proposed scheme. Since, Franco's [62] and Rana's scheme [101] are video watermarking scheme, the embedded video frame(s) are compared with the watermarked image of the proposed scheme.

5.3.1 Visual Quality

In this subsection, the fidelity of the watermarked image of the proposed scheme is evaluated using different visual quality metrics such as PSNR (Peak signal-to-noise ratio), SSIM (Structural Similarity), VIFp (Visual Information Fidelity, pixel domain version) (refer to §2.6) using the VQMT (Video Quality Measurement Tool) [1]. The visual quality performance of the proposed scheme against above metrics for different embedding block size (4×4, 8×8 and 16×16) are tabulated in Table 5.2. The visual quality comparison results of the proposed scheme with the existing Lin & Wu's scheme [55], Kim's scheme [76], Franco's scheme [62] and Rana's scheme [101] are tabulated in Table 5.3.

It is observed from the Table 5.2 that the performances of the proposed scheme against different visual quality metrics like PSNR, SSIM, VIFp, etc. are quite acceptable. It can also be observed from the Table 5.3 that the proposed scheme shows almost comparable (sometimes better) results in comparison with the existing schemes [55, 62, 76] against above visual quality metrics with the same

[1] link: http://mmspg.epfl.ch/vqmt

Table 5.2: *Visual quality measurement of the proposed scheme for different embedding block sizes.*

Image name	4×4 block			8×8 block			16×16 block		
	PSNR	**SSIM**	**VIFp**	**PSNR**	**SSIM**	**VIFp**	**PSNR**	**SSIM**	**VIFp**
Aloe	42.8	0.99012	0.9846	42.1	0.978	0.9813	40.5	0.9412	0.9779
Baby1	47.3	0.9932	0.9927	46.9	0.9843	0.9893	44.8	0.9623	0.9859
Baby2	47.2	0.994	0.9922	46.8	0.9844	0.9888	44.2	0.9488	0.9854
Baby3	49.8	0.9961	0.9938	46.5	0.9882	0.9904	43.7	0.9518	0.987
Bowling1	48.6	0.9927	0.9821	48.6	0.9743	0.9788	44.6	0.9588	0.9754
Bowling2	47.2	0.9938	0.9898	46.5	0.9788	0.9864	43.3	0.9522	0.983
Cloth1	42.5	0.9874	0.9753	42.1	0.9736	0.972	41.6	0.96	0.9686
Cloth2	45.9	0.9924	0.9861	44.8	0.982	0.9828	42.2	0.9442	0.9794
Cloth3	45.6	0.9928	0.9876	42.8	0.9816	0.9843	41	0.9388	0.9809
Cloth4	49.2	0.9908	0.9869	44.3	0.9872	0.9836	41.9	0.9515	0.9802
Flowerpots	50.1	0.9937	0.9898	47.9	0.988	0.9864	44.3	0.9612	0.983
Lampshade1	50.1	0.9962	0.9908	49.5	0.9898	0.9874	44.9	0.9613	0.984
Lampshade2	50.7	0.9951	0.9893	49.4	0.9904	0.9859	46.1	0.9669	0.9825
Midd1	45	0.994	0.9797	44.2	0.9722	0.9764	42	0.92	0.973
Midd2	48.2	0.9926	0.9889	44.3	0.9726	0.9855	42.5	0.9396	0.9821
Monopoly	52.2	0.9929	0.9951	45.2	0.9848	0.9917	43.5	0.9601	0.9883
Plastic	48.7	0.9956	0.9894	47.8	0.9878	0.986	47.4	0.9642	0.9826
Rocks1	46	0.9911	0.9832	43.5	0.9744	0.9799	41	0.9313	0.9765
Rocks2	46.5	0.9917	0.9862	44.1	0.9782	0.9829	41.6	0.9367	0.9795
Wood1	50.5	0.9951	0.9941	47.8	0.9885	0.9907	46.4	0.9711	0.9873
Wood2	48.7	0.9933	0.9878	47.1	0.9865	0.9845	43.5	0.941	0.9811

Table 5.3: *Average PSNR and SSIM comparison result of the proposed scheme with Lin & Wu's scheme [55], Kim's scheme [76], Franco's scheme [62] and Rana's scheme [101].*

		Average PSNR	**Average SSIM**	**Average VIFp**
Proposed scheme	4 × 4	47.7	0.9931	0.9837
	8 × 8	45.8	0.9822	0.9846
	16 × 16	43.4	0.9507	0.9813
Lin's scheme	4 × 4	44.2	0.9942	0.9858
	8 × 8	43.1	0.9901	0.9821
	16 × 16	41.7	0.9822	0.9771
Kim's scheme		43.3	0.9887	0.9798
Kim's scheme*		42.9	0.9868	0.9757
Franco's scheme		41.2	0.9628	0.9519
Rana's scheme		44.2	0.9894	0.9837

payload. Intuitively, an adaptive visual quality threshold (refer to §5.1.5) by carefully choosing the embedding strength (α) may be a reason for maintaining the descent visual quality for the proposed scheme. The embedding strength (α) has been used as a controlling parameter which adaptively handles the visual quality and robustness trade-off.

5.3.2 Robustness

One of the primary goals of this work is to make the proposed scheme invariant to the view synthesis process. Assuming the view synthesis process as an attack (synthesis view attack), robustness of the proposed scheme is mainly evaluated against this attack in this subsection. Moreover, since most of the images are being communicated in compressed format (such as JPEG), the robustness of the proposed scheme against JPEG compression at different quality levels and different noise addition attacks are also analysed in this subsection.

In this scheme, a random bit stream (as a secret massage signal) is used to embed using the key (κ). To show the visually degradation of the watermark, a 64×64 binary image (as shown in Fig.5.7(a)) is used to generate the watermark bit sequence. The extracted watermark at JPEG compression level 50 from left view of the *Aloe* image with the block sizes 4×4, 8×8, and 16×16 are presented in the Fig. 5.7(b), 5.7(c) and 5.7(d) respectively. Fig. 5.7 reveals the accepted perceptual quality of the extracted watermark using the proposed scheme.

In this work, Hamming distance (refer to Eq. 2.10) between the extracted watermark and the original watermark is used as the robustness metric where smaller Hamming distance denotes higher robustness. The proposed scheme is compared with the existing state-of-the-art schemes [55, 62, 76]. For comparison of extracted watermark from the left, right, center view, the left and the right watermarked view (after JPEG compression quality level 75) is used to generate the synthesized center view. Fig. 5.8, 5.9 & 5.10 show the comparison results of the proposed scheme with Lin's scheme [55], Kim's scheme [76], Kim's scheme* [76],

(a) Original watermark (64×64)

(b) Extracted watermark for embedding blocksize 4×4

(c) Extracted watermark for embedding blocksize 8×8

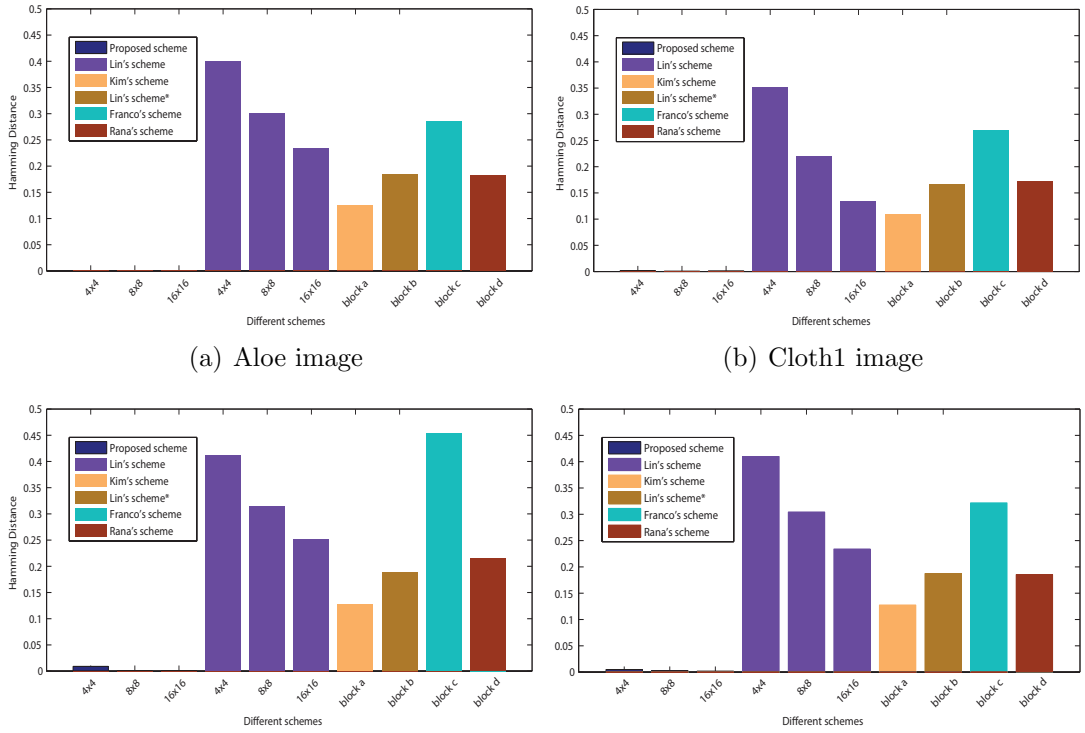
(d) Extracted watermark for embedding blocksize 16×16

Figure 5.7: *(a) Original watermark, (b) (c) (d) Extracted watermark at JPEG compress level 50 from left view of Aloe image.*

Franco's scheme [62] and Rana's scheme [101] for left, right and the center view image. It is to be noted that the center view is the main view while left and right views are the synthesized view for Lin's scheme [55], Kim's and Kim's* scheme [76]. But for the proposed scheme, left and the right views are the main view while the center view is the synthesized view.

From Fig. 5.8, 5.9 & 5.10, it is observed that the proposed scheme mostly outperforms the recent existing schemes [55, 62, 76, 101] with respect to the robustness for all the cases when the watermark is extracted from the left, right or center view. Since the left or right view is the main view (where the watermark is actually embedded) for the proposed scheme, it may be concluded that it performs well against the view synthesis process (or synthesis view attack).

In Fig. 5.11, the robustness of the proposed scheme has been compared with the recent existing schemes [55, 62, 76, 101] against synthesis view attack (i.e. the DIBR based view synthesis process). It is observed that the proposed scheme



(a) Aloe image

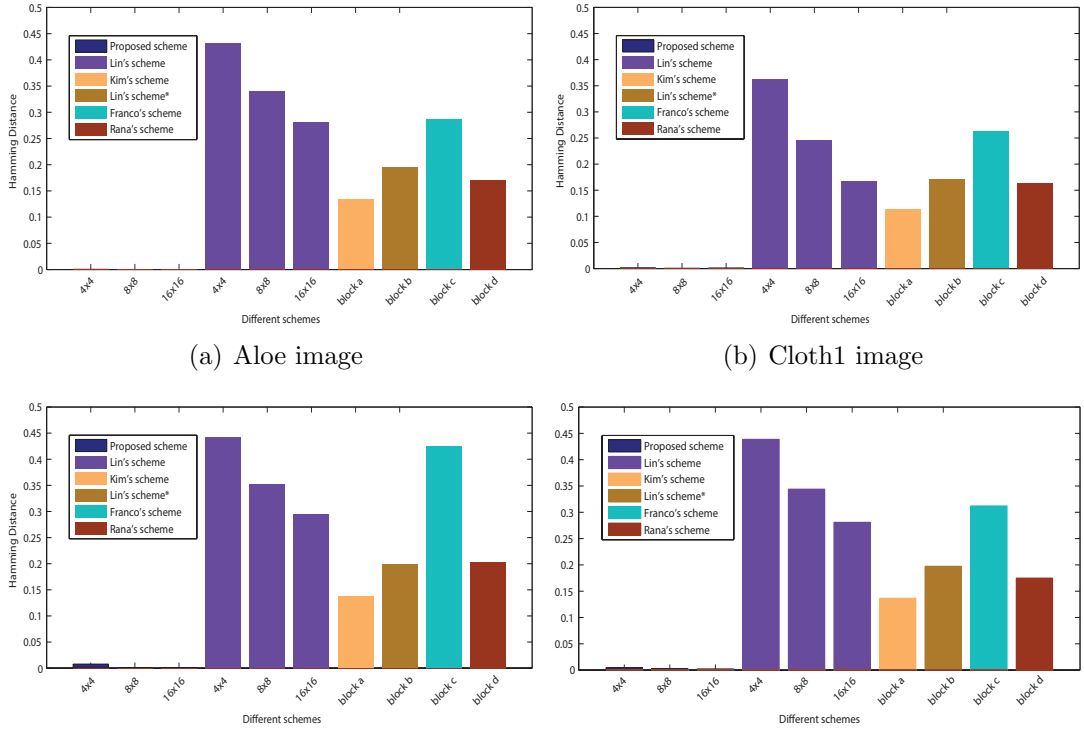
(b) Cloth1 image

(c) Plastic image

(d) Average of Middlebury Stereo 2006 Datasets of 21 image

Figure 5.8: *Hamming distance comparison of proposed scheme with existing schemes [55, 62, 76, 101] for left view at Stereo-JPEG compression at quality 75. ($block_a = (I_w/8) \times (I_h/8)$, $block_b = (I_w/32) \times (I_h/32)$, $block_c$ = linearly embed)*

greatly outperforms the above-mentioned existing schemes where camera view 1 and 5 are taken as the original left and right view, respectively and intermediate views (such as camera views of 1.4, 1.8, ..., 4.6) are the synthesized views. It is noted that camera view 3 is the center view which has been used as embedding (original) view for schemes [55, 76]. So, for camera view 3, the robustness performance of schemes [55, 76] are quite good and comparable to the proposed scheme as there is no view synthesis process is involved in this case for those schemes. The proposed scheme is compared with the existing schemes [55, 62, 76, 101] against Stereo JPEG compression of quality 15 to quality 100 in Fig. 5.12 where it is observed that the proposed scheme mostly outperforms



(a) Aloe image

(b) Cloth1 image

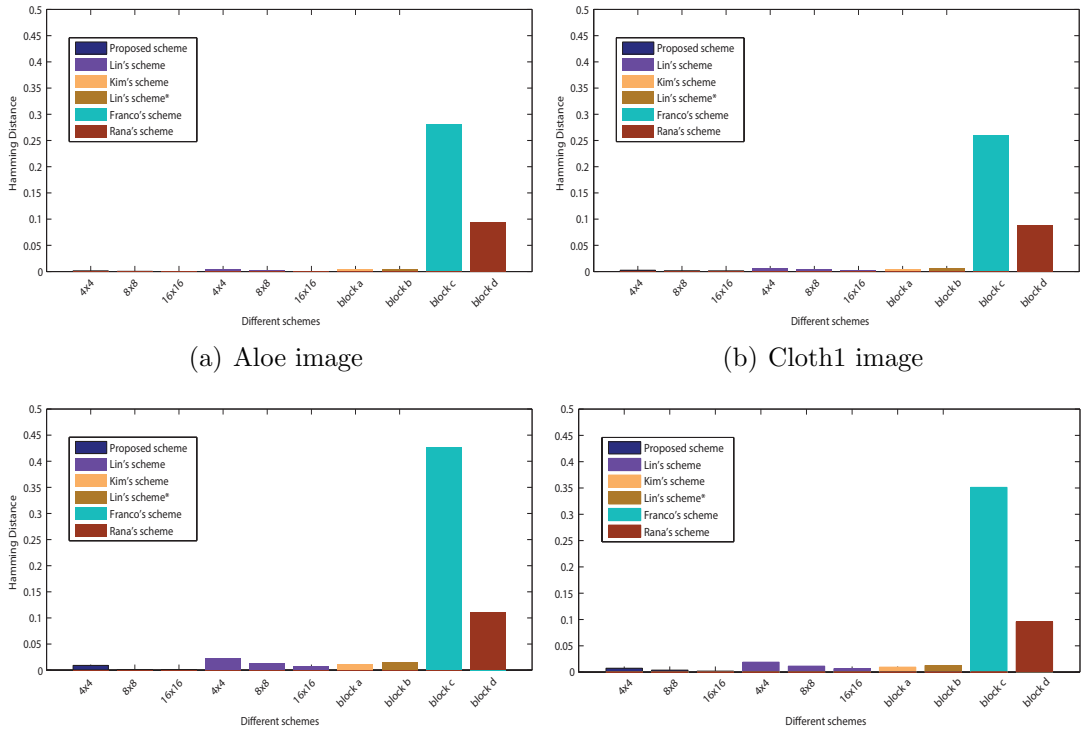
(c) Plastic image

(d) Average of Middlebury Stereo 2006 Datasets of 21 image

Figure 5.9: *Hamming distance comparison of proposed scheme with existing schemes [55, 62, 76, 101] for right view at Stereo-JPEG compression at quality 75.(*$block_a = (I_w/8) \times (I_h/8)$, $block_b = (I_w/32) \times (I_h/32)$, $block_c$ *= linearly embed and* $block_d = 16 \times 256$ *)*

the existing schemes [55, 62, 76, 101]. The similar results against the addition of Gaussian noise (up to variance of 200) and the addition of salt & pepper noise (up to the density of 0.2) are depicted in the Fig. 5.13 and Fig. 5.14 respectively. For both the cases, the proposed scheme shows comparatively better results than the existing schemes.

