Network-adaptive prediction-compensated polyphase multiple description image coding and contrast-guided image interpolation

Wei, Zhe

2012

Wei, Z. (2012). Network-adaptive prediction-compensated polyphase multiple description image coding and contrast-guided image interpolation. Doctoral thesis, Nanyang Technological University, Singapore.

<https://hdl.handle.net/10356/50499>

<https://doi.org/10.32657/10356/50499>

**NANYANG
TECHNOLOGICAL
UNIVERSITY**

**NETWORK-ADAPTIVE PREDICTION-COMPENSATED
POLYPHASE MULTIPLE DESCRIPTION IMAGE CODING
AND CONTRAST-GUIDED IMAGE INTERPOLATION**

**WEI ZHE
SCHOOL OF ELECTRICAL AND ELECTRONIC
ENGINEERING
2012**

**NETWORK-ADAPTIVE PREDICTION-COMPENSATED
POLYPHASE MULTIPLE DESCRIPTION IMAGE CODING
AND CONTRAST-GUIDED IMAGE INTERPOLATION**

WEI ZHE

School of Electrical and Electronic Engineering

A thesis submitted to the Nanyang Technological University
in partial fulfillment of the requirement for the degree of
Doctor of Philosophy

2012

Statement of Originality

I hereby certify that the work embodied in this thesis is the result of original research and has not been submitted for a higher degree to any other University or Institution.

.....

Date

.....

Wei Zhe

Acknowledgments

“With man this is impossible, but with God all things are possible.” —Matthew 19:26

There are many people who deserve more than just an acknowledgement on this page. First, I would like to thank my supervisor, Professor Kai-Kuang Ma, who advised this thesis. It is a privilege to learn from him for five years where he gave me much freedom in doing research. He is very supportive and patient. Most importantly, he not only directed my research work, but also gave me precious suggestions and instructions on the research attitude and effective communications. These valuable experiences will greatly influence my future career and life.

I would also like to thank my earlier teachers without whose support I might never come to this field for a PhD degree, in particular the advisor of my master degree, Professor Canhui Cai in Huaqiao University. They inspired my interests in thinking and learning.

I would like to thank my labmates, Tian Jing, Zeng Huanqiang, Wang Chen and Ong Hui Ting for their valuable advices and discussions on my research. I would also like to thank all of my friends who have made my time at Nanyang Technological University enjoyable.

Finally, I would like to thank my wife, my daughter, and my parents for their unselfish support all the time. Without their love, this work would be impossible.

Contents

Acknowledgments	i
List of Figures	v
List of Tables	xii
Summary	xiv
1 Introduction	1
1.1 Multiple Description Coding (MDC)	1
1.2 Image Interpolation	3
1.3 Main Contributions	4
1.4 Organization	5
2 An Overview of Multiple Description Coding (MDC)	8
2.1 Review of Existing MDC Methods	8
2.2 Polyphase Downsampling MDC (PD-MDC)	13
2.2.1 Review of the PD-MDC	13
2.2.2 Study of Image Downsampling for the PD-MDC	15
2.3 Summary	16

3 Prediction-compensated Polyphase MDC with Adaptive Redundancy Control	17
3.1 Proposed Prediction-compensated Polyphase MDC (PCP-MDC) . .	17
3.1.1 Proposed Polyphase Downsampling Method	17
3.1.2 Framework of the Proposed PCP-MDC	19
3.1.3 Mode-guided Directional Prediction (MDP)	23
3.1.4 Performance of the PCP-MDC	25
3.2 Proposed Adaptive Redundancy Control (ARC)	28
3.2.1 Overview	28
3.2.2 Framework of the Proposed Description-compensated Redundancy Generation (DRG)	30
3.2.3 Proposed Network-aware Optimal Redundancy Allocation (NO-RA)	35
3.3 Proposed Post-processing: Switching Gaussian Filter (SGF)	41
3.4 Simulation Results	43
3.5 Summary	53
4 Contrast-guided Image Interpolation	54
4.1 Review of Interpolation Methods	54
4.2 Fundamental Issues and Ideas	56
4.2.1 Using A Decision Map for Image Interpolation—The Turnkey Solution	56
4.2.2 Edge Versus Contrast	57
4.3 Framework of Contrast-guided Image Interpolation	61
4.3.1 Generation of Directional Variation Fields via Directional Derivative	63

4.3.2	Generation of Contrast-guided Decision Maps via Diffusion Process	65
4.3.3	Image Interpolation Guided by the Contrast-guided Decision Maps	73
4.4	Simulation Results	74
4.5	Summary	91
5	Conclusion and Future Works	92
5.1	Conclusion	92
5.2	Future Works	96
5.2.1	Multiple Description Video Coding	96
5.2.2	Scalable Multiple Description Image Coding	96

List of Figures

1.1	A typical framework of two-channel MDC scheme, which contains three decoders: one central decoder and two side decoders.	2
2.1	Typical framework of the <i>polyphase downsampling multiple description coding</i> (PD-MDC) scheme.	12
2.2	Two ways of partitioning an original image into two descriptions. . .	12
2.3	The pre-processing step proposed in [36] for the redundancy control. . .	14
3.1	Four sub-descriptions generated from an original image by further partitioning each quincunx-downsampled description into two sub-descriptions; the <i>primary sub-description</i> (PS) and the <i>dual sub-description</i> (DS).	19
3.2	The framework of the proposed PCP-MDC scheme.	20
3.3	Illustration of the inter sub-description prediction of a 4×4 block in $\mathbf{X}_{1d}$. The white circles represent the samples of $\mathbf{X}_{1d}$ to be predicted, while the shaded squares denote the samples of the reconstructed sub-description $\tilde{\mathbf{X}}_{1p}$ used for conducting the inter sub-description prediction of $\mathbf{X}_{1d}$. To predict the sample $\mathbf{X}_{1d}(n, m)$ in the block, the inter sub-description prediction exploits its nearest four neighboring samples from the sub-description $\tilde{\mathbf{X}}_{1p}$	22

3.4	Nine prediction modes used in the H.264/AVC intra 8×8 prediction or 4×4 prediction (Mode 2 as the DC mode is not shown).	24
3.5	Illustration of the proposed <i>mode-guided directional prediction</i> (MD-P) algorithm. Depending on the prediction mode determined at the encoding stage of sub-description $\mathbf{X}_{1p}$, the prediction of the sample $\mathbf{X}_{1d}(n, m)$ is to be conducted through one of the three block directions.	25
3.6	Four test images used for simulation experiments.	26
3.7	Objective comparisons between the proposed PCP-MDC and the conventional PD-MDC method [37] on the <i>central distortion</i> and the <i>side distortion</i> in PSNR (dB) versus the total bitrate (bpp), respectively.	27
3.8	Generation of the <i>primary</i> and <i>complementary</i> parts as the input source for the proposed <i>description-compensated redundancy generation</i> (DRG) scheme.	31
3.9	The framework of the proposed <i>adaptive redundancy control</i> (ARC), which consists of two parts: 1) <i>description-compensated redundancy generation</i> (DRG) for quality control, and 2) <i>network-aware optimal redundancy allocation</i> (NORA) for quantity control. Refer to Table 3.2 for the direction of prediction guided by the MDP. The quantizer (Q) controls both quality and quantity.	33
3.10	Comparisons between the ‘actual’ and the proposed fast estimated functions for $R_p(q_p)$ based on the test image “Lena.” The ‘actual’ function is generated by repeatedly running the entire encoder six times.	40

3.11 Comparisons between the ‘actual’ and the proposed fast estimated functions for $R_r(q_r)$ based on the test image “Lena.” The ‘actual’ function is generated by repeatedly running the entire encoder six times.	40
3.12 A portion of the reconstructed “Lena” image coded at 0.25 bpp for two description together with or without exploiting the proposed <i>switching Gaussian filter</i> (SGF).	42
3.13 Eight test images used for simulation experiments.	45
3.14 Objective comparisons among the proposed MDC system and the methods in [37], [38] and [22] in terms of the <i>side distortion</i> versus <i>central distortion</i> , both measured in PSNR (dB). For a fair comparison, the total bitrate for each method is set as 2 bpp (i.e., 1 bpp for each description).	47
3.15 Objective comparisons among the proposed MDC system and the methods in [37], [38] and [22] in terms of the <i>side distortion</i> versus <i>central distortion</i> , both measured in PSNR (dB). For a fair comparison, the total bitrate for each method is set as 1 bpp (i.e., 0.5 bpp for each description).	48
3.16 The average PSNR generated by allocating different amounts of redundancy R_r for the test image “Lena” under three probabilities of channel failure: (a) $\rho = 5\%$, (b) $\rho = 30\%$, and (c) $\rho = 50\%$. The target total bitrate R is 2 bpp. For each case, the amount of redundancy allocated by the proposed NORA method is indicated in the figure by the position of the dashed line; that is, $R_r^* = 0.005$ bpp in (a), $R_r^* = 0.050$ bpp in (b), and $R_r^* = 0.245$ bpp in (c). One can see that they consistently achieve the highest average PSNR result in each case.	49

