Digital Watermarking for Depth-Image-Based Rendering 3D Images and Its Application to Quality Evaluation

by

Lei Chen

A thesis submitted to the Faculty of Graduate and Postdoctoral Studies
in partial fulfillment of the requirements for the degree of

Doctorate of Philosophy
in
Electrical and Computer Engineering

School of Electrical Engineering and Computer Science
Faculty of Engineering
University of Ottawa

Abstract

Due to the rapid development of 3D display market, the protection and authentication of the intellectual property rights of 3D multimedia has become an essential concern. As a consequence, the digital watermarking for 3D image and video is attracting considerable attention. The depth-image-based rendering (DIBR) technique has been playing a critical role in 3D contents representation because of its numerous advantages.

A good digital watermarking algorithm should be robust to various possible attacks, including geometric distortions and compressions. And different from ordinary 2D digital watermarking, there are more specific requirements for 3D watermarking, especially for DIBR 3D image watermarking. Not only the center view, but also the virtual left and right views can be illegally distributed. Therefore, the embedded watermark information should be accurately extracted from these three views individually for content authentication, even under attacks.

In this thesis, we focus on the research of digital watermarking and watermarking based quality evaluation for DIBR 3D images. We first present a 2D image and video watermarking method based on contourlet transform, which is then extended to a robust contourlet-based watermarking algorithm for DIBR 3D images. The watermark is embedded into the center view by quantizing certain contourlet coefficients. The virtual left and right views are synthesized from the watermarked center view and the corresponding depth map. One advantage of our algorithm is its simplicity and practicality. However, the performance on watermark extraction needs to be further improved. As an improvement, a blind watermarking algorithm for DIBR 3D images based on feature regions and ridgelet transform is proposed. The watermarked view

has good perceptual quality under both the objective and subjective image quality measures. Compared with other related and state-of-the-art methods, the proposed algorithm shows superiority in terms of watermark extraction and robustness to various attacks.

Furthermore, as one of the most promising techniques for quality evaluation, a watermarking based quality evaluation scheme is developed for DIBR 3D images. The qualities of the watermarked center view and the synthesized left and right views under distortions can be estimated by examining the degradation of corresponding extracted watermarks. The simulation results demonstrate that our scheme has good performance of quality evaluation for DIBR 3D images under the attacks.

Contents

Abstract **ii**

Contents **iv**

List of Tables **x**

List of Figures **xiii**

List of Acronyms **xxi**

Acknowledgement **xxiii**

Dedication **xxiv**

1 Introduction **1**

1.1 Digital watermarking 1

1.2 3D image watermarking 4

1.3 Depth-image-based rendering system 7

1.3.1 Pre-processing of depth map 8
1.3.2 3D image warping 10
1.3.3 Hole-filling 11
1.4 Evaluation of 3D image watermarking 12
1.5 Applications of 3D image watermarking 14
1.6 Objectives of the thesis 16
1.7 Contributions of the research 18
1.8 Thesis structure 21

2 Fundamental theories and techniques **24**

2.1 Geometric transformations and compressions 24
2.1.1 Geometric transformations 25
2.1.2 Compressions 26
2.2 Common DIBR processing 28
2.2.1 Baseline distance adjustment 28
2.2.2 Depth map variation 28
2.2.3 Different rendering conditions 30
2.3 Contourlet transform 30
2.3.1 Laplacian pyramid 31
2.3.2 Directional filter bank 32
2.3.3 Discrete contourlet transform 33

2.4 Principal component analysis 35
2.5 Scale invariant feature transform 37
2.5.1 Scale space extrema detection 38
2.5.2 Keypoint localization 39
2.5.3 Orientation assignment 40
2.5.4 Keypoint descriptor 40
2.6 Ridgelet transform 42
2.7 Dual-tree complex wavelet transform 44
2.8 Performance criteria and measurements 47
2.8.1 Performance criteria for watermark invisibility 47
2.8.2 Performance measurements for watermark extraction 49
2.8.3 Performance measurements for quality evaluation 51

3 Literature review **54**

3.1 Watermarking algorithms for DIBR 3D images 55
3.2 Watermarking based quality evaluation algorithms 59
3.2.1 Watermarking based 2D image quality evaluation algorithms . 61
3.2.2 3D image quality evaluation algorithms 62
3.3 Contourlet transform based watermarking algorithms 64
3.4 Principal component analysis based watermarking algorithms 67
3.5 Feature based watermarking algorithms 68

3.6 Ridgelet transform based watermarking algorithms 70

3.7 Others 71

4 A contourlet-based 2D watermarking algorithm 74

4.1 The motivation 75

4.2 The proposed image watermarking algorithm 76

4.2.1 Watermark embedding 76

4.2.2 Watermark extraction 78

4.3 Experimental results and analysis 79

4.3.1 Watermark invisibility 80

4.3.2 Robustness against various attacks 81

4.3.3 Performance analysis on robustness to attacks 83

4.4 Extended video watermarking algorithm 85

4.4.1 Watermark embedding 85

4.4.2 Watermark extraction 86

4.4.3 Experimental results 88

4.5 Summary 94

5 A contourlet-based watermarking algorithm for DIBR 3D images 96

5.1 Contourlet transform 97

5.2 Watermark embedding 101

5.3 Watermark extraction 103
5.4 Error analysis 106
5.5 Experimental results and analysis 106
5.5.1 Watermark invisibility 109
5.5.2 Robustness to geometric attacks 112
5.5.3 Robustness to baseline adjustment and depth map variation . 114
5.5.4 Robustness to different DIBR methods 116
5.5.5 Further discussions 116
5.6 Summary 118

6 A novel watermarking algorithm for DIBR 3D images using feature regions 119

6.1 The proposed watermarking algorithm 120
6.1.1 Feature point detection 120
6.1.2 Feature region assignment 122
6.1.3 Watermark embedding 125
6.1.4 Watermark extraction 126
6.1.5 Error probability 129
6.2 Experimental results and analysis 132
6.2.1 Watermark invisibility 132
6.2.2 Robustness to various attacks 134

6.2.3 Robustness to DIBR processing 139
6.2.4 Performance comparison 142
6.2.5 Experimental results on other image pairs 144
6.3 Summary 145
6.4 Some discussions 145

7 Watermarking based quality evaluation for DIBR 3D images 147

7.1 The proposed watermarking based quality evaluation scheme 148
7.1.1 Watermark embedding 151
7.1.2 Watermark extraction 153
7.1.3 Quality evaluation 155
7.2 Experimental results and analysis 156
7.2.1 NC-Quality mapping curve 158
7.2.2 Accuracy of quality evaluation 161
7.2.3 Performance comparison 171
7.3 Summary 173
7.4 Some discussions 174

8 Conclusions and future work 177

Bibliography 180

List of Tables

4.1 Quality measurements of the watermarked images in terms of PSNR, SSIM and MOS. 81

4.2 The NC values of the watermark extracted from the watermarked and un-watermarked images under various attacks. 82

4.3 Target videos and their sizes. 89

4.4 The PSNR values of the watermarked videos under various attacks. . 90

4.5 The NC values of the watermark extracted from the target videos under various attacks. 91

4.6 The NC values of the watermark extracted by the compared video watermarking methods under attacks. 94

5.1 Sizes of test image pairs. 107

5.2 Quality measurements of the watermarked center images for the proposed algorithm in terms of PSNR, SSIM and MOS. 110

5.3 Average NC of the watermark extracted from the center view, the virtual left and right views. 111

5.4 Average PSNR and SSIM, watermark bits and BER of the watermark extracted from the center view, the virtual left and right views. . . . 112

5.5 Comparison of BER and NC for depth map variation. 115

5.6 BER and NC of the extracted watermark for different DIBR methods. 116

5.7 Average PSNR and SSIM of the watermarked images for different subbands. 117

5.8 Average BER of the extracted watermark for different subbands pairing. 117

6.1 The detection accuracy of reference feature points. 129

6.2 Quality measurements of the watermarked center images for the proposed algorithm in terms of PSNR, SSIM and MOS. 134

6.3 Average NC of the watermark extracted from the center view, the virtual left and right views. 135

6.4 Average BER of the extracted watermark for depth map variation and different DIBR methods. 141

6.5 Performance comparison with other related methods. 143

6.6 The watermark extraction accuracy on all the image pairs of the three databases. 144

7.1 Average PSNR (dB) of the center images and the synthesized left and right images for different watermarked subbands. 151

7.2 Value ranges of the control parameters for distortion simulation. . . . 158

7.3 Evaluation accuracy in terms of PSNR and SSIM in RMSE. 167

7.4 Evaluation accuracy in terms of PSNR and SSIM in PLCC. 167

7.5 Evaluation accuracy in terms of PSNR and SSIM in SRCC. 169

7.6 Evaluation accuracy in terms of the 3D quality metrics for the synthesized views in RMSE, PLCC and SRCC. 172

7.7 Performance comparison on quality evaluation accuracy for the center images. 173

7.8 Performance comparison on quality evaluation accuracy for the synthesized 3D images. 173

List of Figures

1.1 A typical watermarking system. 2

1.2 Illegal content distribution at the consumer side. 7

1.3 Depth-image-based rendering system. 8

1.4 (a) The center image; (b) The corresponding depth image; (c) The mapping between intensity and depth value. 9

1.5 Pixel-wise mapping in 3D image warping [29]. 11

1.6 Hole-filling. (a) The right view with holes of image *Ballet*; (b) The corresponding right view after hole-filling. 13

1.7 The requirements for a good digital watermarking system. 14

2.1 Depth map variation. (a) and (b) The original depth images of *Art* and *Books*; (c) and (d) The depth images with Gaussian noise (Variance: 0.1); (e) and (f) The depth images with JPEG compression (Quality factor: 10). 29

2.2 The analysis and synthesis of Laplacian pyramid (LP) [52]. (a) One level of decomposition; (b) The new reconstruction scheme for the LP. 32

2.3 An example of frequency partitioning in DFB, where $l = 3$ and there are $2^3 = 8$ frequency subbands [52]. 33

2.4 The image decomposition in contourlet filter bank (LP and DFB). . . 34

2.5 Contourlet transform of the *Barbara* image. The numbers in the subfigure on bottom left denote the corresponding directional subbands accordingly. 35

2.6 Keypoints selection using SIFT. (a) The original image of *Breakdancers*; (b) The final selected keypoints. 41

2.7 Diagram for the finite ridgelet transform [61]. 42

2.8 DT-CWT for 1-D signal [64]. 45

2.9 2-D impulse responses of the reconstruction filters in the DT-CWT [65]. 46

4.1 Test videos. (a) *Lena* (512×512); (b) *Peppers* (256×256); (c) *Boat* (512×512); (d) *Barbara* (512×512); (e) *House* (256×256). 80

4.2 Estimation of the rotation angle for *Barbara* image. (a) The watermarked *Barbara* image rotated by $-30°$; (b) The corresponding NC values with different rotation angle corrections. 83

4.3 The original *Mobile* video frame in contourlet domain. 86

4.4 The original frame and the corresponding watermarked frame. (a) The first frame of video *Mobile*; (b) The first frame of the watermarked video *Mobile* (PSNR= 41.3966 dB). 87

4.5 The watermarked *Mobile* frames under various video attacks. (a) Rotation (5°); (b) Gaussian noise (Variance: 0.05); (c) Median filtering (5×5); (d) Histogram equalization. 88

4.6 Test videos. (a) *Mobile*; (b) *Stefan*; (c) *Flower*; (d) *Foreman*. 89

4.7 The NC values of the watermark extracted from the target videos under MPEG-2 compression with different bit rates. 93

5.1 The proposed watermark embedding and extraction algorithms. . . . 97

5.2 Contourlet transform of the *Barbara* image. The numbers in the subfigure on bottom left denote the corresponding decomposed subbands accordingly. 98

5.3 Histograms of the contourlet subbands for image *Barbara*. (a) The histogram of subband Y0; (b) The histogram of subband Y3; (c) The histogram of subband Y1; (d) The histogram of subband Y2. 100

5.4 Histograms of the quantization error differences from the un-watermarked images. 107

5.5 Test image pairs of center images and corresponding depth images. (a) Art; (b) Books; (c) Dolls; (d) Moebius; (e) Teddy; (f) Cones; (g) Aloe; (h) Laundry; (i) Orbi; (j) Interview. 108

5.6 Watermark invisibility for *Interview*. (a) The original center view; (b) The depth map; (c) The watermarked center view (PSNR=41.07 dB); (d) The virtual left view rendered from the watermarked center view and the depth map (PSNR=39.26 dB); (e) The virtual right view rendered from the watermarked center view and the depth map (PSNR=39.07 dB). 109

5.7 Average BER of the watermark extracted from the watermarked right views under various attacks. (a) Rotation; (b) Scaling; (c) Gaussian noise; (d) Median filtering; (e) JPEG compression; (f) Rotation and cropping. 113

5.8 Average BER for the watermarked right views under baseline distance adjustment. 114

5.9 Depth map pre-processing. (a) The original depth map of *Teddy*; (b) The depth map with Gaussian noise (Variance: 0.1); (c) The depth map with JPEG compression (Quality factor: 10). 115

6.1 The proposed watermark embedding and extraction algorithms. . . . 121

6.2 The selected feature points of image *Ballet*. (a) The original center image; (b) The selected feature points of the original center image. . . 123

6.3 The selected feature regions for watermark embedding. 124

6.4 The error probability for different values of SNR and r_p. 131

6.5 Test image pairs of center images and corresponding depth images. (a) Ballet (1024 × 768); (b) Breakdancers (1024 × 768); (c) Art (1390 × 1110); (d) Books (1390 × 1110); (e) Dolls (1390 × 1110); (f) Moebius (450 × 375); (g) Teddy (450 × 375); (h) Cones (720 × 576); (i) Orbi (720 × 576); (j) Interview (720 × 576). 133

6.6 Average BER of the watermark extracted from the distorted center view and synthesized left and right views. (a) Rotation; (b) Scaling; (c) Gaussian noise; (d) JPEG compression; (e) JPEG 2000 compression; (f) Combined attacks. 136

6.7 Average BER of the extracted watermark under baseline distance adjusting. 140

6.8 Depth map pre-processing. (a) Original depth map of *Breakdancers*; (b) Depth map with Gaussian noise (Variance: 0.1); (c) Depth map with JPEG compression (Quality factor: 10); (d) Depth map with JPEG 2000 compression (Compression ratio: 50). 141

7.1 The proposed watermarking based image quality evaluation scheme. . 149

7.2 Structure of DT-CWT coefficients. 150

7.3 Visual results of image *Ballet* after watermark embedding. (a) The original center view; (b) The depth map; (c) The watermarked center view (PSNR=42.52 dB); (d) The synthesized left view (PSNR=41.73 dB); (e) The synthesized right view (PSNR=41.16 dB). 154

7.4 Ideal mapping curves under JPEG compression for different number of images. 158

7.5 The generated mapping curves for the watermarked images in terms of PSNR and SSIM under Gaussian noise. (a) NC-PSNR mapping curve of the center images; (b) NC-SSIM mapping curve of the center images; (c) NC-PSNR mapping curve of the synthesized images; (d) NC-SSIM mapping curve of the synthesized images. 160

7.6 The generated mapping curves for the synthesized images in terms of 3D image quality metrics [87] [88] [90] under Gaussian noise. (a) The mapping curve in terms of the 3D quality metric proposed in [87]; (b) The mapping curve in terms of the 3D quality metric proposed in [88]; (c) The mapping curve in terms of the 3D quality metric proposed in [90]. 162

7.7 Quality evaluation for the center view and the synthesized views under Gaussian noise. (a) Quality evaluation for the center view in terms of PSNR; (b) Quality evaluation for the center view in terms of SSIM; (c) Quality evaluation for the synthesized views in terms of PSNR; (d) Quality evaluation for the synthesized views in terms of SSIM; (e) Quality evaluation for the synthesized views in terms of the 3D quality metric proposed in [87]; (f) Quality evaluation for the synthesized views in terms of the 3D quality metric proposed in [88]; (g) Quality evaluation for the synthesized views in terms of the 3D quality metric proposed in [90]. 165

7.8 Quality evaluation for the center view and the synthesized views under JPEG compression. (a) Quality evaluation for the center view in terms of PSNR; (b) Quality evaluation for the center view in terms of SSIM; (c) Quality evaluation for the synthesized views in terms of PSNR; (d) Quality evaluation for the synthesized views in terms of SSIM; (e) Quality evaluation for the synthesized views in terms of the 3D quality metric proposed in [87]; (f) Quality evaluation for the synthesized views in terms of the 3D quality metric proposed in [88]; (g) Quality evaluation for the synthesized views in terms of the 3D quality metric proposed in [90]. 167

7.9 Quality evaluation for the center view and the synthesized views under JPEG 2000 compression. (a) Quality evaluation for the center view in terms of PSNR; (b) Quality evaluation for the center view in terms of SSIM; (c) Quality evaluation for the synthesized views in terms of PSNR; (d) Quality evaluation for the synthesized views in terms of SSIM; (e) Quality evaluation for the synthesized views in terms of the 3D quality metric proposed in [87]; (f) Quality evaluation for the synthesized views in terms of the 3D quality metric proposed in [88]; (g) Quality evaluation for the synthesized views in terms of the 3D quality metric proposed in [90]. 169

7.10 Quality evaluation for the center view and the synthesized views under Gaussian blur. (a) Quality evaluation for the center view in terms of PSNR; (b) Quality evaluation for the center view in terms of SSIM; (c) Quality evaluation for the the synthesized views in terms of PSNR; (d) Quality evaluation for the synthesized views in terms of SSIM; (e) Quality evaluation for the synthesized views in terms of the 3D quality metric proposed in [87]; (f) Quality evaluation for the synthesized views in terms of the 3D quality metric proposed in [88]; (g) Quality evaluation for the synthesized views in terms of the 3D quality metric proposed in [90]. 171

List of Acronyms

BER	Bit Error Rate
CT	Contourlet Transform
DCT	Discrete Cosine Transform
DFB	Directional Filter Bank
DFT	Discrete Fourier Transform
DIBR	Depth-Image-Based Rendering
DoG	Difference of Gaussian
DSCQS	Double-Stimulus Continuous Quality Scale
DT-CWT	Dual-Tree Complex Wavelet Transform
DWT	Discrete Wavelet Transform
FBP	Finite Back Projection
FRAT	Finite Radon Transform
FRIT	Finite Ridgelet Transform
GOP	Group of Pictures
HEVC	High Efficiency Video Coding
HVS	Human Visual System
ISS	Improved Spread Spectrum
JPEG	Joint Photographic Experts Group

KLT	Karhunen-Loeve Transform
LP	Laplacian Pyramid
MOS	Mean Opinion Score
MPEG	Moving Picture Experts Group
NC	Normalized Correlation
NVF	Noise Visibility Function
PCA	Principal Component Analysis
PLCC	Pearson's Linear Correlation Coefficient
PRN	Pseudo Random Number
PSNR	Peak Signal-to-Noise Ratio
RMSE	Root Mean Square Error
RST	Rotation, Scaling and Translation
RT	Ridgelet Transform
SIFT	Scale Invariant Feature Transform
SIR	Stereo Image Recording
SRCC	Spearman's Rank Correlation Coefficient
SSIM	Structural Similarity
ZPS	Zero Parallax Setting

Acknowledgements

First and foremost, I would like to thank my supervisor, Professor Jiying Zhao, for bringing the topic of digital watermarking to me, for his patient guidance and valuable feedbacks during every step of my research. I gratefully acknowledge the China Scholarship Council for funding my work.

I also would like to thank my lab colleagues for their generous help and sincere friendship these years.

Finally, I would like to deeply thank my family for all their love and encouragement, for my parents who raised me and support me all the way. I love you.

This thesis is dedicated to my beloved family.

Chapter 1

Introduction

1.1 Digital watermarking

Due to the rapid development of internet technologies and multimedia applications, the consumers can easily have access to digital information. However, it also leads to the problem of illegal copying and redistribution of digital media [1] [2]. Therefore, the protection and authentication of the intellectual property rights of digital information has become an essential concern.

The conventional cryptographic system permits only valid key holders access to encrypted data. But after the data is decrypted, it would be difficult to trace illegal reproduction. As a consequence, the digital watermarking is attracting considerable attention as an effective complementary technology. It is an identification code that is permanently embedded in the data and remains present within the data after any decryption process. And different from the steganography technique [3] [4], a good digital watermarking scheme can be robust to various possible attacks. Moreover,

there should be no perceptible difference between the watermarked signal and the original signal, and the watermark should be difficult to remove or alter without damaging the host signal [5].

Fig. 1.1 presents a typical watermarking system, which includes a watermark embedding process and a watermark extraction process. The watermark embedder and the watermark detector are denoted by $e(\cdot)$ and $d(\cdot)$ respectively. The inputs to the watermark embedder are the watermark W, the cover signal C and the security key K. The cover or host signal can be an image, video, audio or other media formats. The watermark can be a number sequence or a binary pattern. The security key is used to enhance the security of the whole system. The output of the watermark embedder is the watermarked signal C_w. Therefore, the watermark embedding process can be described as follows:

$$C_w = e(C, W, K) \tag{1.1}$$

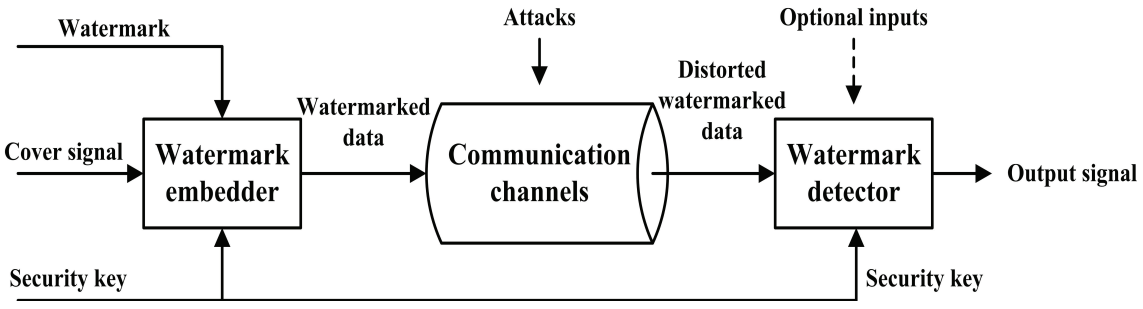


Figure 1.1: A typical watermarking system.

For the watermark detector, the inputs are the distorted watermarked signal C'_w, the security key K, and depending on the approach, the original cover signal C and/or the original watermark W. As discussed in [6], the watermark detector has two

functions. First, it extracts the possible watermark from the watermarked signal; the second function is to justify whether the extracted watermark contains the original watermark or not, and compare the extracted watermark with the original one to obtain a confidence measurement. Therefore, the output of the watermark detector can be the detected existence of the watermark, the extracted watermark W' or a desired measurement. The watermark extraction process can be expressed as:

$$W' = d(C'_w, K, \cdots) \tag{1.2}$$

For a typical digital watermarking system, there are some necessary requirements:

1. The watermark W' can be detected from C_w or C'_w with or without the existence of the cover signal C.

2. C_w should be highly close to C and there is no perceptible difference between them.

3. If C_w is unchanged, the extracted watermark W' exactly matches the original watermark W.

4. For robust watermarking, if C'_w is modified, W' should still be correlated with W so that the existence of the watermark can be clearly judged.

5. For fragile watermarking, W' can indicate the possible tampering to C'_w and provide reflection on the degradation of C'_w.

According to the cover signal that the watermark is embedded into, the digital watermarking can be divided into different categories: digital image watermarking

[7], digital video watermarking [8], digital audio watermarking [9], 3D digital watermarking [10] and others such as text watermarking [11], software watermarking [12], and database watermarking [13]. Digital watermarking can also be categorized into three groups based on the availability of original data: non-blind watermarking [14], semi-blind watermarking [15] and blind watermarking [16]. The non-blind watermarking requires the original data to detect the embedded watermark. The semi-blind watermarking does not require the original data but the watermark, while the blind watermarking requires neither of them for watermark extraction. The blind watermarking is more preferable for its practicality and efficiency.

In this thesis, most of the discussions will be focused on the blind digital 3D image watermarking.

1.2 3D image watermarking

Due to the rapid development of internet technology and three-dimensional (3D) display market, the production of 3D content available for human consumption has increased dramatically. Quantities of recent movies have been produced in 3D form, and it is expected that the 3D image and video applications will be widely commercialized in the near future. Consequently, the protection of the intellectual property rights of 3D multimedia is becoming an urgent need. The 3D watermarking has been considered as an effective technique to solve this problem.

The human visual system (HVS) reconstructs a three-dimensional world from the stereo pair of the retinal images on the 3D display. And the HVS perceives the depth illusion of a 3D image from two different viewpoint images, one from the left-eye

and another from the right-eye. There are two major existing techniques for 3D image representation: stereo image recording (SIR) [17] [18] and depth-image-based rendering (DIBR) [1] [2]. In the SIR system, the left view and right view are recorded simultaneously from two cameras located at the positions of the two eyes. This method offers a high quality viewing for the audience because the two captured devices act like human eyes. However, this method has several drawbacks. It is very difficult and costly to set the two cameras under the same conditions, like the same height and brightness. Moreover, the initial depth condition is fixed so the viewer can not adjust the depth condition they prefer. In addition, the double color images for one scene consume large storage and transmission bandwidth, which is a critical issue for its application [2]. By contrast, the DIBR technique only transmits the center view and the corresponding depth information, and generates the virtual left and right views at the content consumer side. In [19], a new scheme on 3D-TV using DIBR and its actual implementation are described. The studies in "Advanced Three-Dimensional Television System Technologies" (ATTEST) project [20] present the superiority of DIBR 3D image representation over the traditional stereoscopic displays for 3D TV broadcasting. The DIBR technique has attracted more attention due to its low-cost configuration and practicability, and it provides a number of advantages:

1. The depth map consists of gray level pixels and a lot of smooth areas, so the per-pixel depth information without many high frequency components can be compressed efficiently before transmission to save bandwidth.

2. The problems of photometrical asymmetries in terms of brightness, contrast and color between the left and right views are eliminated, because both views are effectively synthesized from the same original center image.

3. The viewers can easily adjust the depth degree by modifying the depth image to suit their preferences in the DIBR system. In other words, the DIBR system offers individualized depth configuration. This is an important feature taken into account the fact that there is a difference in depth appreciation among different viewers [20].

4. The DIBR system is compatible to 2D digital TV so that the center view can also be used as 2D content. And the data representation format of monoscopic image plus associated per-pixel depth information is ideally suited to facilitate 3D post-processing.

Due to the numerous advantages of DIBR technique and its practicality, the digital watermarking for DIBR 3D images has become one of the most popular approaches for preventing the increasing amount of 3D content from being distributed and used illegally. Fig. 1.2 illustrates the illegal re-distribution performed at the content consumer side. A 3D digital watermarking technique should not only consider the situations in which the center view is illegally distributed, but also the circumstances in which either the virtual left view or right view is leaked to the public without permission.

The conversion from 2D to 3D brings new challenges [21]. Different from ordinary 2D digital watermarking [16] [22] [23], there are more specific requirements for 3D watermarking, especially for DIBR 3D image watermarking. Either the center view or the virtual left and right views could be illegally distributed, so the watermark information should be extracted from the three views individually for content authentication. Moreover, besides the center image, the depth map could be distorted by possible attacks during the transmission [24], which has a negative effect on the generation of synthesized left and right images. Therefore, the watermarking tech-

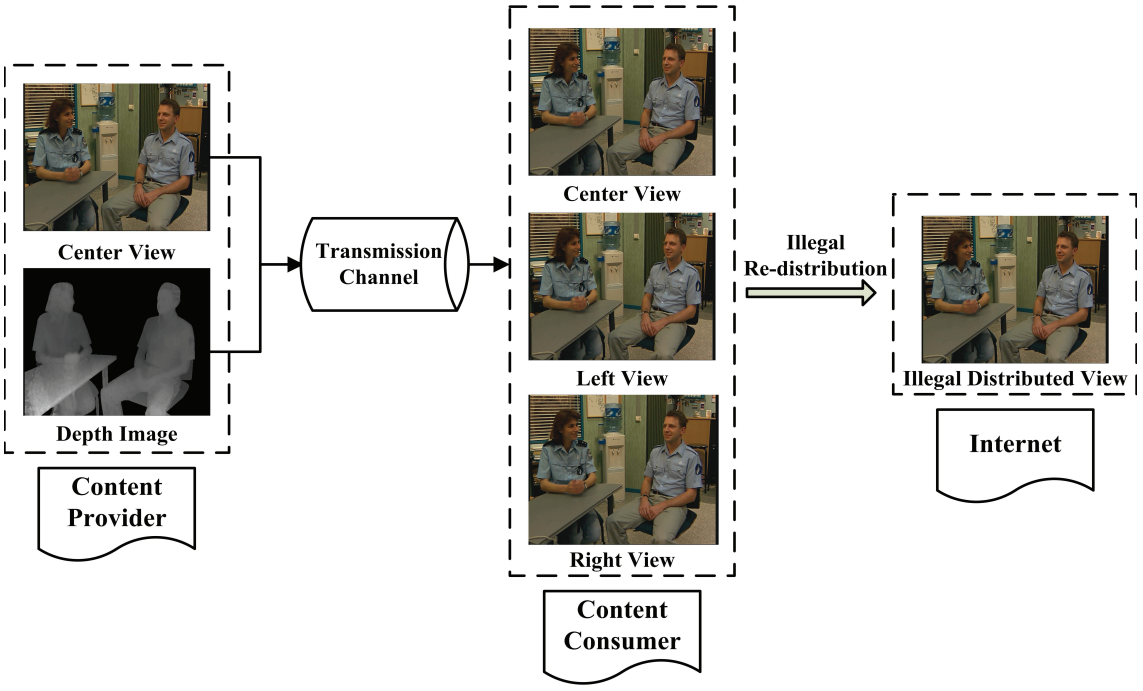


Figure 1.2: Illegal content distribution at the consumer side.

nique for DIBR 3D images should not only be robust to 2D image processing attacks, but also resist some DIBR distortions which will be discussed in the next chapter. However, quantities of 2D watermarking techniques [25] [26] [27] fail to resist 2D to 3D conversion and cannot be directly applied to 3D watermarking. In order to address these problems, the work in this thesis is mainly conducted on the blind digital watermarking for DIBR 3D images.

1.3 Depth-image-based rendering system

In this section, the depth-image-based rendering (DIBR) system is briefly introduced. The DIBR is a process of synthesizing virtual views of a scene from still or moving color images and associated per-pixel depth information [20] [28]. In a DIBR system,

the virtual left view and right view can be rendered by pre-processing of depth map, 3D image warping and hole-filling processes. Note that the hole-filling is not necessary if there are no holes to fill as a result of optimal pre-processing of depth maps [29]. Fig. 1.3 illustrates the major three steps of the DIBR process [20].

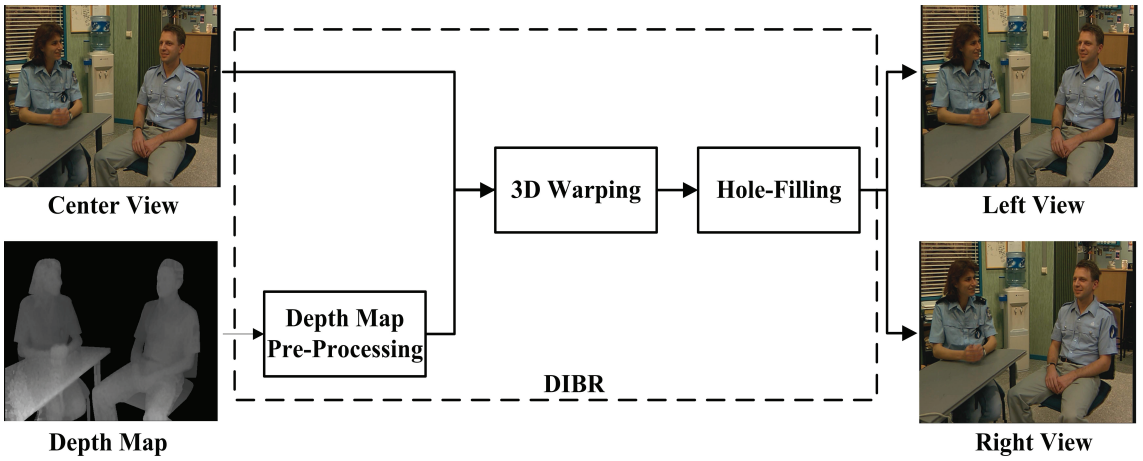


Figure 1.3: Depth-image-based rendering system.

1.3.1 Pre-processing of depth map

As the first step, the pre-processing of depth map by smoothing filter can reduce the hole occurrences in the virtual left and right views, especially in the depth regions with sharp edges and discontinuities. There are two steps in the pre-processing of the depth map: choosing the convergence distance Z_c (zero-parallax setting (ZPS)) and smoothing the depth maps.

Several techniques can be used to establish a ZPS. In the so-called "toed-in" method, the ZPS is selected using a joint inward rotation of the left-eye and right-eye cameras. Different from this one, the ZPS is established by "shifting" the depth

image in the rendering system, where the ZPS plane is expressed by

$$Z_c = \frac{Z_{near} - Z_{far}}{2} \tag{1.3}$$

where Z_{near} is the nearest clipping plane with gray value of 255 and Z_{far} is the farthest clipping plane with gray value of 0 [29]. The zero-parallax setting is chosen to be halfway between Z_{near} and Z_{far}. Fig. 1.4 shows the center image, the corresponding depth image and the mapping between intensity and depth value. In this process, the gray values of an 8-bit depth image lie between Z_{near} and Z_{far}, and all the clipping planes are parallel to the image plane. Then the depth map is further normalized so that the values lie between -0.5 and 0.5.

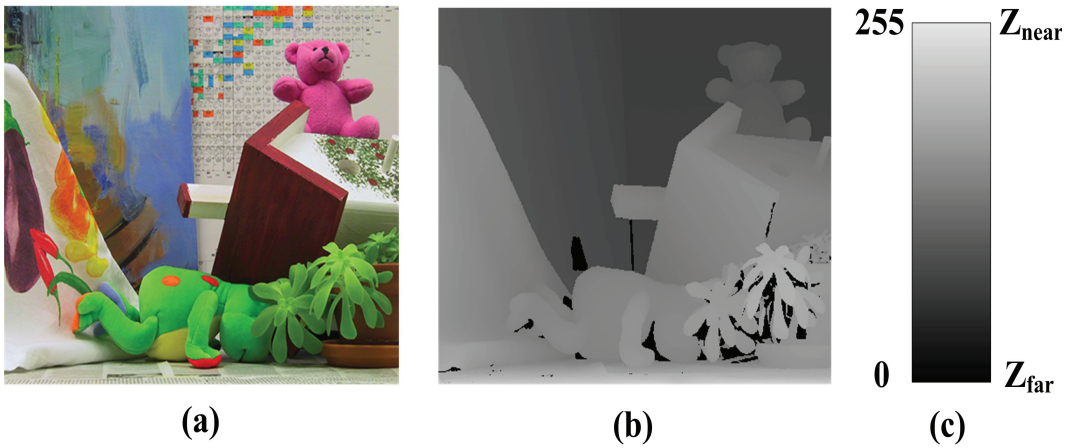


Figure 1.4: (a) The center image; (b) The corresponding depth image; (c) The mapping between intensity and depth value.

The second step in the depth map pre-processing is that different kinds of filters can be adopted to smooth the depth image. For efficient implementation, Gaussian

filter is employed in this thesis because of its adjustable window size [29].

$$g(x,\sigma) = \frac{1}{\sqrt{2\pi\sigma}}\exp\Big(-\frac{x^2}{\sigma^2}\Big),\ -\frac{w}{2} \le x \le \frac{w}{2} \tag{1.4}$$

where w is the filter's window size, and σ is standard deviation, which controls the depth smoothing strength. Different values of w and σ have different effects on the quality of the virtual views generated from the center image.

1.3.2 3D image warping

Three-dimensional image warping is very important in generating the virtual left and right views. In this thesis, we only consider the commonly used parallel camera configuration for the DIBR system. As shown in Fig. 1.5 [29], one point P with the depth Z is mapped on the image plane of three cameras (C_l, C_c, and C_r) at pixels x_l, x_c and x_r, respectively. In the image warping step, the virtual left and right views can be synthesized from the center view and its depth map by the values of baseline distance and focal length. In this case, the vertical coordinate of a pixel in the three views remains the same. From the geometry in Fig. 1.5, the pixel-wise mapping can be performed by the following formulas [29] [30]:

$$x_l = x_c + \frac{t_x}{2}\cdot\frac{f}{Z},\ x_r = x_c - \frac{t_x}{2}\cdot\frac{f}{Z} \tag{1.5}$$

where x_c, x_l and x_r are the viewpoints of the center view, the virtual left and the virtual right views, respectively. f represents the focal length of the camera, t_x denotes the baseline distance and Z is the value of depth map associated with the center view.

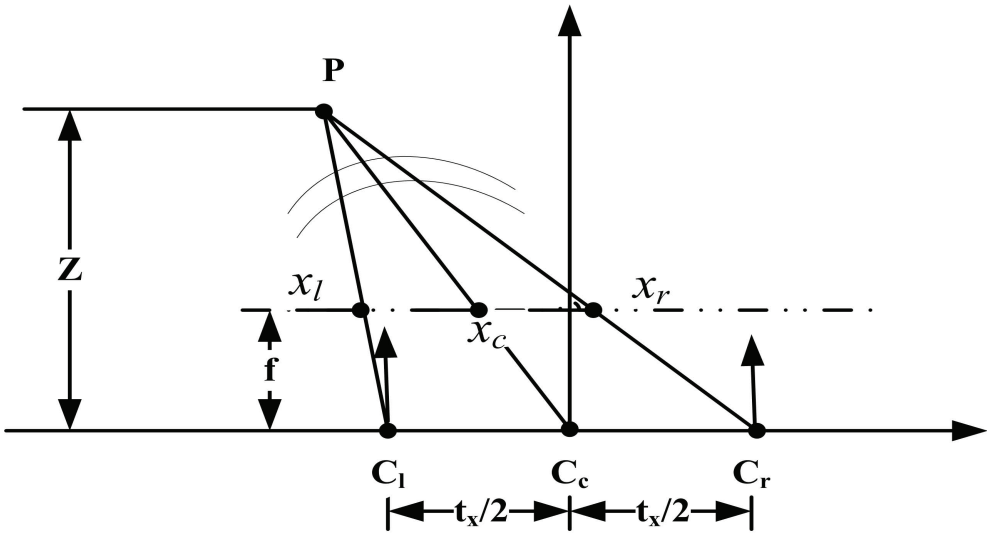


Figure 1.5: Pixel-wise mapping in 3D image warping [29].

According to the depth map pre-processing in Section 1.3.1, the value of baseline distance also indicates the depth range in the synthesized images. From Equ. (1.5), the disparity $x_l - x_r$ is proportional to the baseline distance t_x. And the value of disparity can indicate the distance between the object in the real world and the ZPS. Based on the visibility property of 3D view, the pixels with the farthest depth value are warped first since the closer object can occlude the farther object [1]. The baseline distance t_x can be adjusted by the audience to create a comfortable viewing [31] [32].

1.3.3 Hole-filling

Due to the sharp changes in the depth map and different viewpoints, there are new exposed hole areas revealed in the rendered left and right views after the 3D warping process. Some pixels' information is lost during the pixel-wise mapping and some areas occluded in the original image may become visible in the virtual views. These newly exposed areas, referred to as "disocclusions" in computer graphics, typically

happen at the edge of the foreground objects: when a camera moves rightwards, new areas on the right side of the foreground object comes into view. However, there is no information available in the original view on how these areas look like. The hole occurrences can be reduced by the pre-processing of associated depth map. There are many studies [33] [34] [35] on the hole-filling problem. One straightforward method to fill the holes is the interpolation of neighborhood pixel information. Better qualities of synthesized left and right views can be reconstructed with more complicated extrapolation techniques. Considering the simplicity and practicality, we use linear interpolation to fill the disocclusions. For linear interpolation, for each identified empty pixel in the hole, we search for the three nearest pixels i, j, k and construct a linear plane given their coordinates (x_i, y_i), (x_j, y_j), (x_k, y_k). The empty pixel is interpolated using the constructed plane and its own pixel coordinate. Fig. 1.6 presents an example of the right-eye image with holes and the corresponding image after hole-filling processing.

1.4 Evaluation of 3D image watermarking

For digital watermarking, the following three requirements are commonly used to evaluate the performance of a watermarking scheme. And a good watermarking system should achieve a good trade-off among these requirements, refer to Fig. 1.7.

1. Invisibility. Invisibility means that there should be no perceptible difference between the watermarked signal and the original signal. In other words, to keep the good fidelity of the watermarked signal is required.

2. Robustness. Robustness against various attacks is an important criterion for a

(a)

(b)

Figure 1.6: Hole-filling. (a) The right view with holes of image *Ballet*; (b) The corresponding right view after hole-filling.

digital watermarking technique. For DIBR 3D image watermarking, a good watermarking scheme should be robust to noise addition, image filtering, geometric transformations like rotation, scaling and translation, and lossy compressions

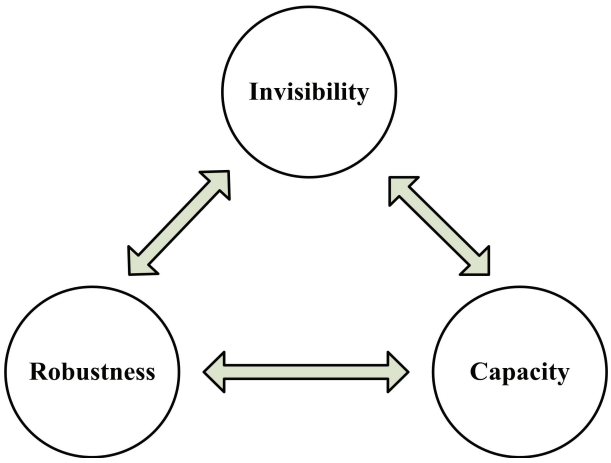


Figure 1.7: The requirements for a good digital watermarking system.

such as JPEG compression and JPEG 2000 compression. Moreover, it should resist the common DIBR processing, such as depth map variation, baseline distance adjustment and different rendering conditions.