In Fig. 5.15, the robustness of the proposed scheme has been compared with the recent existing schemes [55, 62, 76, 101] against collusion attack [75]. It is observed that the proposed scheme outperforms the above-mentioned recent existing schemes against collusion attack [75] where the right view (camera view 5) image is rendered to the left view (camera view 1) and is colluded with the



(a) Aloe image

(b) Cloth1 image

(c) Plastic image

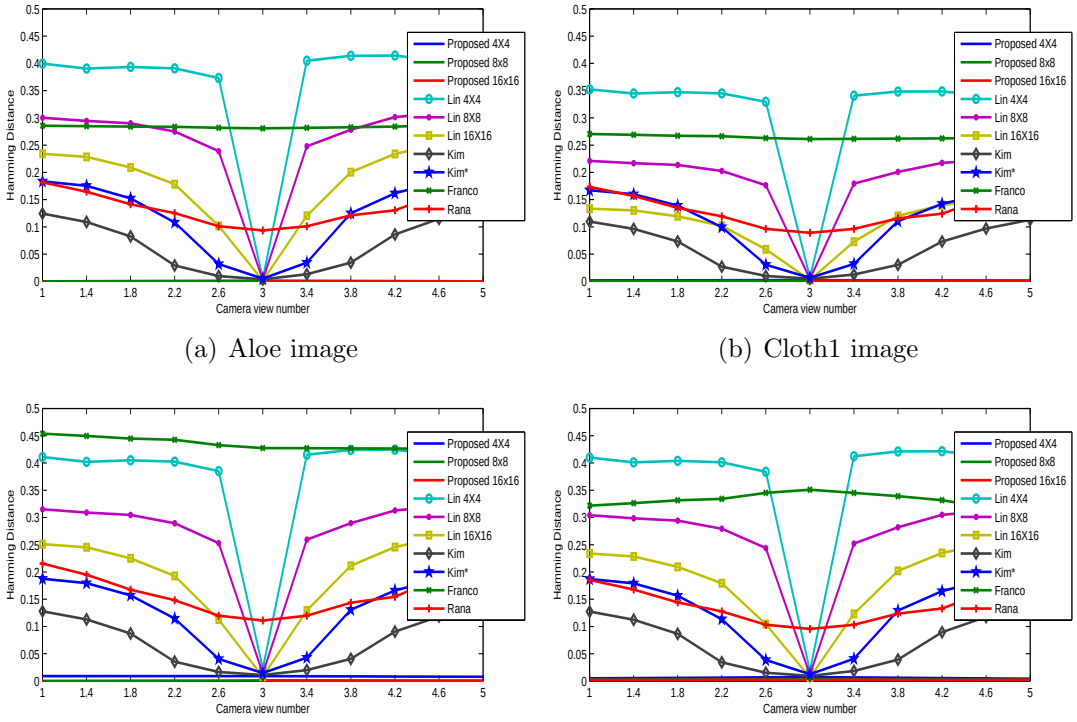
(d) Average of Middlebury Stereo 2006 Datasets of 21 image

Figure 5.10: *Hamming distance comparison of proposed scheme with existing schemes [55, 62, 76, 101] for synthesized center view at Stereo-JPEG compression at quality 75.(*$block_a = (I_w/8) \times (I_h/8)$, $block_b = (I_w/32) \times (I_h/32)$, $block_c$ = *linearly embed and* $block_d = 16 \times 256$*)*

original left view. In this collusion attack, the independent view regions of the synthesized left view (generated using DIBR technique) is filled using hole filling technique [9].

5.3.3 Discussion

In this work, watermark is embedded in the dependent view regions and the robustness of the scheme depends how accurately the embedding regions are detected at the time of extraction. This spatial synchronization has been achieved in this work by selection of the similar dependent view regions from the left and right views with respect to the disparity and original depth information. In



(a) Aloe image

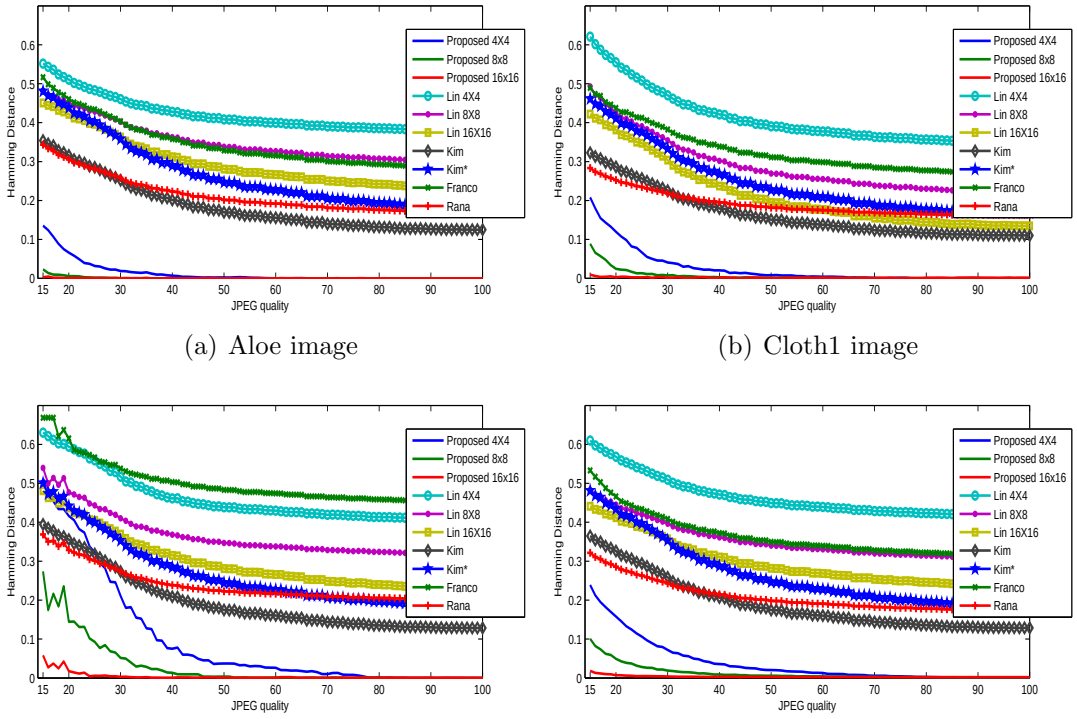
(b) Cloth1 image

(c) Plastic image

(d) Average of Middlebury Stereo 2006 Datasets of 21 image

Figure 5.11: *Hamming distance comparison of proposed scheme with existing schemes [55, 62, 76, 101] against view synthesis attack at Stereo-JPEG compression at quality 75 (where 1 & 5 defines the left and the right views and others are the intermediate synthesized views).*

existing literature, it has been observed that if the embedding view and the extraction view are not same, the watermark signal is degraded. Intuitively, if the size of the embedding block is less than the disparity, the watermark signal when extracted other than embedding view may be degraded. This may be the cause that existing schemes are not performing well for the relatively large disparity values. In the proposed scheme, since the block selection depends on the disparity value (refer to §5.1.4 and 5.2.3), the extracted watermark signal is not degraded even in case of large disparity values. This is the intuitive reason that the proposed scheme outperforms other existing schemes [55, 62, 76, 101] when embedding and extraction views are different as shown in Fig. 8, 9, 10 and 11.



(a) Aloe image

(b) Cloth1 image

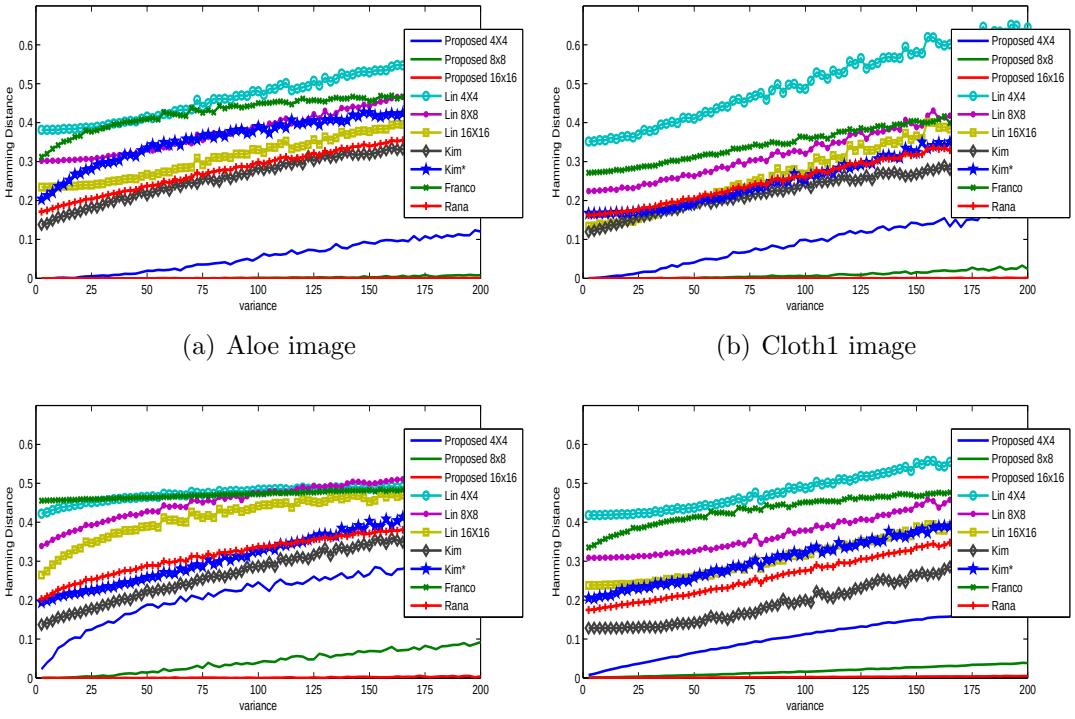
(c) Plastic image

(d) Average of Middlebury Stereo 2006 Datasets of 21 image

Figure 5.12: *Hamming distance comparison of proposed scheme with existing schemes [55, 62, 76, 101] against Stereo-JPEG compression for right view.*

In other words, the selection of the similar dependent view regions from the left and right views and embedding identical watermarks in those similar dependent regions makes the scheme more robust than the existing literature against view synthesis process as well as collusion attacks. Additionally, careful selection of the watermark embedding strength and visual quality threshold makes the scheme robust against Stereo JPEG compression and any other noise addition attacks.

Time Complexity Analysis: In this proposed scheme, the watermark is embedded with the DC coefficients of the dependent regions of both the views. To extract dependent view, DIBR is applied in pixels of the 3D image to generate the synthesis left and right view from the given view. The time complexity for dependent view generation is $\mathcal{O}(I_h \times I_w)$, where I_w and I_h represent the image with



(a) Aloe image

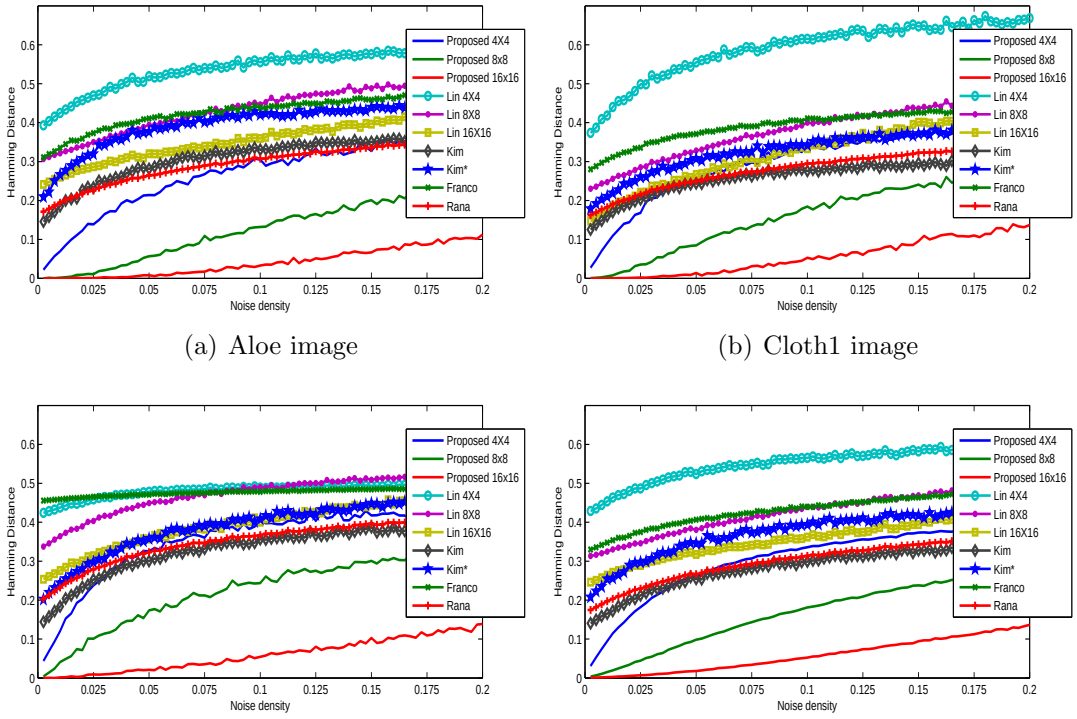
(b) Cloth1 image

(c) Plastic image

(d) Average of Middlebury Stereo 2006 Datasets of 21 image

Figure 5.13: *Hamming distance comparison of proposed scheme with existing schemes [55, 62, 76, 101] against Gaussian noise attack at Stereo-JPEG compression at quality 75 for right view.*

and height respectively. For embedding, $n \times n$ block wise 2D-DCT is used and the DC coefficients are altered as the embedding rule. Using butterfly architecture, 1D-DCT can be computed with complexity of $\mathcal{O}(n \log n)$ for n coefficients. So for a $n \times n$ block, time complexity of 2D-DCT is $\mathcal{O}(n^2 \log n)$ [102]. In this watermarking scheme, some blocks are selected for watermarking (refer to §5.1.5). In extreme case, if all the blocks are selected for watermarking, $(\frac{I_h \times I_w}{n^2})$ times of 2D-DCT need to be done for embedding of watermark in each view image. As a result, the total time complexity can be written as $\mathcal{O}(I_h \times I_w + \frac{I_h \times I_w}{n^2} n^2 \log n) \simeq \mathcal{O}((I_h \times I_w) \log n)$ for embedding of watermark with both the views. In this watermarking method, the watermarking block size is limited to 4×4, 8×8 and 16×16. So the final



(a) Aloe image

(b) Cloth1 image

(c) Plastic image

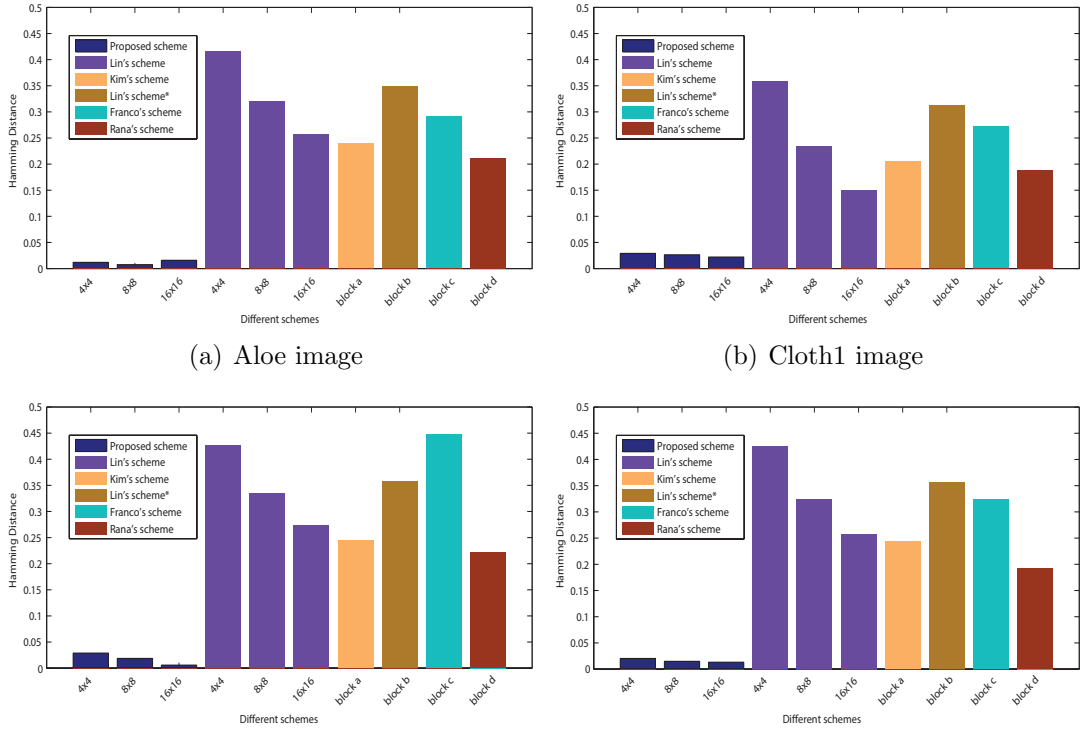
(d) Average of Middlebury Stereo 2006 Datasets of 21 image

Figure 5.14: *Hamming distance comparison of proposed scheme with existing schemes [55, 62, 76, 101] against salt & pepper noise attack at Stereo-JPEG compression at quality 75 for right view.*

time complexity can be written as $\mathcal{O}(I_h \times I_w)$.

5.4 Summary

In this chapter, a 3D image watermarking scheme is presented in multi-view stereo representation. The main aim of the work was to secure the the image content against view synthesis process (synthesis view attack) and collusion attack. In this scheme, horizontally shifted spatially coherent dependent view regions of the left and right view are embedded with identical watermarks to make the scheme robust against synthesis view attack as well as the collusion attack. To improve the robustness of the proposed scheme against image compression and



(a) Aloe image

(b) Cloth1 image

(c) Plastic image

(d) Average of Middlebury Stereo 2006 Datasets of 21 image

Figure 5.15: *Hamming distance comparison result after collusion attack on left image using the right image of the proposed scheme with Lin & Wu's scheme [55], Kim's scheme [76], Franco [62] and Rana's scheme [101] against Stereo-JPEG compression at quality 75 for right view ($block_a = (I_w/8) \times (I_h/8)$, $block_b = (I_w/32) \times (I_h/32)$, $block_c$ = linearly embed and $block_d = 16 \times 256$).*

different noise addition attacks, a secret key is used to select the DC coefficients of the selected 4×4, 8×8 and 16×16 blocks to embed the watermark. A comprehensive set of experiments have been carried out to justify the applicability of the proposed scheme over the existing literature against the different attacks. In the next chapter, the dual tree complex wavelet transformation (DT-CWT) features are used to develop a robust watermarking scheme for the DIBR-3D the image representation.

5. DEPTH-BASED VIEW INVARIANT BLIND 3D IMAGE WATERMARKING FOR MULTI-VIEW REPRESENTATION

Chapter 6

View Invariant Watermarking Using DT-CWT for DIBR-3D Image Representation

It has been observed in §5, for extraction of the watermark, original depth is required as a reference to generate the dependent region. For MVD [13] representation the depth is transmitted with the reference view part, but in other representation the where the reference depth is not available, the watermark may not be extractable.