- 4.1 (a) The presence of an edge can only be *perceived* by the human visual system once a sufficiently-large contrast is formed across two adjacent uniform regions; the larger the pixel-intensity difference across two regions, the stronger their contrast, which leads to a sharper edge being perceived. (b) The resulted binary edge map (or, edge-guided decision map) by exploiting any edge detection method. Note that the detected edge must be drawn on the boundary of either side of region's borderline. (c) An identical image content as that in (a), except with much less contrast being perceived. However, the same edge-guided decision map as shown in (b) will be obtained by exploiting the same edge detection method. 58
- 4.2 (a) Illustration of *confidence band* and its *width* through an image consisting of three uniform regions with three different gray levels. (b) The resulted binary *contrast-guided decision map* of (a), reflecting the local contrast; the higher the contrast, the wider the confidence band. 59
- 4.3 Framework of the proposed contrast-guided image interpolation, which consists of two sequential stages of pixel interpolation with a lattice arrangement as depicted in Fig 4.4. 62
- 4.4 The proposed two-stage image interpolation, where the hollow circles denote the pixels under interpolation for the HR image and the black dots denote available pixels for conducting the interpolation of those pixels represented in the hollow circles: (a) interpolating the 'diagonal' pixels $I(2i, 2j)$ based on the LR image, and (b) interpolating the 'row' pixels $I(2i - 1, 2j)$ and 'column' pixels $I(2i, 2j - 1)$ based on the LR image and the interpolated diagonal pixels $I(2i, 2j)$ obtained in Stage 1. 62

- 4.5 The *mean absolute difference* $MAD(n)$ incurred in the n -th iteration for three test images, “Lena,” “Station,” and “Cameraman,” where the $MAD(n)$ is the difference measured between the diffused 45° -direction DV field of the current iteration and that of the previous iteration. One can see that $MAD(n)$ converges when the number of iterations n increases. In addition, $\epsilon = 0.006$ is set as an empirically-determined threshold to flag that the numerical solution on obtaining the diffused DV field is converged, once $MAD(n) < 0.006$ 69
- 4.6 The DV fields (left column) and the *diffused* DV fields (right column) generated from the test image “Wheel:” (a)-(b) 45° direction and (c)-(d) 135° direction obtained in Stage 1; (e)-(f) 0° direction and (g)-(h) 90° direction obtained in Stage 2. 71
- 4.7 EDMs (left column) and CDMs (right column) generated from the test image “Wheel” by exploiting a thresholding process on the DV fields and the diffused DV fields as shown in Fig. 4.6, respectively: (a)-(b) 45° direction and (c)-(d) 135° direction obtained in Stage 1; (e)-(f) 0° direction and (g)-(h) 90° direction obtained in Stage 2. . . 72
- 4.8 Eight test images used for simulation experiments: (a) “Lena” (256×256), (b) “Station” (256×256), (c) “Cameraman” (256×256), (d) “House” (256×256), (e) “Wheel” (300×160), (f) “Parrot” (768×512), (g) “Butterfly” (480×288), and (h) “Bike” (768×512). . . . 75
- 4.9 A downsampled test image “Wheel” as shown in (a) is interpolated by a factor of 2×2 by using different interpolation methods: (b) bicubic [3], (c) covariance-guided [54], (d) edge-guided, (e) LMMSE-based [55], and (f) the proposed contrast-guided. 77

4.10	A downsampled test image “Station” as shown in (a) is interpolated by a factor of 2×2 by using different interpolation methods: (b) bicubic [3], (c) covariance-guided [54], (d) edge-guided, (e) LMMSE-based [55], and (f) the proposed contrast-guided.	78
4.11	A downsampled test image “Lena” as shown in (a) is interpolated by a factor of 2×2 by using different interpolation methods: (b) bicubic [3], (c) covariance-guided [54], (d) edge-guided, (e) LMMSE-based [55], and (f) the proposed contrast-guided.	79
4.12	A downsampled test image “Cameraman” as shown in (a) is interpolated by a factor of 2×2 by using different interpolation methods: (b) bicubic [3], (c) covariance-guided [54], (d) edge-guided, (e) LMMSE-based [55], and (f) the proposed contrast-guided.	80
4.13	A downsampled color test image “Butterfly” as shown in (a) is interpolated by a factor of 2×2 by using different interpolation methods: (b) bicubic [3], (c) covariance-guided [54], (d) edge-guided, (e) LMMSE-based [55], and (f) the proposed contrast-guided.	82
4.14	A downsampled color test image “Bike” as shown in (a) is interpolated by a factor of 2×2 by using different interpolation methods: (b) bicubic [3], (c) covariance-guided [54], (d) edge-guided, (e) LMMSE-based [55], and (f) the proposed contrast-guided.	83
4.15	A downsampled color test image “Parrot” as shown in (a) is interpolated by a factor of 2×2 by using different interpolation methods: (b) bicubic [3], (c) covariance-guided [54], (d) edge-guided, (e) LMMSE-based [55], and (f) the proposed contrast-guided.	84
4.16	The test images to be enlarged: (a) “Wheel,” (b) “Tripod,” (c) “Butterfly,” (d) “Cloth,” (e) “Roof.”	85

- 4.17 The test image (a) “Wheel” is directly enlarged by a factor of 8×8 via three stages of progressive interpolations by using: (b) bicubic [3], (c) edge-guided, (d) covariance-guided [54], (e) LMMSE-based [55], and (f) the proposed contrast-guided. 86
- 4.18 The test image (a) “Tripod” is directly enlarged by a factor of 8×8 via three stages of progressive interpolations by using: (b) bicubic [3], (c) edge-guided, (d) covariance-guided [54], (e) LMMSE-based [55], and (f) the proposed contrast-guided. 87
- 4.19 The test image (a) “Butterfly” is directly enlarged by a factor of 8×8 via three stages of progressive interpolations by using: (b) bicubic [3], (c) edge-guided, (d) covariance-guided [54], (e) LMMSE-based [55], and (f) the proposed contrast-guided. 88
- 4.20 The test image (a) “Cloth” is directly enlarged by a factor of 8×8 via three stages of progressive interpolations by using: (b) bicubic [3], (c) edge-guided, (d) covariance-guided [54], (e) LMMSE-based [55], and (f) the proposed contrast-guided. 89
- 4.21 The test image (a) “Roof” is directly enlarged by a factor of 8×8 via three stages of progressive interpolations by using: (b) bicubic [3], (c) edge-guided, (d) covariance-guided [54], (e) LMMSE-based [55], and (f) the proposed contrast-guided. 90

List of Tables

2.1	The average Euclidean distance between two adjacent samples . . .	15
3.1	Block direction for each partitioned block in the DS, according to the H.264/AVC prediction mode information of the PS obtained previously.	24
3.2	Mode-guided directional prediction (MDP). The block direction of each partitioned block in the complementary part is determined according to the H.264/AVC prediction modes of the primary part determined previously.	34
3.3	PSNR performance (in dB) with or without using the SGF for the test image “Lena” encoded at different bitrates	43
3.4	Objective comparisons between the proposed MDC system and the method in [38] in terms of the average PSNR (dB) versus the total bitrate (bpp) under probability of channel failure, $\rho = 1\%$. The oversampling factor of the method [38] is fixed at $v_p = 0.5$	50
3.5	Objective comparisons between the proposed MDC system and the method in [38] in terms of the average PSNR (dB) versus the total bitrate (bpp) under four probability of channel failure, $\rho = 10\%$. The oversampling factor of the method [38] is fixed at $v_p = 0.5$	51

3.6	Objective comparisons between the proposed MDC system and the method in [38] in terms of the average PSNR (dB) versus the total bitrate (bpp) under probability of channel failure, $\rho = 20\%$. The oversampling factor of the method [38] is fixed at $v_p = 0.5$	51
3.7	Objective comparisons between the proposed MDC system and the method in [38] in terms of the average PSNR (dB) versus the total bitrate (bpp) under four probability of channel failure, $\rho = 50\%$. The oversampling factor of the method [38] is fixed at $v_p = 0.5$	52
3.8	Comparisons of actual encoding time (per frame). The results were obtained based on the Intel Core 2 Duo 1.86 GHz CPU with 1 GB DDR2 533 MHz RAM memory. Parameter v_p is the oversampling factor.	52
4.1	PSNR results (in dB) of the interpolated HR images using five different interpolation methods, where the LR images are obtained by directly downsampling the HR images. The highest PSNR result achieved in each test case is highlighted in boldface. For the proposed method, its Gain (in dB) over the second-best method is also provided in the parenthesis.	76
4.2	The interpolation times (in seconds) consumed in three different methods. All these methods are implemented in MATLAB 7.8.0. The computer configuration used in our simulation is the Intel Core 2 Duo 1.86 GHz CPU and 4 GB 533 MHz DDR2 RAM.	85

Summary

In this thesis, we propose an efficient *multiple description coding* (MDC) system and a novel image interpolation method. For the MDC system, it consists of two thrust contributions: 1) a new polyphase MDC scheme, called the *prediction-compensated polyphase* MDC (PCP-MDC), for significantly improving coding efficiency, and 2) an *adaptive redundancy control* (ARC) scheme for yielding optimal trade-off between coding efficiency and error resilience as follows.

For the first contribution, the proposed PCP-MDC begins with a simple yet novel approach to first partition each quincunx downsampled description into two sub-descriptions, called the *primary sub-description* (PS) and *dual sub-description* (DS). The PS is encoded by exploiting the H.264/AVC intra coding first, and the DS is subject to prediction coding based on the reconstructed PS. For prediction, the *mode-guided directional prediction* (MDP) algorithm is developed to conduct a fast and accurate prediction for the DS.