3. Capacity. Capacity indicates the maximum amount of data the embedded watermark can carry and those information can be extracted reliably for content authentication.

1.5 Applications of 3D image watermarking

The watermarking technique for DIBR 3D images is developed based on the applications. The main applications are listed as follows:

1. Copyright protection. One of the main applications of 3D digital watermarking is the copyright protection for 3D multimedia products. The copyright owner can prove the ownership by extracting the embedded information. The water-

marks used for this purpose are supposed to be very robust to various attacks intended to remove the watermark [10] [36].

2. Quality evaluation. In this application, the watermark is embedded into the cover signal. The watermarked signal and the watermark suffer the same possible distortions. The quality of distorted signal can be evaluated by examining the degradation of extracted watermark [37] [38] [39].

3. Broadcast monitoring. The watermark pattern can be embedded as the identification information in the broadcast signals and be detected reliably and interpreted correctly. Real-time watermarking allows the live active monitoring for many channels [8].

4. Copy and usage control. It is to prevent the unauthorized copies of copyrighted content. Different payment entitles users to have different privilege (play/copy) on the object. Some systems are expected to have copy and usage control mechanisms, which prevent illegal copy of digital information or limit the number of times of copying.

5. Forensic analysis. Digital rights management (DRM) systems are designed to protect and enforce the copyright control with digital media content. A forensic watermarking is embedded into the master copy of the content, and it allows the owner to identify the illegal or unauthorized use and distribution of copyrighted contents.

1.6 Objectives of the thesis

For the application of copyright protection, 3D digital watermarking is expected to be robust to various attacks. As discussed in Section 1.2, the conversion from 2D digital watermarking to 3D digital watermarking brings new challenges. There are two major categories of attacks for 2D digital watermarking. The first category refers to the geometric attacks including rotation, scaling and translation (RST), cropping, aspect ratio changes. The second category is classified as image/video compressions. JPEG, JPEG 2000, MPEG-2, H.264/AVC and H.265/HEVC are common compressions in digital watermarking. Besides the aforementioned attacks, there are some specific distortions for 3D digital watermarking, especially for DIBR 3D image watermarking. These additional attacks are baseline distance adjustment, depth map variation and different rendering conditions, which are known as common DIBR processing. Therefore, the watermarking technique for DIBR 3D images should not only be robust to the 2D image processing attacks, but also resist the DIBR processing distortions. However, the existing 2D watermarking algorithms can not be directly extended to 3D watermarking since they are not robust to view synthesis process. Either the center view or the synthesized left and right views could be illegal distributed, so the watermark should be extracted from the three views individually for content authentication and copyright protection.

For the application of quality evaluation, the watermark should be embedded into the cover image in a proper way, so that the extracted watermark can accurately reflect the quality changes of the watermarked image under distortions. This is the core of the whole scheme, which determines the accuracy of quality evaluation. And to our best knowledge, there are no existing algorithms that achieve quality evaluation

for DIBR 3D images using the watermarking technique.

Based on these two applications, the thesis has two main objectives:

1. This thesis is aimed to research and propose DIBR 3D image watermarking algorithms robust to various attacks. It is considered the most challenging for a 3D watermarking algorithm to be robust to both 2D image processing attacks and DIBR processing distortions. We will try to propose watermarking algorithms for DIBR 3D images that have good performance in terms of watermark invisibility and robustness to various attacks.

2. This thesis is aimed to research and devise watermarking based quality evaluation schemes for DIBR 3D images. We will try to use the watermarking technique to evaluate the quality of the watermarked images under distortions. The quality of the watermarked center view and the synthesized left view and right view should be estimated using the degradation of corresponding extracted watermarks.

For the envisaged DIBR 3D watermarking algorithms in the thesis, certain directional subbands in the various transforms are preferred for 3D DIBR given the left and right synthesized views created from the center view. As discussed in Section 1.3 and according to Equ. (1.5), the DIBR system maps the pixels of the center image to the virtual left and right images horizontally. Since each pixel moves horizontally, the vertical edges and contours are more easily distorted than the horizontal ones. So we should embed watermark data into the subbands with more information on the horizontal direction. Otherwise, the watermark would probably be destroyed during the process of generating the virtual left and right views, which would lead to

unsuccessful copyright protection and quality evaluation for these DIBR synthesized views.

1.7 Contributions of the research

The contributions of the thesis are on DIBR 3D watermarking. To be specific, they are on watermarking algorithms and watermarking based quality evaluation algorithms for DIBR 3D images. Regarding the contribution on watermarking algorithms for DIBR 3D images, in the thesis, we have compared our proposed watermarking algorithms with the existing state-of-the-art techniques for DIBR 3D image watermarking, and the experimental results show the superiority of our proposed algorithms in terms of robustness to various attacks. Regarding the contribution on watermarking based quality evaluation schemes, our work is the first attempt to evaluate the synthesized left and right views using the degradation of corresponding extracted watermarks. To our best knowledge, there are no algorithms that can evaluate the quality of DIBR 3D images using the watermarking technique. The detailed contributions include the following works:

1. To have a clear understanding of existing watermarking algorithms and develop new watermarking schemes for DIBR 3D images, a detailed literature review on related works has been given.

2. A new blind 2D image watermarking method based on contourlet transform (CT) and principal component analysis (PCA) is first presented. Then it is extended to blind video watermarking. The introduction of the CT produces the

robustness against image and video compressions, and the PCA yields resistance to geometric attacks.

3. Considering the directional multiresolution image representation and convenient tree structures of contourlet transform, a robust contourlet-based watermarking algorithm for DIBR 3D images is developed. The watermark can be detected with low bit error rate (BER) from the center view, the left and right views even when each view is distorted and distributed separately. The proposed algorithm is robust to various image processing attacks and common DIBR processing distortions. However, the performance on watermark extraction still needs to be further improved.

4. As an improvement, a novel blind watermarking algorithm for DIBR 3D images based on feature regions and ridgelet transform (RT) is proposed. The watermarked view has good perceptual quality under both the objective and subjective image quality measures. The proposed algorithm is evaluated against various attacks, and demonstrated to have superiority over other related and state-of-the-art methods in terms of watermark extraction.

5. A watermarking based quality evaluation scheme for DIBR 3D images is devised. The qualities of the watermarked center view and synthesized views can be estimated by examining the degradation of corresponding extracted watermarks. The proposed scheme can evaluate the quality of the watermarked image under various distortions with high accuracy.

Journal articles generated from the research:

1. Lei Chen and Jiying Zhao, "No-reference perceptual quality assessment of stereoscopic images based on binocular visual characteristics", submitted to Signal Processing: Image Communication, under review.

2. Lei Chen and Jiying Zhao, "Quality evaluation of DIBR 3D images based on blind watermarking", submitted to Multimedia Systems, under second review.

3. Lei Chen and Jiying Zhao, "A robust blind watermarking algorithm for depth-image-based rendering 3D images", submitted to Signal Processing: Image Communication, under second review.

4. Lei Chen and Jiying Zhao, "Perceptual quality assessment of stereoscopic images based on local and global visual characteristics", Multimedia Tools and Applications, accepted.

5. Lei Chen and Jiying Zhao, "Contourlet-based image and video watermarking robust to geometric attacks and compressions", Multimedia Tools and Applications, vol. 77, no. 6, pp. 7187-7204, 2018.

6. Lei Chen and Jiying Zhao, "Robust contourlet-based blind watermarking for depth-image-based rendering 3D images", Signal Processing: Image Communication, vol. 54, pp. 56-65, 2017.

Conference papers generated from the research:

1. Lei Chen and Jiying Zhao, "No-reference quality assessment for stereoscopic 3D images based on binocular visual perception," in Proceedings of IEEE International Symposium on Haptic, Audio and Visual Environments and Games, pp. 1-5, 2018.

2. Lei Chen and Jiying Zhao, "Quality assessment of stereoscopic 3D images based on local and global visual characteristics," in Proceedings of IEEE International Conference on Multimedia and Expo Workshops, pp. 61-66, 2017.

3. Lei Chen and Jiying Zhao, "Watermarking based quality assessment for DIBR 3D images," in Proceedings of IEEE International Symposium on Smart Data, pp. 810-814, 2016.

4. Lei Chen and Jiying Zhao, "Robust contourlet-based watermarking for depth-image-based rendering 3D images," in Proceedings of IEEE International Symposium on Broadband Multimedia Systems and Broadcasting, pp. 1-4, 2016.

5. Lei Chen and Jiying Zhao, "Adaptive contourlet-based image watermarking robust to geometric transformations and image compression", in Proceedings of IEEE International Instrumentation and Measurement Technology Conference, pp. 996-1001, 2016.

6. Lei Chen and Jiying Zhao, "Adaptive digital watermarking using RDWT and SVD," in Proceedings of IEEE International Symposium on Haptic, Audio and Visual Environments and Games, pp. 1-5, 2015.

1.8 Thesis structure

The thesis is organized as follows. The fundamental theories and techniques used in this thesis are introduced in Chapter 2. Chapter 3 reviews the existing algorithms on DIBR 3D image watermarking and other related digital watermarking approaches. In

Chapter 4, a blind image watermarking method robust to geometric attacks and compressions is presented, and it is extended to blind video watermarking. In Chapter 5, a robust contourlet-based watermarking algorithm for DIBR 3D images is developed. Before introducing the watermark embedding and extraction steps, we give analysis on why we make similar subbands in a pair, and why we can use coefficient quantization for watermark embedding. We analyze that the watermark should be embedded into the first four subbands instead of other subbands because of DIBR view synthesis, which has been confirmed by the experimental results. The robustness of the proposed algorithm against image processing and common DIBR processing attacks are shown in the experiments. In Chapter 6, a novel digital watermarking algorithm for DIBR 3D images based on feature regions and ridgelet transform is proposed. Before proposing our algorithm, we provide explanations on why we select ridgelet transform for watermark embedding and watermark extraction. The performance of the proposed algorithm is evaluated against various attacks and the simulation results are provided in detail. The analysis and discussions of the performance are addressed as well. In Chapter 7, a watermarking based quality evaluation scheme for DIBR 3D images is devised. Before introducing our proposed scheme, we give discussions on why the DT-CWT is suitable for watermarking based quality evaluation. And we present analysis on the subbands selection. The performance on quality evaluation is given in the experimental results. Finally, the thesis is concluded and some suggestions or ideas for future research work are given in Chapter 8.

Chapter 5, Chapter 6 and Chapter 7 are the main components of the thesis. The schemes in Chapters 5 and 6 are proposed for DIBR 3D image watermarking, and the scheme in Chapter 7 is proposed for watermarking based DIBR 3D image quality

evaluation. All these three chapters are on DIBR 3D watermarking but for different applications which are copyright protection and quality evaluation. The proposed schemes in Chapters 5 and 6 have their own advantages and can be used for different scenarios.

Chapter 2

Fundamental theories and techniques

This chapter introduces the fundamental theories and techniques used in the digital watermarking and watermarking algorithms for depth-image-based rendering (DIBR) 3D images.

2.1 Geometric transformations and compressions

The geometric transformations and compressions are common signal processing attacks for digital watermarking [22]. The geometric transformations, such as rotation, scaling and translation (RST) [40] [41], are considered more difficult to tackle than other types of attacks in image watermarking.

2.1.1 Geometric transformations

Rotation

A two dimensional rotation is operated on an image or video frame by repositioning it along a circular path in the xy plane. The rotation of a point at (x, y) by an angle θ about the origin clockwise can be expressed by the following transformation equations:

$$\begin{cases} x' = xcos\theta + ysin\theta \\ y' = -xsin\theta + ycos\theta \end{cases} \tag{2.1}$$

Scaling

The size of an image can be modified by a scaling transformation. The scaling from the coordinate values (x, y) to the transformed coordinates (x', y') can be obtained by multiplying scaling factors a and b, which is expressed in the following transformation equations:

$$\begin{cases} x' = x \cdot a \\ y' = y \cdot b \end{cases} \tag{2.2}$$

Scaling factors a and b scale images in the x direction and y direction respectively. a and b are usually assigned the same value, where a uniform scaling is produced to maintain relative image proportions.

Translation

A translation (or shift) is operated on an image by repositioning it along a straight line path from one coordinate location to another. The translation of a two dimensional point from the original coordinate position (x, y) to a new position (x', y') can be obtained by adding translation distances, x_0 and y_0:

$$\begin{cases} x' = x + x_0 \\ y' = y + y_0 \end{cases} \tag{2.3}$$

where the translation distance pair (x_0, y_0) is called a translation vector or shift vector.

2.1.2 Compressions

In this section, we briefly discuss the common image and video compressions [42] [43].

JPEG compression and JPEG 2000 compression

JPEG is a frequently used technique of lossy compression for digital images, especially for the images obtained from digital photography. The degree of compression can be adjusted, which allows a selectable trade-off between image quality and storage size [44]. JPEG uses a lossy form of compression based on the discrete cosine transform (DCT).

JPEG 2000 is an image compression standard and coding system, which was created by the Joint Photographic Experts Group committee in 2000. It is possible to store different parts of the same image with different qualities [45]. Due to the

use of discrete wavelet transform (DWT) and a more complicated entropy encoding scheme, JPEG 2000 has a significant superiority over certain modes of JPEG: there is no blocking and the artifacts are less visible.

MPEG-2, H.264/AVC and H.265/HEVC

Video compression removes the spatial and temporal redundancy information from large video data while keeping perceptual quality. It is a trade-off among video quality, disk space, and the cost of hardware [46].

MPEG-2 is widely used for compressing digital TV signals, in either SDTV or HDTV broadcasting and communications. It supports both interlaced and progressive videos and allows three options for chrominance format, 4:2:0, 4:2:2 and 4:4:4. MPEG-2 specifies that the raw frames can be compressed into three types of frames: intra-coded frames (I-frame), predictively coded frames (P-frames), and bidirectional predictively coded frames (B-frames) [47].

H.264 or advanced video coding (AVC) is a block-oriented motion-compensation-based video compression standard. It contains a number of new features, such as multi-picture inter-picture prediction, spatial prediction, lossless macroblock coding, etc. These features enable it to compress video much more efficiently than older standards and provide more flexibility for applications [48].

High efficiency video coding (HEVC), also known as H.265, is the latest video compression standard that can double the data compression ratio at the same level of video quality. The primary changes for HEVC include the expansion of the pattern comparison and difference-coding areas, improved motion compensation filtering, improved motion vector prediction and motion region merging [49].

2.2 Common DIBR processing

Besides the ordinary 2D image processing distortions discussed above, there are some common DIBR processing for the DIBR 3D image system. The virtual left and right views are rendered based on possibly attacked depth images or different baseline distances. In addition, there are three major steps in a DIBR system and the detailed operations can be different for practical applications.

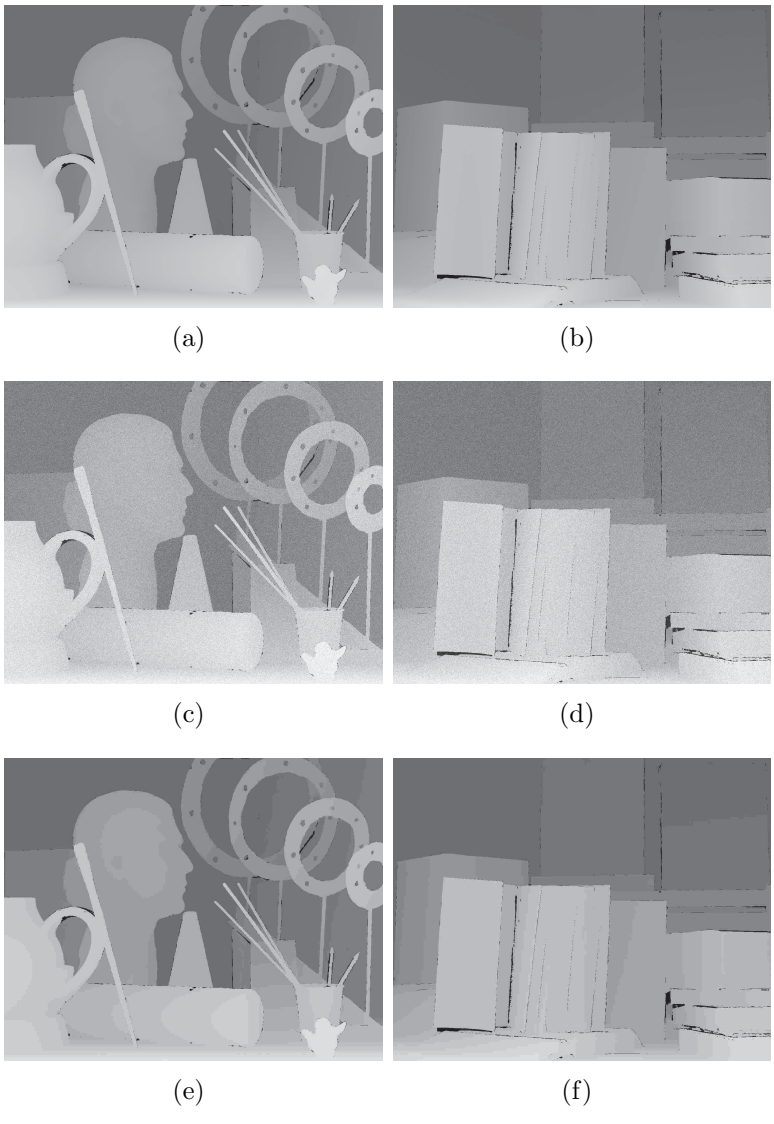
2.2.1 Baseline distance adjustment

When the center image and the depth map are transmitted to the receiver side, the user may change the level of depth perception by adjusting the baseline distance (i.e., t_x in Equ. (1.5)) for rendering. One of the advantages of the DIBR system is the flexibility of depth adjustment. Because personal preference and viewing distance vary, the viewers could adjust the baseline distance within a proper range. The corresponding left-eye and right-eye images will be obtained by rendering with different baseline distances in the DIBR system. However, as discussed in [20] [29], large baseline distances induce discomfort 3D viewing, while small baseline distances make the 3D viewing converging to the traditional 2D viewing.

2.2.2 Depth map variation

As discussed in Chapter 1, the pre-processing of depth image can reduce the hole occurrence to render the left and right views with high quality. However, similar to the center view, the depth map may be compressed or suffer from the channel noise during transmission. Consequently, the receiver side renders the left and right views

based on the modified depth map. And the distortion in a depth map can introduce a position error in the synthesized views [50]. Fig. 2.1 presents the original depth maps of images *Art* and *Books*, and the corresponding distorted images with Gaussian noise and JPEG compression.



(a) (b) (c) (d) (e) (f)

Figure 2.1: Depth map variation. (a) and (b) The original depth images of *Art* and *Books*; (c) and (d) The depth images with Gaussian noise (Variance: 0.1); (e) and (f) The depth images with JPEG compression (Quality factor: 10).

2.2.3 Different rendering conditions

As presented in Chapter 1, there are three major steps in a DIBR system. The detailed operations can be different for different applications. Here we introduce two typical and different rendering methods used in the DIBR system. In [29], the depth maps are pre-processed with an asymmetric filter to smoothen the sharp changes around object boundaries. The smoothing alleviates the disocclusion areas where potential artifacts can occur. The asymmetric nature of the filter reduces the amount of geometric distortion. In [51], the edge dependent depth filter and interpolation are used for the DIBR system. The edge dependent depth filter can decrease the occurrence of big holes, and the edge dependent interpolation is to preserve the edge information of the interpolated area. In addition, the vertical edge rectification provides the depth accuracy along the vertical line.

2.3 Contourlet transform

Natural images contain intrinsic geometrical structures that are key features in visual information. The ordinary wavelets are good at isolating the discontinuities at edge points, but can not capture the smoothness along the contours. Moreover, separable wavelets have limited directional information, which is an important and unique feature for the multidimensional signals. These behaviors indicate that more powerful representations are needed.

Contourlet transform (CT), proposed by Do and Vetterli [52], provides a flexible multiscale and directional decomposition of images. It uses a double filter bank structure to obtain images with smooth contours. There are two major stages in this

double filter bank: Laplacian pyramid (LP) and directional filter bank (DFB). The LP is first applied to capture the point discontinuities, and the DFB is followed to form linear structures from those point discontinuities. The final result is an image expansion using basic elements like contour segments. The contourlets have elongated supports at various scales and directions, which allows contourlets to efficiently approximate a smooth contour at multiple resolutions. In addition, the CT provides a multiscale and directional decomposition in the frequency domain [52].

2.3.1 Laplacian pyramid

The Laplacian pyramid (LP) [53] at each level decomposes a downsampled lowpass version of the original image and the bandpass images. The bandpass image can be considered as the difference between the original image and the lowpass image. The LP decomposition process is described in Fig. 2.2(a), where H and G are the analysis and synthesis filters, and M is the sampling matrix. Note that in multidimensional filter banks, sampling is represented by sampling matrices; for example, downsampling $x[n]$ by M yields $x_d[n] = x[Mn]$, where M is an integer matrix. As can be seen from this figure, given the original signal x, the outputs are a coarse approximation a and a difference b between the original signal and the prediction. This process can be iterated on the coarse signal. The analysis and synthesis filters are time reversal, the use of the optimal linear reconstruction using the dual frame operator is shown in Fig. 2.2(b). The new reconstruction is different from the ordinary approach, where the signal is reconstructed by simply adding back the difference to the prediction from the course signal [52]. Furthermore, it is shown to achieve better performance than the ordinary reconstruction in the presence of noise.

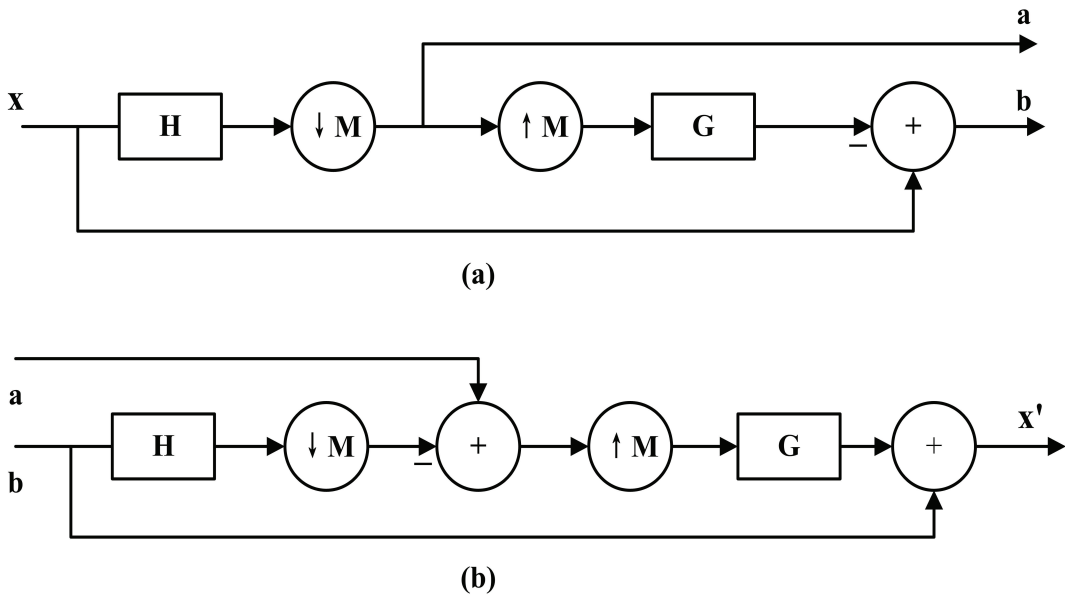


Figure 2.2: The analysis and synthesis of Laplacian pyramid (LP) [52]. (a) One level of decomposition; (b) The new reconstruction scheme for the LP.

Different from the critically sampled wavelet technique, the LP has the property that each pyramid level produces only one bandpass image, and there is no "scrambled" frequencies in this image [52]. In the wavelet filter bank, the frequency scrambling occurs when a highpass channel is folded back into the low frequency band after downsampling, and thus its spectrum is reflected. However, by downsampling the lowpass channel only, this effect can be avoided in the pyramid frames.

2.3.2 Directional filter bank

The directional filter bank (DFB) can be performed with high efficiency via an l-level binary tree decomposition, and there are 2^l subbands with wedge-shaped frequency partitioning as illustrated in Fig. 2.3 [52]. The construction of the DFB in [54] consists of two building blocks. The first building block of the DFB is a two-channel

quincunx filter bank fan filters, which divide a 2D spectrum into horizontal and vertical directions. The second building block is a shearing operator, which amounts to reordering of image samples [52]. The critical issue in the DFB is to combine shearing operators together with quincunx filter banks appropriately to obtain the desired spectrum division shown in Fig. 2.3.

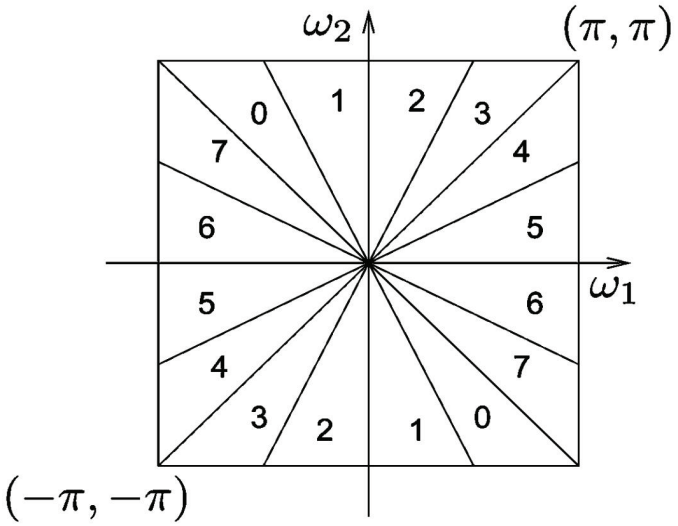


Figure 2.3: An example of frequency partitioning in DFB, where $l = 3$ and there are $2^3 = 8$ frequency subbands [52].

2.3.3 Discrete contourlet transform

Fig. 2.4 describes the image decomposition in the contourlet filter bank. The LP first decomposes the image into the bandpass images and a lowpass image. The lowpass image coveys the low frequencies of the original image. The bandpass image can be considered as the difference between the original image and the lowpass image. Then every bandpass image is fed into the DFB to obtain the directional subbands, which can be adjusted by the users. The scheme can be iterated on the lowpass image to

obtain the images decomposed at different pyramidal levels and multiple directions.

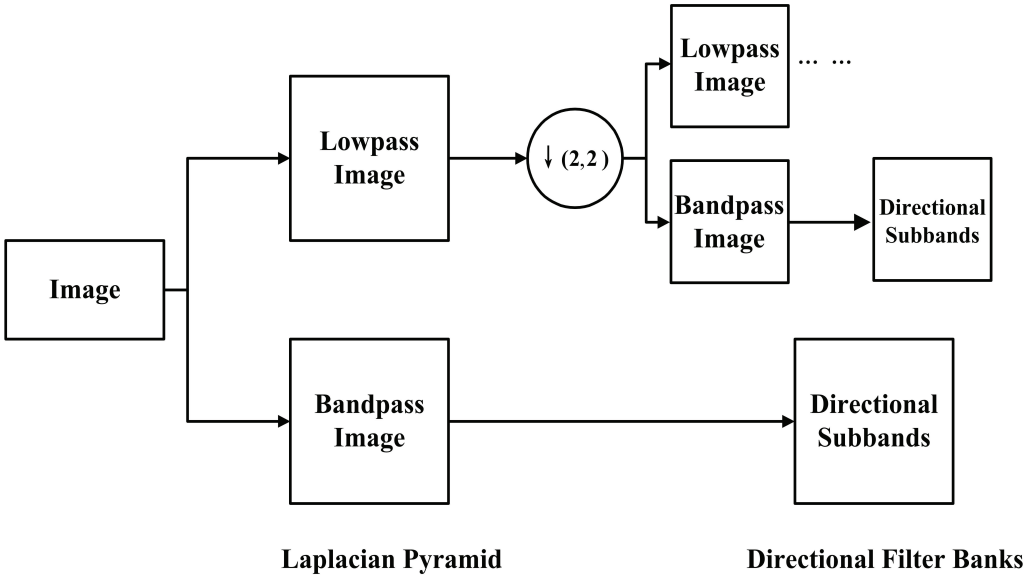


Figure 2.4: The image decomposition in contourlet filter bank (LP and DFB).

Fig. 2.5 gives the final result of the *Barbara* image in contourlet domain, which is the output of Fig. 2.4. The numbers in the subfigure on bottom left denote the corresponding directional subbands accordingly (from left to right, from top to bottom). The original image is first decomposed into two pyramidal levels. The image (A) is the final lowpass image. For the directional subband decomposition in Fig. 2.5, the bandpass images can be decomposed into four directional subbands (B1-B4) and eight directional subbands (C1-C8), respectively. The different images represent different directional subbands at different scales through Laplacian pyramid and directional filter banks. The sizes and shapes of the directional subbands are determined by the iterated contourlet filter banks. The brighter areas indicate larger coefficients while the darker areas represent smaller coefficients [55].

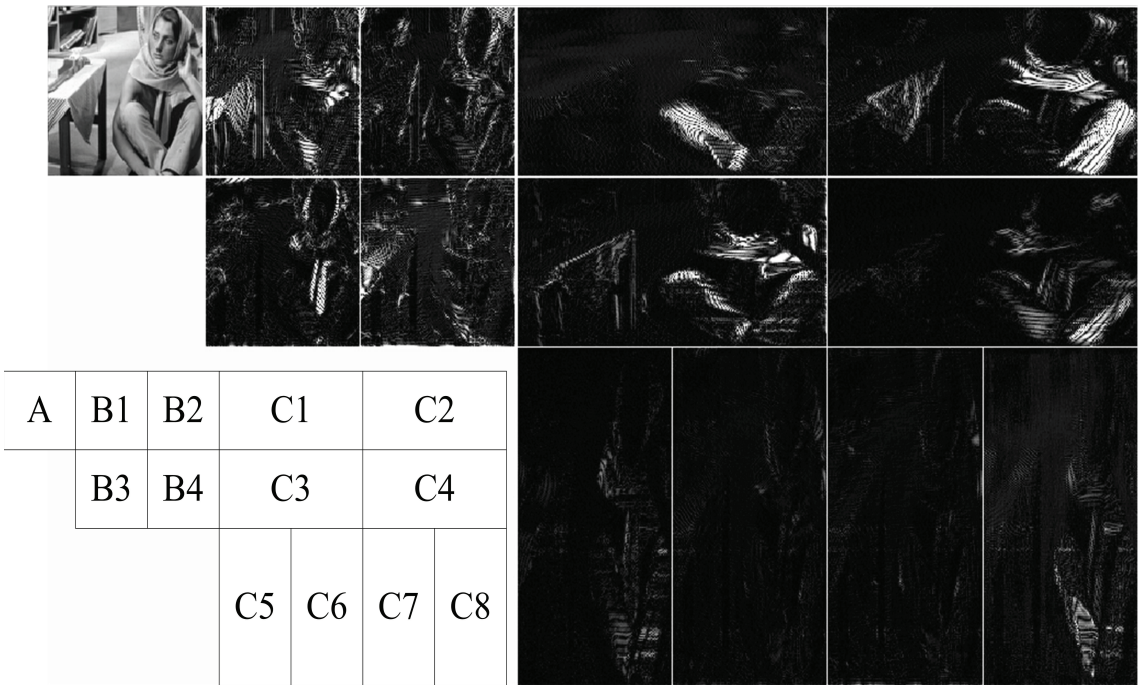


Figure 2.5: Contourlet transform of the *Barbara* image. The numbers in the subfigure on bottom left denote the corresponding directional subbands accordingly.

2.4 Principal component analysis

Principal component analysis (PCA) is defined as an orthogonal linear transformation, which can convert a set of possibly correlated variables into several linearly uncorrelated variables. The number of principal components may be smaller than the number of original variables. In this transformation, the first principal component has the largest possible variance, and each succeeding component in turn has the highest variance under the constraint that it is orthogonal to the preceding components. The resulting vectors are an uncorrelated orthogonal basis set. The PCA is closely related to the Karhunen-Loeve Transform (KLT) and can be seen as an approximation of the KLT. It has good performance in terms of data de-correlation and energy concentration. This technique has been widely applied to many aspects,

such as image processing, pattern recognition and data analysis.

For image processing, the PCA is considered as a linear transform technique to convey most information about an image to principal components [56]. And different from the KLT, most energy of the image is concentrated into the first several principal components and the remaining components are not significant. It is feasible to reconstruct an accurate image by using only the first few principal components. Moreover, the principal components of an image have good stability and do not change significantly when the image is under attack [56] [57].

Let an image be an $M \times N$ matrix, and partition the image into several sub-images denoted by $I = (x_1, x_2, ..., x_m)^T$. Calculate the covariance C_I of the sub-images, the corresponding eigenvalues and eigenvectors.

$$C_I = E[(x_i - m_i)(x_i - m_i)^T] \tag{2.4}$$

where $m_i = E(x_i)$ is the mean vector of each sub-image x_i $(i = 1, 2, ..., m)$, and E denotes calculating averages of the observed quantities over the image block.

By using the eigenvalues and the corresponding eigenvectors of the covariance matrix, each sub-image can be decomposed into uncorrelated coefficients:

$$C_I v_j = \lambda_j v_j \tag{2.5}$$

where λ_j $(j = 1, 2, ..., n)$ is the eigenvalues of the covariance matrix C_I and v_j is the corresponding eigenvector. The orthogonal matrix V formed by the eigenvectors $V = (v_1, v_2, ..., v_n)$ is the basic function of PCA. In order to obtain the principal components of the sub-images, the original image I can be de-correlated by the matrix

V, and the eigen-image P can be calculated using the following equation.

$$P = V^T I = (p_1, p_2, ..., p_n)^T \tag{2.6}$$

The corresponding values p_j $(j = 1, 2, ..., n)$ are the principal components of each sub-image coefficients. Since the decomposition is operated according to the maximum of the variance, the first components contain the significant information of the image and the last components correspond mainly to noise. Also, the original image can be retrieved by

$$I = (V^T)^{-1} P = VP \tag{2.7}$$

2.5 Scale invariant feature transform

Scale invariant feature transform (SIFT) was proposed by D. G. Lowe [58] to extract distinctive invariant features from images. It can be used to perform reliable matching between different views of an object or a scene. In SIFT, the extracted invariant features are represented using some detected interest points, which are well localized in both spatial and frequency domains. These interest points with desirable property of invariance to rotation, scaling and noise addition are selected and defined as keypoints. There are four key stages in the SIFT used to generate the set of image features: scale-space extrema detection, keypoint localization, orientation assignment and keypoint descriptor.

2.5.1 Scale space extrema detection

As the first stage, the detection of scale space extrema is implemented efficiently by using a difference of Gaussian (DoG) function to identify the potential interest points that are invariant to scale and orientation. Detecting locations can be accomplished by searching for stable features across all possible scales. Suppose $I(x, y)$ is the original center image with size of $M \times N$, $L(x, y)$ is the blurred image filtered by the Gaussian filter G_σ defined as:

$$\begin{aligned} G_\sigma &= \frac{1}{\sqrt{2\pi\sigma^2}}\exp[-(x^2+y^2)/(2\sigma^2)] \\ L_1(x,y) &= G_\sigma * I(x,y) \\ L_2(x,y) &= G_{g\sigma} * I(x,y) \end{aligned} \tag{2.8}$$

where g is a constant factor and σ denotes the scale. Here, $L_1(x, y)$ and $L_2(x, y)$ are different in scale by the factor g. The function G_σ is chosen due to its practicality [58]. Then the DoG filtered image can be calculated by

$$\text{DoG} = L_2(x,y) - L_1(x,y) = G_{g\sigma} * I(x,y) - G_\sigma * I(x,y) \tag{2.9}$$

In order to detect the local maxima and minima of DoG, each sample point is compared to its eight neighbors in the image. The sample point is selected as a keypoint candidate only if it is larger than all of these neighbors or smaller than all of them. However, extrema that are close together are quite unstable to small perturbations of the image. The suitable keypoint candidate can be chosen by studying a range of sampling frequencies and using those that provide the most reliable results under a realistic simulation [58].

2.5.2 Keypoint localization

Once a keypoint candidate has been found, the next step is to perform a detailed fit to the nearby data for determining location and scale. This is to select the keypoints with good stability. The candidate points that have low contrast or are poorly localized along an edge are discarded.

For accurate keypoint localization, the approach in [59] calculates the interpolated location of the maximum. The interpolation is implemented using the quadratic Taylor expansion of the DoG function which can be expressed by

$$D(x) = D + \frac{\partial D^T}{\partial x}x + \frac{1}{2}x^T\frac{\partial D^2}{\partial x^2}x \tag{2.10}$$

where D and its derivatives are evaluated at the candidate keypoint. To remove the keypoints with low contrast, the value of the second-order Taylor expansion is calculated and if it is less than a threshold, the candidate keypoint is discarded. In addition, the keypoints with high edge response are eliminated by finding the principal curvatures for the eigenvalues of the second-order Hessian matrix [58]. The principal curvatures can be calculated from a 2×2 Hessian matrix H:

$$H = \begin{bmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{yy} \end{bmatrix} \tag{2.11}$$

where the derivatives can be obtained by calculating the difference of neighboring sample points. Then the ratio of principal curvature is checked as follows:

$$\frac{(D_{xx} + D_{yy})^2}{D_{xx}D_{yy} - (D_{xy})^2} < \frac{(r+1)^2}{r} \tag{2.12}$$

The experiments in [58] use a value of $r = 10$, which rejects keypoints that have a ratio between the principal curvatures greater than 10.

2.5.3 Orientation assignment

A consistent orientation is assigned to each keypoint based on local image properties. The keypoint descriptor can be represented relative to the assigned orientation, which provides invariance to concerned transformations. In the process of orientation assignment, the gradients of the sample points within a region around the keypoint are computed. Then an orientation histogram is created from the gradient. The highest peak in the histogram is detected and assigned as the orientation of the keypoint. In this way, the orientation would not be affected by noise addition and local distortion [58].

2.5.4 Keypoint descriptor

The previous steps have assigned an image location, scale, and orientation to each keypoint. Then a descriptor is computed for the local region that is invariant to the remaining variations like change in illumination. First a set of orientation histograms is calculated from magnitude and orientation values of samples in a 16×16 region around the keypoint. The magnitudes are further weighted by a Gaussian function and the descriptor becomes a vector of all the values of these histograms. This vector is then normalized to unit length in order to enhance invariance to affine changes in illumination.

Fig. 2.6 presents an example of the final result of selected scale-invariant keypoints

using SIFT. These keypoints (marked as "+") are highly distinctive and invariant to image rotation, scaling and noise addition.

(a)

(b)

Figure 2.6: Keypoints selection using SIFT. (a) The original image of *Breakdancers*; (b) The final selected keypoints.

2.6 Ridgelet transform

The ridgelet transform [60] was introduced as a sparse representation of functions on continuous spaces. It was proposed to overcome the problem of the wavelet transform in representing objects with singularities along line [61]. Compared with traditional transforms like discrete wavelet transform (DWT), the ridgelet transform is more effective in sparse image representation, and the most significant information of an image can be represented by the most energetic coefficients. It is a good transform domain for watermark embedding and extraction [62] [63]. Firstly, most of the energy of the image is concentrated in just a few coefficients, which is good for digital watermarking. Moreover, the ridgelet coefficients of an image have good stability and do not change significantly when the image is under possible attacks. In addition, the image can be reconstructed with high accuracy using inverse ridgelet transform.

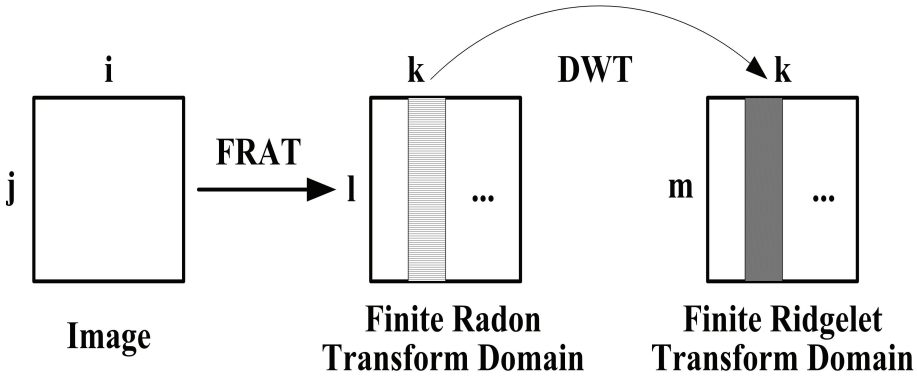


Figure 2.7: Diagram for the finite ridgelet transform [61].

There are different solutions of ridgelet transform for practical applications. In this thesis, we apply the finite ridgelet transform (FRIT) [61] to the watermark embedding regions for its practicality and efficient implementation. As illustrated in Fig. 2.7, there are two major stages in the FRIT. The line singularities of an image is mapped

to point singularities using the finite Radon transform (FRAT). Then the wavelet transform is used on each FRAT projection. The FRAT of a real function f on a finite $p \times p$ grid Z_p^2 ($Z_p = 0, 1, \cdots, p-1$, where p is a prime number) is defined as follows:

$$r_k[l] = \mathrm{FRAT}_f(k,l) = \frac{1}{\sqrt{p}} \sum_{(i,j)\in L_{k,l}} f[i,j] \tag{2.13}$$

where $L_{k,l}$ are a set of points used to form a line on the lattice, k denotes the slope of line and l is the intercept. These lines are obtained by

$$\begin{aligned} &L_{k,l} = \{(i,j) : j = ki + l \pmod p, i \in Z_p\}, 0 \le k < p, \\ &L_{p,l} = \{(l,j) : j \in Z_p\}. \end{aligned} \tag{2.14}$$

Here, $k = p$ corresponds to the case of vertical lines. There exist $p-1$ independent FRAT coefficients in each direction [61]. Then the 1-D wavelet transform is operated on the FRAT projection sequence of each direction to obtain the FRIT coefficients.