In recent time, depth image based rendering (DIBR) based 3D image representation [1, 2] becomes popular due to its compression efficiency. It has been observed in literature [7, 103, 104] that the efficient watermarking system for authenticating DIBR 3D image encoding should consider situations not only where both the virtual left and right views are illegally distributed as 3D content but also where each single view, including the original center view, illegitimately transmitted. Due to certain inherent features like pixel disparity and changes in the depth image etc., the direct extension of existing conventional watermarking schemes for 2D and stereo images [41–60] are not very useful for DIBR [6] based encoding (as discussed in 1.3). In other words, the main challenge is to embed the watermark in such a way that the watermark should resist the view generation

process (it can be treated as a potential attack named as synthesis view attack) of the DIBR technique.

As explained in the previous chapter (refer to §5), recent literatures on the DIBR-3D image watermarking [55, 76], may not resist the view synthesis attack efficiently. In watermarking scheme proposed by Lin and Wu [55], the embedding block size is taken 8×8 and 16×16. Since the average disparity normally is around 200 pixels, embedding in smaller blocks such as 8×8 and 16×16 may not always guarantee the accurate reconstruction of the watermark. In DT-CWT based 3D watermarking scheme proposed by Kim et al. [76], for low resolution images, having block size $(I_w/8 \times I_h/8)$ (where I_w and I_h represent the width and height of the image respectively), may not be able to resist the view synthesis attack specially when the disparity is higher than the block width. Recently, Franco et al. proposed a virtual view invariant frame by frame 3D video watermarking scheme [62] where the watermark is embedded with the coefficients of each row to make the watermarking scheme robust against synthesis view attack. The row pixel positions may change with respective to the left and the right view images due to the presence of independent regions. Franco's scheme [62] didn't handle the above scenario at the time of embedding. Moreover, the row-wise bit embedding may result in visual artefacts.

In summary, it has been observed in the literature (refer to §1.3) that most of the existing schemes are vulnerable to the DIBR based view synthesis process. Since center view image does not have the independent component corresponding to its any of synthesized views [9], the watermark should be extracted from the dependent regions of any synthesized views. So, separating dependent regions from synthesized view, specifically at the time of watermark extraction, is an important task. Another challenge for DIBR-3D representation is that the dependent portion is visible to both the eyes. Hence the tampering of the dependent regions for watermarking embedding will make visual degradation with respect to HVS. In this work, a 3D image watermarking scheme is proposed to insert the watermark bit with the center view images such that the watermark will sustain

against view synthesis attack and different 3D attacks without make any visual degradation with respect to HVS. Also, Using of DT-CWT will improve the robustness by maintaining the decent visual quality. This task is one of the major contributions of this work which is done by an efficient zone selection scheme as described in §6.2.1. A polar angle orientation based watermarking embedding and extraction scheme are presented in §6.2.2 and §6.2.3 respectively. Finally, the experimental results are presented in the §6.3.

6.1 Research Background

6.1.1 Rendering of DIBR-3D-Image

As discussed earlier in §1.1.2, DIBR-3D-image representation [1,2] becomes popular for its compression efficiency. In the time of viewing of the 3D image, the independent region is only visible to either left or right view. So a visual discomfort may occur [3,104] for the existence of the independent view. On other hands, perturbation in the independent regions may not substantially downgrade the image quality with respect to the human vision [65,100]. In this DIBR-3D-Image representation, only centre view image and its depth map have been transmitted to the receiver end where left and right views are rendered with the help of centre view image, its depth and corresponding disparity [1] according to the Eq. 6.1 which is explained earlier in Eq. 1.1

$$\begin{aligned} X_L &= X_C + d_{isp}\, T_L\, \frac{D_{X_C}}{255} \\ X_R &= X_C - d_{isp}\, T_R\, \frac{D_{X_C}}{255} \end{aligned} \tag{6.1}$$

where d_{isp} is the disparity, D_{X_C} is the depth value of the center view image for the location X_C and X_L & X_R are the corresponding shifted locations for the left and the right view respectively. T is the baseline distance between the reference view to the synthesis view. The baseline distance is used to calculate the shift in between the left and the right view pixels. In DIBR-3D representation, the baseline distance of center view is taken as 0. As the disparity represents the

maximum displacement between the left and the right view, the baseline distance between the center view to the left and the right view is represented as $T = 1/2$ for rendering from center view to left and right view.

To obtained the left and the right views, canter view is rendered using the DIBR technique. In this scenario, the common region (say the dependent region) related to the left and the right views are obtained by the rendering process and the independent regions are remained unavailable (due to the partial presence of the independent regions in the center view as discussed in §1.1.2). As a result, the unavailable regions creates holes in left and right views in DIBR-3D image representation and filled up using the hole filling process [9] by the average values of the boundary pixels [1] as shown in Fig. 6.1. It is observed that, due to the

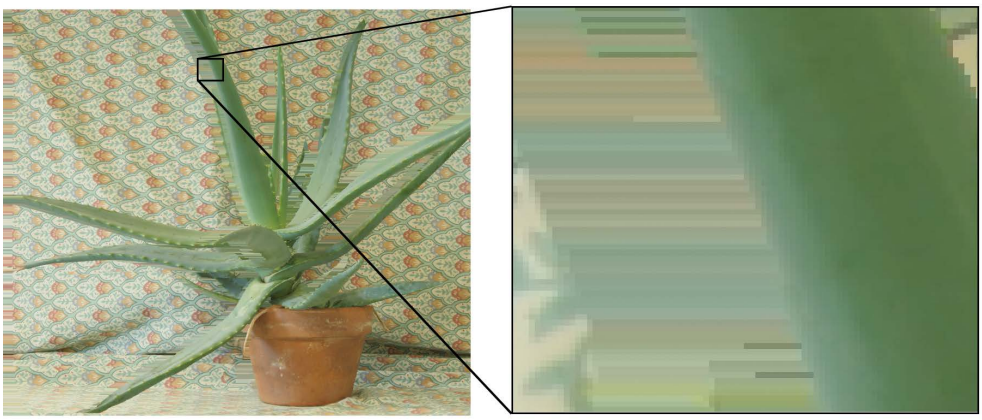


Figure 6.1: *Dependent and independent regions of synthesized left view of Aloe image from center view*

hole filling technique [9], the independent regions gives a horizontal line pattern for the left and the right views of the DIBR-3D-image (refer to Fig. 6.1).

6.1.2 Dual Tree Complex Wavelet Transform (DT-CWT)

The complex wavelet transform [105] is the complex extension of the DWT using a two-dimensional wavelet transform. Kingsbury [89–91] first presented the dual tree complex wavelet transformation (DT-CWT) to analyse it's shift invariant

property. In real analysis of DT-CWT (say dual tree discrete wavelet transformation (DT-DWT)), two independent discrete wavelet transformations (DWTs) are exploited for each tree. Two real dual tree wavelet coefficient (where the second transform is done by picking the opposite samples of the first one) are combined to generate the complex wavelet coefficients and can be represented as $h_a + ih_b$ where h_a and h_b are the real coefficients of the tree 'a' and tree 'b' as explained in §2.4

According to Kingsbury [89–91], it is observed that the DT-CWT is invariant to small shift or geometric distortion. Also, it is observed in the literature [76] that the PSNR between center view and left (or right) view, is relatively better when DT-CWT is used in comparison with DCT or FFT. Moreover, it is experimentally observed that the change of the magnitude of the DT-CWT coefficients due to the DIBR is very less as discussed in §2.4. For the shift of 16 pixels, the rate of change of the energies is less than the ratio of 1.025 : 1 for the DT-CWT coefficients, whereas energies of the DWT coefficient vary by up to 5.45 : 1 [91]. So insertion of the watermark with the DT-CWT coefficients may improve the robustness than DCT of FFT domain watermarking technique against DIBR based view synthesis process. DT-CWT produces six filtered real and complex coefficients at orientation angle 15°, 45°, 75°, −75°, −45°, −15° (let H_1, H_2, H_3, H_4, H_5, H_6). A details description is depicted in §2.4 Fig. 6.2 with the real and complex coefficients of DT-CWT. In the PSNR comparison, it is also observed in the literature [76] that H_1, H_2, H_5, H_6 coefficients of DT-CWT are higher than that of H_3, H_4 coefficients for the 2D-DT-CWT. It is also experimentally observed that the PSNR between coefficient H_2 and H_5 shows highest similarity than any other coefficient pairs for the 2D-DT-CWT [76]. These facts suggest us to embed the watermark using the H_2 and H_5 coefficients i.e of orientation angle 45° and −45° to improve the strength against the horizontal shift of the pixels during DIBR process.

By observing the characteristics of the DT-CWT (as shown in Fig. 6.2), it can be said that the values of H_1, H_6 is maximum and H_3, H_4 is minimum

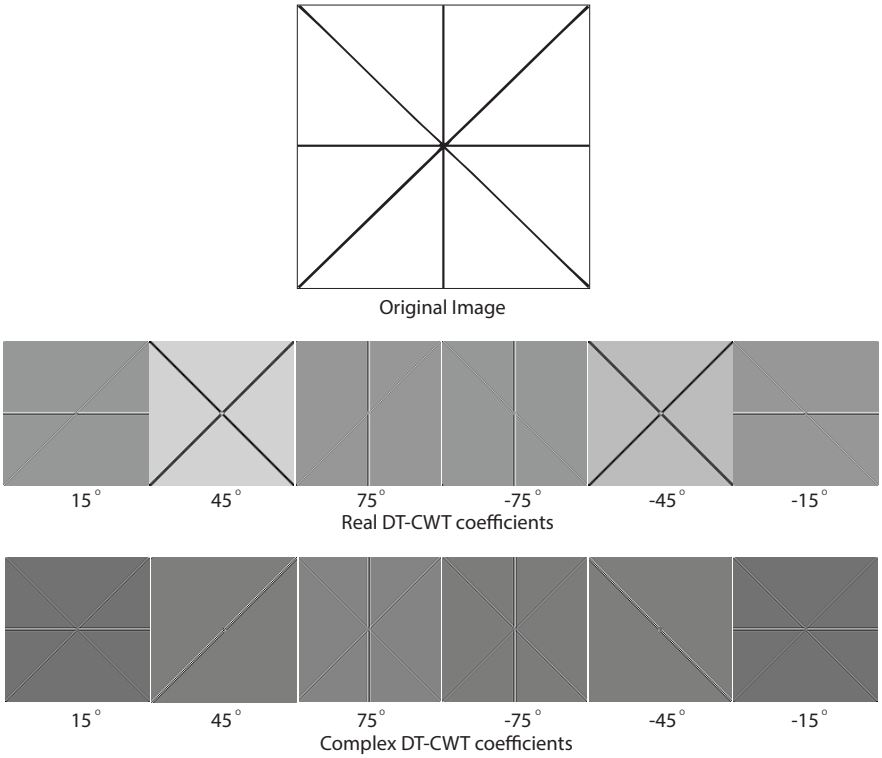


Figure 6.2: *DTCWT Coefficients representation line features*

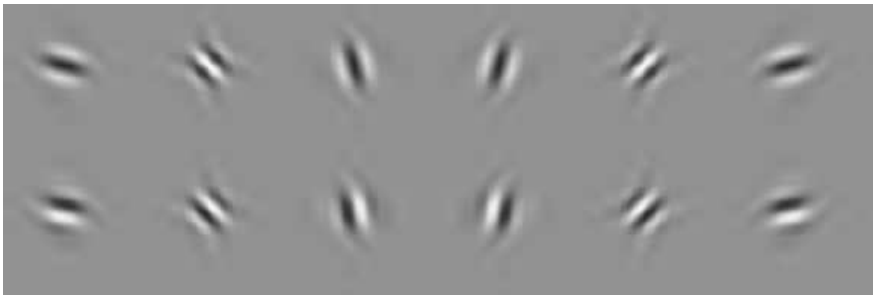
(very close to zero) for real and complex coefficients in a horizontal straight line. Interestingly, an inverse relation has been observed for the vertical lines. Likely an inverse effect is noticed at the original image due to change of the DT-CWT real and complex coefficients which is Gabor like impulse response as shown in Fig. 6.3. Also, by analysing the values of the DT-CWT coefficients as shown in Fig. 6.3, it is observed that H_3, H_4 follow the Eq. 6.2 for horizontal line

$$\left.\begin{array}{l} H_{3R} \approx H_{4R} \\ H_{3C} \approx -H_{4C} \end{array}\right\} \tag{6.2}$$

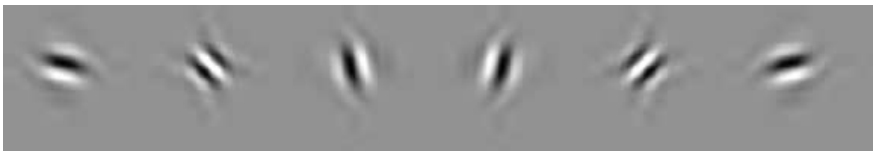
where H_{3R} & H_{4R} defines the real and H_{3C} & H_{4C} defines the complex part of the H_3, H_4 coefficients respectively. Thus it can be said that for the polar orientation, H_3, H_4 coefficients will follow the Eq. 6.3

$$|\theta_{H_3} + \theta_{H_4}| \approx \pi \tag{6.3}$$

where θ_{H_3} and θ_{H_4} are defined the angle θ of the polar orientation, presented as



(a) Inverse DTCWT of real and imaginary part



(b) Inverse DTCWT of combined real and imaginary part

Figure 6.3: *Impact on original image due to change in DT-CWT wavelet coefficients [89]*

$me^{i\theta}$ for the coefficients H_3, H_4 respectively. Moreover, the magnitudes (m) in the polar representation for the H_2, H_3, H_4, H_5 are near to zero (because of the real and complex coefficients are near to zero) where for H_2, H_5 the magnitude is minimum than other coefficients and the θ of the coefficients follow the Eq. 6.4

$$\theta_{H_2} \approx k\frac{\pi}{2} \pm \theta_{H_5} \quad where\ k = -2, -1, 0, 1, 2 \tag{6.4}$$

where θ_{H_2} and θ_{H_5} are defined the angle θ of the polar orientation of the coefficients H_2 & H_5 respectively. So the embedding of watermark with the coefficients H_2 & H_5 should be such a way that the procedure does not match the Eq. 6.4 which helps to remain imperceptible with respect to the human vision.

6.2 Proposed Scheme

In this proposed scheme, the watermark is embedded in the center view image (or the main view), which is then used to render left and right view images for a DIBR-3D representation. One of the main challenges of the proposed scheme is

to make the scheme invariant to the view generation process such that it can be efficiently extracted from any views rendered from the center view using DIBR technique. So the primary goal of this work is to choose the embedding zone such a way that view invariant watermarking can be achieved. In this scheme, the shift invariant property of the DT-CWT [89, 90] transformation has been exploited and selected DT-CWT coefficients of the carefully chosen center view image zones are used for embedding. A zone selection process is illustrated in §6.2.1 and watermark embedding and extraction algorithms are described in subsequent subsections.

6.2.1 Zone Selection

In the DIBR-3D image representation, pixels can move horizontally towards left or right direction for different views. To generate the synthesized left and right views from the center view, depth parameter, and the disparity information are used to detect the movement of each pixel as discussed in the previous section. Due to this horizontal pixel movement during the view synthesis process, the embedded watermark may degrade. In this work, this horizontal pixel movement is handled by exploiting the shift invariant property of the DT-CWT process. In other words, the DT-CWT coefficients of the center image are used for watermark embedding such that the scheme becomes invariant to the horizontal pixel shifting (view synthesis) process. To increase the robustness, 3^{rd} level DT-CWT coefficients are used for embedding. As the watermark is embedded in the center view image, independent regions of the corresponding left or right view image (which essentially does not contain any watermark) should not be used for watermark extraction, otherwise, the robustness of the scheme may degrade. To achieve this, separate embedding and extraction zone selection schemes are explained in the next subsections.

6.2.1.1 Zone Selection for Embedding

To make the scheme robust against view synthesis process, a threshold based embedding coefficient selection process has been employed in this work where robustness threshold (τ) is used to select the 3^{rd} level of 2D-DT-CWT coefficients of orientation angle 45° and -45° (H_2^{l3}, H_5^{l3}). The magnitude (m) of the coefficients are checked with the robustness threshold (τ) to satisfy the Condition 1.

$$\left.\begin{array}{l} if \quad m > \tau \quad then \\ \quad use\ for\ embedding \\ else \\ \quad not\ use\ for\ embedding \end{array}\right\} \textbf{Condition 1}$$

In this work, by experimental observation, threshold (τ) is defined by the Eq. 6.5

$$\tau = \frac{C_{avg}}{32} \tag{6.5}$$

where C_{avg} is the average of all the pixels of the image.

In this proposed scheme, the number of coefficients selected for embedding or extraction of the watermark, may not be always same because some of the coefficients may not available in the synthesized view due to the DIBR view synthesis process. These absent coefficients cause holes in DIBR-3D view generation process. Due to the presence of holes in the synthesized views, watermark quality can be degraded. To handle this problem, embedding coefficients (3^{rd} level of 2D-DT-CWT coefficients) should be filtered such a way that the noisy coefficients are removed. As a result, the number of embedding and extraction coefficients may not be equal. To handle this, 3^{rd} level of 2D-DT-CWT coefficients in each row are successively partitioned into 4 equal parts such that if there exist k number of coefficients in j^{th} row satisfy the Condition 1 then successively each $\frac{k}{4}$ number of coefficients of j^{th} row will be used to accommodate for the unique set of watermark bits.

6.2.1.2 Zone Selection for Extraction

In center view based embedding schemes, the watermark can be extracted either from the original center view or from the synthesized left or right views. In a

case of original center view (where the watermark is actually embedded), the watermark extraction is very straight forward and just reverse of the embedding process. In the case of synthesized views, the extraction process is a bit involved as these views have two components, first, the dependent zone where watermark bits are generally embedded and second is the independent zone where no watermark bits are embedded. So one of the important tasks is to accurately identify the independent regions and removes them from watermark extraction process. Essentially this task is bit easy when the watermark is extracted from the main (center) view but is relatively complicated in synthesized views. In this work, the peculiarity of DT-CWT coefficients has been capitalized to accurately identify the independent zone coefficients. It is already explained (refer to §6.1.2) that in the independent region, the magnitude of the DT-CWT coefficients of orientation angle $45°$ and $-45°$ is very close to zero due to the horizontal lines. In this case, the threshold based approach (refer to §6.2.1.1 Condition 1) is used to remove the independent regions and un-watermarked coefficients. Nevertheless, the proposed approach may not be able to remove all the independent region coefficients as the threshold checking is done at 3^{rd} level of 2D-DT-CWT coefficients.