For the second thrust contribution, two fundamental issues regarding the inserted redundancy are addressed in the proposed ARC scheme: *quality* and *quantity*. For the quality issue, the residuals of the invented *cross* prediction are inserted into each description. For the quantity issue, the amount of redundancy bits allocated to each description are determined according to the network condition, which is consolidated in terms of the probability of channel failure for being incorporated

into our mathematical formulation to derive optimal redundancy allocation condition. To implement the derived optimal condition, a fast estimation algorithm of the rate-distortion function is then developed, followed by exploiting successive approaching algorithm to identify the optimal partition of the target total bitrate splitting between the primary part of the description and its complementary part (i.e., the redundancy part). Furthermore, a special post-processing filter, called the *switching Gaussian filter* (SGF), is developed to facilitate the MDC system for image *enhancement*. Extensive simulation results have consistently shown that the proposed MDC system significantly outperforms other state-of-the-art MDC methods on coding performance with much lower computational complexity.

In addition to the MDC system, the second milestone contribution of this thesis is on the development a novel and generic image interpolation method, called the *contrast-guided* image interpolation. In this work, one of the most important informations to the human visual system—contrast—is optimally incorporated into the interpolation process through the developed *contrast-guided decision map* (CDM), which is a binary map and generated by applying the diffusion process to the directional variation field. The generated CDM is used to guide the interpolation process, and the local contrasts are better preserved compared with that of using the edge map (or, edge-guided decision map) for guiding the image interpolation process. Simulation results have clearly shown that the proposed contrast-guided interpolation method significantly outperforms other state-of-the-art image interpolation methods with much reduced computational complexity.

Chapter 1

Introduction

1.1 Multiple Description Coding (MDC)

With the rapid growth of communications and networking systems, there is an increasing demand for multimedia services such as IPTV and peer-to-peer content distribution. However, one inherent problem with any communications system is that information may be lost during transmission due to channel errors incurred by noise and/or interference. The effect of such information loss is quite devastating for the transport of *compressed* image/video, which is an essential service of multimedia communications nowadays, since a small amount (even a bit) of random alteration on the encoded bitstream during transmission could lead the received bitstream partially or even totally undecodable and cause enormous visual distortion at the receiver. For that, *multiple description coding* (MDC) [1] has been considered as an effective *error-resilient* source coding methodology to combat the issues incurred during the transmission of the encoded bitstreams over unreliable networks.

The basic idea of the MDC is to generate multiple *representations* (or the so-called *descriptions*) from the source data, and these descriptions are then encoded

Figure 1.1. A typical framework of two-channel MDC scheme, which contains three decoders: one central decoder and two side decoders.

and transmitted over different channels independently. The errors incurred in different channels are assumed independent, thus the probability of all channels are in error simultaneously is fairly small. At the destination, if all the descriptions are received correctly, the best quality of the source signal will be reconstructed. On the other hand, if only one or a subset of descriptions are successfully received, the receiver is still able to reconstruct the source signal with a reasonably good or acceptable quality. In other words, the more descriptions the receiver gets, the better quality of the original source can be reconstructed.

A typical MDC system contains only two channels and thus three decoders as shown in Fig. 1.1. If both channels work well without errors, the central decoder attempts to recover the source from the two correctly received descriptions. In this case, any distortion incurred in the reconstructed signal is due to both descriptions, which is thus called the *central* distortion. If one channel is received (say, Channel 1) and the other fails (thus, Channel 2), only the side decoder 1 is able to attempt the reconstruction of the original source signal based on the description in Channel 1 received. In this case, any distortion presented in the reconstructed signal is totally due to the received description from Channel 1; hence, such distortion is called the *side* distortion. In this thesis, the same two-channel MDC system is also assumed as shown in Fig. 1.1. Our major efforts will be on the development of

an efficient multiple description image coding system which will be evaluated on various aspects—namely, coding efficiency (i.e., rate-distortion performance), error resilience, network adaptability and computational complexity.

1.2 Image Interpolation

Spatial resolution is an important metric and attribute of digital image; the higher the spatial resolution, the more image details can be perceived by the human visual system (HVS). Although the sensor technology of digital cameras, camcorders and scanners has achieved tremendous advancement in the past years, the trend of demanding higher spatial resolution images continues across a wide range of image applications in medical imaging, remote sensing, digital photographs, and so on. Image interpolation becomes a practical and feasible solution on achieving the above-mentioned goal through algorithmic and software approach to construct a high-resolution (HR) image from the given, relatively speaking, low-resolution (LR) image. By far, the linear interpolation techniques, such as bilinear interpolation [2], bicubic spline interpolation [3] and bicubic convolution interpolation [4], are the most well-known and widely-exploited image interpolation techniques. These methods have advantages in algorithmic simplicity for fast implementation. However, they were developed mainly from the standpoint of fulfilling mathematical approximation without considering or incorporating any information that is important to the HVS, such as edge, contrast, and so on. Consequently, the interpolated image often suffers from blurring and other annoying artifacts especially introduced around edges. Therefore, for maintaining edge sharpness and improving overall image quality, it is desirable to develop a more efficient image interpolation method, which consider to incorporate the HVS information into the interpolation process.

1.3 Main Contributions

In this thesis, there are two parts: a novel MDC system (Chapter 2 and 3) and an innovative image interpolation method (Chapter 4). In the first part, the proposed MDC system consists of two thrust contributions: (1) a novel *prediction-compensated polyphase* MDC (PCP-MDC) scheme and (2) *adaptive redundancy control* (ARC). For the first contribution, two stages are:

- In PCP-MDC, different from typical methods, each quincunx-downsampled description is further partitioned into two sub-descriptions, called the *primary sub-description* (PS) and *dual sub-description* (DS). The PS is encoded by the H.264/AVC intra coding, while the DS is subject to prediction coding based on the reconstructed PS.
- For conducting the prediction of DS based on PS, the *mode-guided directional prediction* (MDP) algorithm is developed to achieve a fast and accurate prediction.

For the second thrust contribution, an *adaptive redundancy control* (ARC) scheme is developed, which consists of two key components to address two fundamental issues regarding the inserted redundancy information (or bits): *quality* and *quantity*.

- For addressing the quality issue, we propose the *description-compensated redundancy generation* (DRG). The DRG inserts an efficient redundancy information into each description, which is the prediction residuals of one quincunx-downsampled description (i.e., the complementary part), resulted by *cross* prediction based on the other description (i.e., the primary part).

- For addressing the quantity issue, we propose the *network-aware optimal redundancy control* (NORA) scheme. In this scheme, the amount of redundancy bits allocated to each description is determined according to the network condition; for that, the probability of channel failure is considered and incorporated into our mathematical formulation on deriving the optimal redundancy allocation condition. To implement the derived optimal condition, a fast algorithm for the estimation of the rate-distortion function is then developed, followed by exploiting successive approaching algorithm to identify the optimal partition of the target bitrate between the primary part and the complementary part (i.e., the redundancy part) of the description.

We also developed a post-processing method, called the *switching Gaussian filter* (SGF). It is specifically designed for facilitating the proposed MDC system on the aspect of image enhancement. It is able to effectively detect and attenuate the *granular* artifacts that tend to present in the reconstructed image when both descriptions are encoded at *low* bitrates and received in the receiver.

In the second part, the proposed *contrast-guided* image interpolation technique novelly incorporates image *contrast* on the design of an image interpolation algorithm. How to generate a *contrast-guided decision map* (CDM) of the given input image and exploit it to guide the interpolation process is described in this thesis in detail.

1.4 Organization

The outline of this thesis is as follows. In Chapter 2, a comprehensive review of the classical *multiple description coding* (MDC) and *polyphase downsampling* MDC (PD-MDC) algorithms is introduced. Two downsampling methods for the PD-MDC framework are also studied and analyzed.

In Chapter 3, the proposed MDC system is described in detail, which consists of two parts: the *prediction-compensated polyphase* MDC (PCP-MDC) scheme and the *adaptive redundancy control* (ARC) scheme. The framework of PCP-MDC is described in Section 3.1. The *mode-guided directional prediction* (MDP) algorithm for prediction between the *primary* sub-description (PS) and the *dual* sub-description (DS) is developed and described in detail. Extensive simulation results show that the proposed PCP-MDC framework significantly outperforms existing PD-MDC-based methods on the aspects of the *rate-distortion* (R-D) performance and the computational complexity. The ARC method is described in Section 3.2, which is proposed to provide a network-adaptive and optimal redundancy allocation scheme for the PCP-MDC. The frameworks of the *description-compensated redundancy generation* (DRG) and the *network-aware optimal redundancy control* (NORA) are described. A mathematical framework is formulated with incorporation of channel failure probability in order to reflect network conditions. The optimal condition is then derived. A fast R-D function estimation algorithm and a successive approaching algorithm for realizing the derived optimal redundancy allocation are then developed. The *switching Gaussian filter* (SGF) is described in Section 3.3 as a post-processing method for the proposed MDC system.

In Chapter 4, the *contrast-guided* image interpolation is described in detail. In Section 4.1, an overview of existing image interpolation methods is presented. Section 4.2 discusses the fundamental issues and concepts regarding edge and contrast to provide intuitions and justifications on our proposal—using the binary CDM for guiding the image interpolation process. The framework of our proposed contrast-guided algorithm is described in Section 4.3. The generation of CDMs through the diffusion of four directional derivative fields followed by thresholding is then detailed. In Section 4.5, extensive simulations were conducted to evaluate the performance of the proposed interpolation methods to demonstrate the efficiency.