The inverse ridgelet transform is performed by applying inverse wavelet transform to the FRIT coefficient in each direction k to obtain the FRAT coefficients, then the image can be reconstructed with finite back-projection (FBP) operator [61]. FBP is obtained by summation of Radon coefficients of all lines with a point $P_{i,j}$ on them, that is

$$\mathrm{FBP}_r(i,j) = \frac{1}{\sqrt{p}} \sum_{(k,l)\in P_{i,j}} r_k[l],\ (i,j) \in Z_p^2 \tag{2.15}$$

where

$$P_{i,j} = \{(k,l) : l = j - ki \pmod{p}, k \in Z_p\} \cup \{(p,i)\}. \tag{2.16}$$

According to the property of finite geometry Z_p^2, summation of Radon coefficients on all lines which pass through the point (i,j) is equal to the summation of all points in the Z_p^2, except for the points on all $p+1$ lines. This can be shown by substituting Equ. (2.13) into Equ. (2.15) as:

$$\begin{aligned} \mathrm{FBP}_r(i,j) &= \frac{1}{p} \sum_{(k,l) \in P_{i,j}} \sum_{(i',j') \in L_{k,l}} f[i',j'] \\ &= \frac{1}{p} \Big(\sum_{(i',j') \in Z_p^2} f[i',j'] + p \cdot f[i,j] \Big) \\ &= f[i,j] \end{aligned} \tag{2.17}$$

So the back-projection operator indeed computes the inverse FRAT for the images. By using the FRIT, a non-redundant representation of image will be obtained.

2.7 Dual-tree complex wavelet transform

The dual-tree complex wavelet transform (DT-CWT) was first introduced by Kingsbury [64]. This transform has the desirable properties of perfect reconstruction, approximate shift invariance, good directional selectivity, limited redundancy and efficient computation [65]. Unlike the traditional discrete wavelet transform (DWT), two independent DWTs are exploited for the DT-CWT. As shown in Fig. 2.8, a 1-D signal $x[n]$ is decomposed and constructed by two DWTs, each of which is called a

tree and uses different filter banks for perfect reconstruction. The $g0$ and $h0$ are the lowpass filter and highpass filter for Tree A, and $g1$ and $h1$ are the lowpass filter and highpass filter for Tree B. The coefficients produced by these two trees can be combined to form one set of complex coefficients:

$$y_a + jy_b \tag{2.18}$$

where y_a and y_b are real coefficients from these two trees respectively. It can also be denoted in the polar form as $me^{j\theta}$, where $m = \sqrt{y_a^2 + y_b^2}$ and $\theta = arctan(y_b/y_a)$. The dual-tree approach provides wavelet coefficients that are shift invariant. In other words, small shifts in the input signal will not cause significant variations in the DT-CWT coefficients at different scales. Moreover, the filters used in tree B are devised to yield outputs at the sample locations which are discarded in tree A.

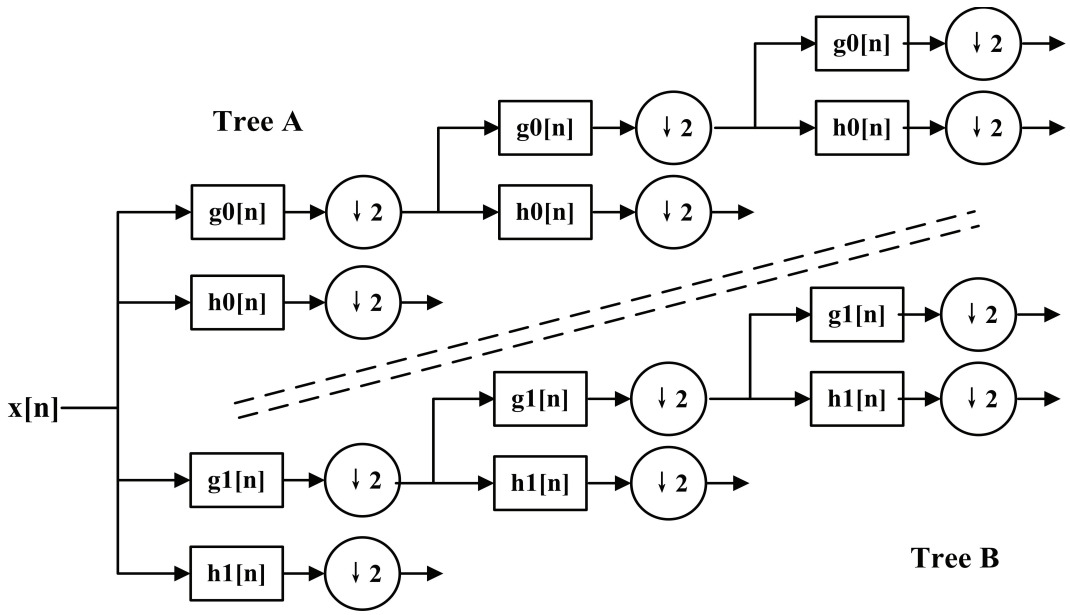


Figure 2.8: DT-CWT for 1-D signal [64].

For 2-D signal, DT-CWT requires a 4:1 increase in the number of coefficients and provides approximate shift invariance in the horizontal and vertical directions. The 2-D DT-CWT produces six subbands that correspond to the outputs of six directional filters oriented at angles of $\pm 15°$, $\pm 45°$ and $\pm 75°$. Fig. 2.9 illustrates the 2-D impulse responses of the reconstruction filters in 2-D DT-CWT. The set of bandpass complex wavelet coefficients can be expressed by

$$y_{s,d}(u_s, v_s) = m_{s,d} e^{j\theta_{s,d}(u_s, v_s)} \tag{2.19}$$

where s and d ($d = 1, 2, \cdots, 6$) represent the level of decomposition and the direction of the filter, respectively. The variables u_s and v_s specify the location of the complex coefficients in each subband.

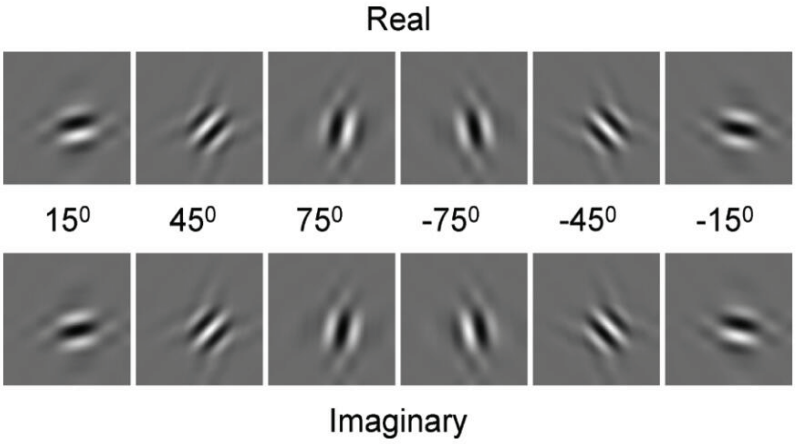


Figure 2.9: 2-D impulse responses of the reconstruction filters in the DT-CWT [65].

2.8 Performance criteria and measurements

In this section, the performance criteria and measurements for watermark invisibility, watermark extraction and quality evaluation are introduced respectively.

2.8.1 Performance criteria for watermark invisibility

For watermark invisibility, we use both objective and subjective quality measures to test the fidelity of watermarked images. In objective image quality evaluation, the peak signal-to-noise ratio (PSNR) and the structural similarity (SSIM) are used. In subjective image quality evaluation, the double-stimulus continuous quality scale (DSCQS) method is adopted.

PSNR

Peak signal-to-noise ratio (PSNR) is most commonly used to measure the quality of image reconstruction. It can be easily defined via the mean square error (MSE).

$$\begin{aligned} \text{MSE} &= \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}(I(i,j)-\hat{I}(i,j))^2 \\ \text{PSNR} &= 20log_{10}\frac{MAX}{\sqrt{\text{MSE}}} \end{aligned} \tag{2.20}$$

where I and $\hat{I}$ are the original image and the distorted image respectively; (i,j) are the coordinates of the current pixel in the image; M and N are the numbers of rows and columns of the image; MN is the total number of pixels in the image; MAX is the maximum possible pixel value of the image. For an 8 bits/pixel grey-scale image, MAX is equal to 255; for a normalized image or video frame, MAX equals to 1.

SSIM

The structural similarity (SSIM) index [66] is a metric for measuring the similarity between two images. It can evaluate the perceived image quality by computing the structural changes between the original image and the distorted image. The SSIM index is proposed to apply on image locally within 8×8 window and mathematically calculated by incorporating the luminance, contrast and structural information. The SSIM between the original image and the distorted image can be expressed as follows:

$$\begin{aligned} \mathrm{SSIM}(x, y) &= f(l(x, y), c(x, y), s(x, y)) \\ &= \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \end{aligned} \tag{2.21}$$

where x and y are the local regions of the original and distorted images respectively; $l(\cdot), c(\cdot), s(\cdot)$ and $f(\cdot)$ are the luminance, contrast, structural functions and the pooling function; C_1 and C_2 are two empirical parameters introduced to guarantee the stability of the SSIM metric. In practice, the mean SSIM (MSSIM) index is used to evaluate the overall image quality.

$$\mathrm{MSSIM}(\hat{I}, I) = \frac{1}{N} \sum_{i=1}^{N} \mathrm{SSIM}(x_i, y_i) \tag{2.22}$$

where x_i and y_i are the image contents at the ith local window; N is the number of local windows of the image. The calculated MSSIM index locates in $[0, 1]$.

DSCQS

Due to the non-linear property of the human visual system (HVS), the PSNR and SSIM are not always the real reflection of perceptual quality for images. So it is necessary to supplement objective measurements with subjective measurements. For subjective image quality evaluation, we use the double-stimulus continuous quality scale (DSCQS) method recommended by the ITU-R [67]. The double-stimulus method is thought to be especially useful when it is not possible to provide test stimulus conditions that exhibit the full range of quality.

The method is cyclic in that the assessor is asked to view a pair of images, each from the same source, but one via the process under examination, and the other one directly from the source. The subject is asked to assess the quality of both. Because most of the subjective methods are sensitive to variations, judgement sessions should include the full ranges of the factors varied. The general viewing conditions for subjective evaluation in laboratory environment are detailed in [67]. The test needs more than 15 observers and they should be non-expert and not experienced assessors. With the criteria of mean opinion score (MOS), the images are rated by trial viewers with five-grade scale from 5 to 1, where 5=Excellent, 4=Good, 3=Fair, 2=Poor and 1=Bad. The high score indicates good perceptual image fidelity after watermarking with subjective evaluation.

2.8.2 Performance measurements for watermark extraction

For watermark extraction, the similarity between the reference watermark pattern and the extracted watermark pattern is used to determine the existence of watermark

information in the target host. The correlation can be computed as different types of inner product of two images.

Linear correlation

The linear correlation is the most basic similarity measure. The linear correlation between two images f and g can be expressed by

$$r_{lc}(f,g) = \frac{1}{M \times N} \sum_x \sum_y f(x,y)g(x,y) \tag{2.23}$$

where $M \times N$ is the size of the images. If f is the reference watermark and g is the watermark extracted from the image, r_{lc} reflects the similarity between the two watermarks. One problem of the linear correlation is that the detection values are highly dependant on the magnitudes of the watermark pattern extracted from the watermarked images.

Normalized correlation

The normalized correlation is used to solve the problem by normalizing the extracted watermark pattern f and the reference pattern g to unit magnitude before computing the inner product between them.

$$r_{nc}(f,g) = \sum_x \sum_y \hat{f}(x,y)\hat{g}(x,y) \tag{2.24}$$

where

$$\begin{cases} \hat{f}(x,y) = \dfrac{f(x,y)}{\sqrt{\sum_i \sum_j f(i,j)^2}} \\ \hat{g}(x,y) = \dfrac{g(x,y)}{\sqrt{\sum_i \sum_j g(i,j)^2}} \end{cases} \tag{2.25}$$

Bit error rate

The bit error rate (BER) is another measure used to evaluate the accuracy of watermark extraction. The BER of extracted watermark can be calculated by

$$\mathrm{BER}(f(i,j), g(i,j)) = \frac{\text{number of } (f(i,j) \neq g(i,j))}{M \times N} \tag{2.26}$$

The values lie between 0 and 1. It is obvious that the lower BER of extracted watermark, the higher accuracy of watermark detection.

2.8.3 Performance measurements for quality evaluation

In this section, the quality metrics root mean square error (RMSE), Pearson's linear correlation coefficient (PLCC) and Spearman's rank-order correlation coefficient (SRCC) are presented. These metrics are used to conduct the accuracy of quality evaluation by comparing the predicted quality and the calculated quality.

Root mean square error

Root mean square error (RMSE) is a frequently used measure of the difference between predicted values and observed values. The RMSE of predicted variable Y with respect

to an observed variable X is defined as

$$\mathrm{RMSE}(X,Y) = \sqrt{\mathrm{MSE}(X,Y)} = \sqrt{\frac{1}{n}\sum_i (X_i - Y_i)^2} \tag{2.27}$$

where n is the number of predicted or observed values, X_i and Y_i are the values of the two variables respectively.

Pearson's linear correlation coefficient

Pearson's linear correlation coefficient (PLCC) is a measure of the linear correlation between two variables X and Y. The value is between +1 and -1, where 1 represents a total positive linear correlation, 0 is no linear correlation, and -1 denotes total negative linear correlation. PLCC can be computed by the covariance and standard deviations of two variables:

$$\rho_{X,Y} = \frac{cov(X,Y)}{\sigma_X \sigma_Y} \tag{2.28}$$

where $cov(X,Y)$ is the covariance of the two variables, and σ_X and σ_Y are the standard deviations.

Spearman's rank-order correlation coefficient

Spearman's rank-order correlation coefficient (SRCC) is a nonparametric measure of rank correlation. It can evaluate the relationship between two variables using a monotonic function. The definition of Spearman correlation is similar to that of Pearson correlation. However, different from the Pearson's correlation used to measure linear relationships, the Spearman's correlation evaluate monotonic relationships [68].

For a sample of size n, the n raw scores X_i, Y_i are transformed to ranks r_{X_i}, r_{Y_i}, and r_s is calculated as

$$r_s = \rho_{X,Y} = \frac{cov(r_X, r_Y)}{\sigma_{r_X}\sigma_{r_Y}} \tag{2.29}$$

where ρ represents the usual Pearson correlation coefficient, but applied to rank variables. $cov(r_X, r_Y)$ is the covariance of the rank variables, σ_{r_X} and σ_{r_Y} are the standard deviations of the rank variables. Only if all n ranks are distinct integers, it can be expressed by

$$r_s = 1 - \frac{6\sum d_i^2}{n(n^2-1)} \tag{2.30}$$

where $d_i = r_{X_i} - r_{Y_i}$ is the difference between the two ranks.

Chapter 3

Literature review

As stated in Chapter 1, a good digital watermarking algorithm should be robust to a variety of possible attacks. And there is a trade-off between imperceptibility and robustness in digital watermarking. In recent years, the blind watermarking for depth-image-based rendering (DIBR) 3D images has attracted considerable attention. Different from ordinary 2D image and video watermarking, there are more specific requirements for 3D watermarking, especially for DIBR 3D image watermarking. Besides the common signal processing attacks like rotation, scaling, noise addition and compression, a good 3D image watermarking algorithm should also resist view synthesis process and some DIBR distortions, such as depth map variation, baseline distance adjustment and different rendering conditions. Based on the theories and techniques that are used, the digital watermarking algorithms in literature are reviewed in this chapter.

3.1 Watermarking algorithms for DIBR 3D images

Due to numerous advantages of the DIBR system, many researches have been conducted on the DIBR technique and some DIBR watermarking algorithms have been proposed.

Fehn *et al.* [19] [20] proposed a DIBR system that allows for an evolutionary introduction of depth perception into the existing 2D digital TV framework. As part of the "Advanced Three-Dimensional Television System Technologies" (ATTEST) project, it is composed of four different functional building blocks: 3D content creation, 3D video coding, transmission and virtual views generation and 3D display. The authors described a new approach on 3D TV using DIBR, which includes the monoscopic color video and the associated depth information. And from these data, one or more virtual views of a real-world scene can be synthesized in real-time at the consumer side. The publications provide a detailed description of the DIBR technique on 3D TV, the comparison with the traditional approaches and a number of implementation details.

Muller *et al.* [69] presented an overview on the 3D video representation using depth data. The proposed format for 3D video representation consists of two or three camera views together with the associated depth information. The depth maps have disparities associated with every sample of the video signal, which can be used to generate arbitrary numbers of additional views via view synthesis. The authors also indicated that efficient compression is required for the transmission of 3D video formats. Tanimoto *et al.* [70] investigated the concept of free-viewpoint TV (FTV)

system and constructed a real-time system from image capture to image display.

Zhu *et al.* [71] provided a watermarking method for the new viewpoint video frame generated by the DIBR technique. The watermark is embedded in the selected blocks in the foreground object of the original video frame. The authors also found that the watermark is better preserved if embedded in the foreground object than the randomly selected embedding location. The method is validated to be robust and secure, but it is not blind because the original frame and the depth information are used for locating the watermark embedding blocks.

Halici *et al.* [72] reported a novel DIBR watermarking algorithm for free viewpoint television systems. The algorithm uses a correlation-based approach, where a watermark pattern is warped for every different view of multi-view video and embedded in the texture maps of those views. First, the projection matrix of the 2D rendered image is estimated. Then with this matrix, a rendered watermark is formed by warping the original watermark pattern. Finally, the symmetric phase only matched filtering is used to find the correlation between the rendered image and the rendered watermark. The simulation results demonstrate the applicability of the proposed algorithm. But this algorithm is not blind since the original images are needed for watermark detection, and it is only applied to images that have one object in them.

Koz *et al.* [73] pointed out the copyright problem for free-view video. In the proposed approach, the watermark is embedded into every frame of multiple views by exploiting the spatial masking properties of the human visual system. The approach is also extended for the case of an unknown virtual camera position and rotation. The transformations on watermark pattern caused by image based rendering operations are analyzed, and camera position and homography estimation techniques are used

for the virtual camera.

Lin *et al.* [1] proposed a digital watermarking scheme for DIBR 3D images, where the watermark embedded in the center view was kept for the left view and right view after rendering. It is a blind watermarking scheme because the watermark detection does not need the original data and the depth image. In this scheme, different reference patterns w_c, w_l and w_r are used to protect the center image I_c, the left-eye image I_l and the right-eye image I_r, respectively. The reference patterns are selected to be pair-wise orthogonal. The watermarks for protecting I_l and I_r are embedded in and generated from I_c. Let $block_{i,l}$ and $block_{i,r}$ denote the i-th blocks of I_l and I_r, respectively. $Render^{-1}(block_{i,l})$ and $Render^{-1}(block_{i,r})$ (the corresponding pixels in I_c for I_l and I_r) can be synthesized by utilizing the knowledge of DIBR. In the embedding process, the left-eye image I_l is divided into N blocks and the watermark bit b_i is embedded into $Render^{-1}(block_{i,l})$ by modulating the reference pattern w_l. The $Render^{-1}(block_{i,l})$ is skipped when the number of holes in the block exceeds a predefined threshold Th. By embedding the watermarks into the pixels of $Render^{-1}(block^{i,l})$ and $Render^{-1}(block_{i,r})$ in I_c, the effect of rendering operation can be eliminated in the content consumer side. I_l and I_r are synthesized by the same corresponding pixels in I_c. In order to detect the source of suspect image, this scheme embeds the watermarks into all the I_c, $Render^{-1}(block_{i,l})$ and $Render^{-1}(block_{i,r})$ by modulating different reference patterns w_c, w_l and w_r respectively. For watermark embedding, a linear approximation of the improved spread spectrum (ISS) technique is used to insert a reference pattern w_p into pixels of the blocks. For watermark detection process, the correlation operator is performed by computing the normalized inner product between the watermarked image and the reference pattern. The proposed

multiple watermarking scheme is effective to reduce the rendering noise and decrease the resultant bit error rate. The experimental results show that this scheme can resist JPEG compression and noise addition. But the frequently occurring geometric attacks like rotation, scaling and translation are not taken into account.

Kim *et al.* [2] developed a robust watermarking algorithm based on dual-tree complex wavelet transform (DT-CWT). In the watermark embedding process, the watermark information is embedded by the quantization process of the carefully selected DT-CWT coefficients groups. In the watermark detection process, the embedded watermark data is extracted using the statistical difference caused by quantization operation. The center image is first divided into subblocks and three-level DT-CWT is applied to each subblock. The coefficients in the level 2 and 3 of subbands are selected for watermark embedding. Then the selected six subbands are grouped into three pairs for each level. The bit sequence of a meaningful signature or text is shuffled to watermark data and embedded into the subbands coefficients. The watermarked center view is reconstructed by inverse DT-CWT from the quantized and unquantized subbands. For watermark detection, since one of the two coefficients in a group is quantized in the embedding process, there are statistical differences between the two subbands in a pair. The quantization error is set according to the pre-defined thresholds. And the watermark data is estimated by comparing the number of quantized rows between the left subband and the right subband. The embedded watermark can be extracted from the center, left and right views under distortions. The proposed algorithm is robust to baseline distance adjustment and depth map configuration. However, the algorithm has poor performance of robustness against combined attacks.

In [74], a perceptual watermarking approach for stereoscopic 3D video using depth information is introduced. The proposed approach utilizes the visual characteristics from the depth information in the DIBR system. To minimize the visual fatigue, the embedded watermark is invisible by exploiting hidden pixel and z axis motion. In [75], a novel unseen visible watermarking (UVW) method for color plus depth map 3D images is proposed, where the embedded watermark data is extracted by changing the rendering conditions. The non-linear property of rendering process is adopted to achieve zero perceptual distortion for virtual view image synthesis under normal rendering conditions. The simulation results show that the watermarked image can be perceived with good fidelity. In [76], a digital watermarking approach based on region of interest for 3D images is presented. Several middle frequency DCT coefficients in the depth perceptual region of interest are adjusted with smaller strength to embed the watermark. In [77], a digital watermarking method for DIBR 3D video is developed to protect each center view, left view and right view. The watermark is embedded into the highest level coefficients of the DT-CWT decomposition of the chrominance channel of the center view. The experimental results show the robustness against lossy JPEG compression and additive noise.

3.2 Watermarking based quality evaluation algorithms

The watermarking based quality evaluation is an important application of digital watermarking. Research on visual quality evaluation can be divided into three categories: full-reference (FR), reduced-reference (RR) and no-reference (NR) models.

Although the full-reference methods [66] [78] give more accurate quality evaluation, they require the presence of the original data to be able to assess. The reduced-reference approaches [79] [80] can evaluate the quality of an image or a video when fractional information of the original data is available. The partial information could be some extracted parameters or extra side data. The no-reference algorithms [81] [82] [83] can provide quality evaluation without accessing any additional information of the original data. Considering the bandwidth saving, the no-reference model is more preferable since the information of the original data do not need to be transmitted to the receiver side for quality evaluation.

The digital watermarking technique has been one of the most promising methods for reduced-reference or no-reference quality evaluation [39]. For watermarking based image quality evaluation technique, the watermark is usually embedded into the cover image based on the characteristics of embedding regions. Both the embedded watermark and the cover image suffer the same attack during transmission. At the receiver side, the quality of watermarked image can be assessed by examining the degradation of extracted watermark [38] [84] [85]. The main objective is to embed the watermark into the cover image in a proper way, so that the extracted watermark can accurately reflect the quality changes of the watermarked image under distortions. This is the core of the whole scheme, which determines the accuracy of quality evaluation.

In this section, the 2D image quality evaluation algorithms based on digital watermarking are first reviewed, and some typical 3D quality evaluation models are introduced subsequently.

3.2.1 Watermarking based 2D image quality evaluation algorithms

In watermarking based quality evaluation schemes, the quality of distorted watermarked image can be estimated using the degradation of corresponding extracted watermark. And they are usually designed to accurately assess the quality of distorted image without the presence of original image.

In [38], an image feature independent watermarking based quality evaluation method is introduced, where the watermark is embedded into the discrete wavelet transform (DWT) domain of the cover image. To achieve a balance between robustness and fidelity, more watermark information is embedded into the blocks with more middle frequency components and fewer watermark information is embedded into the blocks with lower or higher frequency components. The quantization parameters adjustment is used to maintain the high accuracy of quality estimation. The true detection rates (TDR) of the extracted watermark is used to indicate the degradation of the watermark. The quality of degraded image is estimated by mapping the calculated TDR to quality metrics using a respective empirical ideal mapping curve. The quality metrics for JPEG compressed images are calculated and compared with those estimated using the watermarking based approach. The evaluations demonstrate the effectiveness of the proposed method against JPEG image compression. But this method has expensive computation because of the iteration process.

In [86], an improved blind image quality evaluation approach based on spread spectrum watermarking is proposed. The approach incorporates a perceptual model into the spread spectrum watermarking by employing the Watson's just noticeable difference (JND) model of the host image. The locally optimum detector is employed

for watermark detection and used as a quality measure. For each particular image, the watermark is selected from a set of pre-chosen watermarks and the detector parameter is adjusted to approximate the ideal quality curve. This approach can assess the quality of images under JPEG compression but other distortions are not taken into account.

Based on their previous work in [38], the authors [39] subsequently developed a novel image quality evaluation algorithm based on adaptive watermarking and tree structure. The watermarking process is performed in the DWT domain of the cover image. The correlated DWT coefficients across the DWT subbands are grouped into set partitioning in hierarchial trees (SPIHT), which are further decomposed into a set of bitplanes. The watermark is embedded into the selected bit planes of the selected DWT coefficients of the cover image. The watermark embedding strength is assigned by pre-analyzing its content complexity in spatial domain and the perceptual masking effect of the DWT decomposed image in the DWT domain. The ideal mapping curve is obtained from a set of TDR-Quality curves and is used to map the calculated TDR of extracted watermark to a possible quality value. The proposed algorithm is tested in terms of PSNR, wPSNR, Watson JND and SSIM under JPEG compression, JPEG 2000 compression, Gaussian lowpass filtering and Gaussian noise addition. The experimental results show the effectiveness of this algorithm but it has poor performance when applied to 3D image quality evaluation.

3.2.2 3D image quality evaluation algorithms

3D image quality evaluation is very important in 3D multimedia applications. The conversion from 2D to 3D quality evaluation brings many challenges and some research

on quality evaluation for stereoscopic 3D images has been conducted.

In [87], the capabilities of several 2D image quality metrics are investigated in stereoscopic image quality evaluation. A study on the integration of disparity information into quality evaluation is presented. It is concluded that the combinations of the image quality and the disparity quality perform better than using the quality of either the original images or the disparity images solely.

In [88], a full-reference quality evaluation algorithm for stereopairs is proposed. The proposed framework is used for assessing the quality of stereoscopic images that have been afflicted by possible asymmetric distortions. An intermediate image is generated when viewed stereoscopically to have a perceived quality close to that of the cyclopean image, which is synthesized from the stereo image pair, the estimated disparity image and the Gabor filter response. The quality evaluation on the intermediate image yields higher correlations with human subjective judgements. The proposed algorithm outperforms the conventional 2D quality metrics when predicting the quality of stereoscopically viewed images under asymmetrical distortion. To further improve the performance of this algorithm, the authors [89] subsequently reported a no-reference binocular image quality evaluation model that operates on static stereoscopic images. The model deploys 2D and 3D features extracted from stereopairs to evaluate the perceptual quality when viewed stereoscopically. Both symmetrically and asymmetrically distorted stereopairs are handled by accounting for binocular rivalry, which is a perceptual effect that occurs when two eyes view mismatched images at the same retinal locations. The natural scene statistics (NSS) features are used to train a support vector machine (SVM) model to predict the quality of a test stereopair. The simulation results demonstrate the advantages over the

conventional 2D image quality evaluation methods when the proposed algorithm is tested on 3D image quality database.

In [90], how to predict the quality of a stereoscopic 3D image from that of the 2D image is studied. The authors carried out a subjective quality evaluation testing on a database that includes symmetrically and asymmetrically distorted images. They found that the quality prediction bias of the asymmetrically distorted images could lean toward opposite directions, depending on the distortion types and levels. Also, an information content and divisive pooling method is developed for 3D quality evaluation. The results show that this method can eliminate the prediction bias and improve the quality prediction of stereoscopic images.

These proposed quality metrics are all used for quality evaluation of stereoscopic 3D images. In this thesis, different from introducing new models for quality evaluation, we try to use the watermarking technique to evaluate the quality of stereoscopic images. To our best knowledge, there are no existing algorithms that achieve quality evaluation for DIBR 3D images using the watermarking technique.

3.3 Contourlet transform based watermarking algorithms

For digital watermarking, a "good transform" is very important in terms of watermark imperceptibility and robustness to various attacks. In the case of images, the transform should be insensitive to rotation, scaling, noise, filtering and other common signal processing. In addition, with the corresponding inverse transform, the watermarked signal should be reconstructed accurately after the watermark embedding.

The contourlet transform (CT), introduced by Do and Vetterli [52], gives an efficient directional multiresolution image representation. Compared with the DWT and other transforms (DCT, DFT), the main advantages of the contourlet transform are its directional subband selectivity and perfect reconstruction. The contourlet transform allows for a different number of directions at each scale, so that it has better performance in capturing the contours and edges of an image. These contours and edges are very important information and can be utilized for watermark embedding. Due to the directional multiresolution image representation and convenient tree structures, the contourlet transform is good for digital watermarking. In this section, some typical digital watermarking algorithms based on contourlet transform are reviewed.

Song *et al.* [91] presented an adaptive contourlet-based image watermarking algorithm. The watermark is embedded into the contourlet coefficients of the largest detail subbands of the image. Due to the special transform structure of Laplacian pyramid (LP), the watermark information can spread out into all subbands when the watermarked image is reconstructed based on the watermarked contourlet coefficients. Since both the low frequency subbands and the high frequency subbands contain the watermarking data, the proposed algorithm is expected to be robust against both high frequency attacks such as filtering and compression, and low frequency distortions like cropping and histogram equalization. For watermark extraction process, the existence of the watermark is checked by exploiting the unique transform structure of the LP, the correlation between the watermarked image and the spread watermark in all subbands of contourlet domain is calculated. The proposed algorithm is demonstrated to be robust to common attacks such as rotation, noise and compression.

Ranjbar *et al.* [92] developed a two-stage contourlet-based watermark embedding

and extraction method, where the watermark is embedded in the high frequency and low frequency components of an image, respectively. In the first stage, the odd description of the image using multiple description coding is divided into 16×16 non-overlapped blocks and the watermark is embedded into the high frequency component of the CT of the blocks. In the second stage, the watermark data is embedded into the low frequency component of the global CT of the image. So there are two watermark signatures embedded in the image. The main problem of this blind watermarking method is to identify the less affected copy of signatures between the two copies of the embedded watermark. The normalized correlation between these two copies is calculated as an indication of the quality of extracted watermark from the high frequency component. The low frequency embedded signature is used when the quality of the other one is less than a predefined threshold. The simulation results show that the proposed method can resist broader range of attacks than other related methods due to the multiple watermark embedding locations.

Akhaee *et al.* [55] devised an improved multiplicative image watermarking system in contourlet domain. The contourlet coefficients are multiplied by two special functions depending on the value of watermark data. These functions are optimized for the good robustness against additive white Gaussian noise (AWGN) and JPEG compression. The watermark is embedded in the directional subband with the highest energy. The contourlet coefficients are modeled with General Gaussian Distribution (GGD). Based on the Maximum Likelihood (ML) decision rule, an optimal detector by the aid channel side information is used to extract the embedded watermark data. The density function of the noisy contourlet coefficients is analytically calculated. The distribution of these coefficients is approximated with a suitable function to de-

crease the complexity of the receiver. The proposed approach is described in both semi-blind and blind versions. The experimental results show the superiority of this approach on the robustness against common attacks over other compared methods.

3.4 Principal component analysis based watermarking algorithms

As discussed in Chapter 2, the principal component analysis (PCA) is a linear transformation that uses a new coordinate system for the data set. It has advantages of data de-correlation and energy concentration that are suitable for digital watermarking. The PCA has been used in different ways in image and video watermarking methods.

Hien *et al.* [56] introduced a new watermarking method based on the PCA. In the encoding system, the original image is divided into $n \times n$ subblocks, and a set of coefficients in each subblock is selected for watermark embedding. The watermark data consists of pseudo-random number sequence and is embedded into the principal components of the uncorrelated coefficients of each subblock. The watermarked image is obtained by applying the inverse PCA process. In the decoding system, the uncorrelated coefficients are computed by the PCA image basic function and the watermarked coefficients. The correlation value between the reference watermark and possibly corrupted coefficients is computed to detect the watermark data.

Khalilian *et al.* [93] proposed a video watermarking algorithm with empirical PCA-based decoding. In the proposed algorithm, each video frame is divided into the non-overlapping blocks and the 2D DWT transform is applied to each block.

The watermark data is embedded into the LL coefficients of the blocks based on the energy of high frequency subbands and visual saliency. The decoding step makes use of the data projected onto the first principal component. The proposed algorithm is described in two versions: one is semi-blind where the watermark embedding location is stored as side information, and the other is blind. Due to the resilience of the first principal component against attacks, both versions are quite robust to additive noise and compression distortions. The performance of this algorithm could be improved in terms of robustness against desynchronizing attack.

3.5 Feature based watermarking algorithms

Feature based watermarking algorithms are widely used in image watermarking and many studies [94] [95] [96] [97] [98] have been conducted.

A geometrically invariant watermarking algorithm based on feature points is introduced in [94]. The watermark signature is embedded using a classical additive way inside each triangle of the tessellation. The watermark extraction is implemented based on the correlation properties of different triangles. In the proposed algorithm, the watermark is bound with a content descriptor defined by the salient points, and three different kinds of feature points are studied and developed as an improved detector. The simulation results show that the proposed algorithm is robust to JPEG compression, geometrical attacks and transformations. In addition, different perspectives of the content-based watermarking methods are analyzed as an conclusion.

The authors of [95] reported a feature-based robust digital image watermarking approach. It combines image feature and image normalization to resist geometric

attacks. The visually significant points are used as reference points to overcome the geometric synchronization problem. Several copies of a 16-bit watermark sequence are embedded in the original image to improve the robustness. The Mexican hat wavelet scale interaction is adopted for feature extraction. Moreover, the invariance properties of image normalization can significantly reduce the watermark search space. The approach is demonstrated to perform well under mild geometric distortions and common signal processing attacks. One drawback of this approach is that the selected feature points are not very stable.

In [96], a novel 3D watermarking method based on feature points is presented. The SIFT matching is adopted to find the common regions between the center view and the virtual views. The watermark is embedded into the DCT domain of the selected areas in the center view. The SIFT descriptor is used as the synchronization key to extract the watermark from the center image and the synthesized images. The proposed method shows good robustness to some signal distortion attacks such as JPEG compression and median filtering. However, it is not robust to geometric distortions like rotation and scaling.

In [98], the authors aimed to select a non-overlapping feature region set for robust image watermarking. A simulated attacking process and a track-with-pruning procedure are used to evaluate the robustness of feature regions and search a minimal primary feature set. For feature region selection, there are two major stages: the first operational stage is to find out a minimal feature region set under the objective of resisting as many predefined attacks as possible. In the second stage, the primary feature set is extended to enhance the robustness to undefined attacks. This work is formulated as a multidimensional knapsack problem and solved by a genetic based

technique. The proposed method can be robust against various attacks and preserve the high image quality. However, this method consumes too much computation time in measuring the robustness of the feature regions.

3.6 Ridgelet transform based watermarking algorithms

The ridgelet transform [60] was introduced as a sparse representation for functions on continuous spaces. It was proposed to overcome the problem of the wavelet transform in representing objects with singularities along line [61]. Compared with the traditional transform like discrete wavelet transform (DWT), the ridgelet transform is more effective in sparse image representation, and the most significant information of an image can be represented by the most energetic coefficients. It is a good transform domain for watermark embedding and extraction. Firstly, most of the energy of an image is concentrated in just a few coefficients, which is good for digital watermarking. Moreover, the ridgelet coefficients of an image have good stability and do not change significantly when the image is under possible attacks. In addition, the image can be reconstructed with high accuracy using inverse ridgelet transform.

Campisi *et al.* [62] devised a multiplicative watermarking method in the ridgelet transform domain, where the watermark is embedded into the relevant ridgelet coefficients of the edge image. The edge image is obtained by means of a filter bank with the circular harmonic functions, then it is partitioned into small blocks with straight edges. For each block, the direction having the greater energy is selected and the coefficients representing the most significant direction are embedded with random

generated watermark. The detection process is performed by applying dual operations with respect to the ones performed for watermark embedding. The robustness and transparency of the proposed method are proven by the experimental results. But the decoder has to perform full decoding to extract the watermark information. Moreover, the large rotation angles lower the performance of this method.

Kalantari *et al.* [63] developed a robust image watermarking algorithm in the ridgelet domain using universally optimum decoder. The watermark is embedded in the blocks of the host image by modifying the amplitude of ridgelet coefficients. The watermark extraction is performed using a universally optimum decoder. The decoder is dependent of the host signal distribution and is optimized by considering Gaussian noise attack to obtain maximum robustness. A noise variance estimation technique is used in the watermark extraction process. And the error correction is discussed and the error probability is analyzed. The experimental results show the effectiveness of the proposed algorithm and its good robustness against attacks.

3.7 Others

There are also some watermarking methods using other transforms [99] [100] [101] [102], watermarking based quality evaluation methods and image quality evaluation methods.

In [103], we introduced an adaptive digital watermarking approach based on redundant discrete wavelet transform (RDWT) and singular value decomposition (SVD). The video frames and the watermark are decomposed by RDWT, and the watermark is embedded into the subbands of the I-frames. The singular values of the I-frames

are modified with the singular values of the visual watermark. Meanwhile, the watermark embedding strength is adjusted adaptively using the noise visibility function (NVF) with local properties. The experimental results show the effectiveness of the proposed approach, and it is efficient and appropriate for real-time applications.

In [37], we devised a new watermarking based quality evaluation method for DIBR 3D images. The method utilizes the extracted watermark to evaluate the quality of the watermarked image under various distortions. The watermark is embedded into the selected dual-tree complex wavelet transform coefficients of the center view. The virtual left and right views are generated from the watermarked center view and the corresponding depth map. The watermark can be detected from the three views respectively. The relationship between the normalized correlation (NC) of extracted watermarks and the quality metrics of watermarked images under distortions is studied. The novelty of the method is that the authors try to use the watermarking technique to assess the quality of stereoscopic 3D images. The proposed method is demonstrated to be accurate and effective in terms of quality evaluation for DIBR 3D images.

In [104], we presented a novel perceptual quality evaluation algorithm for stereoscopic 3D images by considering the human visual characteristics. After the log-Gabor filter processing, the local amplitude and phase from the left and right views of the reference and distorted 3D images are utilized as features in local quality evaluation. Meanwhile, the global structure changes of the left and right views are also incorporated into the final quality pooling. The overall 3D quality score is obtained by combining the local and global quality indexes together. The effectiveness of the designed metric is verified on three public 3D image quality evaluation databases.

Experimental results show that the proposed algorithm can significantly improve the quality prediction of a stereopair. Moreover, in comparison with other related and typical methods, this algorithm exhibits better performance in terms of consistency with subjective assessment of stereoscopic 3D images.

Chapter 4

A contourlet-based 2D watermarking algorithm

In this chapter, we first propose a 2D image watermarking algorithm robust to geometric attacks and compressions. The proposed blind watermarking algorithm is based on contourlet transform (CT) and principal component analysis (PCA). The watermark is embedded into the principal components of the contourlet coefficients of the most energetic directional subband of the cover image. The normalized correlation (NC) between the original watermark and the watermark extracted from the distorted watermarked image is used as the robustness evaluation criterion. Then we extend this algorithm to blind video watermarking.

4.1 The motivation

The discrete wavelet transform (DWT), as one of the most commonly used transform domain algorithms, has good performance in both spatial and frequency domain. However, ordinary wavelets cannot represent the discontinuities and directional information like contours and edges in an image, which usually have high frequencies. Since the human visual system (HVS) is less sensitive to high frequency components, the watermark data can be embedded in these components to preserve the fidelity of the watermarked image. The contourlet transform (CT), introduced by Do and Vetterli [52], gives an efficient directional multiresolution image representation. Compared with the DWT and other transforms (DCT, DFT), the main advantages of the contourlet transform are its directional subband selectivity and perfect reconstruction. The contourlet transform allows for a different number of directions at each scale, so that it has better performance in capturing the contours and edges of an image. These contours and edges are very important information and can be utilized for watermark embedding. By contrast, the traditional transforms (DCT, DFT and DWT) are only suited for representing one-dimensional discontinuities, and they fail to represent the singularities and capture the directional information in two dimensions [52]. Moreover, the contourlet transform can also reconstruct the image accurately with appropriate reconstruction filters. In addition, the PCA based watermarking methods usually calculate the eigenvectors of the image correlation matrix. It has good performance in terms of energy concentration and de-correlation, which improve the imperceptibility and robustness.

For the digital watermarking techniques, the geometric attacks and compressions are the two most difficult distortions to tackle. Most of the existing image and

video watermarking methods cannot be very robust to both geometric attacks and compressions. Besides, the video watermarking has more requirements since there are a mass of data and inherent frames redundancy in video [8] [105]. It would be impractical to extract the watermark with the original data in every frame of a video. We are motivated to address the limitations of these watermarking techniques.

4.2 The proposed image watermarking algorithm

The proposed contourlet-based blind image watermarking algorithm is formulated in this section.