To remove the remaining independent region coefficients, a further checking has been employed. It is experimentally observed that if 3^{rd} level DT-CWT coefficients follow the Eq. 6.3, 6.4 (the DT-CWT property as described in §6.1.2) the corresponding 2D-DT-CWT coefficients should belong to the independent region. For these case, Eq. 6.3 along with polar magnitudes (m) for H_2, H_3, H_4 and H_5 (the 1^{st} level DT-CWT coefficients) are checked to decide whether the corresponding coefficients belongs to independent zone or not.

6.2.2 Watermark Embedding

A random binary sequence (W) is used as the watermark which is embedded in the selected coefficients of the suitable embedding zone as discussed in the previous section. A pair of two watermark bits (W_p) are embedded into each of

the suitable coefficients of a partition as discussed in §6.2.1.1. For embedding the 3^{rd} level of DT-CWT coefficients of orientation angle 45° and −45° (say H_2^{l3} and H_5^{l3} respectively) are altered such a way that the embedding policy does not match the independent zone selection characteristics. In this scheme, the polar orientation angle θ of each H_2^{l3}, H_5^{l3} coefficients ($\theta_{H_2^{l3}}$ and $\theta_{H_5^{l3}}$) are altered with respect to the watermark bit pair. The corresponding embedding rule is depicted in Fig.6.4. The embedding process is described in Eq. 6.6, 6.7

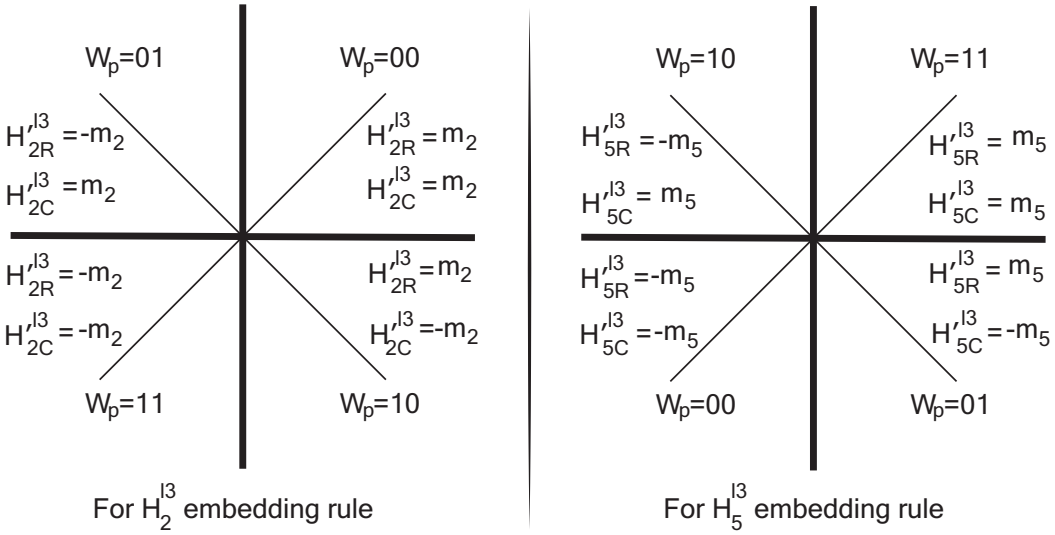


Figure 6.4: *Watermark embedding rule*

$$\left.\begin{aligned} H_{2R}^{\prime l3} &= m_2\,(-1)^{W_{pL}} \\ H_{2C}^{\prime l3} &= m_2\,(-1)^{W_{pM}} \end{aligned}\right\} \tag{6.6}$$

$$\left.\begin{aligned} H_{5R}^{\prime l3} &= -m_5\,(-1)^{W_{pL}} \\ H_{5C}^{\prime l3} &= -m_5\,(-1)^{W_{pM}} \end{aligned}\right\} \tag{6.7}$$

where H_{2R}^{l3} , H_{2C}^{l3} and H_{5R}^{l3} , $H_{5C}^{l3}$are the real and complex part of the coefficient H_2^{l3} and H_5^{l3} respectively. m_2 is defined the magnitude as $m_2 = \sqrt{(H_{2R}^{l3})^2 + (H_{2C}^{l3})^2}$ successively $m_5 = \sqrt{(H_{5R}^{l3})^2 + (H_{5C}^{l3})^2}$ and W_{pM} & W_{pL} are represented as the MSB and LSB of the two consecutive bits of the watermark bit pair (W_p).

In the embedding scheme as depicted in Fig.6.4, the watermark bit pair "00" represents 45°, "01" represents 135°, "10" represents −45°, "11" represents −135°

polar orientation angle of the coefficients H_2^{l3} and opposite quadrant for H_5^{l3} to increase the robustness of the watermarking scheme.

Algorithm 11: Watermark Embedding (I, W).

Input: I: Center view image and W: Watermark bit stream.
Output: W_I:Watermarked center view image.
begin

1. 3^{rd} level of DT-CWT is done on the original center view image (I) to extract the high frequency coefficients.
2. As described in §6.2.1.1.
   (a) Embedding zone selection is done on the 3^{rd} level of DT-CWT coefficients using the 6.2.1.1.
   (b) In each row coefficients are partitioned in to four groups.
3. **for** *Each coefficients partition* **do**
   (a) Using the embedding rule as described in Fig. 6.4 pair of watermark bits are inserted with the selected coefficient using Eq. 6.6, 6.7.
4. Inverse 3^{rd} level of DT-CWT is done with the watermarked coefficients to generate the watermarked center view image (W_I)
5. **return** (W_I)

The step by step embedding scheme is narrated in Algorithm 11 and a block diagram of the overall extraction process is depicted in the Fig. 6.5.

6.2.3 Watermark Extraction

In the proposed watermarking scheme, the watermark can be extracted from any of the original or synthesized views. The extraction scheme is almost the reverse of the embedding scheme. Using the quadrant of the polar orientation of the coefficients, possible watermark bit is calculated for each part as shown in

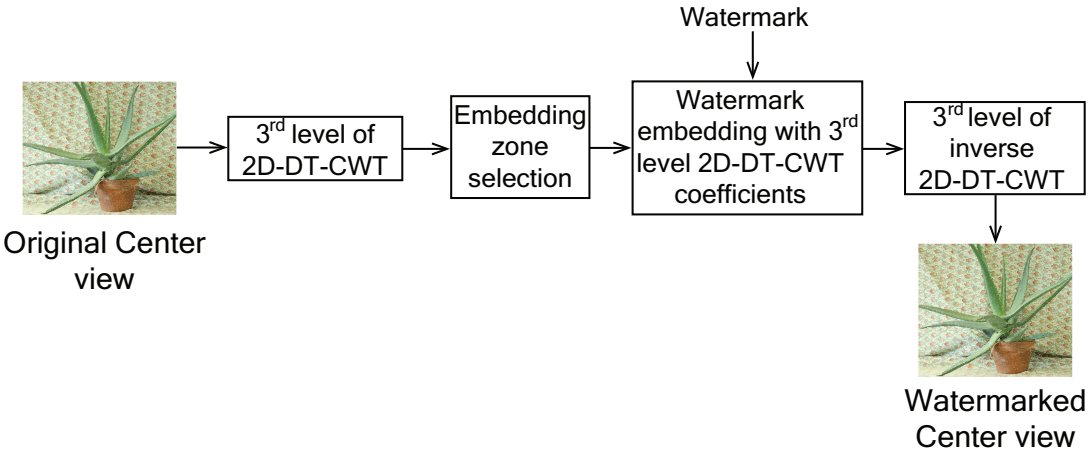


Figure 6.5: *Watermark embedding block diagram*

Eq. 6.8, 6.9

$$\left.\begin{array}{l} if\ \frac{\pi}{2} > \theta'_{H_2^{l3}} \geq 0\ \ then \\ \quad W'_{00} = W'_{00} + 1; \\ if\ \pi > \theta'_{H_2^{l3}} \geq \frac{\pi}{2}\ \ then \\ \quad W'_{01} = W'_{01} + 1; \\ if\ 0 > \theta'_{H_2^{l3}} \geq -\frac{\pi}{2}\ \ then \\ \quad W'_{10} = W'_{10} + 1; \\ if\ -\frac{\pi}{2} > \theta'_{H_2^{l3}} \geq -\pi\ \ then \\ \quad W'_{11} = W'_{11} + 1; \end{array}\right\} \tag{6.8}$$

$$\left.\begin{array}{l} if\ \frac{\pi}{2} > \theta'_{H_5^{l3}} \geq 0\ \ then \\ \quad W'_{11} = W'_{11} + 1; \\ if\ \pi > \theta'_{H_5^{l3}} \geq \frac{\pi}{2}\ \ then \\ \quad W'_{10} = W'_{10} + 1; \\ if\ 0 > \theta'_{H_5^{l3}} \geq -\frac{\pi}{2}\ \ then \\ \quad W'_{01} = W'_{01} + 1; \\ if\ -\frac{\pi}{2} > \theta'_{H_5^{l3}} \geq -\pi\ \ then \\ \quad W'_{00} = W'_{00} + 1; \end{array}\right\} \tag{6.9}$$

where W'_{00}, W'_{01}, W'_{10}, W'_{11} are the possible watermark variable for watermark 00, 01, 10 and 11 successively. $\theta'_{H_2^{l3}}$ and $\theta'_{H_5^{l3}}$ are defined as the θ of the polar orientation of the watermark coefficient.

Using the watermark variables the final watermark is calculated using the

Eq. 6.10

$$\left.\begin{array}{l} if(Max(W'_{00},\ W'_{01},\ W'_{10},\ W'_{11}) == W'_{00})\ \ then \\ W'_p = 00 \\ if(Max(W'_{00},\ W'_{01},\ W'_{10},\ W'_{11}) == W'_{01})\ \ then \\ W'_p = 01 \\ if(Max(W'_{00},\ W'_{01},\ W'_{10},\ W'_{11}) == W'_{10})\ \ then \\ W'_p = 10 \\ if(Max(W'_{00},\ W'_{01},\ W'_{10},\ W'_{11}) == W'_{11})\ \ then \\ W'_p = 11 \end{array}\right\} \tag{6.10}$$

where W'_p is the extracted watermark bit for partition p. Finally the watermark is accumulated to get the bit sequence of the extracted watermark W'.

Algorithm 12: Watermark Extraction (W_I).

Input: W_I: Watermarked Image.
Output: W': Extracted watermark.
begin

1. 1^{st} and 3^{rd} level of DT-CWT is done on the watermarked image (W_I) to extract the high frequency coefficients.
2. As described in §6.2.1.2.
   (a) Extraction zone selection is done on the 3^{rd} level of DT-CWT coefficients using the 6.2.1.1.
   (b) Independent region is removed using the 1^{st} and 3^{rd} level of DT-CWT coefficients using the Eq. 6.3, 6.4.
   (c) In each row coefficients are partitioned in to four groups.
3. **for** *Each coefficients partition* **do**
   (a) Using the embedding rule as described in Fig. 6.4 count the possible pair of watermark bits by comparing the selected coefficient using Eq. 6.8, 6.9.
   (b) Calculate the final watermark bit (W') from the possible pair of watermark bits using Eq. 6.10.
4. **return** (W')

After extraction of the watermark bit sequence, Hamming distance is used to

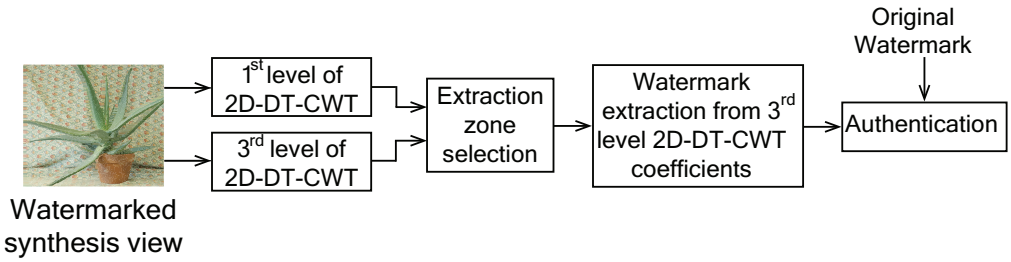


Figure 6.6: *Watermark extraction block diagram*

compare the extracted watermark bits (W') with the original watermark bits (W) as described in §2.6.2 Eq. 2.10. The step by step extraction scheme is narrated in Algorithm 12 and a block diagram of the overall extraction process is depicted in the Fig. 6.6.

6.2.4 Robustness of the Proposed Scheme

The main contribution of the proposed scheme is to make the embedding zone selection process robust against view synthesis process. In this concern, selection of the DT-CWT coefficients makes a major role to impose the view invariance characteristics. As discussed in §6.1.2 the shift of 16 pixels makes an effect of the energies is less than the ratio of 1.025 : 1 for the DT-CWT high coefficients [91]. So using of 3^{rd} level of DT-CWT makes the coefficient invariant to 64 pixels $(16 * 2^{(3-1)})$ and the coefficients can be used for watermark extraction for very less shift (say 64 pixels). For large shift, the location of the coefficients will not be available. To improve the robustness against the large shift, the groups of coefficients are used for embedding purpose (as explained in §6.2.1.1). To make the scheme position invariant watermark is inserted into the group of coefficients using the embedding rule in the §6.2.2 Fig. 6.4. As a result, the watermark can be easily extracted after the view synthesis process (or shift of large pixels).

Another challenge is raised due to the hole-filling [9] process in the synthesis view for DIBR-3D representation [1]. The independent regions of the synthesis views are not available in the synthesis views are creates holes (means the original

reference is not available in time of synthesis process). Those holes are filled by averaging the boundary pixels [9] as explained in §6.1.1. Extraction of the watermark from the synthesis view includes those independent unmarked pixels for extraction and reduce the watermarking strength. in this scenario, the independent regions are removed by analysing the 1^{st} and 3^{rd} level DT-CWT coefficients as explained in §6.2.1. Here the horizontal lines are analysed and removed from the watermarked extraction coefficients (because the hole filling regions creates the horizontal line as explained in §6.1.1).

6.3 Experimental Results

The proposed scheme has been tested with the Middlebury Stereo 2006 Dataset of 21 image [99] (refer to §2.7 Table 2.1). As the scheme is view invariant, the watermark can be extracted from the center view as well as any of the synthesis view. Also, the embedding and the extraction is independent of the depth map. So the scheme can be robust against the baseline distance change of the depth.

In this work, the experiment is carried out using the 21 stereo image dataset, where the view id'1' and view id '5' are taken separately as the center view image to increase the experiment dataset. Here for a single image the watermark bit is inserted is $\frac{I_h}{8} * 4 * 2 = I_h$ (3rd level of DT-CWT, 4 partition and 2 bit for each partition) where I_h is the image height. For a high resolution image the payload is more than 2^{10}. To justify the applicability of the proposed scheme, comparisons are done with the recent existing schemes such as Lin & Wu's scheme [55], Kim's scheme [76], Kim's scheme* [76] and Franco's scheme [62]. In the Kim's scheme*, the embedding sub-block is modified to $(I_w/32) \times (I_h/32)$ to make the similar embedding payload as proposed scheme for comparison purpose.

6.3.1 Visual Quality

Here the degradation of the watermarked image of the proposed scheme is evaluated using different visual quality metrics such as *Peak Signal-to-Noise Ra-*

tio (PSNR), *Structural Similarity* (SSIM) and *Visual Information Fidelity, pixel domain* (VIFp) (as explained in §2.6). The visual quality comparison results of the proposed scheme with the existing Lin & Wu's scheme [55], Kim's scheme, Kim's scheme* [76]and Franco's scheme [62] are given in the Fig. 6.7, 6.8 and 6.9.