Finally, Chapter 5 summarizes this thesis and discusses promising future directions for the multiple description image coding.

Chapter 2

An Overview of Multiple Description Coding (MDC)

2.1 Review of Existing MDC Methods

The concept of multiple descriptions was first proposed by Gersho, Witsenhausen, Wolf, Wyner, Ziv, and Ozarow at the 1979 IEEE Information Theory Workshop [5]. For a two-description and three-decoder case, El Gamal and Cover [5] gave the achievable *rate-distortion* (R-D) performance for a memoryless source. Ozarow demonstrated the tightness of this rate-distortion region for memoryless Gaussian sources [6]. This was followed by other important development of theoretical results on MDC that can be found in [7–9]. In the following, we shall succinctly highlight three classes of existing MDC methods: 1) *multiple description scalar quantization* (MDSQ), 2) frequency-domain (DCT or wavelet) approach, and 3) polyphase downsampling MDC scheme.

The first practical MDC scheme, known as MDSQ, was introduced by Vaishampayan [10, 11]. It constructs the index information of the quantized coefficients to

meet certain MDC's objectives. A simple implementation can be realized by exploiting two quantizers whose decision regions have a shift with respect to each other by one half of the quantization step size. The design of MDSQ was further extended to vector quantization by Vaishampayan *et al.* [12] denoted as the *multiple description lattice vector quantization* (MDLVQ). Bai and Zhu [13] applied the MDLVQ for the wavelet image coding and designed a more efficient multiple description image coder. In this image coder, the directional correlation of the wavelet coefficients in different frequency subbands are taken into account. The parameters of the vector quantizer are then optimized (in the sense of minimizing rate-distortion measurement) based on the varying importance of different subbands. Experimental results have shown that this approach outperforms the optimal MDSQ image coder [14]. More sophisticated quantizers can be designed to improve the coding efficiency. For example, Hafarkhani and Tarokh [15] developed a trellis-coded quantizer to construct multiple descriptions that yields a remarkable performance improvement at the expense of adding a small amount of encoding complexity.

An alternative way of designing the MDC algorithms is on the frequency domain based on the *discrete cosine transform* (DCT) or wavelet transform. In the *multiple description transform coding* (MDTC) [16–20], each description consists of a set of DCT coefficients that are made being correlated to each other through a *correlating transform* [16] (e.g., the *lapped orthogonal transform* [19]) so that the missing coefficients of the lost description can be estimated based on the coefficients located on the same corresponding position and its adjacently available ones from the other description that has been correctly received. In [16], Batllo and Vaishampayan presented a theoretical analysis of the MDTC method for a Gaussian source with memory and a first-order Gauss-Markov source, respectively. In [18], Wang *et al.* proposed an MDTC scheme by using the *pairwise correlating transform*, in which all the DCT coefficients are grouped in pairs, and each pair is multiplied

by a 2×2 transform matrix for introducing correlation (as injected redundancy information). The optimal (in the sense of minimizing rate-distortion measurement) pairing strategy and the transform matrix are derived in [18]. The two transformed DCT coefficients from each pair are then assigned to different descriptions, respectively. Although the correlating transform introduces redundancy information to help the estimation of the lost description when only one description is received, experimental results show that fairly large amount of estimation errors could be incurred at the receiver [18]. To remedy this deficiency, in [21], the prediction residuals are inserted into each descriptions, so that each description includes not only correlating-transformed DCT coefficients but also estimation errors (i.e, prediction residuals). This strategy greatly helps reduce the distortion of the reconstructed image when only one description is received at the decoder.

The descriptions can also be generated in the wavelet domain. The MDTC work proposed in [21] was applied to the JPEG2000 framework as demonstrated in [22], where a two-channel filter bank was exploited to achieve better rate-distortion performance and reduce the computational complexity. Jiang and Ortega [23] proposed an MDC scheme, called the *polyphase transform and selective quantization* (PTSQ), in which all wavelet transform coefficients are grouped into two parts by polyphase downsampling, and each part is then encoded to form the one description. To introduce redundancy into each description for strengthening its error resilience, a more coarsely quantized version of the other description is added into the current description, and vice versa. Compared to the MDSQ and MDTC, the design complexity is much lower, since both polyphase downsampling and selective quantization can be implemented easily. A similar approach is developed in [24] for generating more than two descriptions by exploiting the Set Partitioning in Hierarchical Trees (SPIHT) framework [25]. Cai *et al.* [26] proposed another PTSQ scheme, called the *structure unanimity-based* MDC (SUMDC). In this method, each wavelet coefficient is

downsampled bit-wise to produce two coefficients; one contains odd bits and the other even bits. These coefficients are then encoded independently to form two descriptions. Moreover, a new data packetization scheme was proposed in [26] to further strengthen its error-resilient ability for the packet switching networks. This work was further enhanced in [27] by exploiting the advantages from both SUMD-C and MDSQ. In [28], the *rate-distortion-based* MDC (RD-MDC) was proposed and applied to the JPEG2000 framework for conducting performance evaluation. In this method, the input image is partitioned into blocks, and each block is encoded twice at two different rates. The encoded blocks are then split into two subsets as two descriptions with similar rate distortion characteristics by exploiting the Lagrangian rate allocation function [29]. Simulation results based on the test image “Lena” have shown about 0.5 dB improvement in the side-distortion performance compared to that in [23] and 1 dB improvement to that in [10], respectively, but at the expense of yielding high computational cost. The framework in [30] combines the time-domain lapped transform [31], block level source splitting, Wiener filter based prediction, and the encoding of the prediction residual. The joint optimization of all components is formulated [30]. The framework in [32] is similar to that of the PTSQ [23]. The key improvement of this method is that the coarse quantizers in different descriptions of this method are designed to be mutually refinable, in order to reduce the distortion when some descriptions are lost.

The third class of MDC is called the *polyphase downsampling multiple description coding* (PD-MDC) [33–38] with a general framework as shown in Fig. 2.1. In this approach, descriptions are generated by applying a polyphase downsampling process in the *spatial* domain of the source data using either the *interlaced* downsampling pattern [Fig. 2.2(a)] or the *quincunx* downsampling pattern [Fig. 2.2(b)] to generate two descriptions. The correlations between the corresponding pixels from the two descriptions are utilized by the encoder to improve the robustness of

Figure 2.1. Typical framework of the *polyphase downsampling multiple description coding* (PD-MDC) scheme.

Figure 2.2. Two ways of partitioning an original image into two descriptions.

bitstream transmission against channel errors. Since our developed MDC system belongs to this class, we shall investigate the PD-MDC method further in the next section.

2.2 Polyphase Downsampling MDC (PD-MDC)

2.2.1 Review of the PD-MDC

Early studies have demonstrated some advantages offered by the PD-MDC-based approaches [33–36, 39, 40]; especially, when one description is damaged or lost, the strong *inter-description correlations* between the two descriptions generated from the original source can be well utilized to reconstruct the source image at the decoder. This indicates that the PD-MDC is inherently equipped with strong error-resilience capability. Moreover, compared with other MDC methods, the PD-MDC approach generally yields lower computational complexity. (Note that the computational time spend on the formation of the two descriptions is normally considered as fairly small and thus negligible.) However, the PD-MDC has following drawbacks: (1) degraded coding efficiency, and (2) lack of redundancy control.

Since the downsampling process can greatly deteriorate the sample correlation within each description, the coding efficiency of the PD-MDC gets an obvious reduction. For example, the results presented in [33] based on the test image “Lena” show that, compared with directly encoded with JPEG, the PD-MDC has yielded more than 3 dB loss in coding efficiency. For that, the method in [40] was proposed to improve the rate-distortion performance of the traditional PD-MDC method [33], where the two descriptions are generated by the quincunx downsampling. In addition, the adaptive directional lifting-based transform was proposed to efficiently compress the quincunx-downsampled descriptions.

The redundancy control of the MDC is to adjust the amount of the inter-description correlation that makes a trade-off between the coding efficiency and the error resilience. To be more specific, the increase in the inter-description correlation will improve error resilience capability of the system (i.e., increasing the quality of the decoded signal when only one description is received), but at the

Figure 2.3. The pre-processing step proposed in [36] for the redundancy control.

expense of decreasing the coding efficiency of the MDC system (i.e., decreasing the quality of the decoded signal, if both descriptions are received), and vice versa. Therefore, with an effective redundancy control, the MDC system can be designed in a more flexible and robust way to deal with different networks (e.g., the Internet, mobile networks). In the PD-MDC, the inter-description correlation between two corresponding samples from different descriptions is related to their relative positions (i.e., the distance between them) in the source image. Therefore, due to this non-adjustable distance for a given downsampling method, the inter-description correlation of PD-MDC can not be adjusted. To overcome this problem, oversampling the source image is proposed in [36] as a pre-processing step and incorporated into the traditional PD-MDC scheme as shown in Fig. 2.3. The oversampled image is, in fact, equivalent to the original image digitized at a higher sampling rate. Consequently, the distances between the corresponding samples from different descriptions are shorter, and thus these samples are more correlated. Through that, the amount of redundancy can be controlled by the oversampling factor. The algorithm was further generalized to an arbitrary number of descriptions by Franchi *et al.* [37]. Similar considerations can be found in [38] where the original image is expanded either vertically or horizontally in the *one-dimensional* (1-D) *discrete cosine transform* (DCT) domain, depending on which direction yields less distortion in the reconstructed image at the decoder.