4.2.1 Watermark embedding

Let I and W represent the cover image and the watermark, respectively. The watermark embedding process can be described as follows:

1) The cover image I is transformed into the contourlet domain. The contourlet coefficients $I_{s,d}(i,j)$ of the largest detail or the most energetic directional subband $I_{s,d}$ of the cover image are selected to embed the watermark. $I_{s,d}(i,j)$ is the contourlet coefficients of the cover image at a resolution scale s and frequency direction d. We embed the watermark in the contourlet coefficients of the largest detail subband because these coefficients are related to the contours and edges in the image, which are usually the high frequency components of the image, and human visual system is less sensitive to these components [6]. The

largest detail or the most energetic directional subband can be found as follows:

$$I_{s,d} = \max \left\{ \sum_{i,j} (I_{s,d}(i,j))^2 \right\} \tag{4.1}$$

2) Apply PCA to the chosen directional contourlet subband of the cover image.

3) Use a pseudo random number (PRN) generator to produce a random watermark data sequence W, which consists of both positive and negative values. The length of the watermark sequence equals to the number of selected contourlet coefficients of the cover image. Then we embed the watermark sequence in the contourlet coefficients of the selected subband using the following equation:

$$I'_{s,d}(i,j) = \begin{cases} \dfrac{W \cdot \alpha}{P_{s,d}(i,j)} \cdot I_{s,d}(i,j), & P_{s,d}(i,j) > Th \\ I_{s,d}(i,j), & \text{Otherwise} \end{cases} \tag{4.2}$$

where $I_{s,d}(i,j)$ and $I'_{s,d}(i,j)$ are the original and watermarked contourlet coefficients respectively; $P_{s,d}(i,j)$ denotes the corresponding principal components of $I_{s,d}(i,j)$ before watermarking; the watermark embedding strength α is controlled by the noise visibility function. Here, Th is set empirically as a threshold to avoid too small $P_{s,d}(i,j)$.

4) Obtain the watermarked image I' by using inverse contourlet transform from the watermarked contourlet subband and other un-watermarked contourlet subbands.

In the proposed algorithm, the noise visibility function (NVF) is used to determine the watermark embedding strength. The NVF can be expressed by

$$\text{NVF} = \frac{1}{1 + \sigma_I^2(x, y)} \tag{4.3}$$

where $\sigma_I(x, y)$ is the local variance of the embedding regions. With the NVF, the embedding strength is modulated adaptively based on the characteristics of different local embedding regions [7]. The watermark embedding strength is as follows:

$$\alpha = (1 - \text{NVF}) \cdot \beta \tag{4.4}$$

where β is the predefined embedding strength.

4.2.2 Watermark extraction

The proposed image watermarking algorithm is blind since the watermark extraction does not need the original watermark. Strictly speaking, a blind watermarking scheme does not require any information for watermark extraction. But the community of watermarking usually thinks it is "blind" if only the security key is needed. The watermark extraction is an inverse process of the watermark embedding. The procedure of watermark extraction consists of the following steps:

1) Transform the watermarked image I' into contourlet domain at the same scales and directions as embedding procedure.

2) Apply PCA to the contourlet coefficients of the largest detail or the most energetic directional subband of the watermarked image.

3) Use the PRN generator and the security key to produce the same watermark sequence as the one used in the embedding process.

4) Calculate the gradient of the generated watermark sequence. Then we calculate the normalized correlation (NC) between the gradient of the principal components and the gradient of the generated watermark sequence. The watermark is successfully detected when the value of NC is larger than a predefined threshold.

$$\mathrm{NC} = \frac{\bigtriangledown P'^{T} \times \bigtriangledown W}{\sqrt{(\bigtriangledown P'^{T} \times \bigtriangledown P')(\bigtriangledown W^{T} \times \bigtriangledown W)}} \tag{4.5}$$

where $\times$ denotes matrix multiplication, $\bigtriangledown P'$ and $\bigtriangledown W$ are the gradient of the principal components of watermarked contourlet subbands and the gradient of generated watermark data sequence, respectively.

4.3 Experimental results and analysis

In this section, we conduct experiments to test the performance of the proposed algorithm on the watermark invisibility and the robustness against various attacks, including geometric transformations and image compressions.

For the contourlet transform, with three levels of pyramidal decomposition, we adopt the "9-7" biorthogonal filters for the Laplacian pyramid decomposition stage and "pkva" filters [52] for the directional filter banks stage. They are used because of their efficient implementation. For the watermark embedding, the predefined embedding strength β is initially set as 2.

4.3.1 Watermark invisibility

As a measure of watermark invisibility for the watermarked images, the objective and subjective quality measures are used.

For objective image quality evaluation, we use the PSNR and structure similarity (SSIM) [66] to measure the fidelity of the watermarked images. Fig. 4.1 presents the five images (*Lena*, *Peppers*, *Boat*, *Barbara* and *House*) for testing our algorithm. These images have different characteristics and textures, and they are commonly used for image watermarking. Table 4.1 gives the values of PSNR and SSIM of the five watermarked images for the proposed algorithm. The watermarked image with PSNR above 40 dB or SSIM above 0.95 looks very close to the original one.

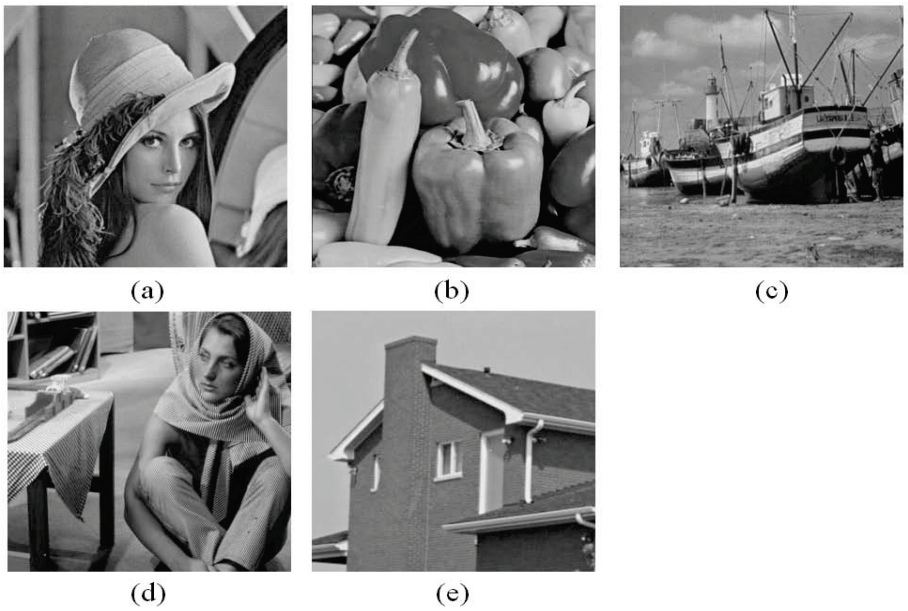


Figure 4.1: Test videos. (a) *Lena* (512×512); (b) *Peppers* (256×256); (c) *Boat* (512×512); (d) *Barbara* (512 × 512); (e) *House* (256 × 256).

For subjective image quality evaluation, we adopt the double-stimulus continuous quality scale (DSCQS) method. 20 viewers (10 males and 10 females) aged from

25 to 50 participated in the experiment. With the criteria of mean opinion score (MOS), the watermarked images were rated by the viewers with five-grade scale from 5 (excellent) to 1 (bad). As can be observed from Table 4.1, the proposed algorithm received relatively high scores in terms of image fidelity. The experimental results demonstrate that the blind watermarking algorithm preserves good perceptual quality of the watermarked images under both objective and subjective quality measures.

Table 4.1: Quality measurements of the watermarked images in terms of PSNR, SSIM and MOS.

Images	PSNR (dB)	SSIM	MOS
Lena	43.5392	0.965	4.3
Peppers	44.3615	0.973	4.5
Boat	44.1383	0.970	4.5
Barbara	45.2730	0.981	4.6
House	46.0317	0.992	4.8

4.3.2 Robustness against various attacks

To validate the robustness of the proposed algorithm, the watermarked image has been tested against commonly used attacks [22]: i) geometric attacks: rotation and scaling; ii) noising attacks: Gaussian noise and salt&pepper noise; iii) filtering attack: median filtering; iv) image compression attacks: JPEG compression and JPEG 2000 compression; v) other attack: histogram equalization (HE).

In this experiment, we have tested this algorithm on 50 images randomly selected from the USC-SIPI image database [106]. The simulation results are presented in Table 4.2. It can be observed that all the NC values of extracted watermarks under various attacks are above 0.71, and the mean of these values is calculated around 0.78.

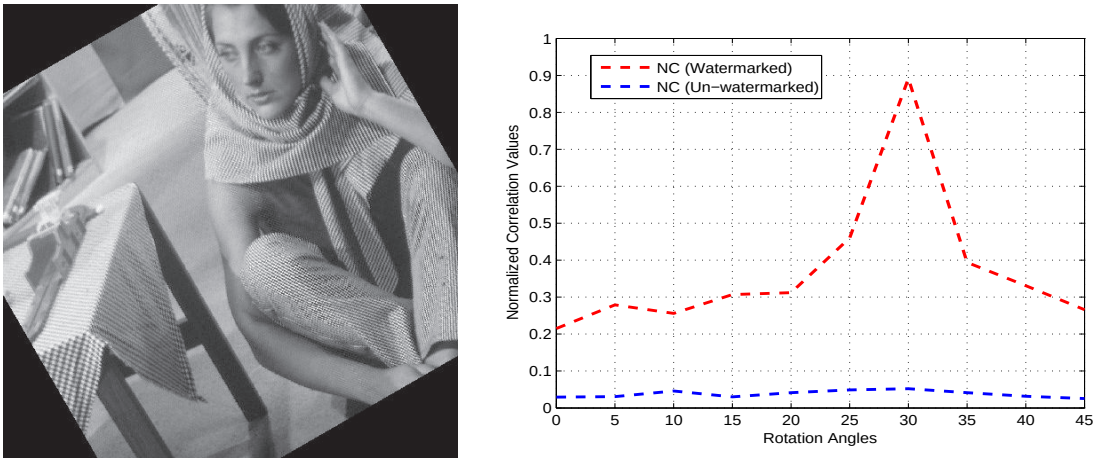
The experimental results demonstrate that the proposed blind image watermarking algorithm has good robustness against attacks. Also, the NC values of the extracted watermarks from un-watermarked images are given. The maximum and minimum values are 0.0521 and 0.0295, with a mean of 0.041. Therefore, there are big and clear separations between the NC values of the watermarked images and the NC values of the un-watermarked images so that the watermark can be easily detected.

Table 4.2: The NC values of the watermark extracted from the watermarked and un-watermarked images under various attacks.

Attacks	NC (watermarked)	NC (un-watermarked)
Rotation (30°)	0.8915	0.0521
Scaling (0.5)	0.8530	0.0490
Median filtering (5×5)	0.8052	0.0413
Gaussian noise (Var. 0.01)	0.7183	0.0295
salt&pepper noise (Density 0.05)	0.7364	0.0307
JPEG compression (QF: 30)	0.8371	0.0460
JPEG 2000 compression (Ratio: 10)	0.8117	0.0399
Histogram equalization	0.7560	0.0414

For the geometric attacks (rotation and scaling), the re-synchronization needs to be performed before watermark extraction. For rotation, the rotation angle is smaller than 30° in the experiment, it is feasible to exhaustively rotate the watermarked image back with 1° or smaller each step depending on applications. After each rotation back, the corresponding NC value is calculated and the angle with the highest NC value is the rotation angle. Fig. 4.2 illustrates the corresponding NC values after different angle corrections for the watermarked and un-watermarked *Barbara* image, respectively. As can be seen from this figure, the NC of watermark extracted from the watermarked image reaches its maximum value when the distorted watermarked image is rotated by 30°. Meanwhile, the NC of watermark from the un-watermarked

image keeps relatively small. For scaling, we assumed that the original size of the image is known at the receiver side. For example, all the tested images are 512×512. Similar to the contourlet-based image watermarking method used in [55], we can resize the attacked image to its original size and then detect the watermark. In fact, the scaling with factors less than one in the experiment can be considered as a low frequency attack, which will be discussed in Section 4.3.3. For other attacks without de-synchronization like Gaussian noise, JPEG compression, we can directly perform watermark extraction on the attacked watermarked image.



(a) (b)

Figure 4.2: Estimation of the rotation angle for *Barbara* image. (a) The watermarked *Barbara* image rotated by $-30°$; (b) The corresponding NC values with different rotation angle corrections.

4.3.3 Performance analysis on robustness to attacks

The good robustness of our proposed algorithm to various attacks mainly comes from the advantages of contourlet transform and principal component analysis (PCA). The various attacks used in this algorithm can be seen as a set of low frequency attacks and

high frequency attacks. The value of normalized correlation (NC) is used to evaluate the performance of watermark extraction. If there is watermark information in the watermarked image even after attacks and it can be extracted with high accuracy, the NC value will be high.

In the proposed algorithm, the watermark is embedded into the largest detail directional subband (the highest frequency subband). So the watermark information is kept in the watermarked image after the low frequency attacks, which just destroy the low frequency components of the image. The proposed algorithm is highly robust against various low frequency attacks, like scaling and histogram equalization. Meanwhile, although the watermark is embedded into the largest detail subband, the watermark can spread out into all subbands when we reconstruct the watermarked image [91] [107]. This is due to the special transform structure of the Laplacian pyramid (LP) [52] in the contourlet transform. So some watermark information can be preserved at the low frequency subbands. There is still watermark data in the watermarked image after the high frequency attacks. And the watermark can be extracted with high accuracy using the proposed algorithm. Thus, the proposed algorithm is also robust to the high frequency attacks, such as JPEG compression and median filtering. In addition, the principal components of an image have good stability and the PCA has sufficient robustness to rotation, image compression, etc. Based on the above analysis and discussions, the proposed algorithm has good robustness to various attacks.

4.4 Extended video watermarking algorithm

In this section, we extend our blind image watermarking algorithm to blind video watermarking. The proposed video watermarking algorithm partitions the raw video into groups of pictures (GOP); applies the contourlet transform [52] to each I-frame of GOP, then embeds and extracts the watermark in the selected contourlet subbands by using the principal component analysis. The video watermark embedding and extraction processes are very similar to those of the blind image watermarking.

4.4.1 Watermark embedding

The watermark is embedded into the blue channel of the chrominance rather than the luminance because the human visual system is less sensitive to this component [8]. In the implementation, each I-frame of GOP is embedded with the same watermark sequence. The watermark embedding process can be described as follows:

1) Segment the raw video into groups of pictures (GOP) with a fixed number of frames. Transform each I-frame of GOP into the contourlet domain and apply PCA to the chosen contourlet subbands of each I-frame of GOP.

2) Generate the watermark sequence by a PRN generator. The watermark sequence consists of both positive and negative values. Embed the watermark sequence in the coefficients of selected subbands of the I-frames using Equ. (4.2). It can be seen from Equ. (4.3) that the watermark embedding strength varies according to different local embedding regions and is adaptive to the video to preserve the fidelity of the watermarked frames.

3) Apply inverse contourlet transform to obtain the watermarked I-frames and generate the watermarked video together with other frames.

Fig. 4.3 is the contourlet transform of an original frame from video *Mobile*. Fig. 4.4 shows the first frame from video *Mobile* and its corresponding watermarked frame. It can be seen that the embedding algorithm preserves good perceptual quality of the watermarked frame.

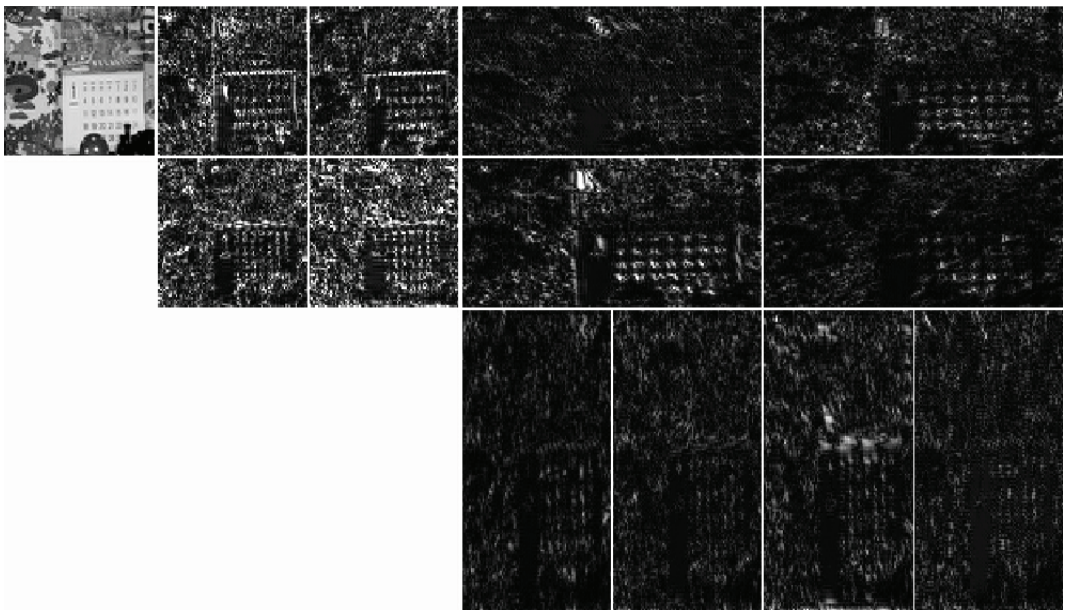


Figure 4.3: The original *Mobile* video frame in contourlet domain.

4.4.2 Watermark extraction

For watermark extraction, the normalized correlation (NC) is used to detect the existence of the watermark. The watermarked video and the watermarking key are available for watermark detection. This is a blind watermarking algorithm since the

(a) (b)

Figure 4.4: The original frame and the corresponding watermarked frame. (a) The first frame of video *Mobile*; (b) The first frame of the watermarked video *Mobile* (PSNR= 41.3966 dB).

process does not need the original video. The watermark extraction is an inverse process of the watermark embedding and consists of the following steps:

1) Segment the watermarked video (possibly attacked) into groups of pictures (GOP) with the same number of frames and select the same contourlet subbands as the embedding process for watermark detection.

2) Apply PCA to the selected contourlet subbands of the watermarked I-frames to obtain the principal components $P'_{s,d}(i,j)$ of the contourlet subbands.

3) Use the PRN generator to generate the same watermark sequence W as the one used in the embedding process. Calculate the NC between the gradient of the principal components $P'_{s,d}(i,j)$ and the gradient of the generated watermark sequence W with Equ. (4.5). The watermark is successfully detected when the value of NC is larger than a predefined threshold.

We take the first frame of video *Mobile* as an example. Fig. 4.5 shows the corresponding watermarked frames after four different kinds of attacks.

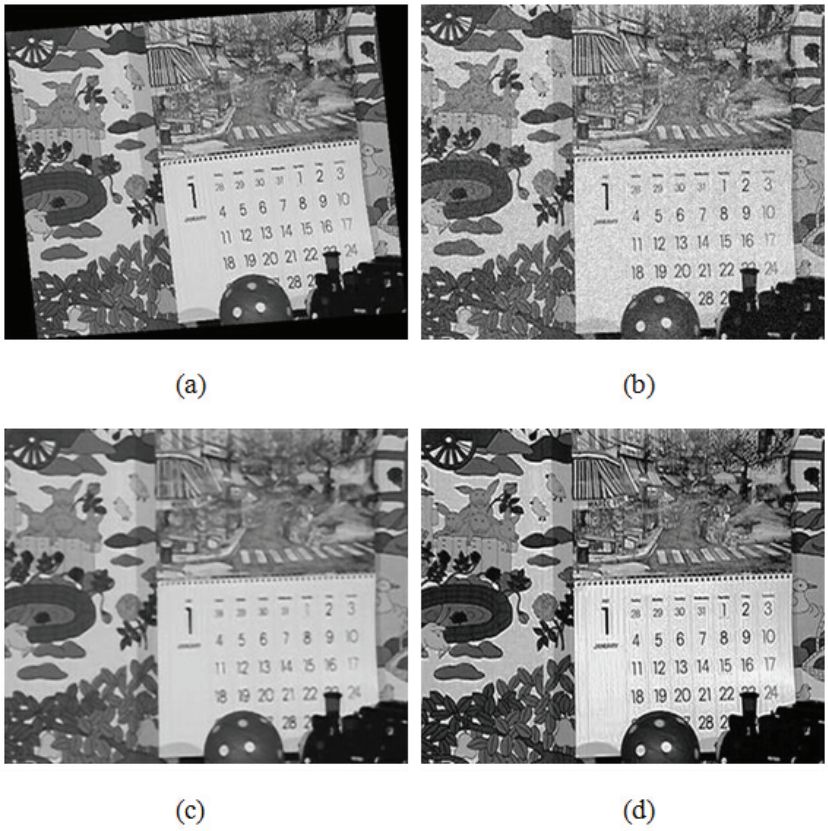


Figure 4.5: The watermarked *Mobile* frames under various video attacks. (a) Rotation (5°); (b) Gaussian noise (Variance: 0.05); (c) Median filtering (5 × 5); (d) Histogram equalization.

4.4.3 Experimental results

In this section, we conduct experiments to test the performance of the proposed algorithm on the robustness against various video attacks. Fig. 4.6 shows the four target video sequences (*Mobile*, *Stefan*, *Flower* and *Foreman*) for testing our algorithm. Their formats and sizes are presented in Table 4.3. In this experiment, we fix

the number of pictures of GOP to 9.

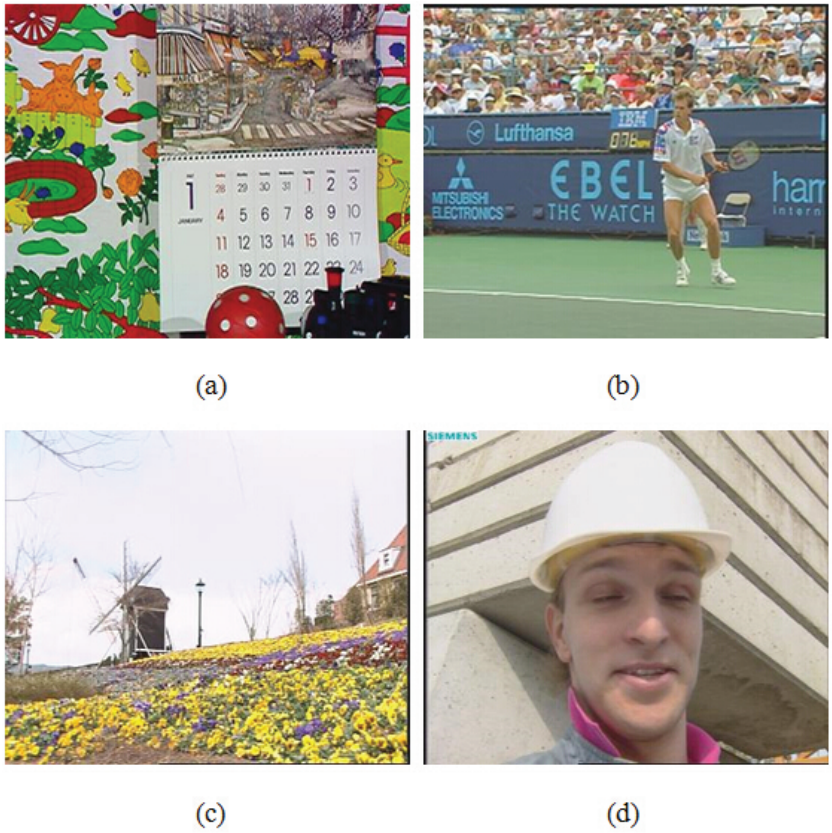


Figure 4.6: Test videos. (a) *Mobile*; (b) *Stefan*; (c) *Flower*; (d) *Foreman*.

Table 4.3: Target videos and their sizes.

Videos	Format
Mobile	CIF (352×288)
Stefan	CIF (352×288)
Flower	CIF (352×288)
Foreman	CIF (352×288)

Watermark invisibility

As a measure for the quality of watermarked frames and the accuracy of extracted watermarks, the PSNR and NC are used. Table 4.4 shows the PSNRs of the four watermarked videos. The video *Flower* has a PSNR of 38.6 dB because it has more details. The watermarked frames with these PSNRs look very close to the original frames and human eyes can not recognize the differences, which demonstrate that the proposed algorithm can keep high quality of the watermarked frames.

Table 4.4: The PSNR values of the watermarked videos under various attacks.

	Mobile	Stefan	Flower	Foreman
PSNR (dB)	40.3061	41.9273	38.6356	42.3155

Rotation and aspect ratio changes

For digital watermarking, geometric transformations are the most common attacks [105] [108]. In case of video, rotation often happens in very slight angles, normally not more than 5°; for frame aspect ratio changes, it converts the watermarked video from 11/9 to 16/9 and is similar to scaling for image watermarking. According to the simulation results in Table 4.5, the NC values of extracted watermarks from the watermarked video under rotations are all above 0.80, and the average NC value of the 4 test videos under aspect ratio changes is above 0.84. It can be seen that the proposed algorithm can resist rotation and frame aspect ratio changes.

Table 4.5: The NC values of the watermark extracted from the target videos under various attacks.

	Mobile	Stefan	Flower	Foreman
Rotation (1°)	0.9039	0.8867	0.8619	0.9123
Rotation (3°)	0.8763	0.8625	0.8381	0.8805
Rotation (5°)	0.8517	0.8363	0.8160	0.8692
Aspect to 16/9	0.8165	0.8352	0.8691	0.8646
Frame averaging	0.9370	0.9643	0.9317	0.9526
Gaussian noise	0.7692	0.7538	0.7260	0.7689
Salt&Pepper noise	0.8362	0.8437	0.8351	0.8564
Median filtering	0.9325	0.9358	0.9519	0.9632
Histogram equalization	0.9725	0.9696	0.9753	0.9668

Frame averaging

Frame averaging means to replace a frame by the average of several consecutive frames of the video. Considering averaging too many frames will degrade video quality, we replace a watermarked I-frame by the average of the watermarked I-frame, its left neighboring frame, and its right neighboring frame. From the experimental results in Table 4.5, the mean value of NC is 0.9464 (between 0.9317 and 0.9643), which shows that the proposed algorithm works effectively against frame averaging.

Gaussian noise and salt&pepper noise

During the transmission, the video could be distorted by noises in the channel. To test the robustness of the proposed algorithm against noise attacks, we added Gaussian noise and salt&pepper noise to the watermarked video at variance 0.01 and density 0.05, respectively. It can be seen from Table 4.5 that the performance under salt&pepper noise is better than that under Gaussian noise. But the high NC values

under the two types of noise distortions demonstrate that our algorithm is robust to both Gaussian noise and salt&pepper noise.

Median filtering and histogram equalization

There are some other kinds of attacks for video watermarking. In the experiment, we take the median filtering and histogram equalization as examples. The watermarked video is processed by the median filter with 5×5 neighborhood. The histogram equalization is used to increase the contrast of the video frames. As can be seen from Table 4.5, the average NC values under median filtering and histogram equalization are 0.9459 and 0.9711 respectively, which show that our algorithm performs well against median filtering and histogram equalization.

MPEG-2 compression, H.264 compression and HEVC/H.265 compression

Fig. 4.7 gives the experimental results for MPEG-2 compression. Each watermarked video is tested with MPEG-2 compression in different bit rates. It can be concluded that higher bit rate produces good video quality and higher PSNR, and yields better watermark extraction. As can be observed from this figure, our video watermarking algorithm has high NC values at different bit rates and can resist MPEG-2 compression.

With a similar PSNR as MPEG-2 compression, when the watermarked video is compressed by H.264 [109] [110] [111], the average normalized correlation (NC) of the detected watermarks is 0.51 (between 0.41 and 0.63). For un-watermarked video under H.264 compression, the average NC is 0.06 (between 0.03 and 0.11). The performance of the proposed video watermarking algorithm is also tested by the advanced

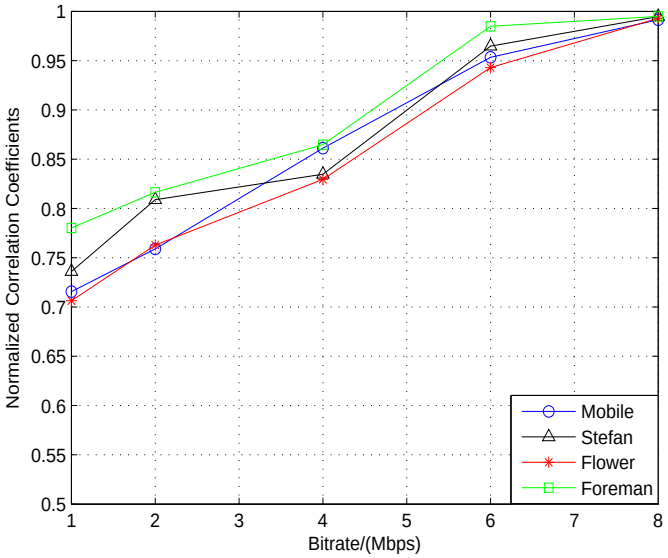


Figure 4.7: The NC values of the watermark extracted from the target videos under MPEG-2 compression with different bit rates.

video compression HEVC/H.265 [112] [113]. For HEVC/H.265 compression, with a similar PSNR as MPEG-2 compression, the average NC of extracted watermarks from the watermarked video is 0.46 (between 0.37 and 0.60) and the average NC from the un-watermarked video is 0.07 (between 0.02 and 0.12). It can be seen that the watermark can be detected under H.264 compression and HEVC/H.265 compression with sufficient robustness.

Performance comparison

In this section, we compare our algorithm with other related methods proposed by Liu *et al.* [8] and Wang *et al.* [114] which are typical video watermarking techniques robust to geometric attacks and video compressions.

According to the experimental results in Table 4.6, for video geometric attacks,

the average NC value of our algorithm under above mentioned attacks is 0.87, and the corresponding NC values of method [114] and method [8] are 0.65 and 0.74, respectively. For H.264 compression, with a PSNR around 40 dB, the NC values of our method are between 0.41 and 0.63 with an average of 0.51. The method in [8] gives the NC values between 0.43 and 0.57 with an average of 0.51. Our algorithm has better performance over the two methods in terms of robustness to geometric attacks, and has comparable performance with the methods in [8] in terms of robustness to H.264 compression.

Table 4.6: The NC values of the watermark extracted by the compared video watermarking methods under attacks.

Attacks	Method [114]	Method [8]	Proposed algorithm
Geometric attacks	0.65	0.74	**0.87**
H.264 compression	N/A	**0.51**	**0.51**

Based on the above experiments and comparisons, we can see that the proposed blind video watermarking algorithm is very robust to video geometric attacks and MPEG-2 compression, and is quite robust to H.264 compression and HEVC/H.265 compression.

4.5 Summary

In this chapter, we first proposed a new blind image watermarking algorithm based on contourlet transform (CT) and principal component analysis (PCA). Different from other related contourlet-based methods, the watermark is embedded in the principal components of selected contourlet coefficients of the cover image. The noise visibil-

ity function (NVF) is used to determine the watermark embedding strength. The experimental results have validated that the proposed algorithm is very robust to various attacks, such as rotation, scaling, noise attacks and image compressions, and performs well in extracting the embedded watermark. Then, the blind image watermarking algorithm is extended to a blind video watermarking algorithm, which can resist video attacks and video compressions. By comparison with other related methods, our proposed watermarking algorithm exhibits better performance in terms of robustness against various attacks.

The main contribution of the proposed algorithm is the robustness to various geometric attacks and compressions. On the whole, this algorithm shows better robustness against attacks in most cases. The introduction of the CT produces robustness against image and video compressions, and the PCA yields resistance to geometric attacks. The proposed algorithms are practical and can be applied for copyright protection and content authentication. Moreover, to our best knowledge, there is no video watermarking method that can be very robust to both geometric attacks and H.264/H.265 compressions now. Our proposed video watermarking algorithm is very robust to video geometric attacks and MPEG-2 compression, and is quite robust to H.264/H.265 compressions. The proposed video watermarking algorithm is practical and can be used for broadcast monitoring.

Chapter 5

A contourlet-based watermarking algorithm for DIBR 3D images

In the previous chapter, a 2D image watermarking algorithm based on contourlet transform (CT) and principal component analysis (PCA) is presented [16] [27]. In this chapter, we extend our work substantially and propose a robust contourlet-based blind watermarking scheme for depth-image-based rendering (DIBR) 3D images [10] [36]. Considering the specific requirements of DIBR watermarking and easy implementation, the PCA is not used in this scheme. As illustrated in Fig. 5.1, the proposed scheme applies the contourlet transform to the center image. We embed the watermark into the selected contourlet subbands of the center image by quantization [115] [116] [117] [118] on certain contourlet coefficients. For watermark extraction, the embedded watermark is detected by the statistical difference between quantized and unquantized contourlet coefficients. In the following, we briefly introduce the contourlet transform for 3D image watermarking, then we describe our proposed wa-

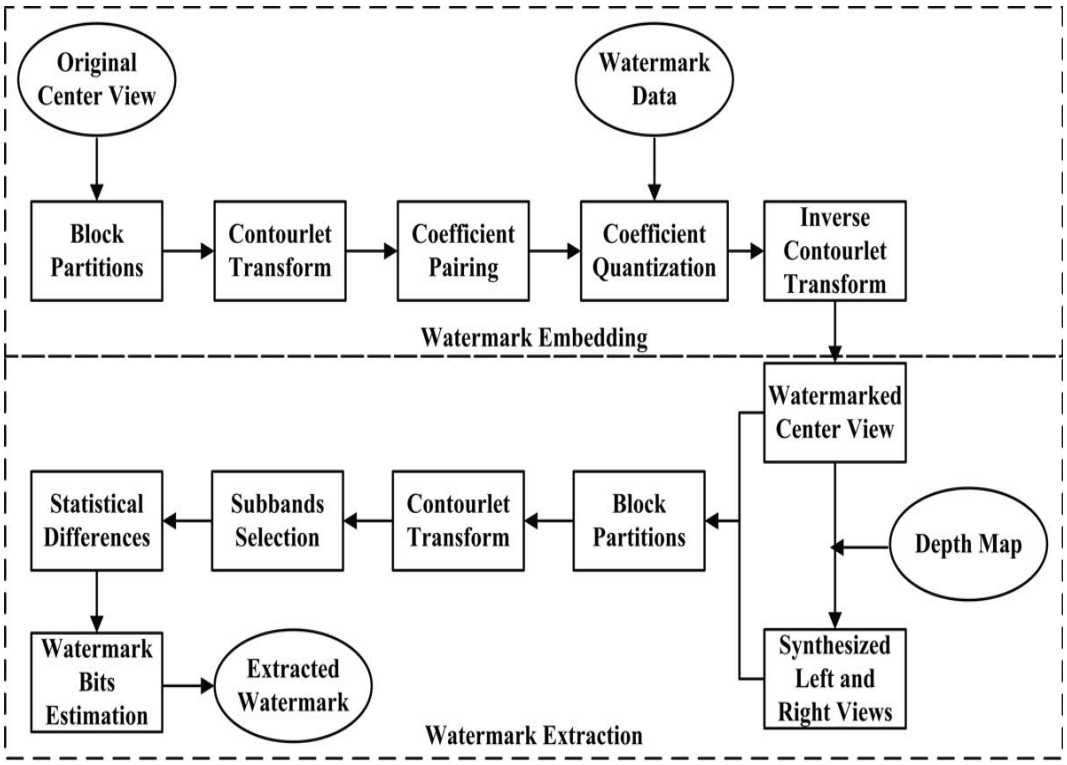


Figure 5.1: The proposed watermark embedding and extraction algorithms.

termark embedding and extraction algorithms.

5.1 Contourlet transform

Contourlet transform [52] uses a double filter bank structure to obtain images with smooth contours. There are two major stages in this double filter bank: Laplacian pyramid (LP) and directional filter bank (DFB). The LP is first applied to capture the point discontinuities, and the DFB is followed to form the linear structures. The LP at each level decomposes the image into bandpass images and a lowpass image. Then every bandpass image is fed into the DFB to obtain the directional subbands. The scheme can be iterated on the lowpass image so that the image can be decomposed

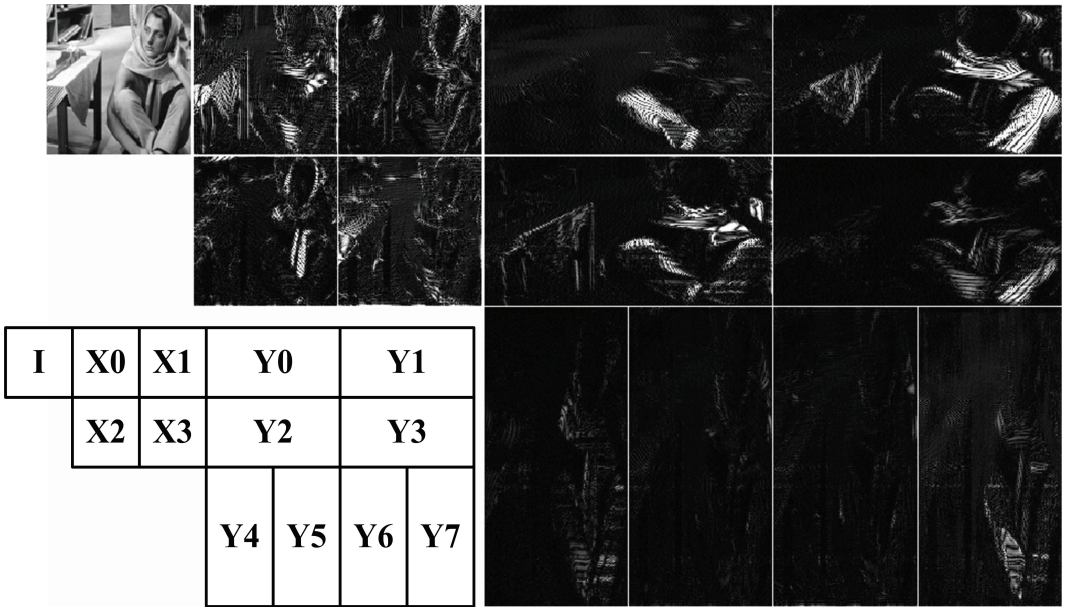


Figure 5.2: Contourlet transform of the *Barbara* image. The numbers in the subfigure on bottom left denote the corresponding decomposed subbands accordingly.

into different directional subbands at multiple scales through LP and DFB [27].

The contourlet transform of *Barbara* image is shown in Fig. 5.2. The original *Barbara* image is first processed by the LP filter bank with a decomposition of level 3. The numbers in the subfigure on bottom left represent the decomposed subbands accordingly (from left to right, from top to bottom). The image (I) is the final lowpass image. The bandpass images obtained from the LP are decomposed into four directional subbands (X0-X3) and eight directional subbands (Y0-Y7) respectively. Take the subband image Y0 for instance, the original image is decomposed into two pyramidal levels, which are then decomposed into eight directional subbands. The subband Y0 is one of the eight frequency partitions of the original image. The four directional subbands and the eight directional subbands provide different details of the original image because they have different contourlet coefficients. The different

images represent different directional subbands at different scales through Laplacian pyramid and directional filter banks. The sizes and shapes of the directional subband images are determined by the iterated contourlet filter banks [52]. The brighter areas indicate larger coefficients while the darker areas represent smaller coefficients [55] [91].

As can be observed from the eight directional subbands in Fig. 5.2, the subbands Y0-Y3 contain more horizontal edges and contours, such as the texture on the scarf, while the subbands Y4-Y7 have more information on vertical direction, like the wrinkles on the trousers. Since the DIBR process maps pixels in the center view horizontally, if the watermark is embedded in the subbands Y4-Y7 with more vertical information, the embedded watermark could be easily destroyed during the 3D image warping process. Therefore, we embed the watermark data in the first four contourlet subbands Y0-Y3.

We have also analyzed the histogram of the four selected contourlet subbands for watermark embedding, and found that there are similar subbands among them. For brevity, we take the image *Barbara* with three-level decomposition for an example. Similar results can be obtained for other images. Fig. 5.3 illustrates the histograms of the first four contourlet subbands (Y0-Y3) selected for watermark embedding and extraction. It can be seen from the figure that the histogram of subband Y0 is similar to that of subband Y3, and the histogram of subband Y1 has is similar to that of subband Y2. Therefore, we can make the similar subbands into a pair: (Y0,Y3) and (Y1,Y2), and embed the watermark into one of subband pair by coefficient quantization. So there are statistical differences between the quantized subbands and unquantized subbands, which can be used for watermark extraction. In the following

sections, we will describe our proposed watermarking scheme in detail.