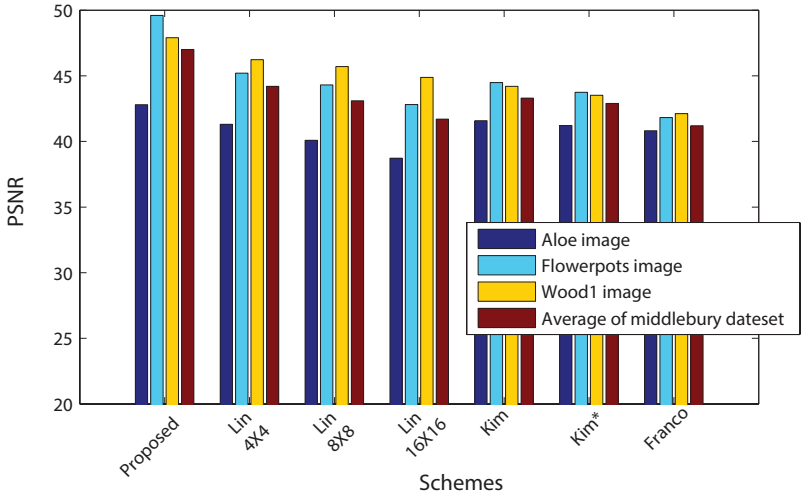


Figure 6.7: *PSNR comparison of the proposed scheme with existing schemes [55, 62, 76] for Middlebury Stereo 2006 Dataset [99]*

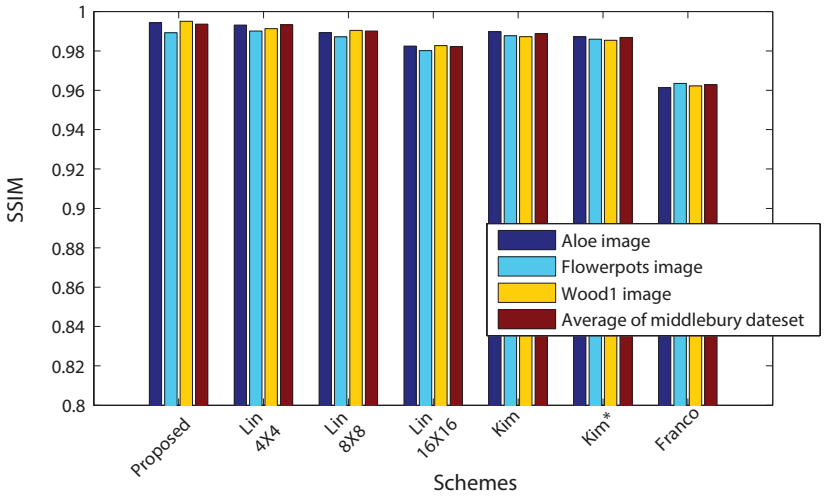


Figure 6.8: *SSIM comparison of the proposed scheme with existing schemes [55, 62, 76] for Middlebury Stereo 2006 Dataset [99]*

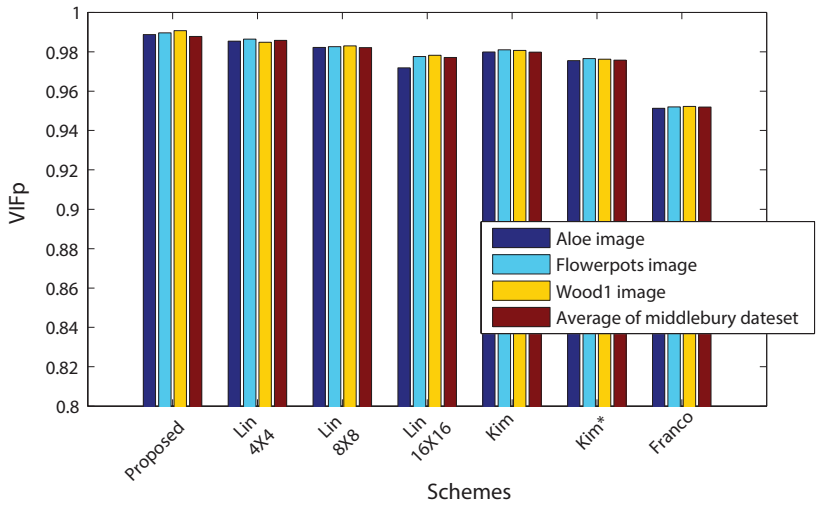


Figure 6.9: *VIFPcomparison of the proposed scheme with existing schemes [55, 62, 76] for Middlebury Stereo 2006 Dataset [99]*

From the Fig. 6.7, 6.8 and 6.9, it is observed that the proposed scheme gives almost comparable (sometimes better) result for the PSNR, SSIM, VIFp against the Lin & Wu's scheme [55], Kim's scheme, Kim's scheme* [76] and Franco's scheme [62] for similar embedding payload. Also it is observed that for same embedding payload, the proposed scheme gives better visual quality than that of the existing schemes.

6.3.2 Robustness

The primary goal of this work is to make the proposed scheme invariant to views generation process from the center view. Assuming the view synthesis using the DIBR technique as an attack, robustness of the proposed scheme is compared with the existing schemes [55, 62, 76]. Moreover, since most of the images are being communicated in compressed format (such as JPEG), the robustness of the proposed scheme against JPEG compression at different quality levels that of the existing schemes and different noise addition attacks are also taken into account in this subsection. In DIBR-3D representation, viewers can adjust the baseline to a proper range (up to 10% of the image width). So the baseline distance change

can be used as an attack for the DIBR-3D image representation. The robustness of the proposed scheme is compared with the existing schemes [55, 62, 76] against baseline distance change attack. Also in DIBR-3D representation, the depth can be preprocessed to a blurred image or using DIBR technique, can be shifted from its original position. For this kind of preprocessing depth attack, the robustness of the proposed scheme is compared with the existing schemes [55, 62, 76].

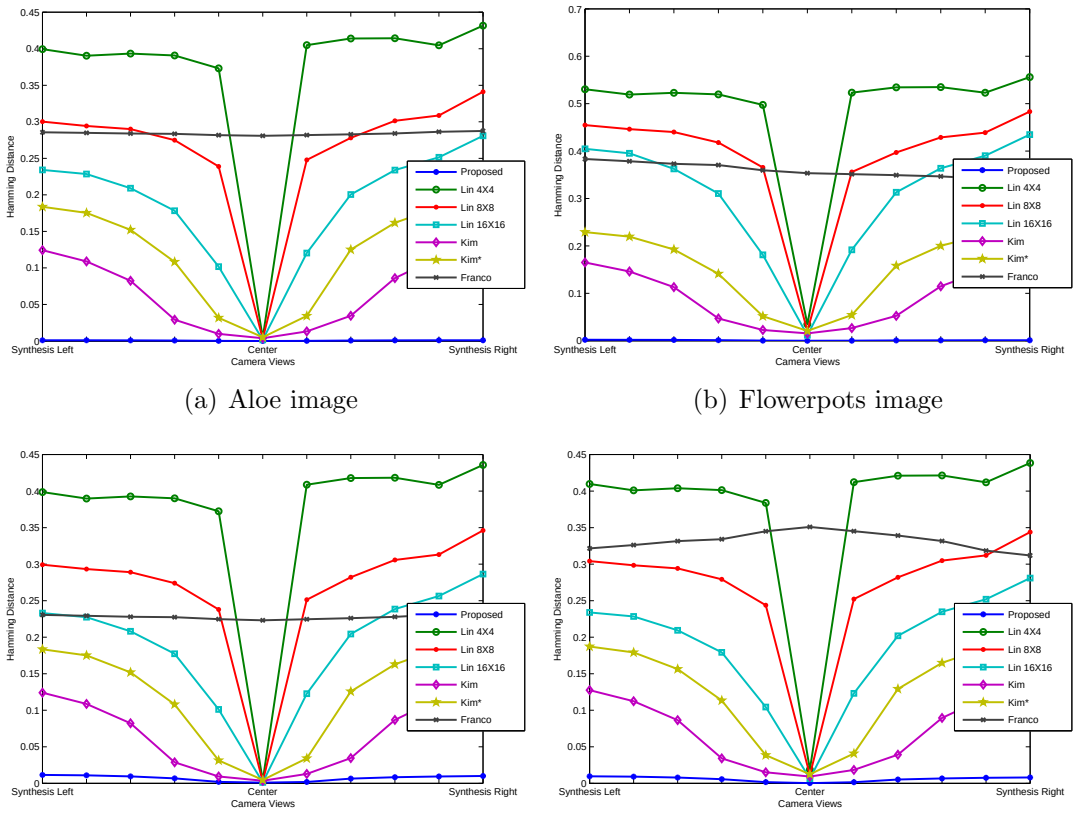
In this scheme, a random bit stream is used to embed with the original center view image. After different attack (view synthesis, JPEG compression, noise addition, baseline distance change [76] and the depth preprocessing [76, 80]) Hamming distance between the extracted watermark and the original watermark is determined which is used as a robustness metric where the smaller hamming distance denotes the better robustness as shown in Fig 6.10-6.16.

From the Fig. 6.10, it is observed that the proposed scheme shows almost negligible hamming distance for the center view as well as for the synthesized view at different positions. Here the experiment is done using view id '1' and view id '5' from the Middlebury Stereo 2006 Dataset as the center view separately. Where left and the right views are synthesized using the given disparity and the depth map. It is also observed that proposed scheme outperforms the existing schemes [55, 62, 76] for different views.

From the Fig. 6.11, it is observed that the proposed scheme gives a better result than the existing schemes [55, 62, 76] for JPEG compression quality from 20 to 100. The similar result can be observed for addition Gaussian noise (up to variance 200) and the addition of salt & pepper noise (up to the density of 0.2) as depicted in Fig. 6.12 and 6.13.

From the Fig. 6.14, it is observed that proposed scheme outperforms Lin's scheme [55] and Franco's scheme [62] and comparable result with the Kim's scheme and Kim's scheme* [76]. For a change of small baseline distance. For larger baseline distance, proposed scheme outperforms the Kim's scheme and Kim's scheme* [76].

Fig. 6.15 and 6.16 shows the result after depth modification attack. After



(a) Aloe image

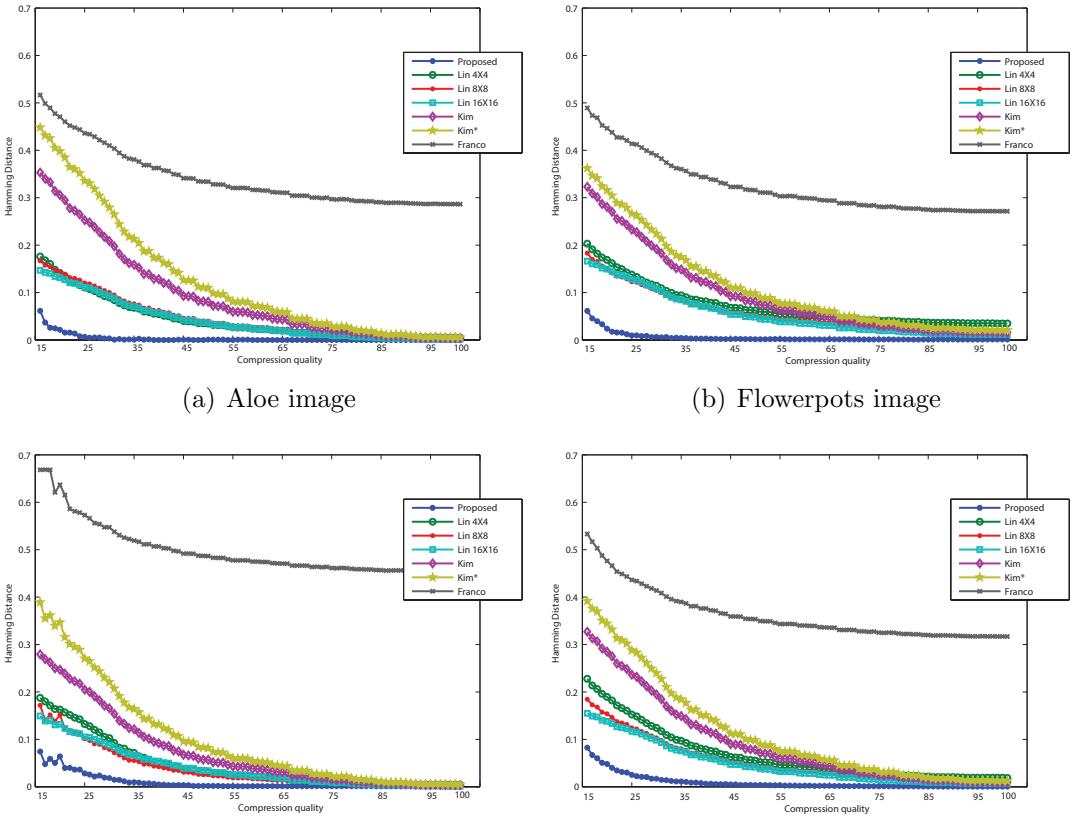
(b) Flowerpots image

(c) Wood1 image

(d) Average of Middlebury Stereo 2006 Dataset

Figure 6.10: *Hamming distance comparison of the proposed scheme with existing schemes [55, 62, 76] against view synthesis attack*

depth modification, the watermark is extracted from the left and the right synthesized views. Here it is observed that the Kim's scheme and Kim's scheme* [76] can somehow handle the is the depth modification attack for the left and the right view. Lin's scheme [55] could not handle the depth modification attack. It is observed that the proposed scheme can handle depth blurring as well as shifted depth using DIBR technique.



(a) Aloe image

(b) Flowerpots image

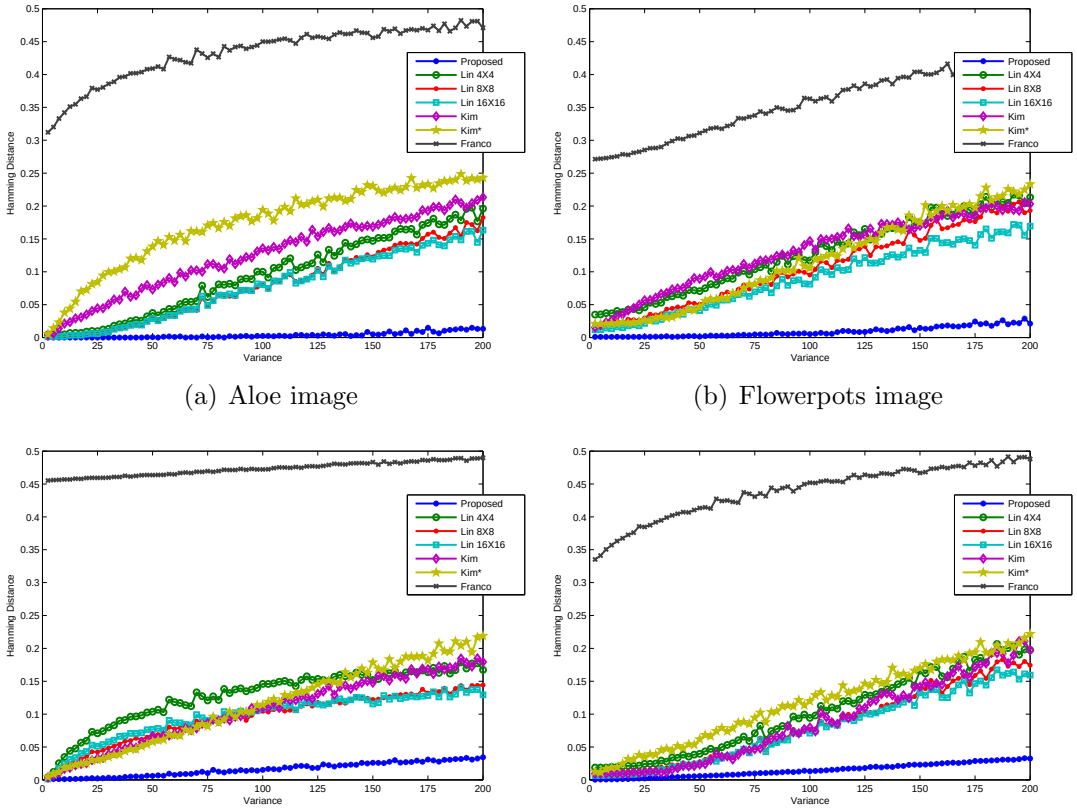
(c) Wood1 image

(d) Average of Middlebury Stereo 2006 Dataset

Figure 6.11: *Hamming distance comparison of the proposed scheme with existing schemes [55, 62, 76] against JPEG compression*

6.3.3 Discussion

In this work, the watermark is embedded with the DT-CWT coefficients to make the watermarking scheme view invariant. The spatial synchronization has been achieved at the time of embedding and extraction of watermark for the shift invariant property of the DT-CWT coefficients. To improve the shift invariance characteristics of the proposed scheme, 3^{rd} level of DT-CWT coefficients are used for embedding. Moreover, in the time of watermark extraction, by analysing the DT-CWT coefficients, hole filled regions (which does not contain watermark) are shorted out to remove the noise of the watermark extraction. In existing



(a) Aloe image

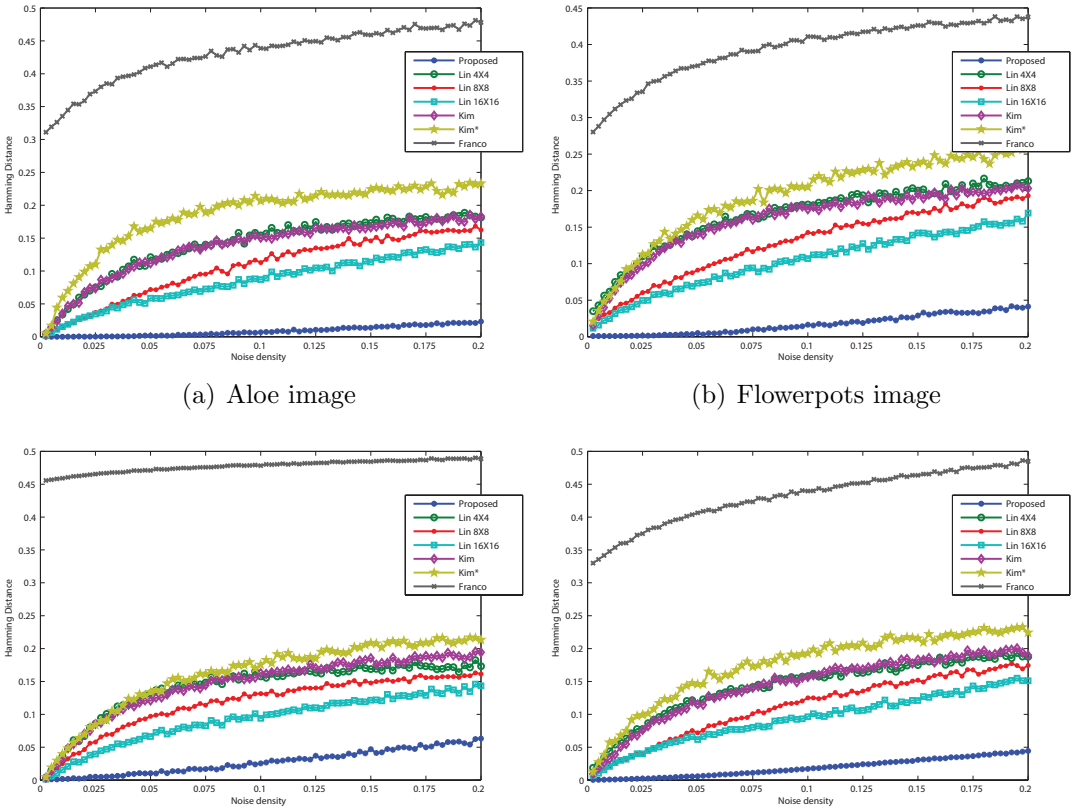
(b) Flowerpots image

(c) Wood1 image

(d) Average of Middlebury Stereo 2006 Dataset

Figure 6.12: *Hamming distance comparison of the proposed scheme with existing schemes [55, 62, 76] against addition of Gaussian noise*

literature, it has been observed that if the embedding view and the extraction view are not same, the watermark signal is degraded. Intuitively, if the size of the embedding block is less than the disparity, the watermark signal when extracted other than embedding view may be degraded. This may be the cause that existing schemes are not performing well for the relatively large disparity values. In this proposed scheme the selection of DT-CWT coefficients which are invariant to the DIBR technique (refer to §6.2.1.1 and §6.2.1.2), the extracted watermark is not degraded in case of large disparity as well as baseline distance change and depth modification attack. Moreover, the robust embedding policy makes the proposed scheme outperforms other existing schemes [55, 62, 76] for different attacks as



(a) Aloe image

(b) Flowerpots image

(c) Wood1 image

(d) Average of Middlebury Stereo 2006 Dataset

Figure 6.13: *Hamming distance comparison of the proposed scheme with existing schemes [55, 62, 76] against addition of salt & pepper noise*

shown in Fig. 6.10 - 6.16.