Table 2.1
THE AVERAGE EUCLIDEAN DISTANCE BETWEEN TWO ADJACENT SAMPLES

Downsampling method	Average distance	
	$l = 4$	$l = 8$
Interlaced	1.50	1.87
Quincunx	1.41	1.71

2.2.2 Study of Image Downsampling for the PD-MDC

There are two different ways to partition an input image $\mathbf{X}$ into two parts $\mathbf{X}_1$ and $\mathbf{X}_2$ in the spatial domain as shown in Fig. 2.2—(a) the interlaced downsampling, and (b) the quincunx downsampling. Since the performance of PD-MDC intimately depends on the downsampling method used, the quality of decoded signal based on two descriptions (i.e., the so-called *central performance*) or that from only one description (i.e., the so-called *side performance*) is separately studied in detail as follows.

The central performance of these two downsampling methods can be evaluated in terms of *intra-description correlation*, which can be quantitatively reflected from the average Euclidean distance [40] measured between the current pixel with respect to its neighboring samples within the same description, since the shorter the distance, the stronger the correlation. The Euclidean distance is defined as,

$$D(n, m) = \frac{1}{l} \sum_{(i, j) \in S} \sqrt{(n - i)^2 + (m - j)^2}, \quad (2.1)$$

where (n, m) is the pixel coordinate of a description's sample under concern, S is a set of neighboring pixels of $x(n, m)$, and l is the total number of the neighboring pixels in S . Fig. 2.2 shows the pixel $x(n, m)$ and its nearest neighboring pixels (i.e., $l = 8$ in this case). The results are documented in Table 2.1, from which we can see that the intra-description correlation of the quincunx downsampling method is stronger than that of the interlaced downsampling method. This suggests that the

quincunx downsampling pattern is more compact for achieving better compression.

For the side performance, the interlaced downsampling method only exploits either the horizontal [as shown in Fig. 2.2 (a)] or the vertical correlations, while the quincunx downsampling method can make better use of the correlations coming from both vertical and horizontal directions simultaneously to recover the missing description.

In conclusion, the above analysis clearly leads us choosing the quincunx downsampling method for the generation of multiple descriptions, which is to be further discussed in the next chapter.

2.3 Summary

In this chapter, a comprehensive review of the existing MDC methods in the literature is presented. In particular, advantages and disadvantages of the PD-MDC methods are highlighted, and two downsampling methods, i.e., the quincunx downsampling and interlaced downsampling, for the PD-MDC method are analyzed and compared, since the performance of PD-MDC intimately depends on the way to conduct downsampling method at its starting point. By computing the average Euclidean distances of the quincunx and the interlaced sampling lattices, it has been observed that the quincunx lattice has a smaller average distance between adjacent samples than that of the interlaced lattice. From the viewpoint of sample correlation measured in terms of consecutive sample distance.

In the next chapter, we will propose the a new polyphase MDC system, which consists of two parts: 1) a new polyphase MDC scheme, called the *prediction-compensated polyphase* MDC (PCP-MDC) for providing a state-of-the-art coding efficiency performance, and 2) an *adaptive redundancy control* (ARC) scheme for yielding optimal trade-off between coding efficiency and error resilience.

Chapter 3

Prediction-compensated Polyphase MDC with Adaptive Redundancy Control

3.1 Proposed Prediction-compensated Polyphase MDC (PCP-MDC)

3.1.1 Proposed Polyphase Downsampling Method

In this thesis, we focus on the polyphase MDC approach (e.g., [33–36, 39, 40]), since they provide attractive coding performance while yielding much lower computational complexity. In the preparation of polyphase source descriptions, there are two kinds of downsampling lattice patterns that can be employed for partitioning the input source image $\mathbf{X}$ into two descriptions $\mathbf{X}_1$ and $\mathbf{X}_2$ over the spatial domain: (a) the *interlaced* downsampling and (b) the *quincunx* downsampling as demonstrated in Fig. 2.2, respectively. In the previous section, we have mathematically show that the average distance of the most adjacently surrounding samples in the

quincunx downsampling lattice is less than that of the interlaced downsampling pattern. This leads to a more compact signal representation for achieving better source compression. Furthermore, when only one description is received, the interlaced downsampling method can only exploit adjacent samples obtained either from the horizontal direction or from the vertical direction. On the other hand, the quincunx downsampling method has advantage of utilizing the correlations coming from both vertical and horizontal directions simultaneously to recover the missing description at the decoder. This motivates us to choose the quincunx downsampling method over the interlaced downsampling approach for the generation of polyphase multiple descriptions.

Note that each description resulted by the quincunx downsampling is not directly suitable for the current image/video encoding system (e.g., the H.264/AVC, the JPEG2000), due to the fact that the samples of each description collectively do not bear a rectangular-shaped image data field. A straightforward way to bypass this problem is to shift all the even rows (or the odd rows) by one-pixel position upward (or downward), for example, as conducted in [36]. However, this process inevitably dislocates the relative positions of the image content originally presented in the source image and creates unwanted high-frequency artifacts in the image. To address this concern, a novel downsampling method is proposed in this thesis, as follows.

Instead of performing pixel shifting operations aforementioned, each quincunx-downsampled description is further partitioned into two disjoint equal-sized sub-descriptions, called the *primary sub-description* (PS) and the *dual sub-description* (DS), respectively, as graphically illustrated in Fig. 3.1; i.e., $\mathbf{X}_1 = \mathbf{X}_{1p} \cup \mathbf{X}_{1d}$ and $\mathbf{X}_2 = \mathbf{X}_{2p} \cup \mathbf{X}_{2d}$ without altering the pixels' original relative positions. One can observe that all the sub-descriptions are now in rectangular shape and ready for exploiting either the H.264/AVC or the JPEG2000 directly.

Figure 3.1. Four sub-descriptions generated from an original image by further partitioning each quincunx-downsampled description into two sub-descriptions; the *primary sub-description* (PS) and the *dual sub-description* (DS).

3.1.2 Framework of the Proposed PCP-MDC

The framework of the proposed *prediction-compensated polyphase* (PCP-MDC) scheme is shown in Fig. 3.2, where the H.264/AVC *intra* coding [41] is chosen as the image codec in our work. Although the H.264/AVC intra coding was originally designed for intra-frame video coding, it has been shown that it outperforms the JPEG and the JPEG2000 in terms of objective evaluation and subjective quality assessment for image coding [42,43]. Moreover, the decoding process of the H.264/AVC decoder is much faster than that of the JPEG2000 decoder [42]. Hence, the H.264/AVC intra coding is adopted here for still image compression.

Figure 3.2. The framework of the proposed PCP-MDC scheme.

Refer to Fig. 3.2, the input image is first partitioned into four sub-descriptions $\mathbf{X}_{1p}$, $\mathbf{X}_{1d}$, $\mathbf{X}_{2p}$, and $\mathbf{X}_{2d}$ as shown in Fig. 3.1. Note that since these sub-descriptions are equally important, without loss of generality, let $\mathbf{X}_{1p}$ and $\mathbf{X}_{2p}$ be the PSs, while $\mathbf{X}_{1d}$ and $\mathbf{X}_{2d}$ be the DSs. Two PSs, $\mathbf{X}_{1p}$ and $\mathbf{X}_{2p}$, are separately encoded by exploiting the H.264/AVC intra encoder. Instead of encoding the DSs, $\mathbf{X}_{1d}$ and $\mathbf{X}_{2d}$, by the H.264/AVC intra encoder directly, the *inter sub-description prediction* is developed and used to encode the DSs as a more efficient coding approach. To be more specific, the reconstructed sub-description $\tilde{\mathbf{X}}_{1p}$ (of $\mathbf{X}_{1p}$) is obtained through a local decoding loop performed at the encoder and then employed for conducting the prediction of $\mathbf{X}_{1d}$; likewise, $\tilde{\mathbf{X}}_{2p}$ (of $\mathbf{X}_{2p}$) is reconstructed for the prediction of $\mathbf{X}_{2d}$. As a result, the *residual* (or prediction error) *sub-descriptions* $\mathbf{E}_{1d}$ (of the DS $\mathbf{X}_{1d}$) and $\mathbf{E}_{2d}$ (of the DS $\mathbf{X}_{2d}$) are obtained. A simplified mathematical expression can be established as

$$\mathbf{E}_{1d} = \mathbf{X}_{1d} - P(\tilde{\mathbf{X}}_{1p}) \quad (3.1)$$

and

$$\mathbf{E}_{2d} = \mathbf{X}_{2d} - P(\tilde{\mathbf{X}}_{2p}), \quad (3.2)$$

where $P(\cdot)$ denotes the inter sub-description prediction process. To illustrate, Fig. 3.3 shows a 4×4 block of $\mathbf{X}_{1d}$ (denoted by the white circles), which is to be predicted based on $\tilde{\mathbf{X}}_{1p}$ (the shaded squares). To predict the pixel $\mathbf{X}_{1d}(n, m)$, the four nearest neighboring samples of the corresponding position on $\tilde{\mathbf{X}}_{1p}$ are employed. To conduct this prediction accurately, a simple and novel prediction method, called the *mode-guided directional prediction* (MDP), is proposed and described in Sub-section 3.1.3.