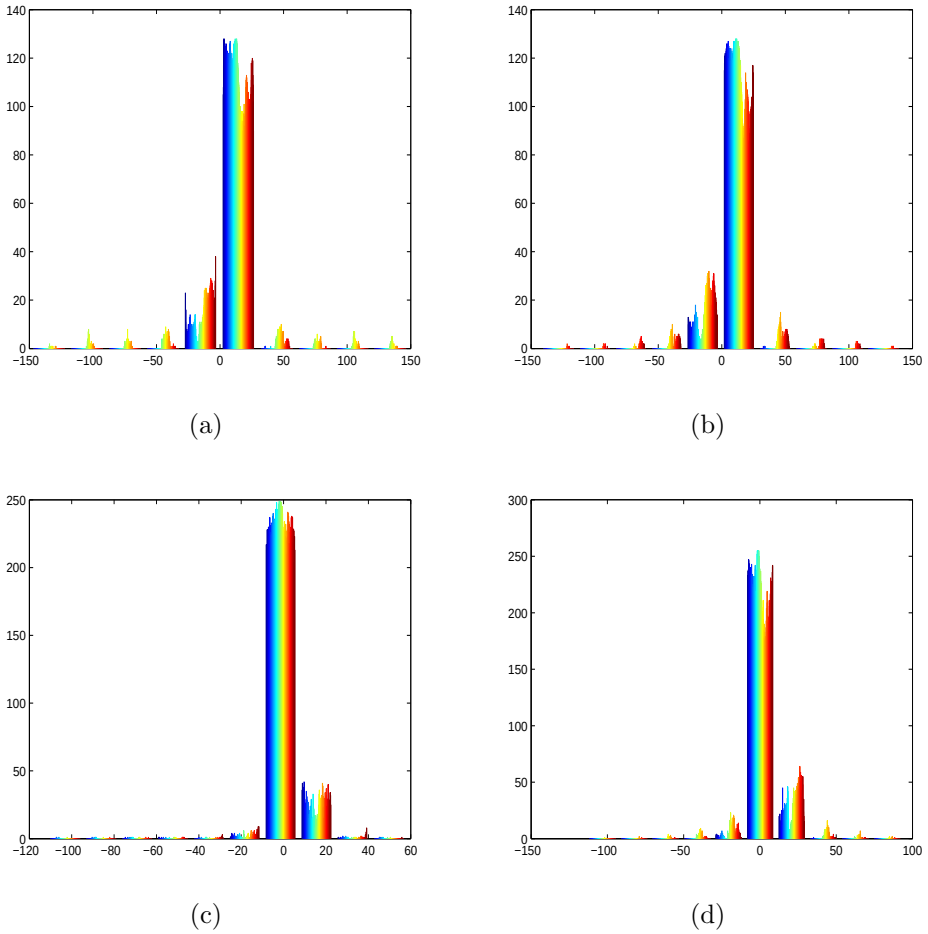


Figure 5.3: Histograms of the contourlet subbands for image *Barbara*. (a) The histogram of subband Y0; (b) The histogram of subband Y3; (c) The histogram of subband Y1; (d) The histogram of subband Y2.

Note that there is a trade-off between imperceptibility and robustness in the level of decomposition. With more numbers of decomposition, the fidelity of the watermarked image will be better preserved, but it leads to less robustness against attacks. In this scheme, we use three-level decomposition of contourlet transform.

5.2 Watermark embedding

The watermark embedding consists of the following steps:

1) The center image I is divided into $M \times N$ subblocks. The size of each block is $(I_w/M) \times (I_h/N)$, where I_w and I_h are the width and height of the center image. Apply contourlet transform to each subblock. We select the coefficients of the contourlet subbands $C_{lv,dr}$ for watermark embedding, where lv and dr are the level of decomposition and the direction of the subbands, respectively. In this scheme, $lv = 3, dr = 0, 1, 2, 3$.

2) The chosen four contourlet subbands $C_{lv,dr}$ in the decomposition of level 3 are divided into two groups based on their similar contourlet coefficients: $(C_{3,0}, C_{3,3})$ and $(C_{3,1}, C_{3,2})$, where each subband has mostly horizontal edges and contours. The similar subbands are paired so that the statistical difference of contourlet coefficients in a group can be used for watermark extraction.

3) Generate a watermark data sequence $W_i, i = 1, 2, \cdots, M \times N$ by using a pseudo random number (PRN) generator, which is spread spectrum consisting of 1 and -1.

4) The watermark data is embedded into the selected four contourlet subbands by coefficient quantization. Here, we quantize subbands $C_{3,0}$ when the watermark data is 1 and quantize subbands $C_{3,3}$ when the watermark data is -1. Subbands $C_{3,1}$ and $C_{3,2}$ are operated with the same process.

If $W_i = 1$,

$$C'^{m,n}_{3,0}(j,k) = \text{round}[C^{m,n}_{3,0}(j,k)/\alpha] \tag{5.1}$$

$$C'^{m,n}_{3,3}(j,k) = C^{m,n}_{3,3}(j,k) \tag{5.2}$$

$$C'^{m,n}_{3,1}(j,k) = \text{round}[C^{m,n}_{3,1}(j,k)/\alpha] \tag{5.3}$$

$$C'^{m,n}_{3,2}(j,k) = C^{m,n}_{3,2}(j,k) \tag{5.4}$$

If $W_i = -1$,

$$C'^{m,n}_{3,3}(j,k) = \text{round}[C^{m,n}_{3,3}(j,k)/\alpha] \tag{5.5}$$

$$C'^{m,n}_{3,0}(j,k) = C^{m,n}_{3,0}(j,k) \tag{5.6}$$

$$C'^{m,n}_{3,2}(j,k) = \text{round}[C^{m,n}_{3,2}(j,k)/\alpha] \tag{5.7}$$

$$C'^{m,n}_{3,1}(j,k) = C^{m,n}_{3,1}(j,k) \tag{5.8}$$

The quantization errors are:

$$E^{m,n}_{3,0}(j,k) = |C'^{m,n}_{3,0}(j,k) - C^{m,n}_{3,0}(j,k)| \tag{5.9}$$

$$E^{m,n}_{3,1}(j,k) = |C'^{m,n}_{3,1}(j,k) - C^{m,n}_{3,1}(j,k)| \tag{5.10}$$

$$E^{m,n}_{3,2}(j,k) = |C'^{m,n}_{3,2}(j,k) - C^{m,n}_{3,2}(j,k)| \tag{5.11}$$

$$E^{m,n}_{3,3}(j,k) = |C'^{m,n}_{3,3}(j,k) - C^{m,n}_{3,3}(j,k)| \tag{5.12}$$

where $C'^{m,n}_{3,0}(j,k)$, $C'^{m,n}_{3,1}(j,k)$, $C'^{m,n}_{3,2}(j,k)$, $C'^{m,n}_{3,3}(j,k)$ and $C^{m,n}_{3,0}(j,k)$, $C^{m,n}_{3,1}(j,k)$, $C^{m,n}_{3,2}(j,k)$, $C^{m,n}_{3,3}(j,k)$ are the coefficient values of the corresponding contourlet subbands after and before quantization, respectively. m and n are the row index and column index of the divided subblock, $m = 1, 2, \cdots, M$, $n = 1, 2, \cdots, N$; j and

k are the row index and column index of the contourlet subband coefficient of each subblock respectively, $j, k = 1, 2, \cdots$; α denotes the watermark embedding strength; *round* represents the rounding operation of a number. $E_{3,0}^{m,n}(j,k)$, $E_{3,1}^{m,n}(j,k)$, $E_{3,2}^{m,n}(j,k)$, $E_{3,3}^{m,n}(j,k)$ are the quantization errors of the contourlet coefficients.

5) The watermarked center view I' can be obtained by inverse contourlet transform from the quantized and unquantized coefficients of the contourlet subbands.

5.3 Watermark extraction

Although the watermark data is embedded into the center view, there is watermark information in the synthesized left and right views after the DIBR process. This is a blind watermarking algorithm since the extraction process does not need the original center view and the depth map. Strictly speaking, a blind watermarking scheme does not require any information for watermark extraction. But the community of watermarking usually thinks it is "blind" if only the security key is needed. The watermark extraction process is the reverse operation of watermark embedding and can be described as follows:

1) Apply contourlet transform to the watermarked center image I' (or virtual left and right views) with the decomposition of level 3. These views may be distorted by various possible attacks.

2) Select the same contourlet subbands as the embedding procedure for watermark detection.

3) Since one of two coefficients in a group is quantized in the embedding process, there is statistical difference between the two subbands. By comparing the quantized and unquantized coefficients in a group, the watermark data $W_i^{'}$ is estimated to be 1 or -1. The quantization error differences are used for watermark extraction:

$$n1(j,k) = \begin{cases} 1, & \text{if } E_{3,0}^{m,n}(j,k) - E_{3,3}^{m,n}(j,k) > Th1 \\ & \text{or } E_{3,1}^{m,n}(j,k) - E_{3,2}^{m,n}(j,k) > Th1; \\ 0, & \text{otherwise.} \end{cases} \tag{5.13}$$

$$n2(j,k) = \begin{cases} 1, & \text{if } E_{3,3}^{m,n}(j,k) - E_{3,0}^{m,n}(j,k) > Th2 \\ & \text{or } E_{3,2}^{m,n}(j,k) - E_{3,1}^{m,n}(j,k) > Th2; \\ 0, & \text{otherwise.} \end{cases} \tag{5.14}$$

Then all the $n1(j,k)$ and $n2(j,k)$ are added together.

$$N1 = \sum_{j,k} n1(j,k) \tag{5.15}$$

$$N2 = \sum_{j,k} n2(j,k) \tag{5.16}$$

Therefore, the extracted watermark can be estimated:

$$W_i^{'} = \begin{cases} 1, & \text{if } N1 > N2; \\ -1, & \text{otherwise.} \end{cases} \tag{5.17}$$

where $n1(j,k)$ and $n2(j,k)$ are used to calculate the numbers of quantized contourlet subbands coefficients; $Th1$ and $Th2$ are the predefined thresholds. Therefore, for each subblock, if the number $N1$ of quantized coefficients in subbands $C_{3,0}$ and $C_{3,1}$ is calculated to be more than the number $N2$ of that in subbands $C_{3,2}$ and $C_{3,3}$, the extracted watermark data $W_i^{'}$ is estimated to be 1; otherwise the extracted watermark data $W_i^{'}$ is -1.

4) Use the PRN generator to produce the same watermark sequence W_i as the one used in the embedding process.

5) The normalized correlation (NC) is calculated between the extracted watermark $W_i^{'}$ and the generated watermark W_i. The NC here is used to detect a watermark. The watermark is successfully detected if the value of the NC is above a predefined threshold.

$$\mathrm{NC} = \frac{\sum_{i=1}^{M\times N} W_i \cdot W_i'}{\sqrt{\sum_{i=1}^{M\times N} W_i^2 \cdot \sum_{i=1}^{M\times N} W_i'^2}} \tag{5.18}$$

We use the bit error rate (BER) to measure the accuracy of watermark extraction for the proposed algorithm. The BER for the extracted watermark sequence $W_i^{'}$ and the generated watermark sequence W_i is defined as

$$\mathrm{BER}(W_i, W_i^{'}) = \frac{\text{number of } (W_i^{'} \neq W_i)}{M \cdot N} \tag{5.19}$$

5.4 Error analysis

The proposed watermark decoding system can yield false positive errors that depend on $Th1$ and $Th2$, discussed in the watermark extraction process. In order to estimate the error rate and set the thresholds properly, we choose 2.5×10^5 coefficients from 10 randomly selected un-watermarked images. The quantization error differences of the two unquantized coefficients in a pair are calculated using the same proposed watermark extraction algorithm. The histograms of the quantization errors from the un-watermarked images are shown in Fig. 5.4. The quantization error difference can be roughly modeled with Gaussian distribution. As can be seen from this figure, there is a slight difference for different embedded watermark information. If the embedded bit W_i is assumed as 1, the probability of quantization error difference would be lower than 0.01 when its value is larger than 1.0. Based on this fact, the threshold $Th1$ should be larger than 1.0. However, if the threshold value is becoming larger, the watermark extraction from the watermarked images would be affected (Some watermark information may not be detected due to the large threshold value). Therefore, the threshold $Th1$ is set to 1.0 to guarantee the high accuracy of watermark detection. The threshold $Th2$ is set to 0.7 with the same process when the embedded watermark data $W_i = -1$.

5.5 Experimental results and analysis

Fig. 5.5 shows 10 experimental pairs of center images and associated depth images obtained from Middlebury Stereo Datasets [119] [120] and HHI (Heinrich-Hertz Institut), Germany [19] [20]). The sizes of the center images and the depth images range

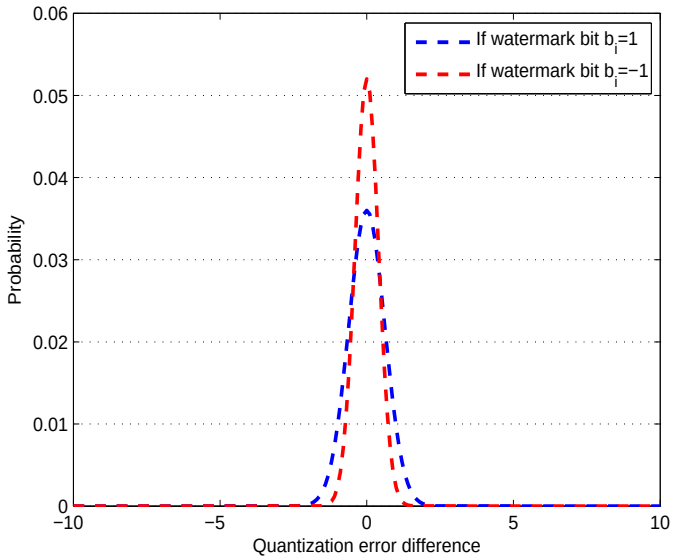


Figure 5.4: Histograms of the quantization error differences from the un-watermarked images.

from 450×375 to 1390×1110, which are given in Table 5.1, and the depth maps are gray scale images of 8-bit level.

Table 5.1: Sizes of test image pairs.

Images	Art	Books	Dolls	Moebius	Teddy
size	1390×1110	1390×1110	1390×1110	1390×1110	450×375
Images	Cones	Aloe	Laundry	Orbi	Interview
size	450×375	1282×1110	1342×1110	720×576	720×576

For the contourlet transform, with two levels of pyramidal decomposition, we use the "9-7" biorthogonal filters for the Laplacian pyramid decomposition stage and "pkva" filters [52] for the directional filter banks stage. They are used because of their efficient implementation.

In this section, we conduct experiments to test the performance of the proposed watermarking algorithm in terms of robustness to various attacks. In this experiment,

Figure 5.5: Test image pairs of center images and corresponding depth images. (a) Art; (b) Books; (c) Dolls; (d) Moebius; (e) Teddy; (f) Cones; (g) Aloe; (h) Laundry; (i) Orbi; (j) Interview.

the baseline distance t_x is set to 5% of the image width for comfortable viewing [1]. We set focal length $f = 1, Z_{far} = t_x/2$ and $Z_{near} = 1$. For watermark embedding and

extraction processes, we set the block indexes $M = N = 8$, the embedding strength $\alpha = 2$, and the thresholds $Th1 = 1.0$ and $Th2 = 0.7$.

5.5.1 Watermark invisibility

Fig. 5.6 presents the visual experimental results of the watermarked center image and the synthesized left and right images for *Interview*. As a measure of watermark invisibility of the watermarked images, subjective and objective quality evaluations are used.

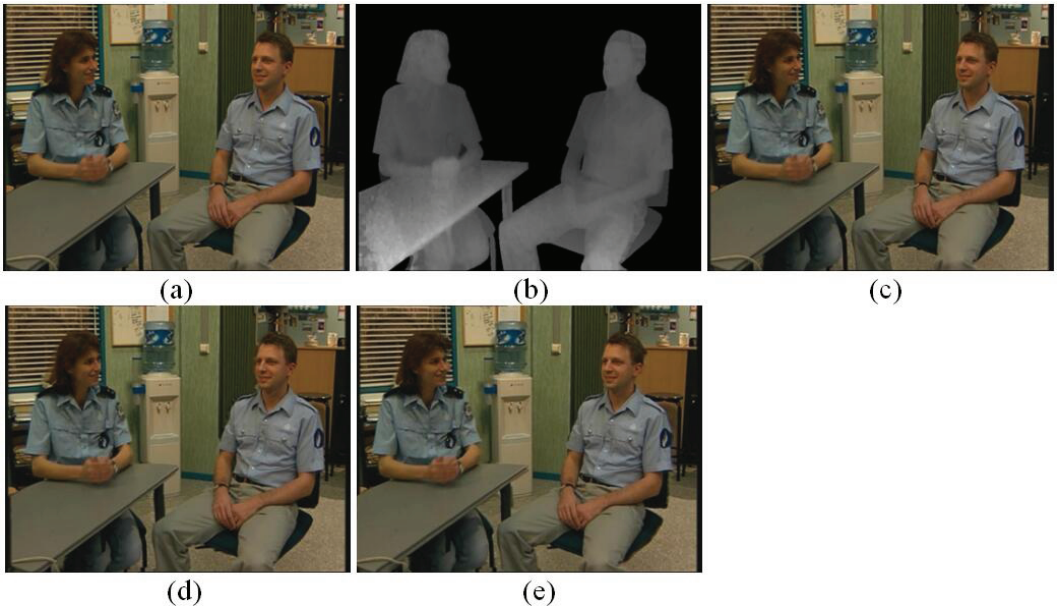


Figure 5.6: Watermark invisibility for *Interview*. (a) The original center view; (b) The depth map; (c) The watermarked center view (PSNR=41.07 dB); (d) The virtual left view rendered from the watermarked center view and the depth map (PSNR=39.26 dB); (e) The virtual right view rendered from the watermarked center view and the depth map (PSNR=39.07 dB).

For objective image quality evaluation, we use peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) [66] to test the fidelity of the watermarked images.

Table 5.2 gives the values of PSNR and SSIM of the watermarked center images for the 10 test image pairs. The watermarked image with PSNR around or above 40 dB or SSIM above 0.95 looks very close to the original one.

Table 5.2: Quality measurements of the watermarked center images for the proposed algorithm in terms of PSNR, SSIM and MOS.

Images	PSNR (dB)	SSIM	MOS
Art	42.71	0.995	4.7
Books	39.65	0.973	4.3
Dolls	41.29	0.986	4.4
Moebius	42.30	0.991	4.6
Teddy	41.43	0.993	4.5
Cones	40.57	0.979	4.4
Aloe	39.52	0.972	4.3
Laundry	40.86	0.985	4.5
Orbi	42.53	0.993	4.8
Interview	41.07	0.971	4.3

Due to the non-linear property of the human visual system (HVS), the PSNR and SSIM are not always the real reflection of the perceptual quality of the images. For subjective image quality evaluation, we adopt the double-stimulus continuous quality scale (DSCQS) method recommended by the ITU-R [67]. There are 15 viewers (8 males and 7 females) aged from 24 to 50 participating in the experiment. With the criteria of mean opinion score (MOS), the images were rated by trial viewers with five-grade scale from 5 (excellent) to 1 (bad). As can be observed from Table 5.2, the proposed algorithm received relatively high scores in terms of image fidelity. The simulation results show that the proposed algorithm preserves good perceptual quality of the watermarked images.

For the 10 test image pairs, the watermark is embedded into the center view, the virtual left and right views are synthesized from the watermarked center view and

Table 5.3: Average NC of the watermark extracted from the center view, the virtual left and right views.

		Center view	Left view	Right view
NC (watermarked)	Maximum	0.969	0.953	0.946
	Minimum	0.938	0.922	0.907
	Average	0.955	0.939	0.930
NC (un-watermarked)	Maximum	0.094	0.078	0.069
	Minimum	0.063	0.047	0.052
	Average	0.070	0.059	0.061

the associated depth map at the receiver side. In the experiments, all the NC values of extracted watermark from watermarked images are above 0.8 while the NC values of extracted watermark from un-watermarked images are very small, lower than 0.1. Table 5.3 presents the NC values of watermark extraction from the watermarked and un-watermarked images for the three views, respectively. From this table we can see that there is a big difference between the NC values from the watermarked and un-watermarked images. So the watermark can be easily detected. In the experiment, the bit error rate (BER) is used to measure the accuracy of watermark extraction. Table 5.4 gives the average values of PSNR, SSIM and watermark bits of the watermarked images, and the BER of the watermark extracted from the center view, the virtual left and right views without any attacks respectively. We have compared our scheme with two other related watermarking algorithms [1] [2] that are typical and state-of-the-art for DIBR 3D images. Table 5.4 demonstrates that there are no significant differences in the values of PSNR, SSIM and watermark bits, but our algorithm can extract the watermark with lower average BER value (0.016) and has competitive performance in terms of watermark extraction from each distributed view.

Table 5.4: Average PSNR and SSIM, watermark bits and BER of the watermark extracted from the center view, the virtual left and right views.

	PSNR	SSIM	Watermark	BER			
				Center view	Left view	Right view	Average BER
Lin *et al.* [1]	41.25	0.989	64	0.007	0.112	0.091	**0.070**
Kim *et al.* [2]	40.73	0.982	64	0.011	0.023	0.036	**0.023**
Proposed	41.14	0.984	64	0.010	0.018	0.021	**0.016**

5.5.2 Robustness to geometric attacks

To validate the robustness of the proposed algorithm, the watermarked images have been tested against various geometric attacks [7] [22] [40]. The performance of the proposed scheme on robustness is compared with other two related watermarking techniques [1] [2] which are typical and state-of-the-art for DIBR 3D images.

For the 10 test image pairs, Fig. 5.7 shows the average BER and corresponding average NC values of the extracted watermark from the synthesized right views under various attacks, including rotation, scaling, Gaussian noise, median filtering and JPEG compression. The values of average NC are all above 0.8. The average BER for the synthesized right views under rotation is shown in Fig. 5.7(a), where the rotation angles are from 0° to 10° in clockwise and anticlockwise directions. Fig. 5.7(b) presents the average BER for the scaled synthesized right views with scaling factors varying from 0.5 to 2.0. By comparing with other watermarking algorithms, the simulation results demonstrate that the proposed algorithm has lower BER and is robust to rotation and scaling.

As can be seen from Fig. 5.7(c) and Fig. 5.7(d), the proposed algorithm shows lower bit error rates when the watermarked views are distorted by Gaussian noise with

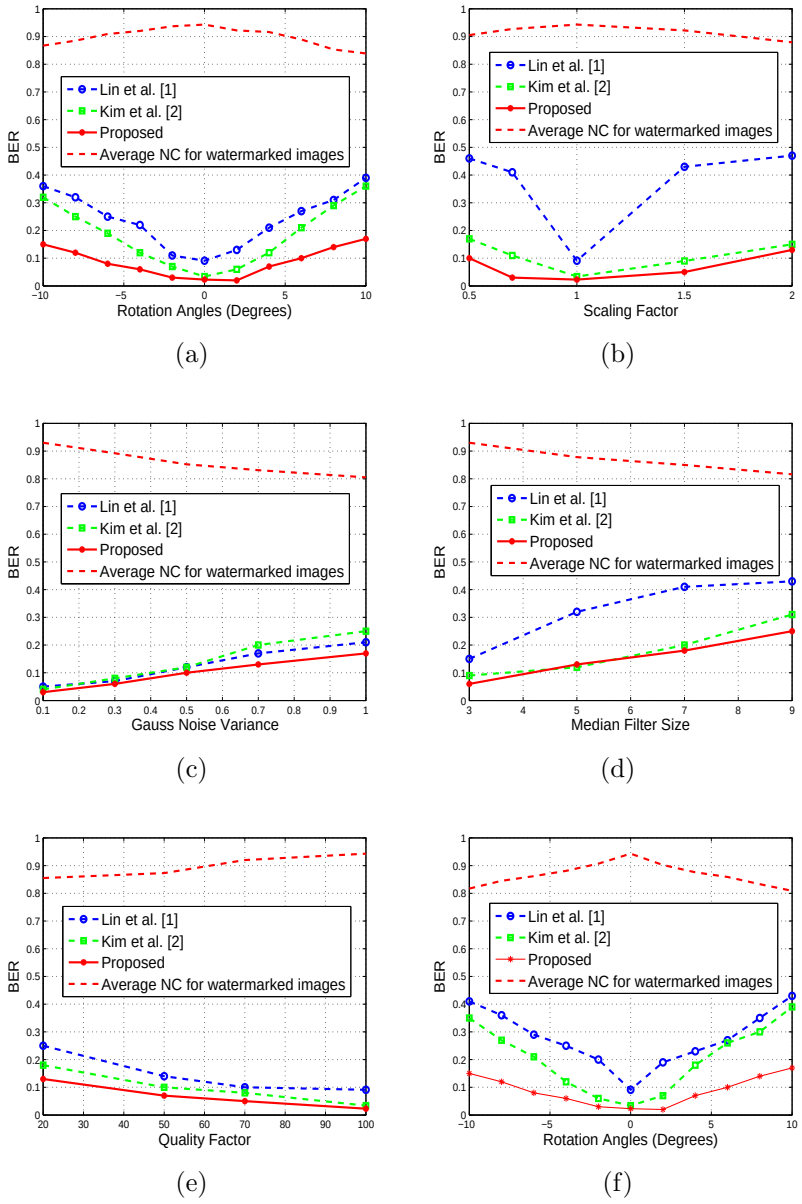


Figure 5.7: Average BER of the watermark extracted from the watermarked right views under various attacks. (a) Rotation; (b) Scaling; (c) Gaussian noise; (d) Median filtering; (e) JPEG compression; (f) Rotation and cropping.

variance from 0.1 to 1.0 and median filtering with different window sizes, respectively. Fig. 5.7(e) gives the average BER under JPEG compression with various quality

factors. Fig. 5.7(f) shows the average BER under combined attacks (cropping after rotation). It can be seen that our algorithm performs well in detecting the watermarks with lower BER under these distortions, especially for combined attacks.

5.5.3 Robustness to baseline adjustment and depth map variation

The baseline distance adjustment and depth map variation are common processing for the DIBR system. The virtual left and right views can be rendered from the center view and the corresponding depth map by the values of different baseline distances. In this experiment, we adjust the baseline distance from 1% to 10% of the center image width. Fig. 5.8 suggests that 5% of the image width is the optimum baseline distance for comfortable viewing, and our algorithm can remain lower BER (less than 0.1) under baseline distance variation.

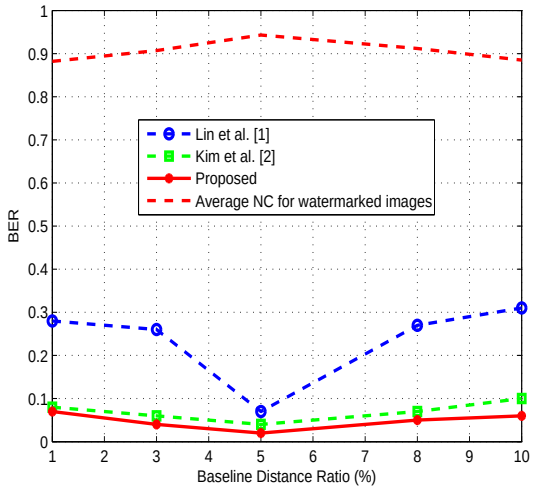


Figure 5.8: Average BER for the watermarked right views under baseline distance adjustment.

As we discussed in Chapter 2, the pre-processing of depth image can reduce the hole occurrence to render the left and right views with high quality. However, similar to the center view, the depth maps can also be distorted by possible attacks. In this experiment, the depth map of the 10 test image pairs are distorted under Gaussian noise with variance 0.1 and JPEG compression with quality factor 10, respectively. The visual results of depth map for image *Teddy* are shown in Fig. 5.9. As can be observed from Table 5.5, compared with other methods, the proposed watermarking algorithm can detect the watermark with lower BER and is robust to depth map variation. This experiment also demonstrates that the depth map can be compressed before transmission to save the bandwidth of the DIBR system.

Figure 5.9: Depth map pre-processing. (a) The original depth map of *Teddy*; (b) The depth map with Gaussian noise (Variance: 0.1); (c) The depth map with JPEG compression (Quality factor: 10).

Table 5.5: Comparison of BER and NC for depth map variation.

	Gaussian noise	JPEG compression	Average NC (watermarked)
Lin *et al.* [1]	0.15	0.17	N/A
Kim *et al.* [2]	0.10	0.12	N/A
Proposed	**0.07**	**0.09**	0.905

5.5.4 Robustness to different DIBR methods

As mentioned in Chapter 2, there are mainly three steps in a DIBR system and the detailed operations can be different for practical applications. In this section, we test our watermark embedding and extraction algorithms on other typical DIBR processes [29] [51]. In [51], the edge dependent depth filter and interpolation are used in depth map processing and hole-filling steps, respectively. In [29], the depth maps are pre-processed with an asymmetric filter. Table 5.6 gives the BER and the corresponding average NC values of extracted watermark from the synthesized left and right views for the two DIBR methods. The virtual left and right views are generated from the watermarked center view and corresponding depth image. So there is watermark information in the synthesized left and right views. The watermark extraction result from the center view remains the same in different DIBR systems. The experimental results show that our watermarking algorithm can extract the watermark with low BER and also has good performance in different DIBR systems.

Table 5.6: BER and NC of the extracted watermark for different DIBR methods.

	Method [51]	Method [29]	Average NC (watermarked)
Virtual left view	0.03	0.02	0.927
Virtual right view	0.04	0.03	0.916

5.5.5 Further discussions

In the above experiments, the watermark is embedded into the contourlet subbands $C_{3,0}$-$C_{3,3}$ instead of subbands $C_{3,4}$-$C_{3,7}$. Table 5.7 presents the average PSNR and SSIM of different watermarked contourlet subbands of the center views and the synthesized views for the 10 test image pairs. It can be seen from the table that the values

of PSNR and SSIM of subbands $C_{3,0}$-$C_{3,3}$ are higher than that of subbands $C_{3,4}$-$C_{3,7}$. Although there is no big difference in the values of PSNR and SSIM for center images, the quality of synthesized images can be significantly different. The average values of PSNR and SSIM for subbands $C_{3,0}$-$C_{3,3}$ are 38.52 dB and 0.956 respectively, while the corresponding values for subbands $C_{3,4}$-$C_{3,7}$ are only 33.93 dB and 0.915. It can be concluded that the subbands with more horizontal information are less affected by the DIBR process, which is discussed in Subsection 5.1 and demonstrated from the experimental results in this subsection.

Table 5.7: Average PSNR and SSIM of the watermarked images for different subbands.

Subbands		$C_{3,0}$-$C_{3,3}$	$C_{3,4}$-$C_{3,7}$
Center images	PSNR (dB)	41.19	40.35
	SSIM	0.983	0.971
Synthesized images	PSNR (dB)	38.52	33.93
	SSIM	0.956	0.915

For the subband pairing, subband $C_{3,0}$ is paired with subband $C_{3,3}$ rather than subband $C_{3,1}$ or $C_{3,2}$ because they have similar contourlet coefficients, which can be used for watermark extraction after quantization. Table 5.8 gives the experimental results of watermark extraction from different subband pairing methods. The bit error rates of extracted watermark are relatively high for the latter two subband pairing methods. Moreover, the value of threshold for quantization error is not easy to set.

Table 5.8: Average BER of the extracted watermark for different subbands pairing.

Subband pairing	$(C_{3,0}, C_{3,3})$ and $(C_{3,1}, C_{3,2})$	$(C_{3,0}, C_{3,1})$ and $(C_{3,2}, C_{3,3})$	$(C_{3,0}, C_{3,2})$ and $(C_{3,1}, C_{3,3})$
Center images	0.01	0.15	0.12
Synthesized images	0.02	0.23	0.27

5.6 Summary

In this chapter, we proposed a blind watermarking algorithm based on contourlet transform for DIBR 3D images. In the proposed algorithm, the watermark data is embedded into the selected contourlet subbands of the center view by quantizing certain contourlet coefficients. The statistical differences between quantized and unquantized contourlet coefficients in a group are used for watermark extraction. The watermark can be detected blindly with low bit error rates from the center view, left view and right view even they are under various attacks, such as geometric transformation, Gaussian noise and JPEG compression. Moreover, the proposed algorithm can be robust to common DIBR processing, such as depth map processing and baseline adjustment. One advantage of the proposed algorithm is its simplicity and practicality. However, the performance of this watermarking algorithm still needs to be further developed. To be specific, the normalized correlation of the extracted watermark is not very high, and its robustness to geometric attacks, Gaussian noise and JPEG compression needs to be improved. In the next chapter, we will continue to conduct research on DIBR 3D image watermarking.

Chapter 6

A novel watermarking algorithm for DIBR 3D images using feature regions

In this chapter, a new digital watermarking algorithm for DIBR 3D images based on feature regions and ridgelet transform (RT) is proposed. The watermark is embedded into the ridgelet coefficients of the selected feature regions. Compared with other transforms, most of the energy of an image is concentrated in just a few of ridgelet coefficients, which is good for watermark embedding. In addition, ridgelet transform has better stability than contourlet transform and the coefficients do not change significantly even when the image is under possible distortions, which contribute to better robustness to attacks. In the following, the proposed watermarking algorithm is presented in detail. The error probability analysis is given and the experimental results are provided. The performance on robustness is compared with that of other

related and state-of-the-art methods.

6.1 The proposed watermarking algorithm

In this section, we present the proposed watermark embedding and extraction algorithms. Fig. 6.1 presents the framework of the proposed algorithm. After finding the reference feature points of the center image, we construct the circular feature regions for watermark embedding and extraction. Then we utilize the improved spread spectrum (ISS) [121] technique to embed the watermark messages into the selected ridgelet coefficients of selected feature regions, and insert the reference pattern into each region. For watermark extraction, the correlation between the selected watermarked coefficients and the reference pattern is used to estimate the watermark data extracted from the feature regions.

6.1.1 Feature point detection

We first detect the feature points of the center image with Gaussian scale model in scale invariant feature transform (SIFT) [58], which is based on the scale space theory to find the feature points invariant to scaling and rotation. In Gaussian scale model, the difference of Gaussian (DoG) is used to select the stable ones among those feature points. Suppose $I(x, y)$ is the original center image with size of $M \times N$, $L(x, y)$ is the

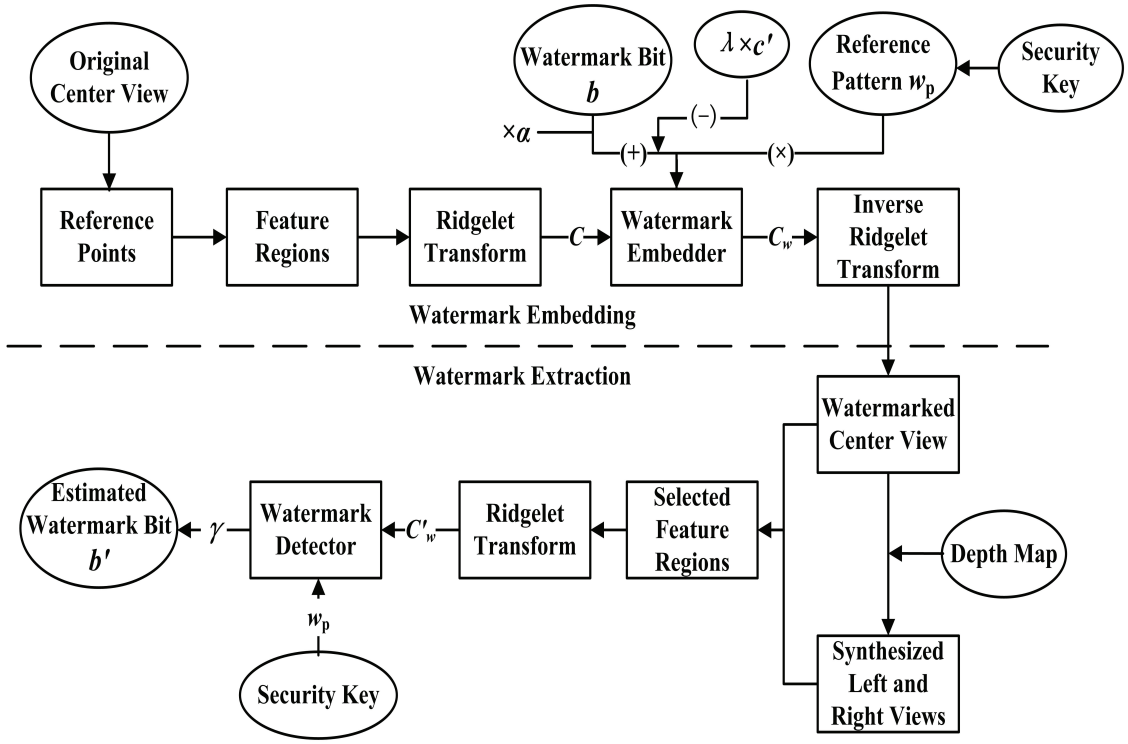


Figure 6.1: The proposed watermark embedding and extraction algorithms.

blurred image filtered by the Gaussian filter G_σ defined as

$$\begin{aligned} G_\sigma &= \frac{1}{\sqrt{2\pi\sigma^2}}\exp[-(x^2+y^2)/(2\sigma^2)] \\ L_1(x,y) &= G_\sigma * I(x,y) \\ L_2(x,y) &= G_{g\sigma} * I(x,y) \end{aligned} \tag{6.1}$$

where g is a constant factor and σ denotes the scale. Here, $L_1(x,y)$ and $L_2(x,y)$ are different in scale by the factor g. Then the DoG filtered image can be calculated by

$$\text{DoG} = L_2(x,y) - L_1(x,y) = G_{g\sigma} * I(x,y) - G_\sigma * I(x,y) \tag{6.2}$$

The DoG can be obtained from the successively filtered images [7]. To reduce the probability of errors and make the feature points still detectable after distortions, the feature points that are very robust to concerned distortions should be selected for watermarking process. The reference points are selected as follows: the feature points with strong edge responses will be removed, because they are usually located near the image edges or the depth map discontinuities and fragile to the view synthesis process. Moreover, in this algorithm, the DoG is iterated on the filtered image to find the stable reference points which are invariant to scaling and rotation. In addition, each feature point has a corresponding scale information. The feature points with either large or small scales are vulnerable to possible attacks and should be discarded. In the implementation, we chose the reference points with scale σ between 6 and 8 empirically. Fig. 6.2 gives the final result of selected reference feature points in the original center image of *Ballet*. From the figure we can see that the selected feature points are not on the object edges, contours of the image, or the discontinuities of corresponding depth map. These feature points can not be easily found from the spatial domain and they are robust against attacks like rotation, scaling and noise addition. Therefore, the selected reference points are very suitable for watermark embedding, and can still be detected for watermark extraction after possible attacks.

6.1.2 Feature region assignment

Once the reference points are selected, we will construct the feature regions for watermark embedding. The reference point is chosen and the disk region centered at this reference point with radius R will be used as watermark embedding location. In the proposed algorithm, we construct the circular feature area using the following

(a)

(b)

Figure 6.2: The selected feature points of image *Ballet*. (a) The original center image; (b) The selected feature points of the original center image.

formula:

$$(x - x_i)^2 + (y - y_j)^2 = R^2 \tag{6.3}$$

where (x_i, y_j) is the selected feature point in image $I(x, y)$, R can control the size of circular region. Some of the circular areas are overlapped and not stable, so these feature regions are removed to make the selected feature areas more suitable for watermark embedding [96]. The circular regions with strong edge responses are also discarded. The final selected feature regions that can be robust to concerned attacks are given in Fig. 6.3.

Figure 6.3: The selected feature regions for watermark embedding.

After that, the orientation assignment [58] and image normalization [7] are employed to make the circular areas rotation and scaling invariant, respectively. For orientation assignment, the gradients of the pixels in a window centered at the reference points are calculated with the first order derivative. Then the histogram of the gradient is computed, where the peak is used as the orientation of the reference point. For image normalization, scaling normalization is incorporated to obtain the scaling invariance for the feature region. It transforms the image to its standard form by adjusting the origin of the image to its centroid.

6.1.3 Watermark embedding

The watermark embedding can be described as follows:

1) Select the reference feature points of the center view, and construct the disk feature regions for watermark embedding. The processes have been detailed in Section 6.1.1 and Section 6.1.2, respectively.

2) Apply ridgelet transform to the selected feature regions of the center view, then embed the watermark data into the ridgelet coefficients of the feature regions with the most energetic direction. The most energetic direction for each feature region can be found using the following formula:

$$k_e = \max_k \left\{ \frac{1}{s} \sum_{m=0}^{s-1} (\mathrm{FRIT}[k, m])^2 \right\} \tag{6.4}$$

The ridgelet coefficients $FRIT[k_e, m]$ are the most energetic coefficients in the selected feature block. m ($m = 0, 1, \cdots, s-1$) denotes the coefficients of ridgelet transform in each direction and s is the number of selected ridgelet coefficients.

3) In this scheme, we use the linear approximation of the improved spread spectrum (ISS) technique [121] to embed the watermark data into the selected ridgelet coefficients. Each watermark data b (1 or -1) is embedded by modifying the amplitude of the selected coefficients as follows

$$c_w = c + (\alpha b - \lambda c')w_p \tag{6.5}$$

where $c = [FRIT[k_e, 0], FRIT[k_e, 1], \cdots]$, c_w and c are the ridgelet coefficients after and before watermark embedding respectively; α is used to control the

embedding strength of the watermark data b; λ represents the reduction ratio of the interference c'; the reference pattern w_p represents a spread spectrum sequence for the embedded watermark data b. A security key is used by the pseudo random number (PRN) generator to produce w_p with zero mean. The ISS technique [121] can improve the robustness of watermarking by reducing the interference c' between the ridgelet coefficients c and the spread spectrum sequence w_p during watermark embedding process. The interference c' is computed as

$$c' = \frac{< c, w_p >}{< w_p, w_p >} \tag{6.6}$$

where $<>$ is the inner product operation.

4) Finally, apply the inverse ridgelet transform to the FRIT coefficients in every selected region and obtain the watermarked image.

6.1.4 Watermark extraction

At the receiver side, the virtual left and right views are synthesized by the watermarked center view and corresponding per-pixel depth map. This is a blind watermarking algorithm since the watermark extraction does not need the original center view and the depth map. Strictly speaking, a blind watermarking scheme does not require any information for watermark extraction. But the community of watermarking usually thinks it is "blind" if only the security key is needed. The watermark data can be determined by checking the sign of the normalized inner product between the watermarked coefficients and the reference pattern w_p. The watermark extraction

process consists of the following steps:

1) Select the reference feature points of the watermarked center view (either the rendered left or right view) and find the feature regions where the watermark data has been embedded for watermark detection.

2) Apply ridgelet transform to the selected watermarked feature regions. The watermarked views are possibly distorted by various attacks n, where the ridgelet coefficients can be expressed as $c'_w = c_w + n$.