Time Complexity Analysis: In this proposed scheme, the watermark is embedded with the DT-CWT coefficients of the center view. For embedding, 3^{rd} layer of DT-CWT is used to increase the robustness with respect to the synthesis view. It is observed that, DT-CWT is a combination of 4 DWT [89–91] (as discussed in §2.4). The time complexity for a 2D-DWT is $\mathcal{O}(I_h I_w)$ where I_w and I_h represent the image with and height respectively. So the time complexity for 3^{rd} layer DT-CWT is $\mathcal{O}(4I_hI_w + 4\frac{I_h}{2}\frac{I_w}{2} + 4\frac{I_h}{4}\frac{I_w}{4}) \simeq \mathcal{O}(I_hI_w)$. in zone selection method (as discussed in §6.2.1), all the coefficients are checked for validate the embedding threshold. For embedding the DT-CWT coefficient values are altered

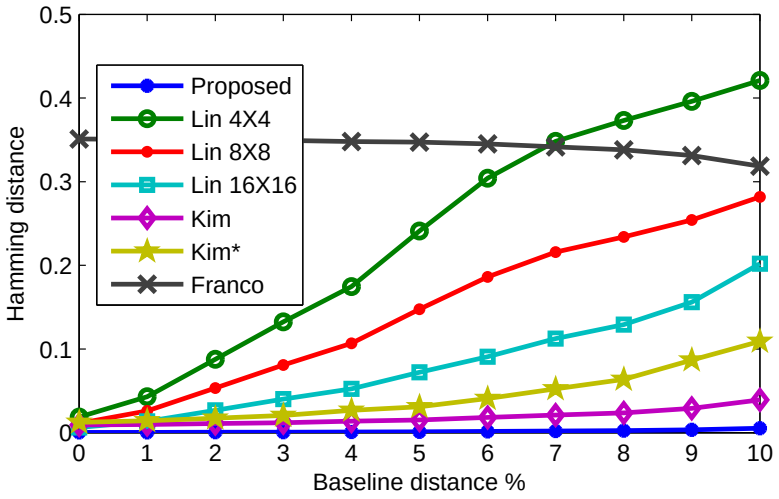
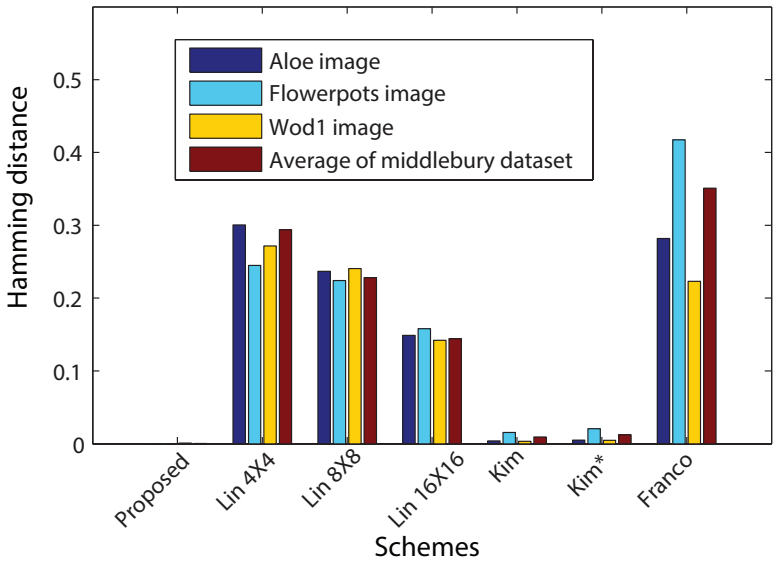


Figure 6.14: *Hamming distance comparison of the proposed scheme with existing schemes [55, 62, 76] against baseline distance change attack*

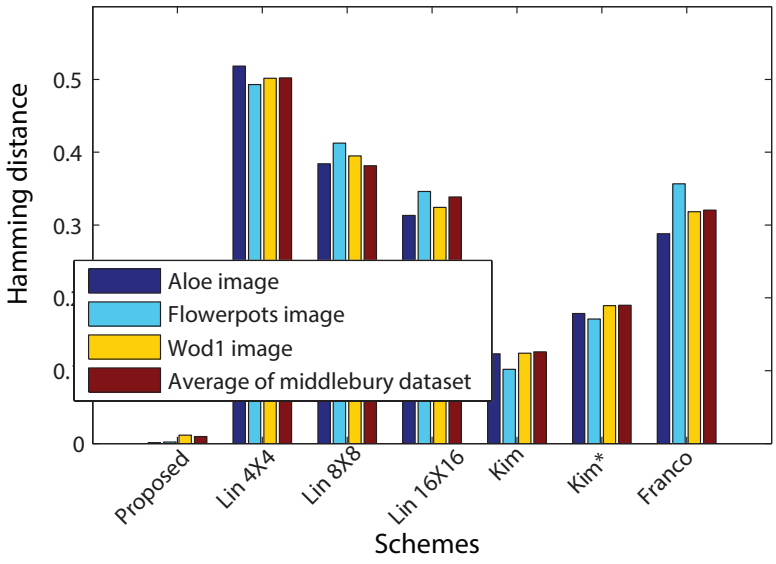
(as explained in §6.2.2) for watermark insertion. For worst case all the coefficients $4\frac{I_h}{8}\frac{I_w}{8}$may participate for embedding. So the total time complexity can be written as $\mathcal{O}((I_hI_w) + 4\frac{I_h}{8}\frac{I_w}{8} + 4\frac{I_h}{8}\frac{I_w}{8}) \simeq \mathcal{O}(I_hI_w)$. For inverse DT-CWT the tome complexity is $\mathcal{O}(I_hI_w)$. Hence the final time complexity can be written as $\mathcal{O}(I_hI_w)$.

6.4 Summary

In this chapter, a DT-CWT based watermarking scheme has been proposed to resist the synthesis view attack during the DIBR-3D representation of the 3D image. Separating dependent coefficients from independent coefficients for synthesized view for correct watermark extraction is one of the main contributions of this chapter. Moreover, a threshold based embedding zone selection has made the scheme more robust against view synthesis process. To increase the robustness, 3^{rd} level 2D DT-CWT has been used for watermark embedding and a novel polar orientation based watermarking embedding and extraction scheme is implemented to make the scheme invariant to coefficient loss during the DIBR process.



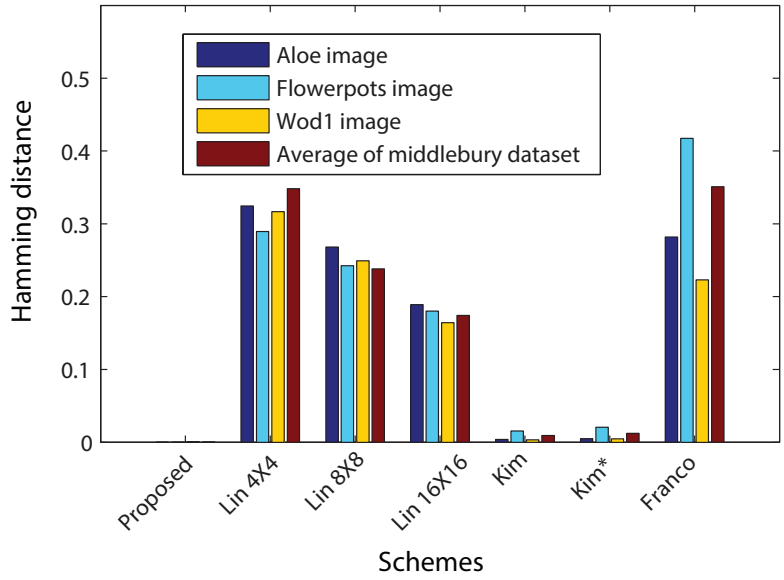
(a) Center view



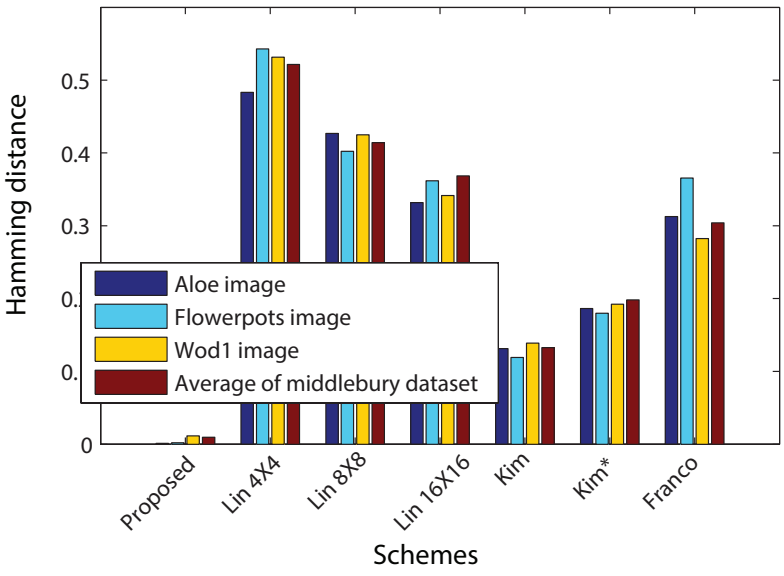
(b) Average of left and right view

Figure 6.15: *Hamming distance comparison of the proposed scheme with existing schemes [55, 62, 76] against blurring depth modification attack*

The experiment results shows the applicability of the proposed scheme over the existing schemes.



(a) Center view



(b) Average of left and right view

Figure 6.16: *Hamming distance comparison of the proposed scheme with existing schemes [55, 62, 76] against depth modification attack using DIBR technique*

Chapter 7

Conclusion and future work

7.1 Summary of the Contributions

With the recent upgradation of the 3D display devices and the advance coding policies, the 3D media become very popular. As a consequence, securing 3D images and video become and important requirement for secure media transmission. In this dissertation, different robust 3D images and video watermarking schemes are addressed for different 3D representations and 3D compressions. A brief summary of contributions made toward these is provided below.

It has been observed in the literature that the existing 3D image and video watermarking schemes fails to meet the three basic requirements in MVD representation, firstly securing the independent regions (say the uncommon regions of the left and the right view) as well as the dependent regions (say the common regions of the left and the right view) of the video sequence secondly, and finally and most importantly the intermediate synthesis views obtained using DIBR technique.

7.1.1 Robust Watermarking for MVD Representation against 3D-HEVC Compression and View Synthesis Attack

In the first contributory chapter of this dissertation, two 3D video watermarking schemes have been proposed against MVD based compression. In the first phase

of this work, a watermarking scheme has been proposed to secure the Z-axis (say independent region) of the 3D video scheme against MVD based 3D-HEVC compression process. The block DCT coefficients of the temporal filtered Z-axis is used to increase the robustness of the watermarking scheme. To justify the applicability of the proposed scheme over the recent existing scheme, a set of experiments has been carried out for different video sequences by maintaining the parameters similar with the 3D-HEVC compression.

In the second phase of contribution, an improvement of the first scheme has been proposed by securing the dependent view and intermediate synthesis views using the shift invariant property of the DT-DWT coefficients. To make the scheme robust against 3D-HEVC compression attack, motion compensated temporal filtering is used over the synthesized center view. A watermark is inserted with the 3^{rd} level 2D DT-DWT coefficients of the temporally filtered center view sequence to achieve the robustness against view synthesis process using the DIBR and collusion attack. The experimental results demonstrate the applicability of the proposed scheme over the existing methods.

7.1.2 Watermarking in Depth Information of 3D Image and Video Sequences against Depth Attack

In the 3D image and video sequences, depth informations have the same importance as the view part. As discussed in the introduction, depth information is used to generate the 3D vision with the view part in MVD and DIBR-3D representation. Hence, depth information is also required to be secured as the view content. As discussed in the literature, the existing depth based watermarking schemes are not sustainable against view synthesis attack.

In this chapter, two 3D depth watermarking schemes are presented for the images and video in the DIBR-3D and MVD representation respectively. In the first work, depth of the DIBR-3D images are watermarked such a way that the embedded locations do not make any visual artefacts with respect to HVS. The foreground object removal using the semantic segmentation method, improves

the visual quality and the robustness of the scheme. A novel SIFT location based watermark embedding policy is used to improve robustness against view synthesis attack. A comprehensive set of experiments has been carried out to justify the applicability of the proposed scheme over the existing literature against the different attacks.

In the second phase of work, the depth image watermarking scheme on DIBR-3D representation is extended for the 3D video sequences in MVD representation. Here the watermark is inserted in the synthesized center view depth to cover the dependent regions of the left and the right depths. The foreground removal is done using the center view to make the embedding scheme imperceptible to HVS. Moreover, SIFT location based watermark embedding policy is used on the embeddable locations of the center view depth to improve robustness against view synthesis attack. Moreover, the connected pixels based embedding after motion compensated temporal filtering, improves the robustness of the scheme against 3D-HEVC compression as well as frame dropping attack. It is experimentally shown that the proposed scheme performs well against the different attacks on the depth sequence and outperforms the existing related scheme.

7.1.3 Depth-based View Invariant Blind 3D Image Watermarking for Multi-view Representation

From the previous contributions and the existing literature, it has been observed that the 3D watermarking scheme can be improved to sustain against view synthesis and different image processing attacks. In the third contribution of this dissertation, a 3D image watermarking scheme is presented in multi-view stereo representation. The main aim of the work was to secure the image content against view synthesis process (synthesis view attack) and collusion attack. In this scheme, horizontally shifted spatially coherent dependent view regions of the left and right view are embedded with identical watermarks to make the scheme robust against synthesis view attack as well as the collusion attack. To improve the robustness of the proposed scheme against image compression and different

noise addition attacks, a secret key is used to select the DC coefficients of the selected 4×4, 8×8 and 16×16 blocks to embed the watermark. The effectiveness of the scheme is experimentally justified against the view synthesis process and different image processing attacks.

7.1.4 View Invariant Watermarking Using DT-CWT for DIBR-3D Image Representation

Like the MVD representation, the robust watermarking scheme should be available for the DIBR-3D representation. In the final contributory chapter of this thesis, a DT-CWT based watermarking scheme has been proposed to resist the synthesis view attack during the DIBR-3D representation of the 3D image. Separating dependent coefficients from independent coefficients to achieve noise free watermark extraction is one of the main contributions of this work. Moreover, a threshold based embedding zone selection scheme has made the scheme more robust against view synthesis process. To increase the robustness, 3^{rd} level 2D DT-CWT is used for watermark embedding and a novel polar orientation based watermarking embedding and extraction scheme is implemented to make the scheme invariant to coefficient loss during the DIBR process. A set of experiments has been carried out to prove the efficiency of the scheme over the existing literature against view synthesis and the depth blurring, depth modification attacks.

A brief summarization of the proposed schemes in each contributory chapter is presented in the following Table 7.1.

7.2 Future Research Scope

The present study is mainly restricted for the uncompressed domain 3D image and video watermarking and thus the proposed schemes are a bit slow because of decoding and further re-encoding. So the equivalent study in the compressed domain may be an interesting future scope. Also, this research never consider image

Table 7.1: *Summarization of the the proposed approaches.*

Contribution	Strategy employed	Attack countered	Performance
MVD based video watermarking	**a)** Embedding of watermark with motion compensated Z-axis coefficients	3D-HEVC compression	Robust than existing scheme in MVD representation
	b) Embedding of watermark with motion compensated center view coefficients	3D-HEVC compression, view synthesis attack, collusion attack	Robust than previous and existing scheme in MVD representation
Depth watermarking of 3D images and video	**a)** Embedding of watermark in less HVS sensitive depth region of center view image	Image compression, depth baseline distance change, depth noise addition attack	Robust than existing scheme in DIBR-3D representation
	a) Embedding of watermark in less HVS sensitive depth region of center view image	3D-HEVC compression, depth baseline distance change, frame dropping attack	Robust than existing scheme in MVD representation
Multi view 3D image watermarking	Embedding of watermark in the common regions of left and right view image	Image compression, noise addition, view synthesis, collusion attack	Robust than existing schemes in multi view image representation
DIBR-3D image watermarking	Embedding of watermark in the DT-CWT coefficients of center view image	Image compression, noise addition, DIBR-view synthesis, depth blurring, depth modification, collusion attack	Robust than existing scheme in DIBR-3D image representation

or video content for watermark zone selection or embedding, so content based 3D image and video watermarking can be an improvement over the present schemes. Furthermore, finding a suitable robust region in n-view representation will be a interesting research topic. Here the watermarking schemes are represented and tested in the environment of 2 views. Later the schemes can be extended in the environment of surround view schemes.