The residual sub-descriptions $\mathbf{E}_{1d}$ and $\mathbf{E}_{2d}$ are subsequently encoded by the DCT, *quantizer* (Q) and *context-adaptive binary arithmetic coding* (CABAC) [44] as the residual encoding conducted in the H.264/AVC standard. The DCT unit used here is a 4×4 integer DCT [45], and the resultant DCT coefficients are

Figure 3.3. Illustration of the inter sub-description prediction of a 4×4 block in $\mathbf{X}_{1d}$. The white circles represent the samples of $\mathbf{X}_{1d}$ to be predicted, while the shaded squares denote the samples of the reconstructed sub-description $\tilde{\mathbf{X}}_{1p}$ used for conducting the inter sub-description prediction of $\mathbf{X}_{1d}$. To predict the sample $\mathbf{X}_{1d}(n, m)$ in the block, the inter sub-description prediction exploits its nearest four neighboring samples from the sub-description $\tilde{\mathbf{X}}_{1p}$.

quantized by using the quantization parameter defined in the H.264/AVC encoder [45]. Finally, the bitstreams of encoded sub-descriptions $\mathbf{X}_{1p}$ and $\mathbf{E}_{1d}$, denoted as $B(\mathbf{X}_{1p})$ and $B(\mathbf{E}_{1d})$ in Fig. 3.2, respectively, are grouped together to form one description. Likewise, the bitstreams of encoded $\mathbf{X}_{2p}$ and $\mathbf{E}_{2d}$ [i.e., $B(\mathbf{X}_{2p})$ and $B(\mathbf{E}_{2d})$, respectively] are grouped together to form another description.

The decoder of the PCP-MDC decodes the bitstreams of the two descriptions independently, if both are received. For each description (say, Description 1 in Fig. 3.2), the decoder partitions it to identify bitstreams $B(\mathbf{X}_{1p})$ and $B(\mathbf{E}_{1d})$ first. The bitstream $B(\mathbf{X}_{1p})$ is decoded by the H.264/AVC intra decoder, and $B(\mathbf{E}_{1d})$ is subsequently decoded by CABAC decoder, *inverse quantizer* (IQ) and *inverse DCT* (IDCT) to recover $\tilde{\mathbf{E}}_{1d}$. The lossy DS $\tilde{\mathbf{X}}_{1d}$ is then obtained based on $\tilde{\mathbf{X}}_{1p}$ and $\tilde{\mathbf{E}}_{1d}$; that is, $\tilde{\mathbf{X}}_{1d} = P(\tilde{\mathbf{X}}_{1p}) + \tilde{\mathbf{E}}_{1d}$, where the $P(\cdot)$ performs the inter sub-description prediction which is the same process as the one performed in the local decoding loop at the encoder. The decoded $\tilde{\mathbf{X}}_{1p}$ and $\tilde{\mathbf{X}}_{1d}$ are then combined to reconstruct $\tilde{\mathbf{X}}_1$ (Fig. 3.1). For the other description (i.e., $\tilde{\mathbf{X}}_2$), it will be conducted in the same way as described above.

3.1.3 Mode-guided Directional Prediction (MDP)

The goal of the proposed MDP is to predict the DS based on the pixels of the *reconstructed* PS by exploiting the *prediction mode* information that has been generated on the H.264/AVC intra coding of PS conducted earlier. Note that the prediction mode information reveals the ‘*block direction*,’ meaning that the majority of edges contained within the block have the same direction as indicated by the mode determined for that block. Therefore, each sample of DS should be predicted along this block direction based on those samples on the reconstructed PS. To be more specific, the way to conduct the block partition for the DS $\mathbf{X}_{1d}$ completely follows the same block partition as conducted in the PS $\mathbf{X}_{1p}$ during the H.264/AVC intra encoding of $\mathbf{X}_{1p}$. For each partitioned block, one prediction mode has been determined for the chosen block size¹— 16×16 , 8×8 , or 4×4 . For a 16×16 block, one of the four prediction modes—the vertical mode (i.e., Mode 0), the horizontal mode (i.e., Mode 1), the DC mode (i.e., Mode 2), and the plane mode (i.e., Mode 3, for 45° direction), will be determined. For an 8×8 or a 4×4 block, there are nine different prediction modes that can be selected from in each case, with eight prediction directions as illustrated in Fig. 3.4 plus the DC prediction mode (i.e., Mode 2), which is not shown in the figure. (In this case, there is no specific direction.) Based on the prediction mode determined for each block of $\mathbf{X}_{1p}$, the corresponding block in $\mathbf{X}_{1d}$ will be predicted by considering three possible edge directions— 45° and 135° and DC (equivalently, 360°) as shown in Table 3.1. Note that when either Mode 0 (for 90° direction) or Mode 1 (for 0° direction) are determined for a block of $\mathbf{X}_{1p}$, the corresponding block in $\mathbf{X}_{1d}$ is supposed to be predicted along 90° or 0° , respectively. However, in these cases, the *immediate* neighboring samples

¹Note that the block size and prediction mode are determined by the exhaustive search [41] only for the establishment of Table 3.1, while a fast block-size decision and mode decision algorithm is normally employed for encoding optimization. [46, 47].

Figure 3.4. Nine prediction modes used in the H.264/AVC intra 8×8 prediction or 4×4 prediction (Mode 2 as the DC mode is not shown).

Table 3.1

BLOCK DIRECTION FOR EACH PARTITIONED BLOCK IN THE DS, ACCORDING TO THE H.264/AVC PREDICTION MODE INFORMATION OF THE PS OBTAINED PREVIOUSLY.

Block size (of PS)		Block direction (of DS)
16×16	8×8 and 4×4	
Mode 3	Modes 3, 7, 8	45°
—	Modes 4, 5, 6	135°
Modes 0, 1, 2	Modes 0, 1, 2	DC ($= 360^\circ$)

along these two directions are not available for the sample under prediction, due to the inherited lattice structure of the quincunx downsampling pattern used for generating sub-descriptions (see Fig. 3.1 and Fig. 3.5). Our experimental results show that in these cases the *2-D bicubic interpolation* method [2] is able to deliver good prediction. Therefore, Mode 0 and Mode 1 are grouped together with Mode 2 as shown in Table 3.1. Note that the block direction 135° is not available for the 16×16 block, since 135° is not concerned in the H.264/AVC for this block size.

To illustrate, Fig. 3.5 shows the prediction of a sample $\mathbf{X}_{1d}(n, m)$ based on the indicated block direction. As shown in Fig 3.5 (a), for the 45° -direction mode, it means that the block contains edges with 45° direction in majority. Therefore,

Figure 3.5. Illustration of the proposed *mode-guided directional prediction* (MDP) algorithm. Depending on the prediction mode determined at the encoding stage of sub-description $\mathbf{X}_{1p}$, the prediction of the sample $\mathbf{X}_{1d}(n, m)$ is to be conducted through one of the three block directions.

the prediction of $\mathbf{X}_{1d}(n, m)$ should be conducted based on four nearest neighboring samples on $\tilde{\mathbf{X}}_{1p}$ along 45° direction by exploiting the 1-D *cubic spline interpolation* method [48]. The prediction for the 135° -direction mode [Fig. 3.5 (b)] will be conducted in the same way as that of the 45° -direction mode. The DC mode [Fig. 3.5 (c)] indicates that no distinct edge direction presents in the block, and thus the 2-D *bicubic interpolation* method should be used. Note that since the edge direction information of our proposed MDP method is totally guided by the prediction mode information generated during the encoding of PS, thus it enjoys a fairly low computational complexity on the prediction process.

3.1.4 Performance of the PCP-MDC

In this thesis, both the *central* rate-distortion performance (i.e., when both descriptions are received) and the *side* rate-distortion performance (i.e., when only one description is received), measured in terms of PSNR (in dB) versus the total bitrate (in bpp), are provided for each method under evaluation and comparison. The simulation results of the proposed PCP-MDC method are compared with that of the traditional PD-MDC method [37], in which the two descriptions are generated by the *interlaced* downsampling. Each description is then independently encoded

(a) Lena

(b) Splash

(c) Peppers

(d) Goldhill

Figure 3.6. Four test images used for simulation experiments.

by exploiting the H.264/AVC intra encoder. Fig. 3.6 displays four test images (with a size of 512×512 each) used in our simulation. From the results as shown in Fig. 3.7, one can observe that although the central rate-distortion performance by using the proposed PCP-MDC method and the traditional PD-MDC method are similar to each other, our proposed method significantly outperforms the traditional PD-MDC approach in the side rate-distortion performance conducted for all the test images. Another advantage of the proposed method is that the computational complexity is fairly low compared with that of the traditional PD-MDC approach. Note that the two descriptions (each with *half* of the original image in

Figure 3.7. Objective comparisons between the proposed PCP-MDC and the conventional PD-MDC method [37] on the *central distortion* and the *side distortion* in PSNR (dB) versus the total bitrate (bpp), respectively.

size) in PD-MDC are encoded by the H.264/AVC intra encoder independently, and each requires an exhaustive search of 22 intra prediction modes. In our proposed PCP-MDC, on the other hand, only two PSs (each with *one-quarter* of the original image in size) are encoded by the H.264/AVC intra encoder. For encoding the two DSs (also one-quarter of the original image in size), the proposed low-complexity MDP algorithm is exploited, and the prediction errors will be encoded by DCT and CABAC.