3) Perform the correlation operation by calculating the normalized inner product between the selected watermarked ridgelet coefficients c'_w and the reference pattern w_p as described below

$$\begin{aligned} \gamma &= \frac{< c'_w, w_p >}{< w_p, w_p >} = \frac{< c_w + n, w_p >}{< w_p, w_p >} \\ &= \frac{< c + (\alpha b - \lambda c')w_p + n, w_p >}{< w_p, w_p >} \\ &= \alpha b + (1 - \lambda)c' + n' \end{aligned} \tag{6.7}$$

4) The reference pattern w_p and the distortion n are statistically uncorrelated (w_p is generated by the PRN generator with a security key, while the distortion is introduced during transmission), thus $n' =< w_p, n > \approx 0$. Since $< c, w_p >$ is much smaller than $< w_p, w_p >$, $c' =< c, w_p > / < w_p, w_p >$ is very small, and the optimum value of λ is close to 1. In addition, α here is positive. Therefore, by checking the sign of γ, we can estimate the extracted watermark data b' [1]

[121],

$$b^{'} = \begin{cases} 1, & \text{if } \gamma > 0; \\ -1, & \text{otherwise.} \end{cases} \tag{6.8}$$

The normalized correlation (NC) is calculated between the estimated watermark $b_i^{'}$ and the embedded watermark b_i, and the NC here is used to detect a watermark. The watermark is successfully detected if the value of the NC is above a predefined threshold.

$$\text{NC} = \frac{\sum_i b_i \cdot b_i'}{\sqrt{\sum_i b_i^2 \cdot \sum_i b_i'^2}} \tag{6.9}$$

We use the bit error rate (BER) to measure the accuracy of the watermark extraction for the proposed algorithm. The BER of extracted watermark b' estimated by using the generated reference pattern is defined as follows:

$$\text{BER}(b, b') = \frac{\text{number of } (b_i^{'} \neq b_i)}{K} \tag{6.10}$$

where K is the number of the watermark data.

As discussed before, the selected feature points are very robust to the concerned attacks, including rotation, scaling and noise addition. So even when the watermarked image is distorted, with the strict rules in Subsection 6.1.1, the feature points used in watermark embedding can be detected accurately for watermark extraction. Table 6.1 lists some experimental results of reference feature points detection under various attacks. In the table, the numerator shows the number of the feature points or feature

regions from which the watermark was successfully detected, and the denominator shows the number of the feature points in which the watermark was actually embedded. It can be seen from this table that the proposed scheme can detect the reference feature points or feature regions with high accuracy under various attacks.

Table 6.1: The detection accuracy of reference feature points.

	Ballet	Art	Books	Teddy	Interview
Without attack	7/7	8/8	8/8	7/7	9/9
Rotation (5°)	7/7	8/8	8/8	7/7	9/9
Rotation (10°)	7/7	8/8	8/8	7/7	9/9
Rotation (15°)	7/7	7/8	8/8	7/7	8/9
Scaling (1.2)	7/7	8/8	8/8	7/7	9/9
Scaling (2.0)	6/7	8/8	8/8	6/7	9/9
Gaussian noise (Var. 0.2)	7/7	8/8	8/8	7/7	9/9
Gaussian noise (Var. 1.0)	6/7	8/8	7/8	7/7	9/9
JPEG compression (QF: 70)	7/7	8/8	8/8	7/7	9/9
JPEG compression (QF: 10)	7/7	8/8	8/8	6/7	8/9
JPEG 2000 compression (Ratio: 20)	7/7	8/8	8/8	7/7	9/9
JPEG 2000 compression (Ratio: 70)	7/7	7/8	8/8	7/7	8/9

6.1.5 Error probability

To compute the error probability for this proposed watermarking algorithm, the selected original coefficient c and the attack noise n are assumed to be modeled as uncorrelated Gaussian random processes.

$$c \sim N(0, \sigma_c^2),\ n \sim N(0, \sigma_n^2) \tag{6.11}$$

If the embedded watermark data $b = 1$, an error occurs when $\gamma < 0$. Similarly, if $b = -1$, an error occurs when $\gamma > 0$. Therefore, the error probability P_0 for a single

circular region can be calculated as

$$P_0 = w_1 Pr(\gamma < 0 | b = 1) + w_2 Pr(\gamma > 0 | b = -1) \tag{6.12}$$

where w_1 and w_2 are the prior probabilities of $b = 1$ and $b = -1$. These two probabilities are defined as $w_1 = w_2 = 1/2$ in this algorithm. In addition, these two cases have the same error probabilities [121]. So the error probability P_0 can be rewritten as

$$P_0 = Pr(\gamma < 0 | b = 1) = Pr(\gamma > 0 | b = -1) \tag{6.13}$$

Let us consider the case when $b = 1$,

$$\begin{aligned} P_0 &= Pr(\gamma < 0 | b = 1) \\ &= \frac{1}{2} erfc\left(\sqrt{\frac{L\sigma_w^2 - \lambda^2\sigma_c^2}{2(\sigma_n^2 + (1-\lambda)^2\sigma_c^2)}}\right) \\ &= \frac{1}{2} erfc\left(\sqrt{\frac{\frac{L\sigma_w^2}{\sigma_c^2} - \lambda^2}{2(\frac{\sigma_n^2}{\sigma_c^2} + (1-\lambda)^2)}}\right) \end{aligned} \tag{6.14}$$

where $erfc(\cdot)$ is the complementary error function, L and σ_w^2 are the length and the energy of the reference pattern w_p respectively. The mathematical deduction of error probability P_0 is provided in detail in [121]. Since the error occurrence of selected feature regions is mutually independent, the error probability of watermark extraction from these feature regions can be obtained. Therefore, the total error probability P_e

for the proposed algorithm can be expressed by

$$P_e = 1 - (1 - P_0)^W \tag{6.15}$$

where W is the number of selected feature regions. Fig. 6.4 presents the average error probability P_e with respect to the values of SNR σ_c^2/σ_n^2 and the relative power $r_p = L\sigma_w^2/\sigma_c^2$. As can be seen from this figure, the larger the SNR values and r_p, the smaller the error probability. This has been confirmed by the experiments. The PSNR values of the watermarked images are all above 40 dB, which will be shown in next section. Accordingly, the average values of calculated SNR and r_p are above 12.8 dB and 10.7. The error probability with respect to these values should be very low according to this figure.

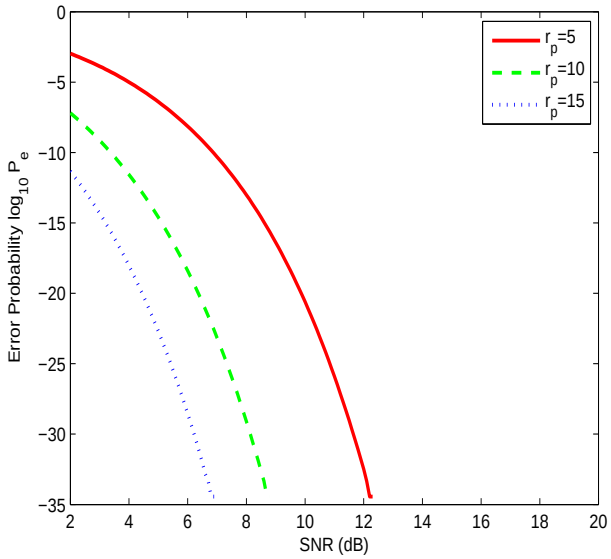


Figure 6.4: The error probability for different values of SNR and r_p.

6.2 Experimental results and analysis

As shown in Fig. 6.5, we have tested our proposed algorithm on 10 experimental pairs of the center images and the associated depth images from Middlebury Stereo Datasets [119] [120] Microsoft Research 3D Video Datasets [122] and HHI (Heinrich-Hertz Institut), Germany [20]. The resolutions of the center images and the depth images range from 450×375 to 1390×1110, and the depth maps are gray scale images of 8-bit level.

In this section, we conduct experiments to evaluate the performance of the proposed watermarking algorithm in terms of watermark invisibility and robustness to various attacks. In the implementation, the focal length $f = 1$ and the baseline distance t_x is set to be 5% of the image width for comfortable viewing. The farthest clipping plane $Z_{far} = t_x/2$ and the nearest clipping plane $Z_{near} = 1$. The watermark embedding strength $\alpha = 2$, the radius R of feature regions is set to 20 and the reduction ratio $\lambda = 1$.

6.2.1 Watermark invisibility

For watermark invisibility, we use both objective and subjective quality measures to test the fidelity of the watermarked images. In objective image quality evaluation, the peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) [66] are adopted. Table 6.2 gives the values of PSNR and SSIM of the watermarked center images for the 10 test image pairs. The mean values of PSNR and SSIM are 41.65 dB and 0.990 respectively. The watermarked image with PSNR around or above 40 dB or SSIM above 0.95 looks very close to the original one.

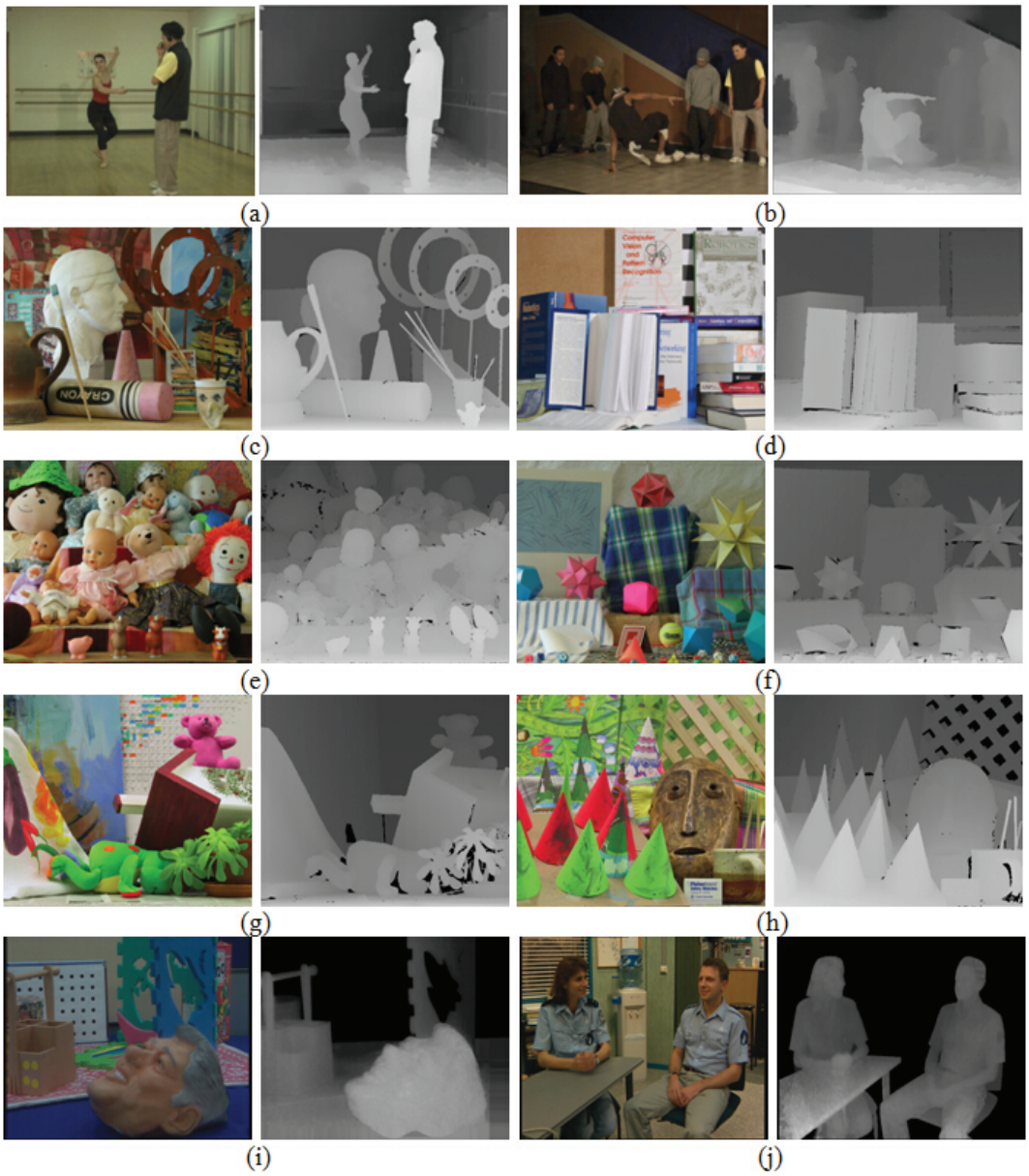


Figure 6.5: Test image pairs of center images and corresponding depth images. (a) Ballet (1024 × 768); (b) Breakdancers (1024 × 768); (c) Art (1390 × 1110); (d) Books (1390 × 1110); (e) Dolls (1390×1110); (f) Moebius (450×375); (g) Teddy (450×375); (h) Cones (720×576); (i) Orbi (720 × 576); (j) Interview (720 × 576).

In subjective image quality evaluation, as recommended by the ITU-R [67], we use the double stimulus continuous quality scale (DSCQS) for judging the imperceptibility

Table 6.2: Quality measurements of the watermarked center images for the proposed algorithm in terms of PSNR, SSIM and MOS.

Images	PSNR (dB)	SSIM	MOS
Ballet	41.15	0.991	4.5
Breakdancers	41.52	0.990	4.4
Art	42.68	0.995	4.7
Books	40.16	0.982	4.3
Dolls	42.04	0.993	4.6
Moebius	42.80	0.997	4.8
Teddy	41.53	0.991	4.6
Cones	40.95	0.985	4.4
Orbi	42.26	0.994	4.5
Interview	41.47	0.989	4.3

of the watermarked images. In this experiment, eight males and seven females (aged between 24 and 50) participated in the quality evaluation. Using the criteria of the mean opinion score (MOS), the participants rated the watermarked images with five-grade evaluation scale, where 5=Excellent, 4=Good, 3=Fair, 2=Poor and 1=Bad. As listed in Table 6.2, the values of MOS are all above 4.3 and the high scores reflect the good perceptual quality of the watermarked images. Therefore, the experimental results demonstrate that the proposed algorithm has good performance in terms of image fidelity under the objective and subjective quality measures.

6.2.2 Robustness to various attacks

To evaluate the performance of the proposed watermarking algorithm comprehensively, we have tested it on the robustness to various distortions [22] [40], including geometric transformations like rotation and scaling, Gaussian noise, JPEG compression, JPEG 2000 compression and combined attacks.

For each of the 10 test image pairs, the watermark is embedded into the center image, the virtual left and right images are rendered from the watermarked center image and the corresponding depth map at the receiver side. Table 6.3 provides the average normalized correlation (NC) values of watermark extraction from the the center view and the virtual left and right views respectively, where NC1 is the NC value from watermarked images and NC2 is the corresponding value from un-watermarked images. As can be observed from this table, there are big and clear separations between NC1 and NC2 under various attacks. So the watermark can be easily detected from the three views individually for content authentication. In the experiment, we evaluate the accuracy of watermark extraction from the three views with bit error rate (BER) when the watermarked center view is under the above mentioned attacks. Fig. 6.6 gives the results of BER from the three watermarked views and the corresponding average NC values.

Table 6.3: Average NC of the watermark extracted from the center view, the virtual left and right views.

	Center View		Left View		Right View	
	NC1	NC2	NC1	NC2	NC1	NC2
Without attack	0.952	0.086	0.937	0.109	0.925	0.097
Rotation (5°)	0.906	0.143	0.892	-0.092	0.879	0.051
Rotation (15°)	0.881	-0.072	0.853	0.104	0.826	-0.047
Scaling (0.5)	0.894	-0.016	0.860	-0.035	0.867	0.023
Scaling (1.5)	0.912	0.077	0.893	0.095	0.885	-0.040
Gaussian noise (Var. 0.2)	0.920	0.061	0.904	-0.064	0.896	-0.103
Gaussian noise (Var. 0.8)	0.898	-0.085	0.832	0.029	0.818	0.056
JPEG compression (QF: 10)	0.875	-0.110	0.819	-0.068	0.822	-0.037
JPEG compression (QF: 30)	0.904	0.096	0.867	-0.082	0.856	0.090
JPEG 2000 compression (Ratio: 20)	0.907	-0.026	0.861	-0.063	0.864	-0.095
JPEG 2000 compression (Ratio: 70)	0.882	0.021	0.855	-0.043	0.851	0.036

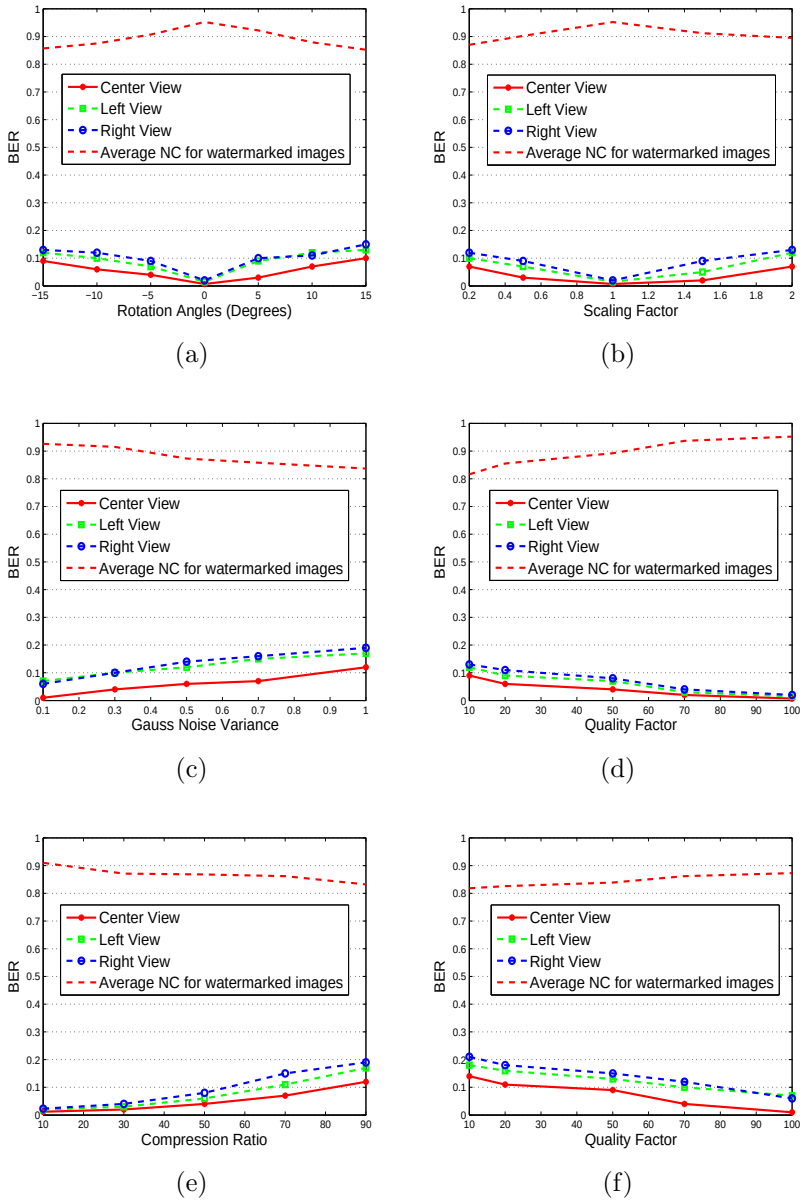


Figure 6.6: Average BER of the watermark extracted from the distorted center view and synthesized left and right views. (a) Rotation; (b) Scaling; (c) Gaussian noise; (d) JPEG compression; (e) JPEG 2000 compression; (f) Combined attacks.

Geometric attacks

Rotation and scaling are common geometric attacks. In this algorithm, as discussed in Section 6.1.1, the Gaussian scale model is used to select the geometrical transform

invariant feature points and the constructed feature regions are also rotation and scaling invariant. For rotation, the watermarked center images are rotated from 0° to 15° in clockwise and counterclockwise directions with a step of 5°. Fig. 6.6(a) presents the average BER of the watermark information extracted from the rotated watermarked center view and the synthesized left and right views. The horizontal axis indicates the rotation angles and the vertical axis denotes the values of BER. For scaling, the proposed watermarking algorithm is tested with scaling factors varying from 0.2 to 2.0. The corresponding results are illustrated in Fig. 6.6(b). The mean values of BER for rotation and scaling are 0.11 and 0.09, respectively. The experimental results show that the proposed algorithm is highly robust to rotation and scaling.

Gaussian noise

The center image may suffer from the channel noise during transmission. In this experiment, the proposed algorithm is evaluated under Gaussian noise addition. The virtual left and right images are rendered based on the Gaussian noised center image. As can be seen from Fig. 6.6(c), the watermarked center images are distorted by Gaussian noise with variance from 0.1 to 1.0. The mean value of BER of the watermark extracted from the center image and synthesized images is 0.14. Note that the image is seriously polluted when the variance reaches 1.0. Therefore, the low BER values demonstrate that the proposed algorithm works effectively against Gaussian noise.

JPEG compression

The performance of the proposed algorithm under JPEG compression is further investigated in this subsection. The watermarked images are attacked by JPEG compression with quality factors from 10 to 100. According to the experimental results from the Fig. 6.6(d), when the quality factor becomes smaller, the value of BER increases. However, all the values of BER are below 0.12, which show that the proposed 3D image watermarking algorithm performs well under JPEG compression.

JPEG 2000 compression

The performance of the proposed algorithm under JPEG 2000 compression is also tested in the experiment. The watermarked images are distorted by JPEG 2000 compression with compression ratios from 10 to 90. The higher compression ratio indicates more severe distortion to the image. As can be observed from Fig. 6.6(e), the value of BER is still acceptable even when the compression ratio reaches 90, which demonstrates that the proposed algorithm has good performance in terms of watermark extraction under JPEG 2000 compression.

Combined attacks

To test the performance of the proposed algorithm under combined attacks. The effects of both JPEG compression and Gaussian noise added on the center image are studied. The variance of Gaussian noise is 0.5 and the quality factor of JPEG compression varies from 10 to 100. Compared with one kind of attack in Fig. 6.6(c) and 6.6(d), there is only slightly increase of BER under combined attacks illustrated

in Fig. 6.6(f). The mean value of BER under combined attacks is 0.17, which implies that the proposed algorithm is relatively robust to combined attacks with JPEG compression and Gaussian noise.

6.2.3 Robustness to DIBR processing

In addition to the geometric attacks, the watermarking technique for DIBR 3D images should also resist DIBR processing. The depth map variation and baseline distance adjustment are common distortions in the DIBR system. The virtual left and right views are rendered based on possibly attacked depth map or different baseline distances. Moreover, although there are three major steps in a DIBR system, the detailed operations can be different for practical applications. However, the result of watermark extraction from the center view remains the same under depth map variation, baseline adjustment and different rendering conditions.

Baseline distance adjustment

In a DIBR system, the center view and the depth map are transmitted to the receiver side to synthesize the virtual left and right views. As discussed in Chapter 2, the user can adjust the viewing by modifying the baseline distance. But too large or small baseline distance will introduce uncomfortable viewing [20] [29]. In the experiment, we adjust the baseline distance from 1% to 10% of the center image width. Fig. 6.7 suggests that 5% of the image width is the optimum baseline distance and the bit error rates go up when it diverges to 1% and 10%. However, the BERs of watermark extracted from the watermarked center view and new generated left and right views keep relatively low under different baseline distances.

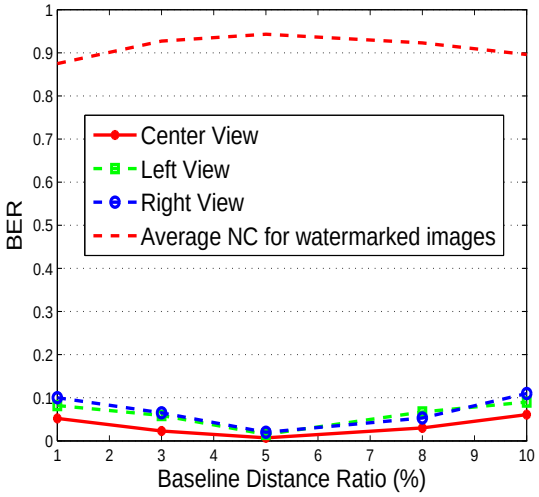


Figure 6.7: Average BER of the extracted watermark under baseline distance adjusting.

Depth map variation

Similar to the center view, the depth map may be distorted by different attacks during the transmission to the consumer side. In this experiment, the depth maps of the 10 test image pairs are distorted under Gaussian noise with variance 0.1, JPEG compression with quality factor 10 and JPEG 2000 compression with compression ratio 50, respectively. Fig. 6.8 gives the original depth map of image *Breakdancers* and the distorted depth images. The average bit error rates of extracted watermark from the synthesized left and right views are given in Table 6.4. The values of BER are all below 0.12, which show that the proposed algorithm is robust to depth map variation. This simulation result also demonstrates that the depth map can be compressed before transmission, which is an advantage of the DIBR system over other 3D display techniques.

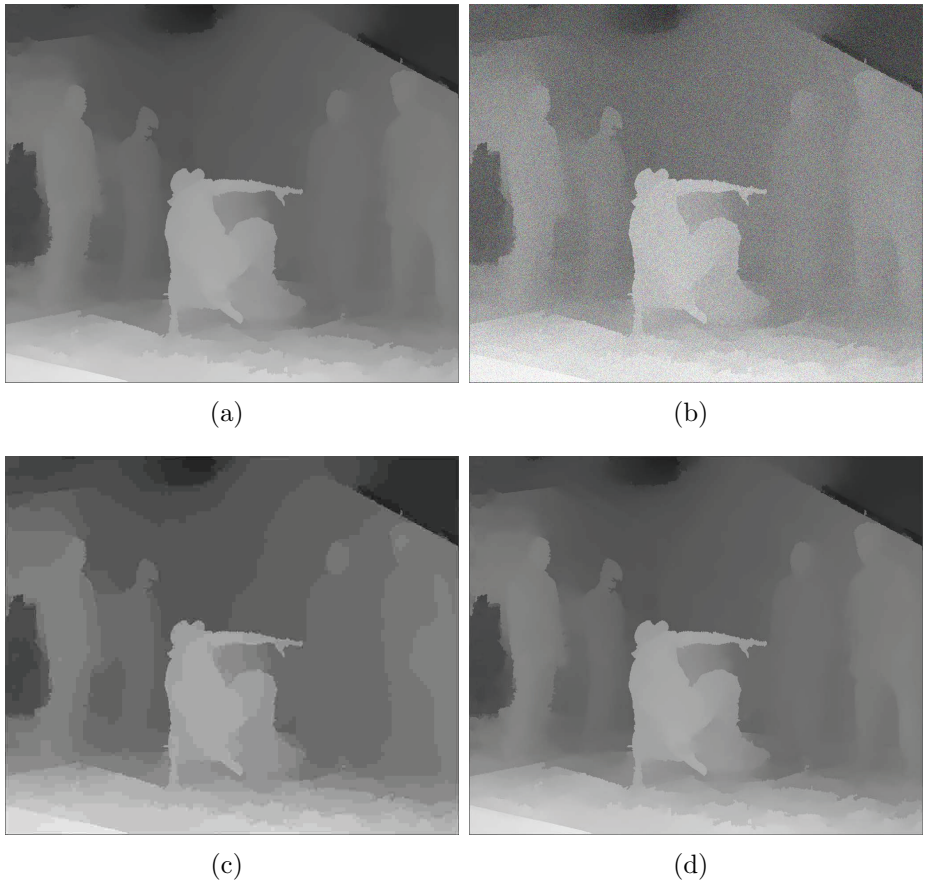


Figure 6.8: Depth map pre-processing. (a) Original depth map of *Breakdancers*; (b) Depth map with Gaussian noise (Variance: 0.1); (c) Depth map with JPEG compression (Quality factor: 10); (d) Depth map with JPEG 2000 compression (Compression ratio: 50).

Table 6.4: Average BER of the extracted watermark for depth map variation and different DIBR methods.

	Gaussian Noise	JPEG	JPEG 2000	Method [29]	Method [51]
Virtual left view	0.08	0.10	0.06	0.05	0.03
Virtual right view	0.11	0.12	0.08	0.06	0.05

Different DIBR methods

In the experiment, we have tested our watermarking algorithm under different rendering conditions [29] [51]. In [29], the depth map is pre-processed with an asymmetric

filter to smoothen the sharp changes around object boundaries. In [51], the edge dependent depth filter and interpolation are used in depth processing and hole-filling, respectively. Table 6.4 also presents the BER of extracted watermark data from the virtual views for the two DIBR methods. The values of BER are all below 0.06, which show that the proposed watermarking algorithm is highly robust to different rendering processes.

6.2.4 Performance comparison

In this subsection, our proposed algorithm is compared with other related watermarking algorithms proposed by Lin *et al.* [1], Kim *et al.* [2] and Chen *et al.* [10]. Method [1] and method [2] are chosen for performance comparison because they are the typical and state-of-the-art techniques for DIBR 3D watermarking. Method [10] is our proposed watermarking scheme in Chapter 5. Table 6.5 presents the experimental results of performance comparison in terms of watermark invisibility and robustness against various attacks. The accuracy of watermark detection is measured using the average BER value of watermark extracted from the watermarked center view and the synthesized views.

For watermark invisibility, compared with other existing methods, our algorithm has the highest values of PSNR and MOS, and there are no significant differences in SSIM among the four methods. For the watermark extraction without any distortions, the average BER of our method is much lower, only 0.008, which demonstrates that the proposed algorithm can detect the watermark with higher accuracy. The performance comparison on robustness to various attacks is also tested. Compared with other state-of-the-art methods [1] [2] and our previous method [10], the proposed

Table 6.5: Performance comparison with other related methods.

	Method [1]	Method [2]	Method [10]	Proposed method
PSNR (dB)	41.52	41.07	41.16	**41.65**
SSIM	**0.991**	0.985	0.987	0.990
MOS	4.4	4.3	4.4	**4.5**
BER without attacks	0.071	0.025	0.017	**0.008**
BER under attacks				
Rotation (10°)	0.43	0.31	0.16	**0.12**
Scaling (0.5)	0.35	0.19	0.15	**0.08**
Gaussian noise (Var. 0.5)	0.15	0.12	0.14	**0.10**
JPEG compression (QF: 10)	0.38	0.33	0.21	**0.11**
Combined attacks	0.47	0.41	0.26	**0.17**
Baseline distance ratio (10%)	0.29	0.15	0.12	**0.09**
Depth map with Gaussian noise (Var. 0.1)	0.24	0.10	**0.07**	**0.07**
Depth map with JPEG compression (QF: 10)	0.18	0.12	0.09	**0.08**

algorithm has the lowest values of BER and exhibits better performance of robustness against rotation, scaling, Gaussian noise, JPEG compression and combined attacks. Take the rotation for an instance, when the watermarked center image is rotated by 10°, the average BER values of the extracted watermarks from the three views in method [1], method [2] and method [10] are 0.43, 0.31 and 0.16 respectively. The corresponding BER value for this proposed algorithm is 0.12, which shows better robustness against rotation and higher accuracy of watermark extraction. Furthermore, the BERs of this algorithm are lower than those of the methods [1] [2] and comparable to those of the method [10] in terms of depth map variation and baseline distance adjustment. On the whole, as can be seen from this table, the proposed algorithm has superiority over other related and state-of-the-art methods on the robustness against image processing attacks and common DIBR processing.

6.2.5 Experimental results on other image pairs

In this subsection, in order to demonstrate that the effectiveness of the proposed algorithm is not limited, we have applied our algorithm to all the image pairs in the three publicly available databases for DIBR 3D images. Table 6.6 presents the performance of the proposed algorithm in terms of watermark invisibility and robustness against various attacks. It can be observed from this table that the average values of PSNR and SSIM are relatively high and the average BERs are quite low. The experimental results are similar to those for the 10 test image pairs in the previous experiments. This fact shows that the proposed algorithm is effective in watermark extraction, and the effectiveness is independent of specific image pairs in DIBR 3D image databases.

Table 6.6: The watermark extraction accuracy on all the image pairs of the three databases.

	Proposed method
PSNR (dB)	41.61
SSIM	0.992
BER without attacks	0.009
BER under attacks	
Rotation (10°)	0.12
Scaling (0.5)	0.10
Gaussian noise (Var. 0.5)	0.09
JPEG compression (QF: 10)	0.13
JPEG 2000 compression (Ratio: 50)	0.08
Combined attacks	0.16
Baseline distance ratio (10%)	0.10
Depth map with Gaussian noise (Var. 0.1)	0.06
Depth map with JPEG compression (QF: 10)	0.09
Depth map with JPEG 2000 compression (Ratio: 50)	0.05

6.3 Summary

In this chapter, we proposed a blind DIBR 3D image watermarking algorithm in the ridgelet domain using feature regions for copyright protection. In the proposed algorithm, the reference feature points are selected with Gaussian scale model and the disk feature regions are constructed for watermark embedding and extraction. The watermark data is embedded into the selected ridgelet coefficients of the selected feature regions. The simulation results show that the proposed algorithm exhibits good performance in terms of both watermark invisibility and robustness to image processing attacks and DIBR processing. Moreover, our algorithm has lower values of BER and higher accuracy of watermark extraction than other related DIBR 3D image watermarking methods.

For future work, there are several issues to be further studied. The image modeling could be well exploited with possible distribution such as generalized Gaussian distribution or Markov random field. The watermarking technique for multi-view synthesis or free-view synthesis is worthy of further investigation. We will continue to conduct research on 3D image watermarking and try to extend our algorithm to DIBR 3D video watermarking.

6.4 Some discussions

In this section, we give some discussions on the proposed watermarking algorithms in Chapter 4, Chapter 5 and Chapter 6. In Chapter 4, a contourlet-based 2D image watermarking algorithm is presented, and it is extended to video watermarking. In Chapter 5, a contourlet-based watermarking algorithm for DIBR 3D images is de-

veloped. In Chapter 6, a novel digital watermarking algorithm for DIBR 3D images using feature regions is proposed. There are some relationships among the algorithms in these three chapters:

1. The three chapters are all about the application of digital watermarking on copyright protection. The algorithm in Chapter 4 is proposed for 2D digital watermarking, while the algorithms in Chapter 5 and Chapter 6 are proposed for DIBR 3D watermarking.

2. The algorithm in Chapter 4 is based on contourlet transform and principal component analysis. Considering the specific requirements of DIBR 3D image watermarking, the algorithm in Chapter 5 is developed based on contourlet transform. However, the performance on watermark extraction needs to be further improved. The algorithm in Chapter 6 is proposed to tackle the limitation of extraction accuracy. The experimental results demonstrate its superiority over other state-of-the-art methods, including the algorithm in Chapter 5.

3. The algorithms in Chapter 5 and Chapter 6 have their own merits. One of the most important advantages for the former is its simplicity, practicality and easy implementation. The latter is more accurate in watermark extraction and has better robustness against attacks. Therefore, they can be used for different scenarios and applications.

Chapter 7

Watermarking based quality evaluation for DIBR 3D images

In this chapter, a novel quality evaluation scheme based on blind watermarking for DIBR 3D images is devised. The scheme utilizes the watermarking technique to evaluate the quality of watermarked images under various distortions. In this scheme, the watermark is embedded into the selected coefficients of dual-tree complex wavelet transform (DT-CWT) sub-bands of the center view. The virtual left and right views are synthesized from the watermarked center view and the associated depth map using the DIBR technique at the receiver side. The watermark can be detected from the three views individually, and the quality of these views under various attacks can be estimated by examining the degradation of corresponding extracted watermarks. In addition, the quality evaluation is made possible by checking the generated mapping curve, which maps the normalized correlation (NC) of the extracted watermark to the quality measure of the watermarked image under distortions.

Dual-tree complex wavelet transform is suitable for watermarking based quality evaluation. First, it has some subbands with more information on horizontal direction, which has robustness to view synthesis process. More importantly, the DT-CWT has certain sensitivities to distortions: the changes of the coefficients can reflect the strength of the distortion, which is good for quality evaluation. If a transform is too robust to attacks, the coefficients may not change when the watermarked image is under a small distortion. In this case, the degradation of extracted watermark would not accurately reflect the quality changes of the watermarked image under distortions. In the following, the watermarking scheme is described in detail and the corresponding simulation results and analysis are presented.

7.1 The proposed watermarking based quality evaluation scheme

The proposed watermarking based image quality evaluation scheme is illustrated in Fig. 7.1. The watermark embedding and extraction are performed in the dual-tree complex wavelet transform (DT-CWT) domain [123] [124]. The DT-CWT has desirable properties of shift invariance and perfect reconstruction, which can be used for image watermarking. Moreover, there are some DT-CWT subbands with more information on horizontal direction, which has robustness to the DIBR synthesis process [2] [77]. In addition, the DT-CWT has directional selectivity and the changes of the coefficients can reflect the strength of the distortion, which are good for quality evaluation.

Considering the trade-off between imperceptibility and robustness, we use three-

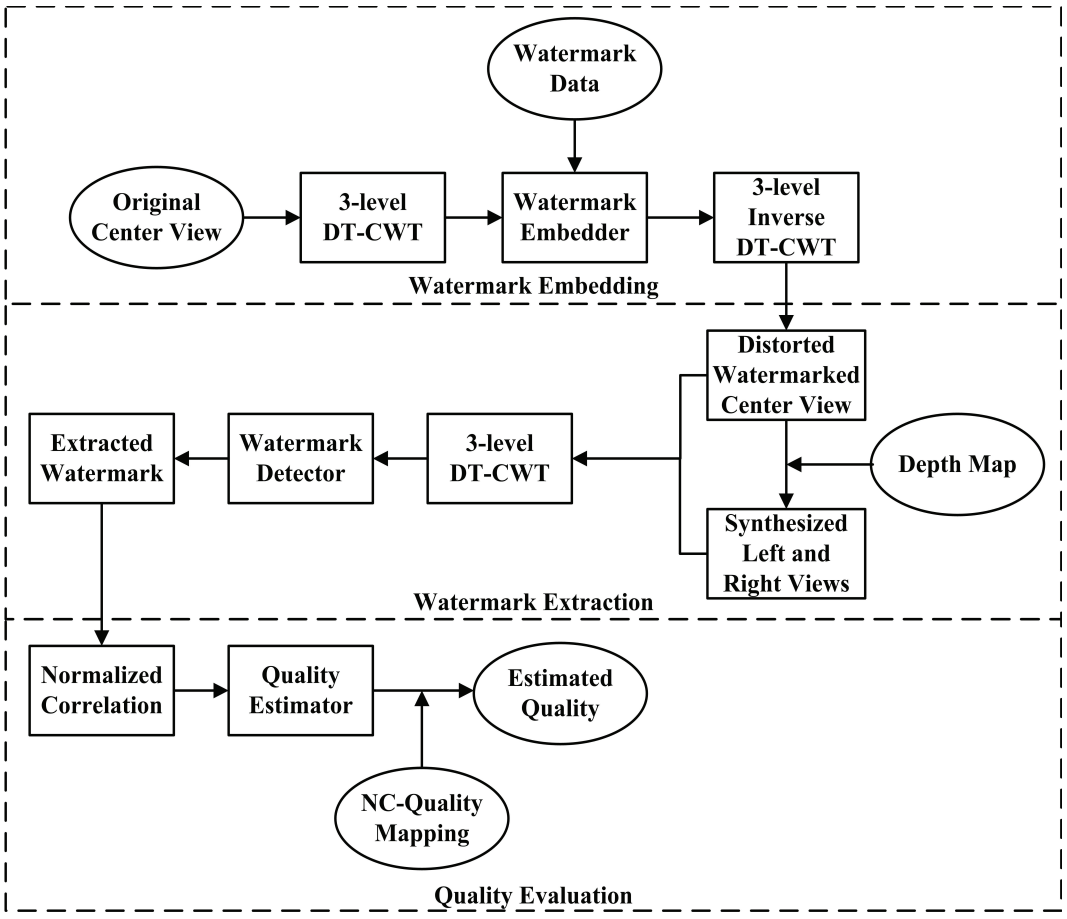


Figure 7.1: The proposed watermarking based image quality evaluation scheme.

level decomposition of DT-CWT. Fig. 7.2 presents the six directional subbands $C_{lv,dr}$ in each level oriented at angles $+15°$, $+45°$, $+75°$, $-75°$, $-45°$ and $-15°$, where the level index $lv = 1, 2, 3$, the direction index $dr = 1, 2, 3, 4, 5, 6$. Subbands $C_{lv,1}$ and subbands $C_{lv,6}$ have more information on horizontal direction, while subbands $C_{lv,3}$ and subbands $C_{lv,4}$ contain more vertical edges than other subbands. Since the DIBR process maps pixels of the center image horizontally [29] [32], if we embed the watermark data in the subbands with more vertical information, the watermark data could be easily destroyed during the DIBR process. So the watermark is embedded

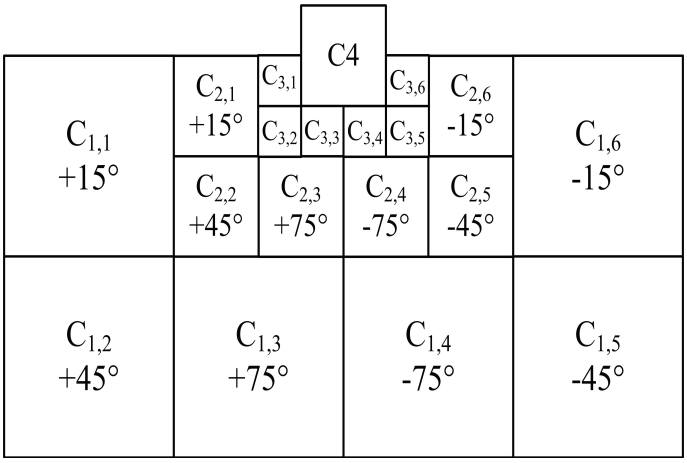


Figure 7.2: Structure of DT-CWT coefficients.

in the subbands $C_{lv,1}, C_{lv,2}, C_{lv,5}$ and $C_{lv,6}$. Meanwhile, because the human visual system (HVS) is more sensitive to the changes of low frequencies, in order to keep the image fidelity and watermark invisibility, we use the level 2 and level 3 of DT-CWT subbands for watermark embedding [125]. Therefore, there are 8 decomposed DT-CWT subbands used for watermarking based quality evaluation in this scheme.