References

[1] L. Zhang and W. J. Tam, "Stereoscopic image generation based on depth images for 3d tv," *Broadcasting, IEEE Transactions on*, vol. 51, no. 2, pp. 191–199, June 2005. [Pg.1], [Pg.3], [Pg.4], [Pg.67], [Pg.111], [Pg.127], [Pg.129], [Pg.130], [Pg.141]

[2] C. Fehn, "Depth-image-based rendering (dibr), compression, and transmission for a new approach on 3d-tv," pp. 93–104, 2004. [Online]. Available: http://dx.doi.org/10.1117/12.524762 [Pg.1], [Pg.3], [Pg.67], [Pg.97], [Pg.127], [Pg.129]

[3] R. Patterson, "Human factors of 3-d displays," *Journal of the Society for Information Display*, vol. 15, no. 11, pp. 861–871, 2007. [Online]. Available: http://dx.doi.org/10.1889/1.2812986 [Pg.1], [Pg.16], [Pg.38], [Pg.54], [Pg.68], [Pg.69], [Pg.81], [Pg.129]

[4] P. Kauff, N. Atzpadin, C. Fehn, M. Mller, O. Schreer, A. Smolic, and R. Tanger, "Depth map creation and image-based rendering for advanced 3dtv services providing interoperability and scalability," *Signal Processing: Image Communication*, vol. 22, no. 2, pp. 217 – 234, 2007, special issue on three-dimensional video and television. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0923596506001329 [Pg.1], [Pg.5], [Pg.95]

[5] B. Ozkalayci and A. Alatan, "3d planar representation of stereo depth images for 3dtv applications," *Image Processing, IEEE Transactions on*,

vol. 23, no. 12, pp. 5222–5232, Dec 2014. [Pg.1], [Pg.5], [Pg.95], [Pg.96], [Pg.98], [Pg.114]

[6] C. Fehn and R. Pastoor, "Interactive 3-dtv-concepts and key technologies," *Proceedings of the IEEE*, vol. 94, no. 3, pp. 524–538, March 2006. [Pg.1], [Pg.9], [Pg.13], [Pg.14], [Pg.17], [Pg.28], [Pg.35], [Pg.37], [Pg.39], [Pg.53], [Pg.81], [Pg.95], [Pg.97], [Pg.98], [Pg.107], [Pg.111], [Pg.127]

[7] D. M. Hoffman, A. R. Girshick, K. Akeley, and M. S. Banks, "Vergence–accommodation conflicts hinder visual performance and cause visual fatigue," *Journal of vision*, vol. 8, no. 3, p. 33, 2008. [Pg.1], [Pg.36], [Pg.38], [Pg.53], [Pg.67], [Pg.69], [Pg.95], [Pg.97], [Pg.127]

[8] G. Tech, H. Schwarz, K. Muller, and T. Wiegand, "3d video coding using the synthesized view distortion change," in *Picture Coding Symposium (PCS), 2012*, May 2012, pp. 25–28. [Pg.3], [Pg.4], [Pg.39]

[9] Y.-C. Fan and T.-C. Chi, "The novel non-hole-filling approach of depth image based rendering," in *3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video, 2008*, May 2008, pp. 325–328. [Pg.4], [Pg.53], [Pg.71], [Pg.120], [Pg.128], [Pg.130], [Pg.141], [Pg.142]

[10] R. Brott and J. Schultz, "16.3: Directional backlight lightguide considerations for full resolution autostereoscopic 3d displays," in *SID Symposium digest of technical papers*, vol. 41, no. 1. Wiley Online Library, 2010, pp. 218–221. [Pg.4], [Pg.95]

[11] P. D. Panabaker and S. S. Cho, "Quality autostereoscopic displays," in *SMPTE International Conference on Stereoscopic 3D for Media and Entertainment*, July 2010, pp. 1–10. [Pg.4], [Pg.95]

[12] H. Schwarz and T. Wiegand, "Inter-view prediction of motion data in multiview video coding," in *2012 Picture Coding Symposium*, May 2012, pp. 101–104. [Pg.5]

[13] S. Bosse, H. Schwarz, T. Hinz, and T. Wiegand, "Encoder control for renderable regions in high efficiency multiview video plus depth coding," in

2012 Picture Coding Symposium, May 2012, pp. 129–132. [Pg.5], [Pg.12], [Pg.13], [Pg.28], [Pg.67], [Pg.79], [Pg.127]

[14] Y. Chen, G. Tech, K. Wegner, and S. Yea, *Test Model 8 of 3D-HEVC and MV-HEVC*, jct3v-h1003 ed., Joint Collaborative Team on 3D Video Coding Extension Development of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11, Mar.-Apr. 2014. [Pg.xix], [Pg.5], [Pg.7], [Pg.12], [Pg.13], [Pg.28], [Pg.35], [Pg.36], [Pg.53], [Pg.54], [Pg.67], [Pg.98]

[15] G. C. Langelaar, I. Setyawan, and R. L. Lagendijk, "Watermarking digital image and video data. a state-of-the-art overview," *IEEE Signal Processing Magazine*, vol. 17, no. 5, pp. 20–46, Sep 2000. [Pg.8]

[16] R. B. Wolfgang and E. J. Delp III, "Fragile watermarking using the vw2d watermark," in *Electronic Imaging'99*. International Society for Optics and Photonics, 1999, pp. 204–213. [Pg.8]

[17] 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.8]

[18] J. Haitsma and T. Kalker, "A watermarking scheme for digital cinema," in *Proceedings 2001 International Conference on Image Processing (Cat. No.01CH37205)*, vol. 2, Oct 2001, pp. 487–489 vol.2. [Pg.8]

[19] I. J. Cox, M. Miller, and J. Bloom, "Watermarking applications and their properties," in *Information Technology: Coding and Computing, 2000. Proceedings. International Conference on*, 2000, pp. 6–10. [Pg.8]

[20] S.-J. Lee and S.-H. Jung, "A survey of watermarking techniques applied to multimedia," in *Industrial Electronics, 2001. Proceedings. ISIE 2001. IEEE International Symposium on*, vol. 1, 2001, pp. 272–277 vol.1. [Pg.8]

[21] X. Chang, W. Wang, J. Zhao, and L. Zhang, "A survey of digital video watermarking," in *Natural Computation (ICNC), 2011 Seventh International Conference on*, vol. 1, July 2011, pp. 61–65. [Pg.8]

[22] M. Alghoniemy and A. H. Tewfik, "Geometric invariance in image watermarking," *IEEE Transactions on Image Processing*, vol. 13, no. 2, pp. 145–153, Feb 2004. [Pg.9]

[23] M. Barni, "Effectiveness of exhaustive search and template matching against watermark desynchronization," *IEEE Signal Processing Letters*, vol. 12, no. 2, pp. 158–161, Feb 2005. [Pg.9]

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

[25] G. Caner, A. M. Tekalp, G. Sharma, and W. Heinzelman, "Local image registration by adaptive filtering," *IEEE Transactions on Image Processing*, vol. 15, no. 10, pp. 3053–3065, Oct 2006. [Pg.9]

[26] J. L. Dugelay, S. Roche, C. Rey, and G. Doerr, "Still-image watermarking robust to local geometric distortions," *IEEE Transactions on Image Processing*, vol. 15, no. 9, pp. 2831–2842, Sept 2006. [Pg.9]

[27] C.-T. Hsu and J.-L. Wu, "Hidden digital watermarks in images," *IEEE Transactions on Image Processing*, vol. 8, no. 1, pp. 58–68, Jan 1999. [Pg.9]

[28] S. Xiang, H. J. Kim, and J. Huang, "Invariant image watermarking based on statistical features in the low-frequency domain," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, no. 6, pp. 777–790, June 2008. [Pg.9]

[29] W. Bender, D. Gruhl, N. Morimoto, and A. Lu, "Techniques for data hiding," *IBM Systems Journal*, vol. 35, no. 3.4, pp. 313–336, 1996. [Pg.9], [Pg.103], [Pg.104]

[30] Y. Zhao and R. L. Lagendijk, "Video watermarking scheme resistant to geometric attacks," in *Proceedings. International Conference on Image Processing*, vol. 2, 2002, pp. II–145–II–148 vol.2. [Pg.9]

[31] Y. Shao, L. Zhang, G. Wu, and X. Lin, "A novel frequency domain watermarking algorithm with resistance to geometric distortions and copy attack," in *Circuits and Systems, 2003. ISCAS '03. Proceedings of the 2003 International Symposium on*, vol. 2, May 2003, pp. II–940–II–943 vol.2. [Pg.9]

[32] D. He and Q. Sun, "A rst resilient object-based video watermarking scheme," in *Image Processing, 2004. ICIP '04. 2004 International Conference on*, vol. 2, Oct 2004, pp. 737–740 Vol.2. [Pg.9]

[33] Y. Liu and J. Zhao, "Rst invariant video watermarking based on 1d dft and radon transform," in *2008 5th International Conference on Visual Information Engineering (VIE 2008)*, July 2008, pp. 443–448. [Pg.9]

[34] ——, "Rst invariant video watermarking based on log-polar mapping and phase-only filtering," in *2010 IEEE International Conference on Multimedia and Expo*, July 2010, pp. 1305–1310. [Pg.9]

[35] M. Asikuzzaman, M. J. Alam, A. J. Lambert, and M. R. Pickering, "A blind digital video watermarking scheme with enhanced robustness to geometric distortion," in *2012 International Conference on Digital Image Computing Techniques and Applications (DICTA)*, Dec 2012, pp. 1–8. [Pg.9]

[36] H. Zhang, J. Li, and C. Dong, "Multiple video zero-watermarking based on 3d dft to resist geometric attacks," in *2012 2nd International Conference on Consumer Electronics, Communications and Networks (CECNet)*, April 2012, pp. 1141–1144. [Pg.9]

[37] X. C. Yuan and C. M. Pun, "Feature based video watermarking resistant to geometric distortions," in *2013 12th IEEE International Conference on Trust, Security and Privacy in Computing and Communications*, July 2013, pp. 763–767. [Pg.9]

[38] M. Asikuzzaman, M. J. Alam, A. J. Lambert, and M. R. Pickering, "A blind high definition videowatermarking scheme robust to geometric and temporal synchronization attacks," in *2013 Visual Communications and Image Processing (VCIP)*, Nov 2013, pp. 1–6. [Pg.9]

[39] Z. Li, L. F. Liu, and C. X. Jiang, "A self-adaptive video dual watermarking based on the motion characteristic and geometric invariant for ubiquitous multimedia," in *2015 IEEE International Conference on Smart City/SocialCom/SustainCom (SmartCity)*, Dec 2015, pp. 28–32. [Pg.9]

[40] R. Maharjan, A. Alsadoon, P. W. C. Prasad, A. M. S. Rahma, A. Elchouemi, and S. A. Senanayake, "A proposed robust video watermarking algorithm: Enhanced extraction from geometric attacks," in *2016 First International Conference on Multimedia and Image Processing (ICMIP)*, June 2016, pp. 45–50. [Pg.9]

[41] D. Zheng, S. Wang, and J. Zhao, "Rst invariant image watermarking algorithm with mathematical modeling and analysis of the watermarking processes," *IEEE Transactions on Image Processing*, vol. 18, no. 5, pp. 1055–1068, May 2009. [Pg.9], [Pg.15], [Pg.68], [Pg.96], [Pg.127]

[42] A. Chammem, M. Mitrea, and F. Prêteux, "Dwt-based stereoscopic image watermarking," in *IS&T/SPIE Electronic Imaging*. International Society for Optics and Photonics, 2011, pp. 786 326–786 326. [Pg.9], [Pg.15], [Pg.68], [Pg.96], [Pg.127]

[43] D. C. Hwang, K. H. Bae, and E.-S. Kim, "Stereo image watermarking scheme based on discrete wavelet transform and adaptive disparity estimation," pp. 196–205, 2004. [Online]. Available: http://dx.doi.org/10.1117/12.506616 [Pg.9], [Pg.15], [Pg.68], [Pg.96], [Pg.127]

[44] C. Deng, X. Gao, X. Li, and D. Tao, "A local tchebichef moments-based robust image watermarking," *Signal Processing*, vol. 89, no. 8, pp. 1531 – 1539, 2009. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0165168409000474 [Pg.9], [Pg.15], [Pg.68], [Pg.96], [Pg.127]

[45] A. V. Subramanyam, S. Emmanuel, and M. S. Kankanhalli, "Robust watermarking of compressed and encrypted jpeg2000 images," *IEEE Transactions on Multimedia*, vol. 14, no. 3, pp. 703–716, June 2012. [Pg.9], [Pg.10], [Pg.15], [Pg.68], [Pg.96], [Pg.127]

[46] S. Wang, D. Zheng, J. Zhao, W. J. Tam, and F. Speranza, "Adaptive watermarking and tree structure based image quality estimation," *IEEE Transactions on Multimedia*, vol. 16, no. 2, pp. 311–325, Feb 2014. [Pg.9], [Pg.10], [Pg.15], [Pg.68], [Pg.96], [Pg.127]

[47] P. Korus, J. Biaas, and A. Dziech, "Towards practical self-embedding for jpeg-compressed digital images," *IEEE Transactions on Multimedia*, vol. 17, no. 2, pp. 157–170, Feb 2015. [Pg.9], [Pg.10], [Pg.15], [Pg.68], [Pg.96], [Pg.127]

[48] F. Sheng-Li, Y. Mei, J. Gang-Yi, S. Feng, P. Zong-Ju, and F. Sheng-li, "A digital watermarking algorithm based on region of interest for 3d image," in *2012 Eighth International Conference on Computational Intelligence and Security*, Nov 2012, pp. 549–552. [Pg.9], [Pg.15], [Pg.68], [Pg.96], [Pg.127]

[49] Y. H. Lin and J. L. Wu, "Unseen visible watermarking for color plus depth map 3d images," in *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 2012, pp. 1801–1804. [Pg.9], [Pg.15], [Pg.68], [Pg.96], [Pg.127]

[50] A. Al-Haj and Y. Salman, "A watermarking scheme for dibr images," in *2015 International Conference on 3D Imaging (IC3D)*, Dec 2015, pp. 1–6. [Pg.9], [Pg.15], [Pg.68], [Pg.96], [Pg.127]

[51] Y. Zhang, K. Khan, L. Lv, and P. Cosman, "Binocular suppression based visual masking model for stereo image watermarking," in *2015 IEEE 16th International Conference on Communication Technology (ICCT)*, Oct 2015, pp. 32–36. [Pg.9], [Pg.15], [Pg.68], [Pg.96], [Pg.127]

[52] P. Campisi, "Object-oriented stereo-image digital watermarking," *Journal of Electronic Imaging*, vol. 17, no. 4, pp. 043 024–043 024–5, 2008. [Online]. Available: http://dx.doi.org/10.1117/1.3009554 [Pg.9], [Pg.15], [Pg.68], [Pg.96], [Pg.127]

[53] E. Halici and A. Alatan, "Watermarking for depth-image-based rendering," in *Image Processing (ICIP), 2009 16th IEEE International Conference on*, Nov 2009, pp. 4217–4220. [Pg.9], [Pg.10], [Pg.15], [Pg.68], [Pg.96], [Pg.127]

[54] F. Sheng-li, Y. Mei, J. Gang-yi, S. Feng, and P. Zong-ju, "A digital watermarking algorithm based on region of interest for 3d image," pp. 549–552, Nov 2012. [Pg.9], [Pg.10], [Pg.15], [Pg.68], [Pg.96], [Pg.127]

[55] Y.-H. Lin and J.-L. Wu, "A digital blind watermarking for depth-image-based rendering 3d images," *Broadcasting, IEEE Transactions on*, vol. 57, no. 2, pp. 602–611, June 2011. [Pg.xxi], [Pg.xxii], [Pg.xxiii], [Pg.xxviii], [Pg.9], [Pg.10], [Pg.11], [Pg.15], [Pg.68], [Pg.96], [Pg.113], [Pg.114], [Pg.115], [Pg.116], [Pg.117], [Pg.118], [Pg.119], [Pg.120], [Pg.121], [Pg.122], [Pg.123], [Pg.124], [Pg.125], [Pg.127], [Pg.128], [Pg.142], [Pg.143], [Pg.144], [Pg.145], [Pg.146], [Pg.147], [Pg.148], [Pg.149], [Pg.150], [Pg.151], [Pg.152]

[56] S. Jaipuria, "Watermarking for depth map based 3d images using wavelet transform," in *Communications and Signal Processing (ICCSP), 2014 International Conference on*, April 2014, pp. 181–185. [Pg.9], [Pg.11], [Pg.15], [Pg.68], [Pg.96], [Pg.127]

[57] D. Trick, W. Berchtold, M. Schafer, and M. Steinebach, "3d watermarking in the context of video games," in *Multimedia Signal Processing (MMSP), 2013 IEEE 15th International Workshop on*, Sept 2013, pp. 418–423. [Pg.9], [Pg.11], [Pg.15], [Pg.37], [Pg.56], [Pg.68], [Pg.96], [Pg.127]

[58] K. Arun and P. Poul, "Protection of depth-image-based rendering 3d images using blind watermarking," in *Computing, Communications and Networking Technologies (ICCCNT),2013 Fourth International Conference on*, July 2013, pp. 1–6. [Pg.9], [Pg.11], [Pg.12], [Pg.15], [Pg.68], [Pg.96], [Pg.127]

[59] M. Asikuzzaman, M. Alam, A. Lambert, and M. Pickering, "A blind watermarking scheme for depth-image-based rendered 3d video using the dual-tree complex wavelet transform," in *Image Processing (ICIP), 2014 IEEE International Conference on*, Oct 2014, pp. 5497–5501. [Pg.9], [Pg.11], [Pg.15], [Pg.68], [Pg.96], [Pg.127]

[60] Y. Fu, "Robust image watermarking scheme based on 3d-dct," in *Fuzzy Systems and Knowledge Discovery, 2009. FSKD '09. Sixth International*

Conference on, vol. 5, Aug 2009, pp. 437–441. [Pg.9], [Pg.11], [Pg.15], [Pg.68], [Pg.96], [Pg.127]

[61] E. Garcia and J. Dugelay, "Texture-based watermarking of 3d video objects," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 13, no. 8, pp. 853–866, Aug 2003. [Pg.9], [Pg.15], [Pg.37], [Pg.38], [Pg.68]

[62] J. Franco-Contreras, S. Baudry, and G. Doerr, "Virtual view invariant domain for 3d video blind watermarking," in *Image Processing (ICIP), 2011 18th IEEE International Conference on*, Sept 2011, pp. 2761–2764. [Pg.xxi], [Pg.xxii], [Pg.xxiii], [Pg.xxviii], [Pg.9], [Pg.12], [Pg.15], [Pg.68], [Pg.113], [Pg.114], [Pg.115], [Pg.116], [Pg.117], [Pg.118], [Pg.119], [Pg.120], [Pg.121], [Pg.122], [Pg.123], [Pg.124], [Pg.125], [Pg.128], [Pg.142], [Pg.143], [Pg.144], [Pg.145], [Pg.146], [Pg.147], [Pg.148], [Pg.149], [Pg.150], [Pg.151], [Pg.152]