Although the proposed PCP-MDC can achieve attractive performance on coding efficiency, it would be highly desirable, especially from the practical point of view, to adaptively allocate proper amount of redundancy (for error resilience) based on the network/channel conditions feedbacked from the network, so as to provide *optimal* trade-off between error-resilience and coding efficiency. To tackle this issue, we shall present another thrust contribution: *adaptive redundancy control* (ARC) method in the following section. The developed ARC will be incorporated with the proposed PCP-MDC to form the entire MDC system for conducting extensive simulations on test images (see Section 3.4).

3.2 Proposed Adaptive Redundancy Control (ARC)

3.2.1 Overview

The purpose of allocating redundancy to all the descriptions is to enhance the capability of *error resilience* of the MDC system but at the expense of sacrificing its *coding efficiency*. Consequently, when only one description is received, the amount of redundancy bits inserted will be beneficial to the quality of the decoded signal.

Therefore, it would be highly desirable to have an optimal and adaptive redundancy allocation mechanism for the MDC system. Furthermore, from the practical point of view, network condition should be considered in the design of redundancy allocation algorithm. To our best knowledge, this is completely missing in the existing spatial-domain-based PD-MDC works. For that, the probability of channel failure is incorporated in our mathematical formulation and derivation, which is to be detailed in Section 3.2.3.

It has been observed that the redundancy of the traditional PD-MDC framework solely relies on the inter-description correlation which is limited by the pixel distance in the original source image. Consequently, the inter-description correlation of PD-MDC is inherently fixed, since the pixel distance is fixed for a given downsampling pattern. To overcome this problem, oversampling the source image is proposed in [36] as a pre-processing step for the traditional PD-MDC scheme. The oversampled image, in fact, can be regarded as the original image digitized at a denser sampling grid. Consequently, the distance between two corresponding samples from different descriptions is reduced, making the samples more correlated and thus increasing their inter-description correlation; thus, the amount of redundancy can be controlled by the oversampling factor imposed. The algorithm was further generalized to an arbitrary number of descriptions by Franchi *et al.* [37]. Similar considerations can be found in [38] where the original image is oversampled either in the vertical direction or in the horizontal direction over the corresponding *one dimensional* (1-D) DCT domain. In order to determine the best oversampling direction, the yielded reconstruction errors computed for both directions are compared at the encoder side.

The above-mentioned oversampling approach only controls the introduced redundancy amount (quantity wise) to some extent, since the imposed oversampling

factor is the only control mechanism exploited for this purpose. Besides, the redundancy information (quality wise) introduced through oversampling is also limited and inevitably incurs unwanted interpolation error. Also, the oversampling approach incurs higher computational complexity at the encoder. Furthermore, none of these schemes were designed with consideration of dynamic network conditions. Note that the dynamic network conditions have been considered in some frequency-domain MDC systems. For example, in [22], the probability of channel failure is taken into consideration, the optimal redundancy allocation is derived by minimizing the side distortion based on the proposed filter banks. However, the development of a spatial-domain optimal redundancy allocation by considering the dynamic network conditions is completely missing. To address these deficiencies, a redundancy allocation approach, called *adaptive redundancy control* (ARC), is proposed in this thesis and described in the following sub-sections.

3.2.2 Framework of the Proposed Description-compensated Redundancy Generation (DRG)

The proposed ARC method addresses two fundamental issues: 1) how to generate more effective redundancy information—*quality* issue; and 2) how to incorporate dynamic network conditions into consideration for *optimally* determining the amount of redundancy to be inserted—*quantity* issue. For the former, we propose the *description-compensated redundancy generation* (DRG) scheme, in which we begin with duplicating the original source image $\mathbf{X}$ as shown in Fig. 3.8, and each $\mathbf{X}$ is quincunx-downsampled into two parts, as follows.

Let the two parts, $\mathbf{X}_1^p$ and $\mathbf{X}_2^p$, partitioned from one source image $\mathbf{X}$ be the *primary* parts, while the $\mathbf{X}_1^c$ and $\mathbf{X}_2^c$ partitioned from the same source image $\mathbf{X}$ be the *complementary* parts. The primary parts $\mathbf{X}_1^p$ and $\mathbf{X}_2^p$ are encoded by exploiting

Figure 3.8. Generation of the *primary* and *complementary* parts as the input source for the proposed *description-compensated redundancy generation* (DRG) scheme.

the proposed PCP-MDC as described in Section 3.1. The encoding of $\mathbf{X}_1^c$ and $\mathbf{X}_2^c$ is subject to predictive coding, similar to that of DSs in the PCP-MDC scheme described previously, except that the prediction is performed in a *cross* prediction manner. That is, the reconstructed primary part $\widetilde{\mathbf{X}}_1^p$ of Description 1 is employed for conducting the prediction of the complementary part $\mathbf{X}_2^c$ of Description 2; likewise, $\widetilde{\mathbf{X}}_2^p$ is employed for the prediction of $\mathbf{X}_1^c$. The cross-prediction *residual* of the complementary part $\mathbf{X}_2^c$ (denoted as $\mathbf{E}_2^c$) is further partitioned into two rectangular-shaped sub-descriptions, and each one is further encoded by the DCT, quantizer, and CABAC. Finally, the bitstreams of the encoded $\mathbf{X}_1^p$ and of the encoded $\mathbf{E}_2^c$, denoted as $B(\mathbf{X}_1^p)$ and $B(\mathbf{E}_2^c)$ in Fig. 3.9, respectively, are grouped together to form the bitstream of Description 1 and transmitted over a communications channel. Likewise, the bitstreams of the encoded $\mathbf{X}_2^p$ and of the encoded $\mathbf{E}_1^c$ (i.e., the cross-prediction residual of the complementary part $\mathbf{X}_1^c$) are grouped together to form the bitstream of Description 2 [i.e., $B(\mathbf{X}_2^p) + B(\mathbf{E}_1^c)$] for being transmitted over another channel. At the receiver, if only one description is received, besides the primary part of the current description, it has the complementary part that was generated from the other description through a prediction process at the encoder and served as the redundancy part for reconstructing the original image with improved image quality

at the decoder. Note that the complementary (or redundancy) part is introduced in each description, which can be flexibly controlled by adjusting the quantization step size, rather than tempering on the oversampling factor as employed in conventional PD-MDC discussed previously.

Figure 3.9. The framework of the proposed adaptive redundancy control (ARC), which consists of two parts: 1) *description-compensated redundancy generation* (DRG) for quality control, and 2) *network-aware optimal redundancy allocation* (NORA) for quantity control. Refer to Table 3.2 for the direction of prediction guided by the MDP. The quantizer (Q) controls both quality and quantity.

Table 3.2

MODE-GUIDED DIRECTIONAL PREDICTION (MDP). THE BLOCK DIRECTION OF EACH PARTITIONED BLOCK IN THE COMPLEMENTARY PART IS DETERMINED ACCORDING TO THE H.264/AVC PREDICTION MODES OF THE PRIMARY PART DETERMINED PREVIOUSLY.

Block size used for the primary part		Block directions determined for the complementary part
16×16	8×8 and 4×4	
Mode 1	Modes 1, 6, 8	0°
Mode 0	Modes 0, 5, 7	90°
Modes 2, 3	Modes 2, 3, 4	DC

It is important to note that although the prediction process conducted in Fig. 3.9 also exploits the proposed MDP algorithm, the possible direction modes are only limited to 0° , 90° and DC (in case of ‘no clear direction’), as shown in Table 3.2, since the available neighboring samples around a pixel under prediction are all situated along either 0° direction or 90° direction; otherwise, the DC mode will be assumed.

At the receiver, the aim is to reconstruct the compressed version of the original source; i.e., $\tilde{\mathbf{X}}$. If both descriptions are correctly received and decoded, only the primary parts (i.e., $\tilde{\mathbf{X}}_1^p$ and $\tilde{\mathbf{X}}_2^p$) will be decoded and used for the reconstruction of $\tilde{\mathbf{X}}$, and the two complementary parts (i.e., $\tilde{\mathbf{X}}_1^c$ and $\tilde{\mathbf{X}}_2^c$) will be simply discarded without decoding. On the other hand, if only one description is received (say, Description 1), then $\tilde{\mathbf{X}}$ can only be built up based on the received bitstream of Description 1, containing the bitstreams of $\tilde{\mathbf{X}}_1^p$ and $\tilde{\mathbf{E}}_2^c$ (i.e., $B(\tilde{\mathbf{X}}_1^p)$ and $B(\tilde{\mathbf{E}}_2^c)$).

In the remaining sub-sections, we shall propose the *network-aware optimal redundancy allocation* (NORA) scheme (Fig. 3.9) to address the second fundamental issue of the proposed ARC algorithm—redundancy allocation guided by network condition.