Here we provide some analysis on subbands selection. 20 experimental images are randomly selected from the databases in the experiment and tested for watermark invisibility. Table 7.1 gives the average PSNR of the center images and the synthesized left and right images for different watermarked directional subbands. As can be observed from this table, although there are no big differences in the PSNRs of center images, the qualities of synthesized images are significantly different among different quantized subbands. The PSNR values of subbands $C_{lv,3}$ and $C_{lv,4}$ $(lv = 2, 3)$ are much lower than that of other directional subbands, which demonstrate that these four subbands are not suitable for watermark embedding.

In the following subsections, we will present the proposed watermarking based

Table 7.1: Average PSNR (dB) of the center images and the synthesized left and right images for different watermarked subbands.

Quantized subbands	Center images	Synthesized images
$(C_{lv,1}, C_{lv,6})$	41.31	39.95
$(C_{lv,2}, C_{lv,5})$	42.07	40.80
$(C_{lv,3}, C_{lv,4})$	40.62	35.16

image quality evaluation scheme, which consists of watermark embedding, watermark extraction and quality evaluation.

7.1.1 Watermark embedding

Different images have different frequency distributions. The watermark embedding process is implemented in the DT-CWT domain, because the DT-CWT can decompose an image into different frequency subbands, and different frequency components have different sensitivities to the distortions, so the watermark can be embedded in a proper way to reflect the quality changes of the distorted watermarked images. The watermark embedding process is described in detail in this subsection.

The center image I is divided into $M \times N$ subblocks. The size of each block is $(I_w/M) \times (I_h/N)$, where I_w and I_h are the width and height of the center image. Apply DT-CWT transform to each subblock. We select the coefficients of the DT-CWT subbands $C_{lv,dr}$ for watermark embedding, where lv and dr are the level of decomposition and the direction of the subbands, respectively. In this scheme, based on previous discussion, $lv = 2, 3$ and $dr = 1, 2, 5, 6$.

We embed the watermark data into the coefficients of the selected DT-CWT subbands which represent the energetic direction. For the DT-CWT transform, most of the energy of an image is concentrated in the energetic coefficients and the changes

of these coefficients can reflect the strength of the attacks, which contributes to the accuracy of quality evaluation. The energetic direction for each chosen subband can be found using the following formula:

$$j_e = \max_j \left\{ \frac{1}{s} \sum_{k=0}^{s-1} (C_{lv,dr}^{m,n}(j,k))^2 \right\} \tag{7.1}$$

where $C_{lv,dr}^{m,n}(j,k)$ is the coefficient magnitude of the selected DT-CWT subbands. m and n are the row and column indexes of the divided subblock, $m = 1, 2, \cdots, M$, $n = 1, 2, \cdots, N$; j and k are the row and column indexes of the DT-CWT subband coefficient of each subblock, $j, k = 0, 1, \cdots$; s is the number of selected DT-CWT coefficients. The DT-CWT coefficients $C_{lv,dr}^{m,n}(j_e,k)$ are the energetic coefficients in the selected subband.

In this scheme, we use the linear approximation of the improved spread spectrum (ISS) technique [121] to embed the watermark data $b = (b_1, b_2, \cdots)$ into the selected DT-CWT coefficients. The watermark data b consisting of 1 and -1 are embedded by modifying the amplitude of the selected coefficients as follows:

$$c_w = c + (\alpha b - \lambda c')w_p \tag{7.2}$$

where $c = [C_{lv,dr}^{m,n}(j_e,0), C_{lv,dr}^{m,n}(j_e,1), \cdots]$, c_w and c are the DT-CWT coefficients after and before watermark embedding respectively; α is used to control the embedding strength of the watermark data b; λ represents the reduction ratio of the interference of c'; the reference pattern w_p is a "chip sequence" produced by the pseudo random

number (PRN) generator; c' can be generated by c and computed as

$$c' = \frac{< c, w_p >}{< w_p, w_p >} \tag{7.3}$$

where <> is the inner product operation.

Finally, the watermarked center image I' can be reconstructed from the watermarked and un-watermarked DT-CWT subbands by inverse DT-CWT transform. Together with the corresponding depth map, the watermarked center image is transmitted to the receiver side to generate the watermarked left and right images.

Fig. 7.3 gives the visual experimental results of the watermarked center image and the synthesized left and right images for *Ballet*. The peak signal-to-noise ratio (PSNR) values of these images are all above 40 dB. It can be seen that the proposed scheme preserves good perceptual quality of the watermarked images.

7.1.2 Watermark extraction

Although the watermark data is embedded into the center view, there is watermark information in the virtual left and right views after the DIBR process. This is a blind watermarking scheme since the extraction process does not need the original center view and the depth map. Strictly speaking, a blind watermarking scheme does not require any information for watermark extraction. But the community of watermarking usually thinks it is "blind" if only the security key is needed. The watermark extraction is the inverse procedure of watermark embedding.

Apply DT-CWT transform to the watermarked center view I' (or the virtual left and right views) with the decomposition of level 3. Select the same DT-CWT

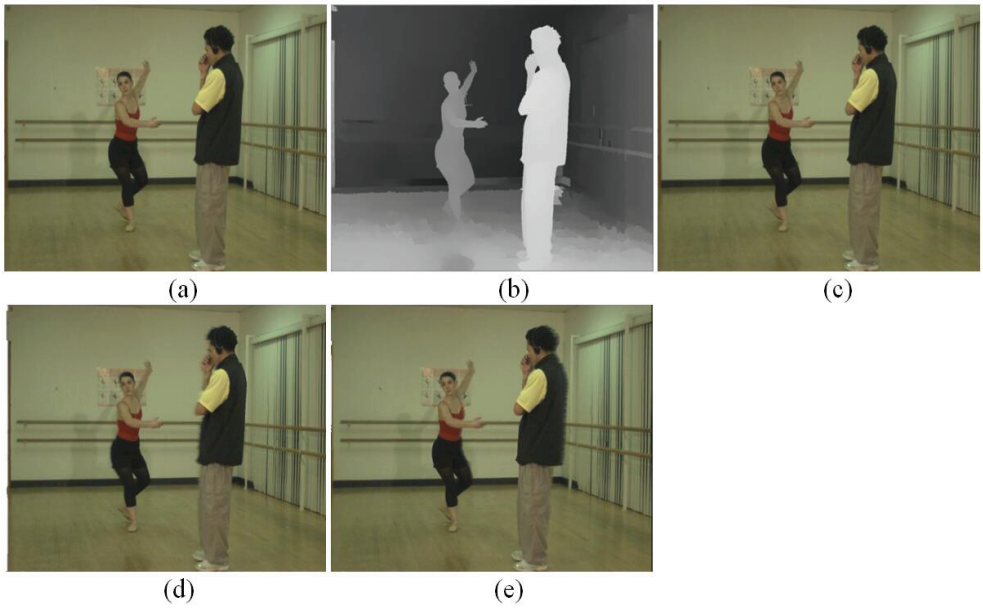


Figure 7.3: Visual results of image *Ballet* after watermark embedding. (a) The original center view; (b) The depth map; (c) The watermarked center view (PSNR=42.52 dB); (d) The synthesized left view (PSNR=41.73 dB); (e) The synthesized right view (PSNR=41.16 dB).

subbands in the level 2 and level 3 as the embedding process for watermark extraction. The watermarked views are possibly distorted by various attacks n, where the DT-CWT coefficients can be expressed as $c'_w = c_w + n$. Perform the correlation operation by calculating the normalized inner product between the selected watermarked DT-CWT coefficients c'_w and the reference pattern w_p as described below.

$$\begin{aligned}\gamma &= \frac{< c'_w, w_p >}{< w_p, w_p >} = \frac{< c_w + n, w_p >}{< w_p, w_p >} \\ &= \frac{< c + (\alpha b - \lambda c')w_p + n, w_p >}{< w_p, w_p >} \\ &= \alpha b + (1 - \lambda)c' + n'\end{aligned} \tag{7.4}$$

The reference pattern w_p and the distortion n are statistically uncorrelated, thus

$n' =< w_p, n > \approx 0$. Since $< c, w_p >$ is much smaller than $< w_p, w_p >$, $c' =< c, w_p > / < w_p, w_p >$ is very small, and the optimum value of λ is close to 1. In addition, α here is positive. Therefore, by checking the sign of γ, we can estimate the extracted watermark data b' [121],

$$b^{'} = \begin{cases} 1, & \text{if } \gamma > 0; \\ -1, & \text{otherwise.} \end{cases} \tag{7.5}$$

7.1.3 Quality evaluation

When the watermarked image is distorted, the image quality will be degraded. Since the watermark is embedded into the cover image and they suffer the same attack during transmission, the quality of distorted watermarked image can be evaluated by examining the degradation of extracted watermark [39]. The watermark degradation can be measured using the correlation between the distorted watermark and the original watermark. Here, the normalized correlation (NC) is used to indicate the degradation of extracted watermark.

$$\text{NC} = \frac{\sum_i b_i \cdot b'_i}{\sqrt{\sum_i b_i^2 \cdot \sum_i b_i'^2}} \tag{7.6}$$

where b'_i and b_i are the extracted watermark and the original reference watermark, respectively.

The quality of possibly distorted watermarked image can be measured using various full-reference quality metrics, such as PSNR, structural similarity (SSIM) [66] and 3D quality metrics. The NC value of extracted watermark can be used to reflect the

quality of watermarked image. With more severe degradation of watermarked image, the NC value decreases monotonically according to the experimental results in the following sections. Therefore, it is feasible to evaluate the quality of distorted watermarked image using a respective mapping between the NC values and the quality metrics.

$$\hat{Q} = f(\mathrm{NC}) \tag{7.7}$$

where $\hat{Q}$ is the estimated quality of the watermarked image, $f(\bullet)$ is the mapping function and it is represented by the mapping curve in the proposed algorithm. The generation of mapping curve will be presented in the next section.

7.2 Experimental results and analysis

In this section, we conduct experiments to test the performance of the proposed scheme. The Middle Stereo Datasets [20], HHI (Heinrich-Hertz Institut), Germany [120] and LIVE 3D Image Quality Database [88] [89] [126] are used to verify the performance of our proposed watermarking based quality evaluation scheme. In the experiment, the baseline distance t_x is set to 5% of the image width for comfortable viewing. We set the focal length $f = 1, Z_{far} = t_x/2$ and $Z_{near} = 1$. For the watermark embedding and extraction, we set the block indexes $M = N = 8$, the embedding strength $\alpha = 2$ and the reduction ratio $\lambda = 1$. For the quality evaluation, without loss of generality, 30 image pairs are randomly selected from these databases to generate the mapping curves. Another set of 50 image pairs is randomly chosen to test the accuracy of quality evaluation.

The proposed scheme is evaluated under JPEG compression, JPEG 2000 compression, Gaussian noise and Gaussian blur, which are typical and commonly used distortions for quality evaluation. For each of the four different kinds of distortions, the watermarked image is distorted with various strengths listed in Table 7.2. The qualities of watermarked center images are evaluated in terms of the 2D quality metrics including PSNR and SSIM. The qualities of virtual left and right images are assessed in terms of the 2D quality metrics and some 3D quality metrics [87] [88] [90]. Since the left and right images are symmetrically distorted in the experiment, the quality of a stereo-pair can be calculated using the average quality of the left and right images. In this scheme, the watermark itself degrades the images and the effect of this degradation is also considered in the quality evaluation.

The proposed scheme is evaluated under JPEG compression, JPEG 2000 compression, Gaussian noise and Gaussian blur. The watermarked image is attacked with various strengths listed in Table 7.2. The qualities of watermarked center images are evaluated in terms of the 2D quality metrics including PSNR and SSIM. The qualities of virtual left and right images are assessed using the 2D quality metrics and some 3D quality metrics. Since the left and right images are symmetrically distorted in the experiment, the quality of a stereo-pair is taken to be the average quality of the left and right images. In this thesis, the watermark itself degrades the images and the effect of this degradation in the process is also considered in the performance evaluation.

Table 7.2: Value ranges of the control parameters for distortion simulation.

Distortion	Control parameter	Range
JPEG	Quality factor	[100:-5:10]
JPEG 2000	Compression rate	[1:-0.05:0.01]
Gaussian noise	Noise standard deviation	[0:0.5:10]
Gaussian blur	Variance of Gaussian	[0:1:20]

7.2.1 NC-Quality mapping curve

The NC-Quality mapping curve is used to reflect the relationship between the NC value of extracted watermark and the quality of distorted watermarked image. The ideal mapping curve [39] is generated by averaging a set of NC-Quality curves and one ideal mapping curve needs to be obtained in terms of one quality metric under one kind of distortion.

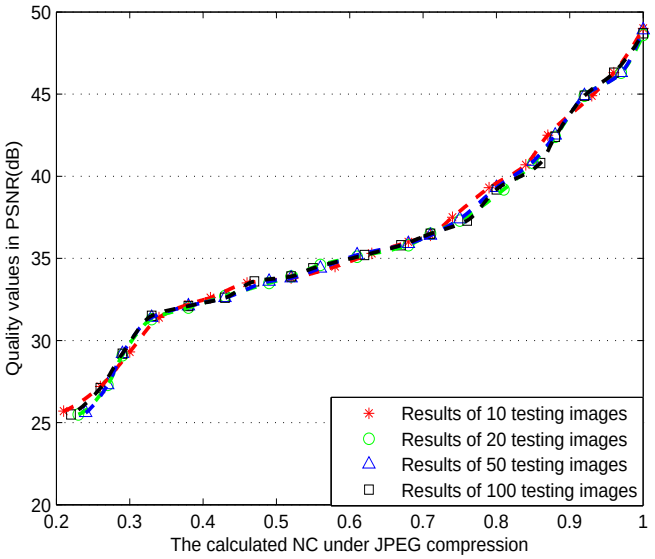


Figure 7.4: Ideal mapping curves under JPEG compression for different number of images.

Fig. 7.4 gives the ideal mapping curves obtained respectively using 10, 20, 50

and 100 different watermarked center images in terms of PSNR under JPEG compression respectively. Each watermarked center image is compressed with JPEG quality factors from 100 to 10 with a step size of -5. For each point, it represents a correspondence between the NC value of extracted watermark and the calculated PSNR of watermarked image. Then the ideal mapping curve is obtained by locally averaging the NC-Quality mapping curves [39]. Take the ideal mapping curve obtained from 20 images as an example. The PSNR axis is divided into small intervals $[max(0, i-0.5), i+0.5)]$ $(i = 0, 1, 2, 3 \cdot \cdot\cdot)$. There are several points whose PSNR values fall into the same one interval on the PSNR axis. Then the mean PSNR value of these points is computed for interval i and defined as $\overline{PSNR_i}$. So the corresponding pair $(i, \overline{PSNR_i})$ represents one point on the ideal mapping curve. After all the points (one for each interval) on the mapping curve are determined, the ideal mapping curve is obtained. As can be seen from Fig. 7.4, the four generated ideal mapping curves are almost identical. Our experimental results show that there is no significant difference with more images. In this scheme, the NC-Quality mapping curves are generated using the same set of 30 image pairs. The mapping curves under other distortions can be obtained with similar process. Once the mapping curve is generated, it is stored at the receiver side and no transmission is needed.

In this experiment, we evaluate the quality of the watermarked center images in terms of PSNR and SSIM under the aforementioned four types of distortions. Fig. 7.5 shows the generated mapping curves of the watermarked images under Gaussian noise. Fig. 7.5(a) and 7.5(b) are the mapping curves in terms of PSNR and SSIM for the center images. As can be observed from the figure, when the NC value of extracted watermark changes from 0.52 to 0.61, the calculated PSNR increases from

29.7 dB to 38.6 dB, and the corresponding calculated SSIM increases from 0.85 to 0.91. In addition, it can be noticed that with the increasing distortion strength, the NC values of extracted watermark and corresponding calculated quality metrics decrease monotonically, which also demonstrates that the quality of watermarked image can be estimated using the extracted watermark.

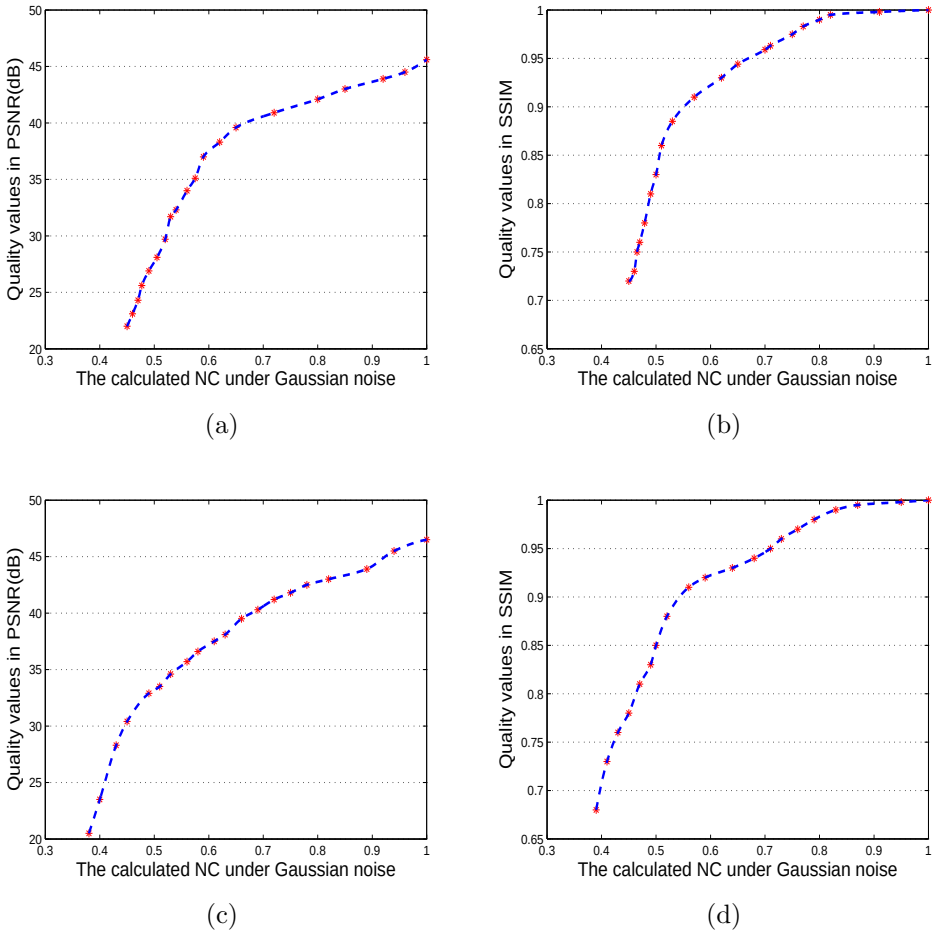


Figure 7.5: The generated mapping curves for the watermarked images in terms of PSNR and SSIM under Gaussian noise. (a) NC-PSNR mapping curve of the center images; (b) NC-SSIM mapping curve of the center images; (c) NC-PSNR mapping curve of the synthesized images; (d) NC-SSIM mapping curve of the synthesized images.

For the synthesized left and right images, we evaluate them in terms of both the 2D quality metrics and 3D quality metrics under the four kinds of distortions, respectively. Since the left and right images are symmetrically distorted in this scheme, the NC and the calculated PSNR and SSIM for the stereo-pairs are taken to be the average values from those of the left and right views. Fig. 7.5(c) and 7.5(d) give the generated mapping curves for the synthesized left and right images in terms of PSNR and SSIM under Gaussian noise. Compared with the mapping curves of the watermarked center images in Fig. 7.5(a) and 7.5(b), it can be stated that the NC value of extracted watermark is slightly lower when they have the same values of PSNR or SSIM. Furthermore, we adopt the 3D quality metrics for the quality evaluation of the synthesized images. The 3D quality metrics proposed by You [87], Chen [88] and Wang [90] are the typical and state-of-the-art quality evaluation methods for stereoscopic 3D images. Fig. 7.6 presents the mapping curves between the NC values of extracted watermark and the quality scores of synthesized images in terms of these 3D quality metrics. It can be seen that with the increasing NC value, the 3D quality score for the stereoscopic images increases monotonically. Based on this property, the quality of 3D images can be predicted using these mapping curves in terms of the 3D quality metrics.

7.2.2 Accuracy of quality evaluation

Once the mapping curve is generated, the quality of the watermarked image can be evaluated by the NC value of the extracted watermark. In this section, a set of 50 image pairs is used to test the accuracy of quality evaluation for the proposed scheme. The correlations between the calculated qualities and the predicted qualities

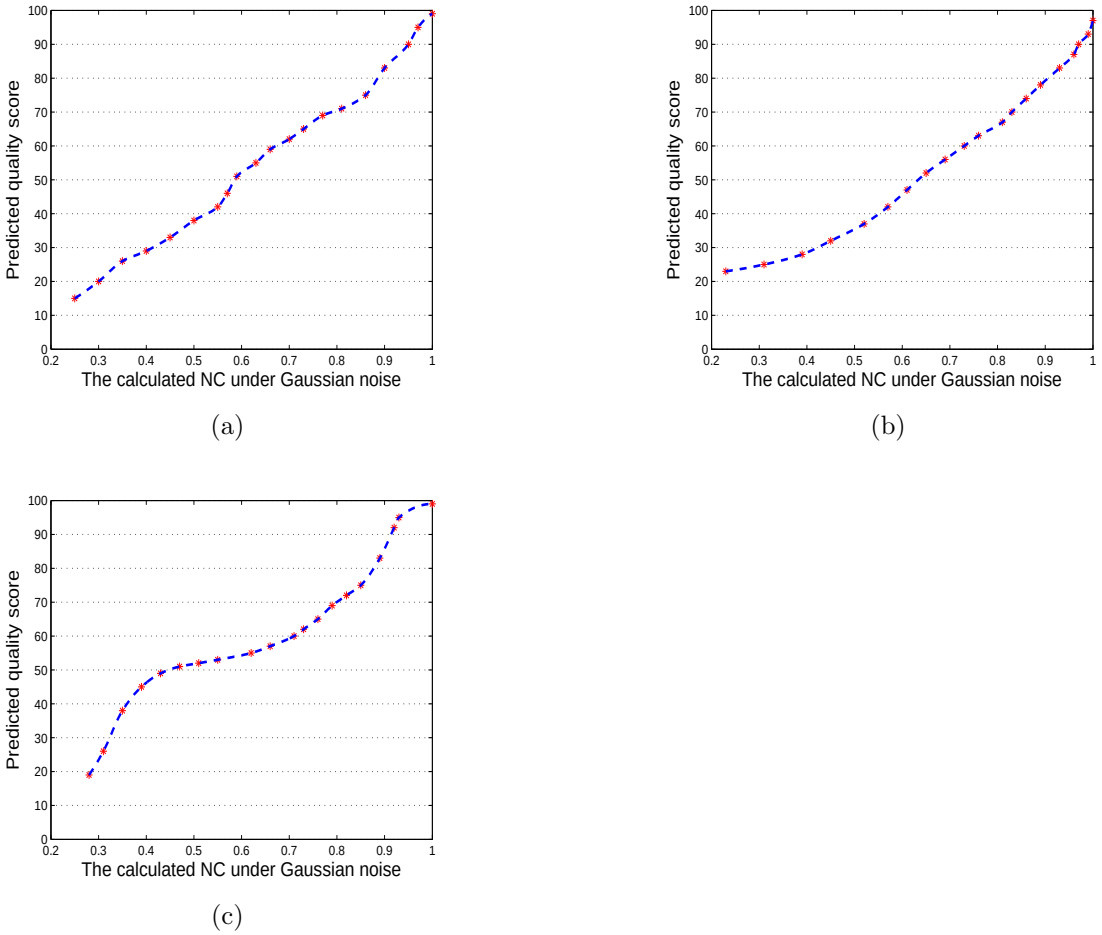


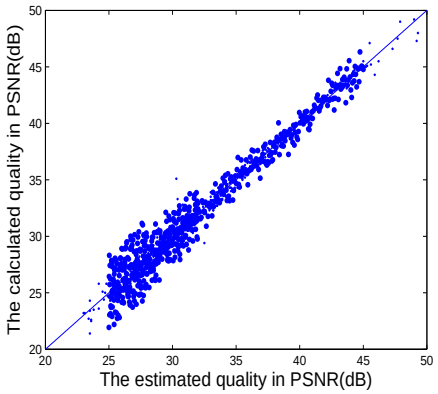
Figure 7.6: The generated mapping curves for the synthesized images in terms of 3D image quality metrics [87] [88] [90] under Gaussian noise. (a) The mapping curve in terms of the 3D quality metric proposed in [87]; (b) The mapping curve in terms of the 3D quality metric proposed in [88]; (c) The mapping curve in terms of the 3D quality metric proposed in [90].

are computed for evaluation accuracy. Fig. 7.7 gives the experimental results of quality evaluation for the center view and synthesized views in terms of the 2D quality metrics and 3D quality metrics under Gaussian noise. As can be seen from this figure, the horizontal axis and the vertical axis denote the predicted quality and the corresponding calculated quality, respectively. The solid line with a 45-degree angle

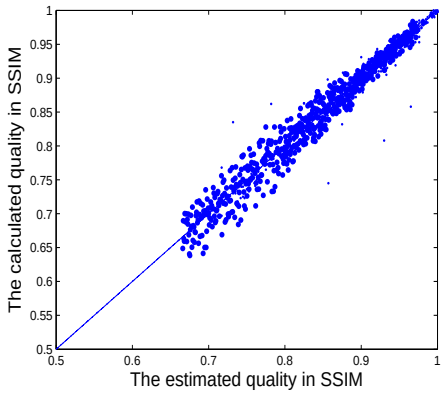
indicates that the estimated quality is equal to the calculated quality. The closer the scattered points to the solid line, the more accurate the quality evaluation. Fig. 7.7(a) and 7.7(b) are the quality evaluation results for the center view in terms of PSNR and SSIM, which show high correlations between the calculated qualities and the predicted qualities. However, according to Fig. 7.7(c) and 7.7(d), the distributions of scatter points for the synthesized views are not so convergent to the solid line as those points in Fig. 7.7(a) and 7.7(b). It can be concluded that the 2D quality metrics are not very optimum for 3D quality evaluation. So the 3D quality models are employed for the quality evaluation of the synthesized 3D images. Fig. 7.7(e), 7.7(f) and 7.7(g) give the quality evaluation results for the stereoscopic images in terms of 3D quality scores [87] [88] [90]. The simulation results show that there are high correlations between the estimated quality scores obtained from the corresponding mapping curves and the 3D image quality scores calculated using the 3D quality metrics.

The corresponding performances of quality evaluation for the proposed scheme under JPEG compression, JPEG 2000 compression and Gaussian blur are provided in Fig. 7.8, Fig. 7.9 and Fig. 7.10, respectively. From the experimental results, we can observe that the predicted quality is in line with the calculated quality in reality, and the proposed scheme exhibits high accuracy of quality evaluation for DIBR 3D images.

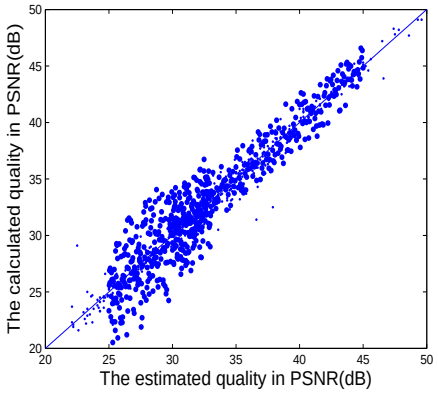
To measure the accuracy of quality evaluation for the proposed scheme, the correlation between the calculated quality and the estimated quality is calculated using RMSE, PLCC and SRCC. The low value of RMSE and high values of PLCC and SRCC reflect the high accuracy of quality evaluation. Tables 7.3, 7.4 and 7.5 give the accuracy of quality evaluation in terms of the 2D quality metrics in RMSE, PLCC


The calculated quality in PSNR(dB)
The estimated quality in PSNR(dB)

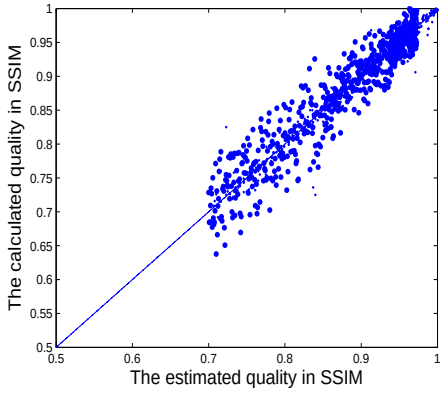
(a)


The calculated quality in SSIM
The estimated quality in SSIM

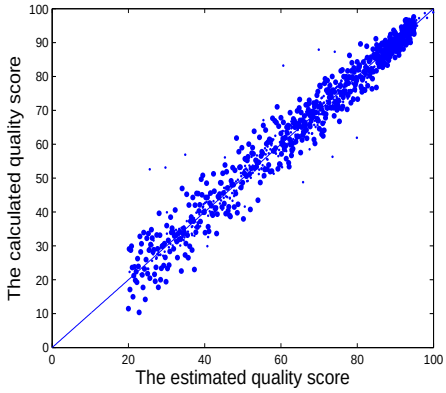
(b)


The calculated quality in PSNR(dB)
The estimated quality in PSNR(dB)

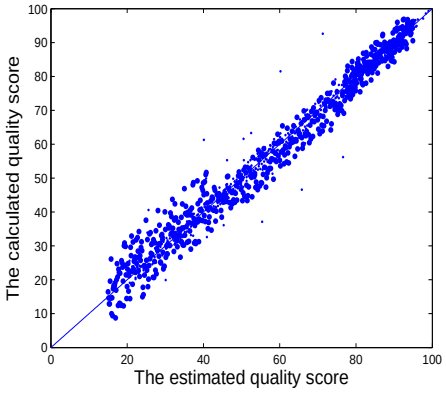
(c)


The calculated quality in SSIM
The estimated quality in SSIM

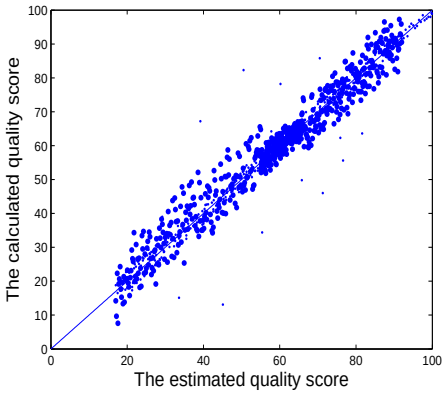
(d)


The calculated quality score
The estimated quality score

(e)


The calculated quality score
The estimated quality score

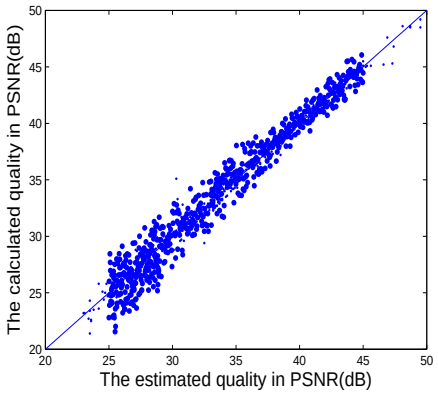
(f)



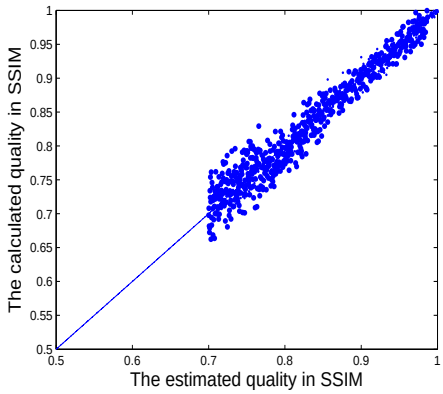
(g)

Figure 7.7: Quality evaluation for the center view and the synthesized views under Gaussian noise. (a) Quality evaluation for the center view in terms of PSNR; (b) Quality evaluation for the center view in terms of SSIM; (c) Quality evaluation for the synthesized views in terms of PSNR; (d) Quality evaluation for the synthesized views in terms of SSIM; (e) Quality evaluation for the synthesized views in terms of the 3D quality metric proposed in [87]; (f) Quality evaluation for the synthesized views in terms of the 3D quality metric proposed in [88]; (g) Quality evaluation for the synthesized views in terms of the 3D quality metric proposed in [90].

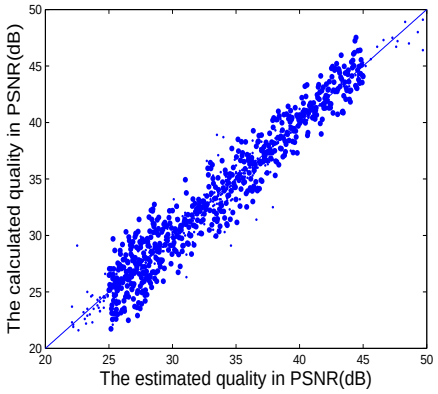
and SRCC, respectively. Under the four different groups of distortions, the calculated RMSE in terms of PSNR and SSIM for the center view are lower than 0.856 and 0.764 in Table 7.3. According to the experimental results in Table 7.4 and Table 7.5, the corresponding PLCC and SRCC for the center view are all higher than 0.952 and 0.943, which show the estimated qualities are highly correlated with the calculated qualities. For the synthesized views in terms of the 2D quality metrics, the evaluation accuracies in RMSE, PLCC and SRCC are not as good as those of the center view. But the results of quality evaluation accuracy are still acceptable. It can be seen that the proposed scheme can evaluate the quality of center view with high accuracy and assess the quality of synthesized views with relatively high accuracy in terms of the 2D quality metrics under various distortions.


The calculated quality in PSNR(dB)
The estimated quality in PSNR(dB)

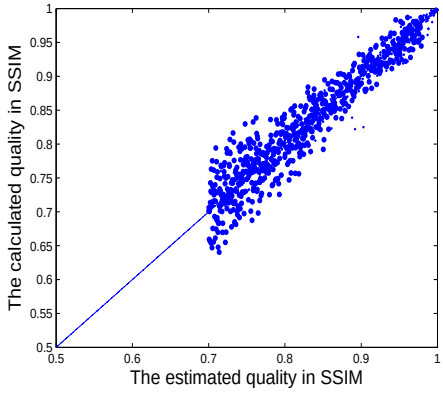
(a)


The calculated quality in SSIM
The estimated quality in SSIM

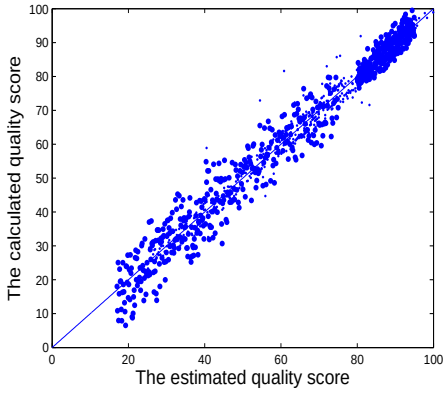
(b)


The calculated quality in PSNR(dB)
The estimated quality in PSNR(dB)

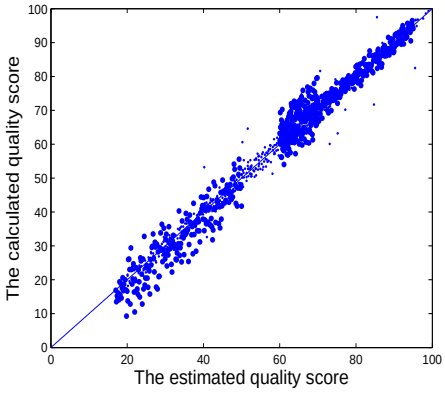
(c)


The calculated quality in SSIM
The estimated quality in SSIM

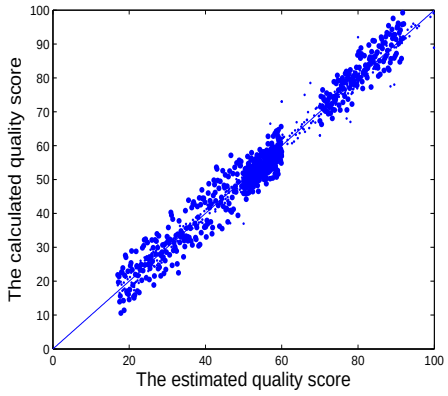
(d)


The calculated quality score
The estimated quality score

(e)


The calculated quality score
The estimated quality score

(f)



(g)

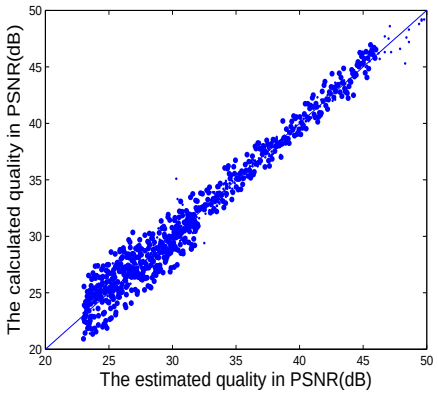
Figure 7.8: Quality evaluation for the center view and the synthesized views under JPEG compression. (a) Quality evaluation for the center view in terms of PSNR; (b) Quality evaluation for the center view in terms of SSIM; (c) Quality evaluation for the synthesized views in terms of PSNR; (d) Quality evaluation for the synthesized views in terms of SSIM; (e) Quality evaluation for the synthesized views in terms of the 3D quality metric proposed in [87]; (f) Quality evaluation for the synthesized views in terms of the 3D quality metric proposed in [88]; (g) Quality evaluation for the synthesized views in terms of the 3D quality metric proposed in [90].

Table 7.3: Evaluation accuracy in terms of PSNR and SSIM in RMSE.

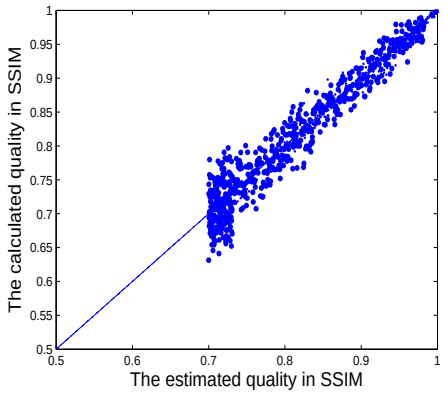
		JPEG	JPEG 2000	Gaussian noise	Gaussian blur
Center view	PSNR	0.521	0.703	0.856	0.870
	SSIM	0.370	0.565	0.429	0.764
Synthesized views	PSNR	3.642	5.793	4.746	6.185
	SSIM	1.935	3.128	2.754	3.601

Table 7.4: Evaluation accuracy in terms of PSNR and SSIM in PLCC.

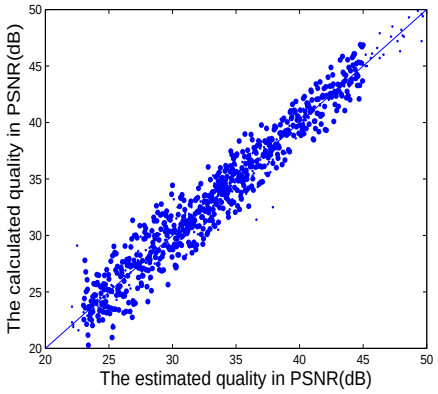
		JPEG	JPEG 2000	Gaussian noise	Gaussian blur
Center view	PSNR	0.981	0.963	0.969	0.952
	SSIM	0.987	0.976	0.980	0.974
Synthesized views	PSNR	0.915	0.882	0.903	0.857
	SSIM	0.924	0.890	0.909	0.852


The calculated quality in PSNR(dB)
The estimated quality in PSNR(dB)

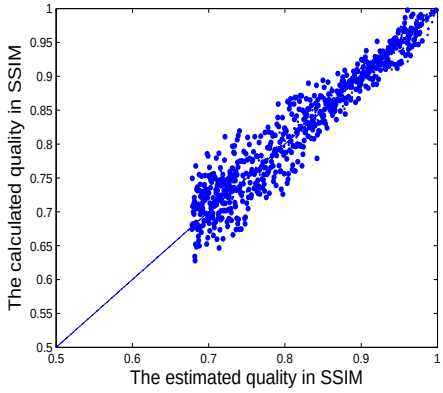
(a)


The calculated quality in SSIM
The estimated quality in SSIM

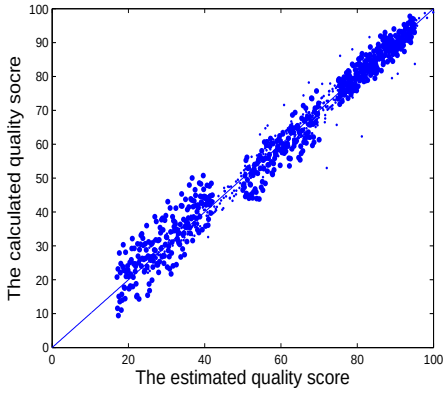
(b)


The calculated quality in PSNR(dB)
The estimated quality in PSNR(dB)

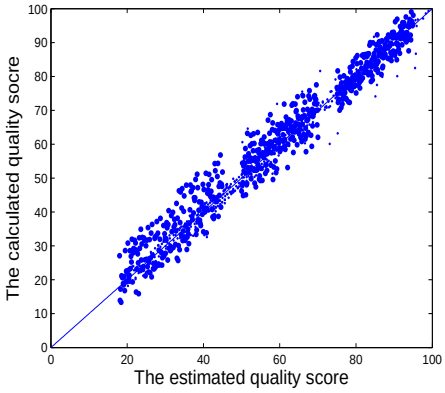
(c)


The calculated quality in SSIM
The estimated quality in SSIM

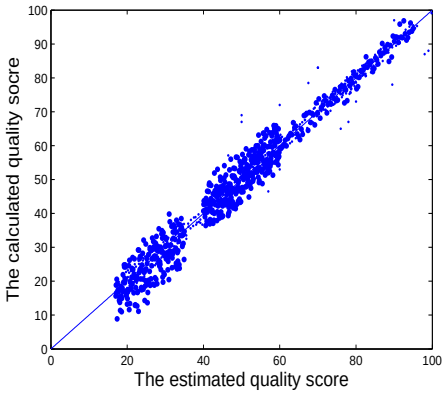
(d)


The calculated quality socre
The estimated quality score

(e)


The calculated quality socre
The estimated quality score

(f)



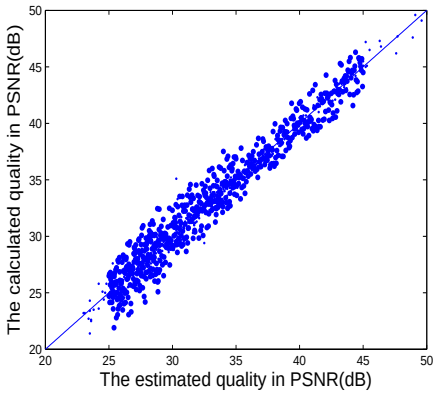
(g)

Figure 7.9: Quality evaluation for the center view and the synthesized views under JPEG 2000 compression. (a) Quality evaluation for the center view in terms of PSNR; (b) Quality evaluation for the center view in terms of SSIM; (c) Quality evaluation for the synthesized views in terms of PSNR; (d) Quality evaluation for the synthesized views in terms of SSIM; (e) Quality evaluation for the synthesized views in terms of the 3D quality metric proposed in [87]; (f) Quality evaluation for the synthesized views in terms of the 3D quality metric proposed in [88]; (g) Quality evaluation for the synthesized views in terms of the 3D quality metric proposed in [90].