[63] G. Louizis, A. Tefas, and I. Pitas, "Copyright protection of 3d images using watermarks of specific spatial structure," in *Multimedia and Expo, 2002. ICME '02. Proceedings. 2002 IEEE International Conference on*, vol. 2, 2002, pp. 557–560 vol.2. [Pg.9], [Pg.12], [Pg.15], [Pg.37], [Pg.56], [Pg.68]

[64] K. Ntalianis, N. Doulamis, A. Doulamis, and S. Kollias, "Automatic stereoscopic video object-based watermarking using qualified significant wavelet trees," in *Consumer Electronics, 2002. ICCE. 2002 Digest of Technical Papers. International Conference on*, June 2002, pp. 188–189. [Pg.9], [Pg.13], [Pg.15], [Pg.37], [Pg.56], [Pg.68]

[65] M.-J. Lee, J.-w. Lee, and H.-K. Lee, "Perceptual watermarking for 3d stereoscopic video using depth information," in *Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP), 2011 Seventh International Conference on*, Oct 2011, pp. 81–84. [Pg.xx], [Pg.xxvii], [Pg.9], [Pg.12], [Pg.14], [Pg.15], [Pg.38], [Pg.48], [Pg.49], [Pg.50], [Pg.51], [Pg.54], [Pg.60], [Pg.62], [Pg.63], [Pg.64], [Pg.68], [Pg.77], [Pg.129]

[66] C. Wu, K. Yuan, M. Cheng, and H. Ding, "Differential watermarking scheme of stereo video," in *Communication Technology (ICCT), 2012 IEEE*

14th International Conference on, Nov 2012, pp. 744–748. [Pg.9], [Pg.13], [Pg.15], [Pg.37], [Pg.38], [Pg.56], [Pg.68]

[67] Y. H. Chen and H.-C. Huang, "A robust watermarking scheme for stereoscopic video frames," in *Consumer Electronics (ISCE), 2013 IEEE 17th International Symposium on*, June 2013, pp. 295–296. [Pg.9], [Pg.13], [Pg.14], [Pg.15], [Pg.37], [Pg.56], [Pg.68]

[68] ——, "A wavelet-based image watermarking scheme for stereoscopic video frames," in *Intelligent Information Hiding and Multimedia Signal Processing, 2013 Ninth International Conference on*, Oct 2013, pp. 25–28. [Pg.9], [Pg.13], [Pg.14], [Pg.15], [Pg.37], [Pg.56], [Pg.68]

[69] J. Franco-Contreras, S. Baudry, and G. Dorr, "Virtual view invariant domain for 3d video blind watermarking," in *2011 18th IEEE International Conference on Image Processing*, Sept 2011, pp. 2761–2764. [Pg.9], [Pg.15], [Pg.37], [Pg.56], [Pg.68]

[70] Y. H. Chen and H. C. Huang, "A wavelet-based image watermarking scheme for stereoscopic video frames," in *2013 Ninth International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, Oct 2013, pp. 25–28. [Pg.9], [Pg.15], [Pg.37], [Pg.56], [Pg.68]

[71] Z. Yao, L. Chen, R. Pan, S. Yang, and K. Curran, "Improved blind watermarking of 3d objects based on partition," in *2009 Fifth International Conference on Natural Computation*, vol. 2, Aug 2009, pp. 307–311. [Pg.9], [Pg.15], [Pg.37], [Pg.56], [Pg.68]

[72] Y. H. Chen and H. C. Huang, "A robust watermarking scheme for stereoscopic video frames," in *2013 IEEE International Symposium on Consumer Electronics (ISCE)*, June 2013, pp. 295–296. [Pg.9], [Pg.15], [Pg.37], [Pg.56], [Pg.68]

[73] M. Asikuzzaman, M. J. Alam, A. J. Lambert, and M. R. Pickering, "A blind watermarking scheme for depth-image-based rendered 3d video using the dual-tree complex wavelet transform," in *2014 IEEE International*

Conference on Image Processing (ICIP), Oct 2014, pp. 5497–5501. [Pg.9], [Pg.14], [Pg.15], [Pg.37], [Pg.56], [Pg.68]

[74] ——, "Robust dt cwt-based dibr 3d video watermarking using chrominance embedding," *IEEE Transactions on Multimedia*, vol. 18, no. 9, pp. 1733–1748, Sept 2016. [Pg.9], [Pg.12], [Pg.14], [Pg.15], [Pg.37], [Pg.56], [Pg.68]

[75] 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.10], [Pg.11], [Pg.14], [Pg.15], [Pg.28], [Pg.36], [Pg.96], [Pg.97], [Pg.119]

[76] H.-D. Kim, J.-W. Lee, T.-W. Oh, and H.-K. Lee, "Robust dt-cwt watermarking for dibr 3d images," *Broadcasting, IEEE Transactions on*, vol. 58, no. 4, pp. 533–543, Dec 2012. [Pg.xxi], [Pg.xxii], [Pg.xxiii], [Pg.xxviii], [Pg.11], [Pg.12], [Pg.14], [Pg.15], [Pg.16], [Pg.27], [Pg.29], [Pg.54], [Pg.57], [Pg.96], [Pg.113], [Pg.114], [Pg.115], [Pg.116], [Pg.117], [Pg.118], [Pg.119], [Pg.120], [Pg.121], [Pg.122], [Pg.123], [Pg.124], [Pg.125], [Pg.128], [Pg.131], [Pg.142], [Pg.143], [Pg.144], [Pg.145], [Pg.146], [Pg.147], [Pg.148], [Pg.149], [Pg.150], [Pg.151], [Pg.152]

[77] A. Tefas, G. Louizis, and I. Pitas, "3d image watermarking robust to geometric distortions," in *Acoustics, Speech, and Signal Processing (ICASSP), 2002 IEEE International Conference on*, vol. 4, May 2002, pp. IV–3465–IV–3468. [Pg.12]

[78] Y. Han, "Geometric algorithms for least squares estimation of 3-d information from monocular image," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 15, no. 2, pp. 269–282, Feb 2005. [Pg.15], [Pg.68]

[79] Y. Guan, Y. Zhu, X. Liu, G. Luo, Z. Sun, and L. Zhang, "A digital blind watermarking scheme based on quantization index modulation in depth map for 3d video," in *Control Automation Robotics Vision (ICARCV), 2014 13th International Conference on*, Dec 2014, pp. 346–351. [Pg.xxvii], [Pg.xxviii],

[Pg.15], [Pg.68], [Pg.76], [Pg.77], [Pg.78], [Pg.79], [Pg.89], [Pg.90], [Pg.91], [Pg.92], [Pg.93]

[80] L. Zhang and W. J. Tam, "Stereoscopic image generation based on depth images for 3d tv," *IEEE Transactions on Broadcasting*, vol. 51, no. 2, pp. 191–199, June 2005. [Pg.16], [Pg.145]

[81] 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.18], [Pg.26], [Pg.71], [Pg.97]

[82] 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, special Issue on Subband/Wavelet Interframe Video Coding. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0923596504000372 [Pg.23]

[83] S.-J. Choi and J. W. Woods, "Motion-compensated 3-d subband coding of video," *IEEE Transactions on Image Processing*, vol. 8, no. 2, pp. 155–167, Feb 1999. [Pg.23]

[84] 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 *Image Processing, 2004. ICIP '04. 2004 International Conference on*, vol. 5, Oct 2004, pp. 2845–2848 Vol. 5. [Pg.23]

[85] 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.23]

[86] S. Rana, N. Sahu, and A. Sur, "Robust watermarking for resolution and quality scalable video sequence," *Multimedia Tools and Applications*, pp. 1–30, 2014. [Online]. Available: http://dx.doi.org/10.1007/s11042-014-2023-1 [Pg.23], [Pg.24], [Pg.39], [Pg.56], [Pg.81]

[87] S. Zheng, S. Jayasumana, B. Romera-Paredes, V. Vineet, Z. Su, D. Du, C. Huang, and P. H. S. Torr, "Conditional random fields as recurrent neural networks," in *International Conference on Computer Vision (ICCV)*, 2015. [Pg.25], [Pg.69], [Pg.70], [Pg.74], [Pg.82]

[88] E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PP, no. 99, pp. 1–1, 2016. [Pg.26]

[89] R. Anderson, N. Kingsbury, and J. Fauqueur, "Determining multiscale image feature angles from complex wavelet phases," in *Image Analysis and Recognition*, ser. Lecture Notes in Computer Science, M. Kamel and A. Campilho, Eds. Springer Berlin Heidelberg, 2005, vol. 3656, pp. 490–498. [Online]. Available: http://dx.doi.org/10.1007/11559573_61 [Pg.xxii], [Pg.26], [Pg.27], [Pg.55], [Pg.57], [Pg.130], [Pg.131], [Pg.133], [Pg.134], [Pg.149]

[90] N. Kingsbury, "Complex wavelets for shift invariant analysis and filtering of signals," *Applied and Computational Harmonic Analysis*, vol. 10, no. 3, pp. 234 – 253, 2001. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1063520300903439 [Pg.26], [Pg.27], [Pg.130], [Pg.131], [Pg.134], [Pg.149]

[91] ——, "The dual-tree complex wavelet transform: A new technique for shift invariance and directional filters," pp. 319–322. [Pg.27], [Pg.130], [Pg.131], [Pg.141], [Pg.149]

[92] J. G. Proakis, "Digital communications. 1995," *McGraw-Hill, New York*. [Pg.28]

[93] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, April 2004. [Pg.29], [Pg.30]

[94] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *IEEE Transactions on Image Processing*, vol. 15, no. 2, pp. 430–444, Feb 2006. [Pg.29], [Pg.31]

[95] Y. L. Xiaopeng Fan, Wen Gao and D. Zhao, "Flicking reduction in all intra frame coding," *Joint Video Team of ISO/IEC MPEG and ITU-T VCEG. JVT-E070*, October 2002. [Pg.29], [Pg.31]

[96] A. D.Vatolin, M.Smirnov and V.Yoockin, "Msu video quality measurement tool," MSU Graphics and Media Lab, Tool, 2001-2008. [Online]. Available: http://www.compression.ru/video/ [Pg.29], [Pg.31]

[97] K. Egiazarian, J. Astola, N. Ponomarenko, V. Lukin, F. Battisti, and M. Carli, "New full-reference quality metrics based on hvs," in *CD-ROM proceedings of the second international workshop on video processing and quality metrics, Scottsdale, USA*, vol. 4, 2006. [Pg.29], [Pg.30]

[98] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *The Thrity-Seventh Asilomar Conference on Signals, Systems Computers, 2003*, vol. 2, Nov 2003, pp. 1398–1402 Vol.2. [Pg.29], [Pg.31]

[99] D. Scharstein and C. Pal, "Learning conditional random fields for stereo," in *Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on*, June 2007, pp. 1–8. [Pg.xxii], [Pg.32], [Pg.75], [Pg.113], [Pg.142], [Pg.143], [Pg.144]

[100] S. Rana and A. Sur, "Blind 3d video watermarking based on 3d-hevc encoder using depth," in *Proceedings of the Ninth Indian Conference on Computer Vision, Graphics and Image Processing*, ser. ICVGIP '14. New York, NY, USA: ACM, 2014. [Online]. Available: http://dx.doi.org/10.1145/2683483.2683535 [Pg.xxvii], [Pg.54], [Pg.61], [Pg.62], [Pg.63], [Pg.64], [Pg.81], [Pg.129]

[101] ——, "3D video watermarking using DT-DWT to resist synthesis view attack," in *23rd European Signal Processing Conference (EUSIPCO) (EUSIPCO 2015)*, Nice, France, Aug. 2015. [Pg.xxi], [Pg.xxii], [Pg.xxviii],

[Pg.96], [Pg.97], [Pg.113], [Pg.114], [Pg.115], [Pg.117], [Pg.118], [Pg.119], [Pg.120], [Pg.121], [Pg.122], [Pg.123], [Pg.124], [Pg.125]

[102] M. Jridi, Y. Ouerhani, and A. Alfalou, "Low complexity dct engine for image and video compression," pp. 86 560F–86 560F–9, 2013. [Online]. Available: http://dx.doi.org/10.1117/12.2006174 [Pg.123]

[103] V. D. Silva, A. Fernando, S. Worrall, H. K. Arachchi, and A. Kondoz, "Sensitivity analysis of the human visual system for depth cues in stereoscopic 3-d displays," *IEEE Transactions on Multimedia*, vol. 13, no. 3, pp. 498–506, June 2011. [Pg.127]

[104] H. Kim and S. Lee, "Transition of visual attention assessment in stereoscopic images with evaluation of subjective visual quality and discomfort," *IEEE Transactions on Multimedia*, vol. 17, no. 12, pp. 2198–2209, Dec 2015. [Pg.127], [Pg.129]

[105] J.-M. Lina and L. Gagnon, "Image enhancement with symmetric daubechies wavelets," pp. 196–207, 1995. [Online]. Available: http://dx.doi.org/10.1117/12.217575 [Pg.130]

Appendix: A

Summary of Publications

Related to Ph.D.

- **Rana, S.**, Sur, A.: ***"View Invariant DIBR-3D Image Watermarking Using DT-CWT".***: (Under Review)
- **Rana, S.**, Sur, A.: ***"Depth-Based View-Invariant Blind 3D Image Watermarking".***: ACM Trans. Multimedia Comput. Commun. Appl. 12, 4, Article 48 (August 2016), 23 pages. DOI: 10.1145/2957751 URL: http://dx.doi.org/10.1145/2957751
- **Rana, S.**, Gaj, S., Sur, A.: ***"View Invariant 3D Video Watermarking using Depth Based Embedding".*** 2016 International Conference on Digital Image Computing: Techniques and Applications (DICTA). DOI: 10.1109/DICTA.2016.7797095 URL: http://dx.doi.org/10.1109/DICTA.2016.7797095
- **Rana, S.**, Gaj, S., Sur, A., Bora, P.: ***"Segmentation Based 3D Depth Watermarking using SIFT".*** 2016 IEEE 18th International Workshop on Multimedia Signal Processing (MMSP), Montreal, QC, 2016, pp. 1-6. doi: 10.1109/MMSP.2016.7813367 URL: http://dx.doi.org/10.1109/MMSP.2016.7813367
- **Rana, S.**, Sur, A.: ***"3D video watermarking using DT-DWT to resist synthesis view attack".***: in European Signal Processing Confer-

ence 2015 (EUSIPCO 2015), Nice, France, Aug. 2015. DOI: 10.1109/EUSIPCO.2015.7362342 URL http://dx.doi.org/10.1109/EUSIPCO.2015.7362342

- **Rana, S.**, Sur, A.: ***"Blind 3D Video Watermarking Based on 3D-HEVC Encoder Using Depth".*** ICVGIP '14: Proceedings of the Nineth Indian Conference on Computer Vision, Graphics and Image Processing. New York, NY, USA: ACM, 2014. DOI: 10.1145/2683483.2683535. URL: http://dx.doi.org/10.1145/2683483.2683535

Spin-off to Ph.D.

- **Rana, S.**, Gaj, S., Sur, A.: ***"Detection of Fake 3D Video".*** (Communicated)

- Gaj, S., **Rana, S.**, Sur, A., Bora, P.: ***"A Robust Watermarking Scheme against Frame Blending, Projection and Content Adaptation Attacks".*** 2016 International Conference on Digital Image Computing: Techniques and Applications (DICTA). DOI: 10.1109/DICTA.2016.7796996 URL: http://dx.doi.org/10.1109/DICTA.2016.7796996

- **Rana, S.**, Gaj, S., Sur, A., Bora, P.: ***"Detection of Fake 3D Video Using CNN".*** 2016 IEEE 18th International Workshop on Multimedia Signal Processing (MMSP), Montreal, QC, 2016, pp. 1-6. doi: 10.1109/MMSP.2016.7813368 URL: http://dx.doi.org/10.1109/MMSP.2016.7813368

- Gaj, S., **Rana, S.**, Sur, A., Bora, P.: ***"A Drift Compensated Reversible Watermarking Scheme for H.265/HEVC".*** 2016 IEEE 18th International Workshop on Multimedia Signal Processing (MMSP), Montreal, QC, 2016, pp. 1-6. doi: 10.1109/MMSP.2016.7813358 URL: http://dx.doi.org/10.1109/MMSP.2016.7813358

- **Rana, S.**, Mattu, V., Sur, A.: ***"SIFT Based View Invariant Watermarking for DIBR 3D Image".***: in Proceedings of the Fifth National Conference on Computer Vision, Pattern Recognition, Image Pro-

cessing and Graphics 2015. DOI: 10.1109/NCVPRIPG.2015.7490030 URL: http://dx.doi.org/10.1109/NCVPRIPG.2015.7490030

- Gaj, S., **Rana, S.**, Lekharu, A., Sur, A., Bora, P.: ***"RST Invariant Multi View 3D Image Watermarking Using DWT and SVD".*** in Proceedings of the Fifth National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics 2015. DOI: 10.1109/NCVPRIPG.2015.7490066 URL: http://dx.doi.org/10.1109/NCVPRIPG.2015.7490066

- **Rana, S.**, Sahu, N., Sur, A.: ***"Robust watermarking for resolution and quality scalable video sequence".*** Multimedia Tools and Applications Volume 74, Issue 18, pp. 7773-7802 (2015). DOI 10.1007/ s11042-014-2023-1. URL: http://dx.doi.org/10.1007/s11042-014-2023-1

- Sahu, N, **Rana, S.**, Sur, A..: ***"MCDCT-TF based video watermarking resilient to temporal and quality scaling".*** Multimedia Tools and Applications pp. 1-26 (2015). DOI: 10.1007/s11042-015-2949-y. URL: http://dx.doi.org/DOI10.1007/s11042-015-2949-y

- Sur, A., Krishna, S., Sahu, N., **Rana, S.**: ***"Detection of Motion Vector Based Video Steganography".*** Multimedia Tools and Applications Volume 74, Issue 23, pp. 1047910494 (2015). DOI: 10.1007/s11042-014-2181-1. URL: http://dx.doi.org/10.1007/s11042-014-2181-1

Under Preparation

- **Rana, S.**, Kamra, R., Sur, A.: ***"Motion Vector Based Video Steganography using Similar Region Selection".***

- **Rana, S.**, Sur, A.: ***"3D Video Watermarking Using View and Depth Embedding".***