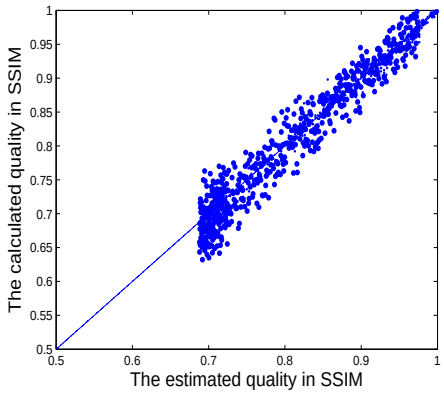
Table 7.5: Evaluation accuracy in terms of PSNR and SSIM in SRCC.

		JPEG	JPEG 2000	Gaussian noise	Gaussian blur
Center view	PSNR	0.961	0.943	0.961	0.945
	SSIM	0.973	0.950	0.968	0.952
Synthesized views	PSNR	0.886	0.864	0.880	0.850
	SSIM	0.911	0.873	0.894	0.852

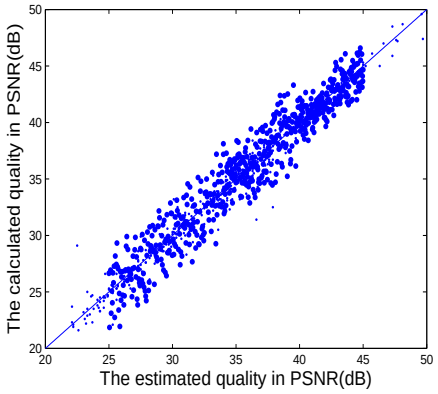
Moreover, for the quality evaluation of the synthesized views, the accuracies in terms of the 3D quality metrics in RMSE, PLCC and SRCC are also calculated. As can be observed from Table 7.6, the values of RMSE are between 0.735 and 2.364 with an average of 1.255, the mean values of PLCC and SRCC are 0.957 and 0.949 respectively, which show high correlations between the estimated qualities and the calculated qualities. In addition, the accuracies in terms of the 3D quality metrics


The calculated quality in PSNR(dB)
The estimated quality in PSNR(dB)

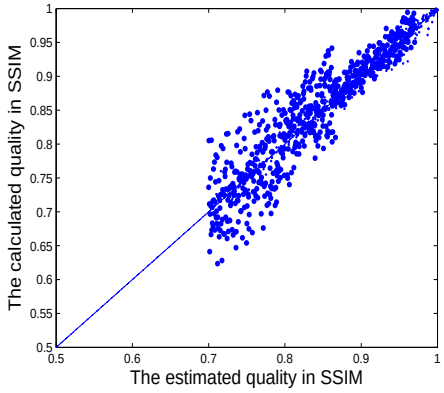
(a)


The calculated quality in SSIM
The estimated quality in SSIM

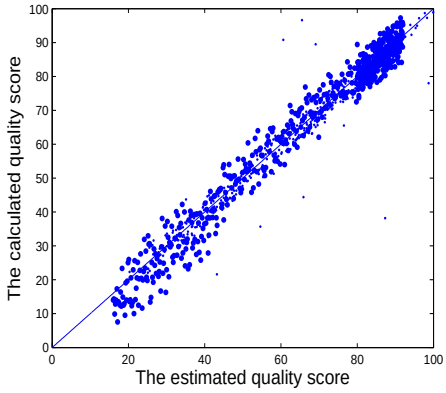
(b)


The calculated quality in PSNR(dB)
The estimated quality in PSNR(dB)

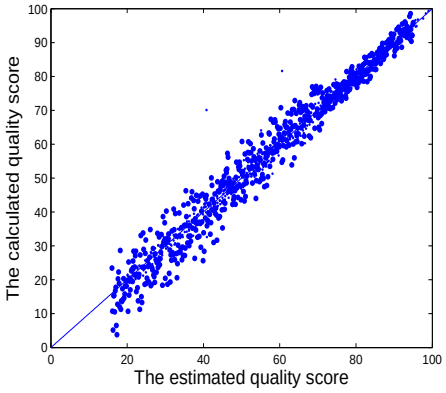
(c)


The calculated quality in SSIM
The estimated quality in SSIM

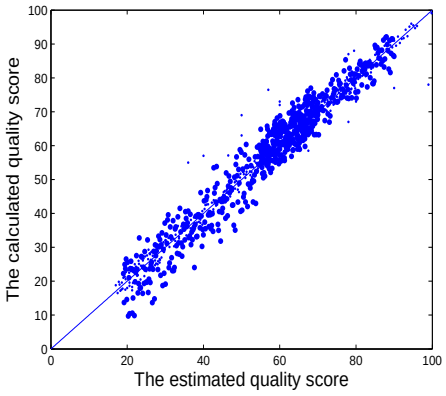
(d)


The calculated quality score
The estimated quality score

(e)


The calculated quality score
The estimated quality score

(f)



(g)

Figure 7.10: Quality evaluation for the center view and the synthesized views under Gaussian blur. (a) Quality evaluation for the center view in terms of PSNR; (b) Quality evaluation for the center view in terms of SSIM; (c) Quality evaluation for the the synthesized views in terms of PSNR; (d) Quality evaluation for the synthesized views in terms of SSIM; (e) Quality evaluation for the synthesized views in terms of the 3D quality metric proposed in [87]; (f) Quality evaluation for the synthesized views in terms of the 3D quality metric proposed in [88]; (g) Quality evaluation for the synthesized views in terms of the 3D quality metric proposed in [90].

are higher than those in terms of the 2D quality metrics for the synthesized views. According to the experimental results, it can be concluded that our proposed scheme is highly accurate in quality evaluation for the center view in terms of the 2D quality metrics and for the synthesized views in terms of the 3D quality models under the four types of distortions.

7.2.3 Performance comparison

In this section, our proposed scheme is compared with other related and state-of-the-art quality evaluation methods based on digital watermarking. The methods in [38] and [39] are used for 2D image quality evaluation. To the best of our knowledge, there are no quality evaluation algorithms for DIBR 3D images using the watermark-

Table 7.6: Evaluation accuracy in terms of the 3D quality metrics for the synthesized views in RMSE, PLCC and SRCC.

		You [87]	Chen [88]	Wang [90]
JPEG	RMSE	1.405	0.979	1.230
	PLCC	0.949	0.936	0.971
	SRCC	0.943	0.952	0.964
JPEG 2000	RMSE	2.364	1.915	0.909
	PLCC	0.940	0.963	0.982
	SRCC	0.935	0.952	0.974
Gaussian noise	RMSE	0.963	0.747	1.132
	PLCC	0.958	0.972	0.987
	SRCC	0.945	0.946	0.952
Gaussian blur	RMSE	1.915	0.735	0.761
	PLCC	0.940	0.929	0.960
	SRCC	0.925	0.943	0.951

ing technique. In the experiment, the same set of center images and synthesized 3D images in Section 7.2.2 is tested for performance comparison. According to the previous discussion, the quality of a stereo-pair can be calculated using the average quality of the left and right images for the compared methods.

Table 7.7 and Table 7.8 present the comparison results for the center images and synthesized 3D images under the four kinds of distortions respectively. The quality evaluation accuracy is measured in terms of PSNR and SSIM in RMSE, PLCC and SRCC. For the quality evaluation of the center images, the average values of RMSE, PLCC and SRCC are given in Table 7.7. As can be seen from this table, our scheme provides comparable evaluation accuracy with the two related methods for the center images. However, the experimental results of quality evaluation for the synthesized 3D images are significantly different. According to the experimental results in Table 7.8, the two compared algorithms have poor performance when applied to quality evaluation for stereoscopic images. The evaluation accuracies in RMSE, PLCC and

SRCC are inferior to that of our proposed scheme. The average values of PLCC and SRCC for the two compared methods are both below 0.7359 and 0.7018, while the corresponding values for the proposed algorithm are above 0.8893 and 0.8700 respectively. Based on the above experimental results and analysis, on the whole, our proposed scheme exhibits high evaluation accuracy for both the center image and stereoscopic 3D images.

Table 7.7: Performance comparison on quality evaluation accuracy for the center images.

	Method [38]			Method [39]			Proposed method		
	RMSE	PLCC	SRCC	RMSE	PLCC	SRCC	RMSE	PLCC	SRCC
PSNR	0.5211	0.9872	0.9704	0.8648	0.9897	0.9527	0.7375	0.9663	0.9525
SSIM	0.7393	0.9516	0.9530	0.3726	0.9468	0.9610	0.5320	0.9793	0.9608

Table 7.8: Performance comparison on quality evaluation accuracy for the synthesized 3D images.

	Method [38]			Method [39]			Proposed method		
	RMSE	PLCC	SRCC	RMSE	PLCC	SRCC	RMSE	PLCC	SRCC
PSNR	12.1314	0.6858	0.6905	9.1463	0.7359	0.7018	5.0915	0.8893	0.8700
SSIM	10.6375	0.7190	0.6341	7.5220	0.6794	0.6402	2.8545	0.8938	0.8825

7.3 Summary

In this chapter, we presented a watermarking based quality evaluation scheme for DIBR 3D images. The proposed scheme is designed to evaluate image quality in terms of the existing 2D and 3D quality metrics under the distortions. In the scheme, the watermark is embedded into the selected coefficients of selected DT-CWT sub-bands of the center view. The watermark can be extracted from the watermarked center view and the synthesized left and right views individually. Then we use the

degradation of the extracted watermarks to evaluate the qualities of the center image and synthesized 3D images. The NC-Quality mapping curves are generated and used for quality evaluation. The correlation between the estimated quality and the calculated quality is calculated for evaluation accuracy. The experimental results show that our proposed scheme is highly accurate and effective for the quality evaluation of DIBR 3D images.

There are several issues to be addressed, the left image and the right image are symmetrically distorted in the experiment. To improve the performance of quality evaluation for the asymmetrically distorted 3D images, we think that the relationship between the degradation of extracted watermark and the quality of stereoscopic images needs to be studied. In addition, the proposed scheme can be extended to estimate the quality of an image under multiple distortions. The quality evaluation for 3D image and video is still in infancy and needs to be further developed. For future work, we will continue to conduct research on 3D image quality evaluation and try to explore new techniques for 3D video quality evaluation.

7.4 Some discussions

In this section, we give some explanations on why contourlet transform, ridgelet transform and dual tree complex wavelet transform are used in Chapter 5, Chapter 6 and Chapter 7 respectively.

For DIBR 3D watermarking, there are two specific requirements: 1. The watermark should be extracted with high accuracy from the virtual left and right views respectively. 2. The watermarking scheme should be robust to various attacks, in-

cluding common DIBR processing distortions. For watermarking based quality evaluation, there are also two requirements: 1. The watermark should be extracted from the virtual left and right views respectively. 2. The extracted watermark should reflect the quality changes of the watermarked image under various distortions, instead of robustness to distortions.

In Chapter 7, the dual-tree complex wavelet transform (DT-CWT) is used for watermarking based quality evaluation: Some DT-CWT subbands have more information on horizontal direction, and have robustness to DIBR view synthesis process (DIBR process maps pixels of the center image horizontally). The DT-CWT has directional selectivity and the changes of coefficients can reflect the strength of distortion, which are good for quality evaluation. To be specific, the DT-CWT can decompose an image into different frequency subbands, and different frequency components have different sensitivities to the distortions, so the watermark can be embedded in a proper way to reflect the quality changes of the distorted watermarked images. This is different from contourlet transform used in Chapter 5. Some contourlet subbands have more information on horizontal direction, and have robustness to view synthesis process. But the directional subbands (the decomposition of bandpass image) for watermark embedding only contain higher frequency components of an image. Moreover, the contourlet subbands have some stability and small distortions rarely affect the changes of coefficients. These facts on the two aspects show that contourlet transform is not suitable for quality evaluation.

In Chapter 5, contourlet transform is used for DIBR watermarking: It meets the above-mentioned two requirements of DIBR 3D watermarking. The advantage is its simplicity and practicality. The four selected subbands can be divided into two pairs

based on similar coefficients. So coefficient quantization can be used for watermark embedding, and statistical difference can be used for watermark extraction. These operations are easy to implement. In Chapter 6, the watermark is embedded into the ridgelet coefficients of the selected feature regions. Compared with other transforms, most of the energy of an image is concentrated in just a few of ridgelet coefficients, which is good for watermark embedding. In addition, it has better stability than contourlet transform and the coefficients do not change significantly even when the image is under possible distortions, which contribute to better robustness to attacks.

Based on the above discussions, dual-tree complex wavelet transform is suitable for watermarking based quality evaluation, while contourlet transform and ridgelet transform are good for DIBR watermarking. The proposed watermarking scheme in Chapter 5 is simple and practical, which can be applied to real-time watermarking. The proposed watermarking scheme in Chapter 6 has better robustness to various attacks, and can be used in scenarios where high accuracy of watermark extraction is required.

Chapter 8

Conclusions and future work

The thesis is focused on the digital watermarking and watermarking based quality evaluation for depth-image-based rendering (DIBR) 3D images. After introducing the fundamental theories and techniques of 3D image watermarking, we reviewed the existing and related watermarking methods for DIBR 3D images. It provides a solid basis for further research in this field.

Firstly, a blind 2D image watermarking algorithm robust to geometric attacks and compressions was introduced. The proposed algorithm is based on contourlet transform (CT) and principal component analysis (PCA). The watermark is embedded into the principal components of the contourlet coefficients of the largest detail directional subband of the cover image. The watermarking strength is adjusted adaptively to preserve the perceptual image quality using the noise visibility function (NVF). The simulation results demonstrate that this algorithm is robust to a variety of attacks, such as rotation, scaling and image compressions. Then the algorithm was extended to blind video watermarking. The performance is evaluated against video attacks like

rotation, frame averaging and video compressions.

Considering the directional multiresolution image representation and convenient tree structure of contourlet transform, a robust blind watermarking scheme for DIBR 3D images in contourlet domain was proposed. After applying contourlet transform to the center view at the content provider side, the watermark is embedded into the selected contourlet subbands by quantizing certain contourlet coefficients. The virtual left and right views are generated from the watermarked center view and the associated depth map using DIBR technique at the receiver side. The statistical differences between quantized and unquantized contourlet coefficients are used for watermark extraction. One of the most important advantages is its simplicity and practicality. However, the performance on watermark detection needs to be further improved.

As an improvement, a novel blind watermarking algorithm for DIBR 3D images based on ridgelet transform and feature regions was developed. After selecting the reference points of the center view, the feature regions are constructed for watermark embedding. Then the watermark is embedded into the amplitudes of ridgelet coefficients of the most energetic direction. The watermarked view has good perceptual quality under both the objective and subjective image quality measures. The error probability is analyzed and confirmed by the experiments. Compared with other related and state-of-the-art methods, the proposed algorithm exhibits better performance in watermark invisibility and robustness against attacks.

The watermarking based quality evaluation has been a promising technology for 3D image quality evaluation. A quality evaluation scheme for DIBR 3D images was devised. The proposed scheme utilizes the watermarking technique to evaluate the

quality of watermarked images under various distortions. The watermark can be detected from the watermarked center images and synthesized images, and the quality of these images can be estimated by examining the degradation of corresponding extracted watermarks. The experimental results show the effectiveness of the proposed scheme in terms of quality evaluation for DIBR 3D images.

Digital watermarking for 3D image and video is still in its infancy and needs to be further developed. For future work, several possible approaches could be taken. The watermark embedding regions could be modeled mathematically with possible distributions. Other watermark embedding techniques can also be taken into consideration. By exploring these properties, the capacity and error correction ability of the watermarking system should be improved. In addition, the watermarking based quality evaluation technique is worthy of further investigation. In a word, this is an exciting and promising area, which could lead to many innovative ideas and applications and benefit the multimedia industry.

Bibliography

[1] Y. H. Lin and J. L. Wu. A digital blind watermarking for depth-image-based rendering 3D images. *IEEE Trans. Broadcast.*, 57(2):602–611, 2011.

[2] H. D. Kim, J. W. Lee, T. W. Oh, and H. K. Lee. Robust DT-CWT watermarking for DIBR 3D images. *IEEE Trans. Broadcast.*, 58(4):533–543, 2012.

[3] S. Katzenbeisser and F. Petitcolas. *Information techniques for steganography and digital watermarking.* Artech House, Norwood, MA, 2000.

[4] I. J. Cox, M. L. Miller, J. A. Bloom, J. Fridrich, and T. Kalker. *Digital watermarking and steganography.* Morgan Kaufman, San Francisco, CA, 2007.

[5] C. Podilchuk and E. Delp. Digital watermarking: algorithms and applications. *IEEE Signal Process. Mag.*, 18(4):33–46, 2001.

[6] I. J. Cox, M. L. Miller, and J. A. Bloom. *Digital watermarking.* Morgan Kaufman, San Francisco, CA, 2001.

[7] D. Zheng, S. Wang, and J. Zhao. RST invariant image watermarking algorithm with mathematical modeling and analysis of the watermarking processes. *IEEE Trans. Image Process.*, 18(5):1055–1068, 2009.

[8] Y. Liu and J. Zhao. A new video watermarking algorithm based on 1D DFT and Radon transform. *Signal Process.*, 90(2):626–639, 2010.

[9] P. Bassia, I. Pitas, and N. Nikolaidis. Robust audio watermarking in the time domain. *IEEE Trans. Multimed.*, 3(2):232–241, 2001.

[10] L. Chen and J. Zhao. Robust contourlet-based blind watermarking for depth-image-based rendering 3D images. *Signal Process.: Image Commun.*, 54:56–65, 2017.

[11] D. Huang and H. Yan. Interword distance changes represneted by sine waves for watermarking text images. *IEEE Trans. Circuits Syst. Video Technol.*, 11(12):1237–1245, 2001.

[12] C. S. Collberg and C. Thomborson. Watermarking, temper-proofing, and obfuscation-tools for software protection. *IEEE Trans. Software Engineering*, 28(8):735–746, 2002.

[13] G. Gupta and J. Pieprzyk. Database relation watermarking resilient against secondary watermarking attacks. *in: Proceedings of International Conference on Information Systems Security*, pages 222–236, 2009.

[14] P. Rasti, S. Samiei, M. Agoyi, S. Escalera, and G. Anbarjafari. Robust non-blind color video watermarking using QR decomposition and entropy analysis. *J. Vis. Commun. Image Represent.*, 38:838–847, 2016.

[15] R. Ni, Q. Ruan, and H. D. Cheng. Secure semi-blind watermarking based on iteration mapping and image features. *Pattern Recog.*, 38(3):357–368, 2005.

[16] L. Chen and J. Zhao. Contourlet-based image and video watermarking robust to geometric attacks and compressions. *Multimed. Tools Appl.*, pages 1–18, 2017.

[17] D. C. Hwang, K. H. Bae, M. H. Lee, and E. S. Kim. Real-time stereo image watermarking using discrete cosine transform and adaptive disparity maps. *in: Proceedings of SPIE Multimedia Systems and Applications VI*, 5241:233–242, 2003.

[18] D. C. Hwang, K. H. Bae, and E. S. Kim. Stereo image watermarking scheme based on discrete wavelet transform and adaptive disparity estimation. *in: Proceedings of SPIE Mathematics of Data/Image Coding, Compression, and Encryption VI, with Applications*, 5208:196–205, 2004.

[19] C. Fehn. A 3D-TV approach using depth-image-based rendering (DIBR). *Visualization, Imaging, and Image Processing*, pages 482–487, 2003.

[20] C. Fehn. Depth-image-based rendering (DIBR), compression and transmission for a new approach on 3D-TV. *in: Proceedings of SPIE Conference on Stereoscopic Displays and Virtual Reality Systems XI*, 5291:93–104, 2004.

[21] W. J. Tam and L. Zhang. 3D-TV content generation: 2D-to-3D conversion. *in: Proceedings of IEEE International Conference on Multimedia and Expo*, pages 1869–1872, 2006.

[22] D. Zheng, Y. Liu, and J. Zhao. A survey of RST invariant image watermarking algorithms. *ACM Computing Surveys*, 39(2):1–91, 2007.

[23] Y. Liu, D. Zheng, and J. Zhao. An image rectification scheme and its applications in RST invariant digital image watermarking. *Multimed. Tools Appl.*, 34(1):57–84, 2007.

[24] P. Kauff, N. Atzpadin, C. Fehn, M. Muller, O. Schreer, A. Somlic, and R. Tanger. Depth map creation and image-based rendering for advanced 3DTV services providing interoperability and scalability. *Signal Process.: Image Commun.*, 22(2):217–234, 2007.

[25] D. Zheng, J. Zhao, and A. Saddik. RST-invariant digital image watermarking bsed on log-polar mapping and phase correlation. *IEEE Trans. Circuits Syst. Video Technol.*, 13(8):753–765, 2003.

[26] D. Zheng, Y. Liu, and J. Zhao. RST invariant digital watermarking based on a new phase-only filtering method. *Signal Process.*, 85(12):2354–2370, 2005.

[27] L. Chen and J. Zhao. Adaptive contourlet-based image watermarking robust to geometric transformations and image compression. *in: Proceedings of IEEE International Instrumentation and Measurement Technology Conference*, pages 996–1001, 2016.

[28] C. Fehn, R. de la Barre, and S. Pastoor. Interactive 3-D-TV-Concepts and key technologies. *Proc. IEEE*, 94(3):524–538, 2006.

[29] L. Zhang and W. J. Tam. Stereoscopic image generation based on depth images for 3D TV. *IEEE Trans. Broadcast.*, 51(2):191–199, 2005.

[30] L. Zhang, C. Vazquez, and S. Knorr. 3D-TV content creation: automatic 2D-to-3D video conversion. *IEEE Trans. Broadcast.*, 57(2):372–383, 2011.

[31] M. Lambooij, W. IJsselsteijn, M. Fortuin, and I. Heynderickx. Visual discomfort and visual fatigue of stereoscopic displays: a review. *J. Imaging Sci. Technol*, 53(3):030201–1–030201–14, 2009.

[32] W. J. Tam, F. Speranza, S. Yano, K. Shimono, and H. Ono. Stereoscopic 3D-TV: visual comfort. *IEEE Trans. Broadcast.*, 57(2):335–346, 2011.

[33] P. J. Lee and Effendi. Nongeometric distortion smoothing approach for depth map preprocessing. *IEEE Trans. Multimed.*, 13(2):246–254, 2011.

[34] M. Solh and G. AlRegib. Hierarchical hole-filling for depth-image-based view synthesis in FTV and 3D video. *IEEE J. Sel. Topics Signal Process.*, 6(5):495–504, 2012.

[35] C. Yao, T. Tillo, Y. Zhao, J. Xiao, H. Bai, and C. Lin. Depth map driven hole filling algorithm exploiting temporal correlation information. *IEEE Trans. Broadcast.*, 60(2):394–403, 2014.

[36] L. Chen and J. Zhao. Robust contourlet-based watermarking for depth-image-based rendering 3D images. *in: Proceedings of IEEE International Symposium on Broadband Multimedia Systems and Broadcasting*, pages 1–4, 2016.

[37] L. Chen and J. Zhao. Watermarking based quality assessment for DIBR 3D images. *in: Proceedings of IEEE International Symposium on Smart Data*, pages 810–814, 2016.

[38] S. Wang, D. Zheng, J. Zhao, W. J. Tam, and F. Speranaza. An image quality evaluation method based on digital watermarking. *IEEE Trans. Circuits Syst. Video Technol.*, 17(1):98–105, 2007.

[39] S. Wang, D. Zheng, J. Zhao, W. J. Tam, and F. Speranaza. Adaptive watermarking and tree structure based image quality estimation. *IEEE Trans. Multimedia*, 16(2):311–325, 2014.

[40] C. Y. Lin, M. Wu, J. Bloom, I. Cox, M. Miller, and Y. Lui. Rotation, scale and translation resilient watermarking for images. *IEEE Trans. Image Process.*, 10(5):767–782, 2001.

[41] J. J. K. O Ruanaidh and T. Pun. Rotation, scale and translation invariant spread spectrum digital image watermarking. *Signal Process.*, 66(3):303–317, 1998.

[42] A. C. Bovik. *Hand book of image and video processing.* Elsevier Academic, Burlington, MA, 2010.

[43] A. M. Tekalp. *Digital video processing.* Prentice-Hall, Upper Saddle River, NJ, 2015.

[44] G. K. Wallace. The JPEG still picture compression standard. *IEEE Trans. Consumer Electron.*, 38(1):xviii–xxxiv, 1992.

[45] A. Skodras, C. Christopoulos, and T. Ebrahimi. The JPEG 2000 still image compression standard. *IEEE Signal Process. Mag.*, 18(5):36–58, 2001.

[46] Y. Wang, J. Ostermann, and Y. Q. Zhang. *Digital video processing and communications.* Prentice-Hall, Englewood Cliffs, NJ, 2001.

[47] S. Biswas, S. R. Das, and E. M. Petriu. An adaptive compressed MPEG-2 video watermarking shcems. *IEEE Trans. Instrum. Meas.*, 54(5):1853–1861, 2005.

[48] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra. Overview of the H.264/AVC video coding standard. *IEEE Trans. Circuits Syst. Video Technol.*, 13(7):560–576, 2003.

[49] G. J. Sullivan, J. Ohm, W. J. Han, and T. Wiegand. Overview of the high efficiency video coding (HEVC) standard. *IEEE Trans. Circuits Syst. Video Technol.*, 22(12):1649–1668, 2012.

[50] W. S. Kim, A. Ortega, P. Lai, D. Tian, and C. Gomila. Depth map distortion analysis for view rendering and depth coding. *in: Proceedings of IEEE International Conference on Image Processing*, pages 721–724, 2009.

[51] W. Y. Chen, Y. L. Chang, S. F. Lin, L. F. Ding, and L. G. Chen. Efficient depth image based rendering with edge dependent depth filter and interpolation. *in: Proceedings of IEEE International Conference on Multimedia and Expo*, pages 1314–1317, 2005.

[52] M. N. Do and M. Vetterli. The contourlet transform: an efficient directional multiresolution image representation. *IEEE Trans. Image Process.*, 14(12):2091–2106, 2005.

[53] P. J. Burt and E. H. Adelson. The Laplacian pyramid as a compact image code. *IEEE Trans. Commun.*, COM-31(4):532–540, 1983.

[54] M. N. Do. *Directional multiresolution image representations.* Ph.D. dissertation, School Comput. Commun. Sci., Swiss Fed. Inst. Technol., Lausanne, Switzerland, 2001.

[55] M. A. Akhaee, S.M. E. Sahraeian, and F. Marvasti. Contourlet-based image watermarking using optimum detector in a noisy environment. *IEEE Trans. Image Process.*, 19(4):967–980, 2010.

[56] T. D. Hien, Y. W. Chen, and Z. Nakao. PCA based digital watermarking. *Knowledge-Based Intelligence Information and Engineering Systems*, 2773:1427–1434, 2003.

[57] Q. Zhao and H. Liu. PCA-based web page watermarking. *Pattern Recog.*, 40(4):1334–1341, 2007.

[58] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.*, 60(2):91–110, 2004.

[59] M. brown and D. G. Lowe. Invariant features from interest point groups. *in: Proceedings of British Machine Vision Conference*, pages 656–665, 2002.

[60] E. J. Candès. *Ridgelets: theory and applications.* Ph.D. dissertation, Dept. Statistics, Standard Univ., Stanford, CA, 1998.

[61] M. N. Do and M. Vetterli. The finite ridgelet transform for image representation. *IEEE Trans. Image Process.*, 12(1):16–28, 2003.

[62] P. Campisi, D. Kundur, and A. Neri. Robust digital watermarking in the ridgelet domain. *IEEE Signal Process. Lett.*, 11(10):826–830, 2004.

[63] N. K. Kalantari, S. M. Ahadi, and M. Vafadust. A robust image watermarking in the ridgelet domain using universally optimum decoder. *IEEE Trans. Circuits Syst. Video Technol.*, 20(3):396–406, 2010.

[64] N. Kingsbury. The dual-tree complex wavelet transfrom. *in: Proceedings Eighth IEEE DSP Workshop*, 1998.

[65] N. Kingsbury. Complex wavelets for shift invariant analysis and filtering of signals. *Appl. Comput. Harmon. Anal.*, 10(3):234–253, 2001.

[66] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Process.*, 13(4):600–612, 2004.

[67] Rec. ITU-R BT.500-11. Methodology for the subjective assessment of the quality of television pictures. *International Telecommunication Union*, 2002.

[68] J. L. Myers, A. D. Well, and R. F. Lorch. *Research design and statistical analysis*. Routledge, New York, 2010.

[69] K. Muller, P. Merkle, and T. Wiegand. 3D video representation using depth maps. *Proc. IEEE*, 99(4):643–656, 2011.

[70] M. Tanimoto, M. P. Tehrani, T. Fujii, and T. Yendo. Free-viewpoint TV. *IEEE Signal Process. Mag.*, 28(1):67–76, 2011.

[71] N. Zhu, G. Ding, and J. Wang. A novel digital watermarking method for new viewpoint video based on depth map. *in: Proceedings of International Conference on Intelligent Systems Design and Applications*, pages 3–7, 2008.

[72] E. Halici and A. Alatan. Watermarking for depth-image-based rendering. *in: Proceedings of IEEE International Conference on Image Processing*, pages 4217–4220, 2009.

[73] A. Koz, C. Cigla, and A. Alatan. Watermarking of free-view video. *IEEE Trans. Image Process.*, 19(7):1785–1797, 2010.

[74] M. J. Lee, J. W. Lee, and H. K. Lee. Perceptual watermarking for 3D stereoscopic video using depth information. *in: Proceedings of International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, pages 81–84, 2011.

[75] Y. H. Lin and J. L. Wu. Unseen visible watermarking for color plus depth map 3D images. *in: Proceedings of International Conference on Acoustics, Speech and Signal Processing*, pages 1801–1804, 2012.

[76] S. Fan, M. Yu, G. Jiang, F. Shao, and Z. Peng. A digital watermarking algorithm based on region of interest for 3D image. *in: Proceedings of International Conference on Computational Intelligence and Security*, pages 549–552, 2012.

[77] M. Asikuzzaman, A. J. Lambert M. J. Alam, and M. R. Pickering. A blind watermarking scheme for depth-image-based rendering 3D video using the dual-tree complex wavelet transform. *in: Proceedings of IEEE International Conference on Image Processing*, pages 5497–5501, 2014.

[78] H. R. Sheikh, M. F. Sabir, and A. C. Bovik. A statistical evaluation of recent full reference image quality assessment algorithms. *IEEE Trans. Image Process.*, 15(11):3441–3452, 2006.

[79] Z. Wang and E. P. Simoncelli. Reduced-reference image quality assessment using a wavelet domain natural image statistic model. *in: Proceedings of SPIE Conference on Human Vision and Electronic Imaging*, 5666:149–159, 2005.

[80] Q. Li and Z. Wang. Reduced-reference image quality assessment using divisive normalization-based image representation. *IEEE J. Select. Topics Signal Process.*, 3(2):201–211, 2009.

[81] Z. Wang, H. R. Sheikh, and A. C. Bovik. No-reference perceptual quality assessment of JPEG compressed images. *in: Proceddings of IEEE International Conference on Image Processing*, 1:477–480, 2002.

[82] A. Mittal, A. K. Moorthy, and A. C. Bovik. No-reference image quality assessment in the spatial domain. *IEEE Trans. Image Process.*, 21(12):4695–4708, 2012.

[83] M. A. Saad, A. C. Bovik, and C. Charrier. Blind image quality assessment: a natural scene statistics approach in the DCT domain. *IEEE Trans. Image Process.*, 21(8):3339–3352, 2012.

[84] P. Campisi, M. Carli, G. Giunta, and A. Neri. Blind quality assessment system for multimedia communications using tracing watermarking. *IEEE Trans. Signal Process.*, 51(4):996–1002, 2003.

[85] S. Wang, D. Zheng, J. Zhao, W. J. Tam, and F. Speranaza. A digital watermarking and perceptual model based video quality measurement. *in: Proceedings of IEEE International Instrumentation and Measurement Technology Conference*, pages 1729–1734, 2005.

[86] E. Nezhadarya, Z. Wang, and R. Ward. Image quality monitoring using spread spectrum watermarking. *in: Proceedings of IEEE International Conference on Image Processing*, pages 2233–2236, 2009.

[87] J. You, L. Xing, A. Perkis, and X. Wang. Perceptual quality assessment for stereoscopic images based on 2D image quality metrics and disparity analysis. *in: Proceedings of International Workshop on Video Processing and Quality Metrics for Consumer Electronics*, pages 61–66, 2010.

[88] M.-J. Chen, C.-C. Su, D.-K. Kwon, L. K. Cormack, and A. C. Bovik. Full-reference quality assessment of stereopairs accounting for rivalry. *Signal Process.: Image Commun.*, 28(9):1143–1155, 2013.

[89] M.-J. Chen, L. K. Cormack, and A. C. Bovik. No-reference quality assessment of natural stereopairs. *IEEE Trans. Image Process.*, 22(9):3379–3391, 2013.

[90] J. Wang, A. Rehman, K. Zeng, S. Wang, and Z. Wang. Quality prediction of asymmetrically distorted stereoscopic 3D images. *IEEE Trans. Image Process.*, 24(11):3400–3414, 2015.

[91] H. Song, S. Yu, X. Yang, L. Song, and C. Wang. Contourlet-based image adaptive watermarking. *Signal Process.: Image Commun.*, 23(3):162–178, 2008.

[92] S. Ranjbar, F. Zargari, and M. Ghanbari. A highly robust two-stage contourlet-based digital image watermarking method. *Signal Process.: Image Commun.*, 28(10):1526–1536, 2013.

[93] H. Khalilian and I. V. Bajic. Video watermarking with empirical PCA-based decoding. *IEEE Trans. Image Process.*, 22(12):4825–4840, 2013.

[94] P. Bas, J. Chassery, and B. Macq. Geometrically invariant watermarking using feature points. *IEEE Trans. Image Process.*, 11(9):1014–1028, 2002.

[95] C. W. Tang and H. M. Hang. A feature-based robust digital image watermarking scheme. *IEEE Trans. Signal Process.*, 51(4):950–959, 2003.

[96] S. Wang, C. Cui, and X. Niu. Watermarking for DIBR 3D images based on SIFT feature points. *Measurement*, 48:54–62, 2014.

[97] C. Cui, S. Wang, and X. Niu. A novel watermarking for DIBR 3D images with geometric rectification based on feature points. *Multimed. Tools Appl.*, pages 1–29, 2015.

[98] J. S. Tsai, W. B. Huang, and Y. H. Kuo. On the selection of optimal feature region set for robust digital image watermarking. *IEEE Trans. Image Process.*, 20(3):735–743, 2011.

[99] V. Solachidis and I. Pitas. Circularly symmetric watermarking embedding in 2D DFT domain. *IEEE Trans. Image Process.*, 10(11):1741–1753, 2001.

[100] X. Kang, J. Huang, Y. Q. Shi, and Y. Lin. A DWT-DFT composite watermarking scheme robust to both affine transform and JPEG compression. *IEEE Trans. Circuits Syst. Video Tehcnol.*, 13(8):776–786, 2003.

[101] M. Cancellaro, F. Battisti, M. Carli, G. Boato, F. G. B. De Natale, and A. Neri. A communtative digital image watermarking and encryption method in the tree structured Haar transform domain. *Signal Process.: Image Commun.*, 26(1):1–12, 2011.

[102] C. Wang, X. Wang, and Z. Xia. Geometrically invariant image watermarking based on fast radial harmonic fourier moments. *Signal Process.: Image Commun.*, 45:10–23, 2016.

[103] L. Chen and J. Zhao. Adaptive digital watermarking using RDWT and SVD. *in: Proceedings of IEEE International Symposium on Haptic, Audio and Visual Environments and Games*, pages 1–5, 2015.

[104] L. Chen and J. Zhao. Quality assessment of stereoscopic 3D images based on local and global visual characteristics. *in: Proceedings of IEEE International Conference on Multimedia Expo*, 2017.

[105] L. Wang, H. Ling, F. Zou, and Z. Lu. Real-time compressed-domain video watermarking resistance to geometric distortions. *IEEE Trans. Multimed.*, 19(1):70–79, 2012.

[106] University of South California: The usc-sipi image database. http://sipi.usc.edu/database/.

[107] S. Khalighi, P. Tirdad, and H. R. Rabiee. A contourlet-based image watermarking scheme with high resistance to removal and geometrical attacks. *EURASIP Journal on Advances in Signal Processing*, 2010(Article No. 21), 2010.

[108] M. Asikuzzaman, M. J. Alam, A. J. Lambert, and M. R. Pickering. Imperceptible and robust blind video watermarking using chrominance embedding: a set of approaches in the DT CWT domain. *IEEE Trans. Inf. Forensics Secur.*, 9(9):1502–1517, 2014.

[109] A. Mansouri, A. M. Aznaveh, F. T. Azar, and F. Kurugollu. A low complexity video watermarking in H.264 compressed domain. *IEEE Trans. Inf. Forensics Secur.*, 5(4):649–657, 2010.

[110] M. Fallahpour, M. Semsarzadeh, S. Shirmohammadi, and J. Zhao. A real-time spatio-temporal watermarking scheme for H.264/AVC. *in: Proceedings of IEEE International Instrumentation and Measurement Technology Conference*, pages 872–875, 2013.

[111] Y. Tew and K. Wong. An overview of information hiding in H.264/AVC compressed video. *IEEE Trans. Circuits Syst. Video Technol.*, 24(2):305–319, 2014.

[112] K. Ogawa and G. Ohtake. Watermarking for HEVC/H.265 stream. *In: Proceedings of International Conference on Consumer Electronics*, pages 102–103, 2015.

[113] T. Dutta and H. P. Gupta. A robust watermarking framework for high efficiency video coding (HEVC)-encoded video with blind extraction process. *J. Vis. Commun. Image Represent.*, 38:29–44, 2016.

[114] Y. Wang and A. Pearmain. Blind MPEG-2 video watermarking robust against geometric attacks: a set of approaches in DCT domain. *IEEE Trans. Image Process.*, 15(6):1536–1543, 2006.

[115] B. Chen and G. W. Wornell. Quantization index modulation: a class of provably good methods for digital watermarking and information embedding. *IEEE Trans. Inf. Theory*, 47(4):1423–1443, 2001.

[116] S. H. Wang and Y. P. Lin. Wavelet tree quantization for copyright protection watermarking. *IEEE Trans. Image Process.*, 13(2):154–165, 2004.

[117] W. H. Lin, S. J. Horng, T. W. Kao, P. Fan, C. L. Lee, and Y. Pan. An effi-

cient watermarking method based on significant difference of wavelet coefficient quantization. *IEEE Trans. Multimed.*, 10(5):746–757, 2008.

[118] H. Y. Huang, C. H. Yang, and W. H. Hsu. A video watermarking technique based on Pseudo-3-D DCT and quantization index modulation. *IEEE Trans. Inf. Forensics Secur.*, 5(4):625–637, 2010.

[119] D. Scharstein and C. Pal. Learning conditional random fields for stereo. *in: Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*, pages 1–8, 2007.

[120] H. Hirschmuller and D. Scharstein. Evaluation of cost functions for stereo matching. *in: Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*, pages 1–8, 2007.

[121] H. S. Malvar and D. A. F. Florencio. Improved spread spectrum: a new modulation techinque for robust watermarking. *IEEE Trans. Signal Process.*, 51(4):898–905, 2003.

[122] C. L. Zitnick, S. B. Kang, M. Uyttendaele, S. Winder, and R. Szeliski. High-quality video view interpolation using a layered representation. *ACM Trans. Graphics*, 23(3):600–608, 2004.

[123] P. Loo and N. Kingsbury. Digital watermarking using complex wavelets. *in: Proceedings of IEEE International Conference on Image Processing*, pages 29–32, 2000.

[124] I. W. Selesnick, R. G. Baraniuk, and N. G. Kingsburry. The dual-tree complex wavelet transform. *IEEE Signal Process. Mag.*, 22(6):123–151, 2005.

[125] L. E. Coria, M. R. Pickering, P. Nasiopoulos, and R. K. Ward. A video watermarking scheme based on the dual-tree complex wavelet transform. *IEEE Trans. Inf. Forensics Secur.*, 3(3):466–474, 2008.

[126] A. K. Moorthy, C.-C. Su, A. Mittal, and A. C. Bovik. Subjective evaluation of stereoscopic image quality. *Signal Process.: Image Commun.*, 28(8):870–883, 2012.