3.2.3 Proposed Network-aware Optimal Redundancy Allocation (NORA)

3.2.3.1 Formulation of Rate-distortion Functions

Consider the transmission of two descriptions over two different channels (see Fig. 3.9), and assume that each channel has a feedback information from the network regarding the probability of channel failure², ρ . Further let D_c and D_s be the *central distortion* and the *side distortion*, respectively. These distortions are measured in terms of the *mean squared error* (MSE), which is the distortion of the reconstructed source yielded at the receiver. Hence, our aim is to minimize the *average total distortion* $\overline{D}$ incurred at the receiver, which can be formulated as

$$\overline{D} = (1 - \rho)^2 D_c + 2\rho(1 - \rho) D_s. \quad (3.3)$$

Denote D_p and D_r as the MSE caused by the lossy compression of the primary part and of the complementary part incurred in each description, respectively. Note that the D_p from both descriptions are considered being equal, since they are generated from the quincunx-downsampled description of the same input source and processed by the same encoding algorithm. Likewise, this is also true for D_r . As mentioned earlier, if both descriptions were received, only the primary parts from both descriptions will be used for source reconstruction, while the complementary parts will be discarded without decoding. In this case, the *central* distortion D_c is caused by the lossy compression on the primary parts; that is, $D_c = 2D_p$. On the other hand, if only one description was received, both the primary part and the complementary part of the received description will be decoded and used for

²Note that a more general setting is to have different failure probabilities ρ_1 and ρ_2 for the feedback channels 1 and 2, respectively. However, for the simplicity of mathematical derivation, we set $\rho_1 = \rho_2 = \rho$ in this work by reasonably assuming that these channels are from the *same* network and therefore undergoing the same conditions in the long-run stationary sense, such as available channel bandwidth, network traffics, and so on.

source reconstruction; therefore, the *side* distortion D_s becomes $D_s = D_p + D_r$. Substituting D_c and D_s into (3.3), we have

$$\begin{aligned}\overline{D} &= 2(1 - \rho)^2 D_p + 2\rho(1 - \rho)(D_p + D_r) \\ &= 2(1 - \rho)D_p + 2\rho(1 - \rho)D_r.\end{aligned}\tag{3.4}$$

Denote R_p and R_r (in bpp) as the bitrates of the primary part and of the complementary part incurred in each description, respectively. Furthermore, it is reasonable to assume that the bitrates generated from both descriptions are considered being equal. Hence, the total bitrate R of the two encoded descriptions can be computed as

$$R = 2R_p + 2R_r.\tag{3.5}$$

Therefore, (3.4) can be expressed in terms of the *rate-distortion* (R-D) functions as

$$\overline{D}(R) = 2(1 - \rho)D_p(R_p) + 2\rho(1 - \rho)D_r(R_r),\tag{3.6}$$

where $D_p(R_p)$ and $D_r(R_r)$ are the R-D functions resulted by encoding the primary part and the complementary part at the bitrates of R_p and R_r , respectively.

3.2.3.2 Derivation of Optimal Condition

We assume that the above-mentioned R-D functions in (3.6), $\overline{D}(R)$, $D_p(R_p)$ and $D_r(R_r)$ are convex, strictly decreasing, and differentiable everywhere. Note that this assumption is quite reasonable and has been widely exploited in the image and video coding applications, such as JPEG2000 [29] and H.264/AVC [49]. Based on this assumption, the global minimum of $\overline{D}(R)$ can be found by optimally determining the value of R_p (or R_r). Hence, by taking the derivative of $\overline{D}(R)$ with respect to R_p in (3.6) and using the chain rule, it can be arrived at

$$\frac{\partial \overline{D}(R)}{\partial R_p} = 2(1 - \rho) \frac{\partial D_p(R_p)}{\partial R_p} + 2\rho(1 - \rho) \frac{\partial D_r(R_r)}{\partial R_r} \frac{\partial R_r}{\partial R_p}.\tag{3.7}$$

By setting $\frac{\partial \bar{D}(R)}{\partial R_p} = 0$ in (3.7) and exploiting (3.5), the condition of optimality can be arrived at

$$D'_p(R_p) = \rho D'_r(R_r). \quad (3.8)$$

It is interesting to observe that the network condition, in terms of the probability of channel failure ρ , is equivalent to the ratio of two first-derivative R-D functions for the primary part and the complementary part. Note that the derivative of an R-D function, $D'(R)$, at each quantization parameter q can be approximated by computing the slope of a secant line passing through two adjacent R-D performance points $[D(q-1), R(q-1)]$ and $[D(q+1), R(q+1)]$. Therefore, the derivatives in (3.8) can be approximated by

$$D'_p(R_p) \approx \frac{D_p(q_p+1) - D_p(q_p-1)}{R_p(q_p+1) - R_p(q_p-1)}, \quad (3.9)$$

$$D'_r(R_r) \approx \frac{D_r(q_r+1) - D_r(q_r-1)}{R_r(q_r+1) - R_r(q_r-1)}. \quad (3.10)$$

The functions $D_p(q_p)$ and $R_p(q_p)$ and $D_r(q_r)$ and $R_r(q_r)$ in (3.9) and (3.10), in general, can be straightforwardly established by repeatedly executing the encoding process with different values (integers only) of the quantization parameters q_p and q_r (say, with six pre-determined values for each parameter). The obtained rate and distortion results are then used to establish the R-D functions by connecting these performance points. Obviously, such repeated executions would inevitably incur much higher computational complexity. In the following, we shall present a fast estimation method to establish these functions with much less computation.

3.2.3.3 Fast Estimation of $D_p(q_p)$, $R_p(q_p)$, $D_r(q_r)$, and $R_r(q_r)$

As mentioned earlier, $D_p(q_p)$ and $D_r(q_r)$ can be easily obtained, since the distortion is only introduced by the quantization process without involving other parts of the encoding system. On the other hand, the estimation of $R_p(q_p)$ and $R_r(q_r)$ are much

more difficult, since the resultant bitrate can only be obtained on the completion of the entire encoding process. In fact, the relationship between the bitrate R and the quantization q is *nonlinear*. To tackle this issue, He and Mitra [50] have demonstrated that, for a typical image/video transform coding system, the relationship between the bitrate R and the percentage of zero DCT coefficients (denoted as parameter α in [50]) after quantization is approximately a *linear* function. Obviously, the measurement of parameter α is fairly straightforward and easy by simply counting how many zero DCT coefficients appeared in the quantized DCT domain. Furthermore, since the parameter α monotonically increases with the quantization parameter q , there exists a one-to-one mapping between them. Therefore, instead of directly estimating the $R(q)$ function, they proposed to establish the $R(\alpha)$ function first, and then set up a lookup table between α and q for linking $R(\alpha)$ to $R(q)$.

According to [50], we find that such a linear relationship also exists in the encoding of the primary part and the complementary part; that is, the bitrates R_p and R_r are linearly proportionated with the percentages of non-zero DCT coefficients, $(1 - \alpha_p)$ and $(1 - \alpha_r)$, respectively. One might consider that $R_p(\alpha_p)$ should pass through a performance point (100%, 0), since when the percentage of zero DCT coefficients is 100% (i.e., $\alpha_p = 100\%$), the total bitrate R_p yielded should be zero. However, we find that: 1) by quantizing all DCT coefficients to zero, the resultant total bitrate is *not* necessarily to be exactly zero but around 0.01 bpp on average due to the overhead bits introduced for the formation of the bitstream; and 2) the amount of these overhead bits is nearly the same for different pictures. Therefore, the straight line is modified by passing through the performance point (100%, 0.01) instead.

To determine another performance point on this line, we only need to encode the primary part *once* by using a fixed quantization parameter q_p . For that, $q_p = 23$ is chosen in this work, and the resultant α_p and R_p can be obtained subsequently.

With the two performance points (100%, 0.01) and (α_p, R_p) , we can immediately construct the linear function $R_p(\alpha_p)$, followed by setting up a one-to-one lookup table between q_p and α_p ; hence, $R_p(q_p)$ is obtained. Note that, the lookup table can be easily built up by repeatedly running the quantizer part only, since for each quantization parameter q_p , the corresponding α_p can be obtained by simply counting the number of zero DCT coefficients. Since conducting the quantization only needs a division operation applied to each coefficient, therefore the computational complexity is greatly reduced when compared with repeated-execution approach that needs to complete the entire encoding process on each run. The function $D_p(q_p)$ can be formulated by conducting quantizing and inverse quantizing of the DCT coefficients at different values of q_p . For each value of q_p , the total distortion D_p can be obtained by simply computing the MSE between the DCT coefficients before and after conducting the quantization process.

The estimations of $R_r(q_r)$ and $D_r(q_r)$ are conducted in the same way as the estimation of $R_p(q_p)$ and $D_p(q_p)$ described previously, except that the $R_r(\alpha_r)$ function is modeled as a straight line passing the point (100%, 0). This is because that when all the DCT coefficients are quantized to zero, we only need to transmit a flag to inform the decoder that there is no complementary part. To find another point for establishing the straight line $R_r(\alpha_r)$, we use $q_r = 32$ in this work to encode the complementary part to obtain its resulted α_r and R_r .

To appreciate how accurate of the estimation of $R_p(q_p)$ and $R_r(q_r)$ using our proposed fast estimation method, a test image “Lena” is experimented for demonstration as presented in Fig. 3.10 and Fig. 3.11. One can see that the estimated functions $R_p(q_p)$ and $R_r(q_r)$ using the proposed method are very close to the actual functions obtained by exploiting the repeated-execution approach, which is computational extensive. Finally, we shall use our estimated functions to realize the condition of optimality (3.8) in the following.

Figure 3.10. Comparisons between the ‘actual’ and the proposed fast estimated functions for $R_p(q_p)$ based on the test image “Lena.” The ‘actual’ function is generated by repeatedly running the entire encoder six times.

Figure 3.11. Comparisons between the ‘actual’ and the proposed fast estimated functions for $R_r(q_r)$ based on the test image “Lena.” The ‘actual’ function is generated by repeatedly running the entire encoder six times.

3.2.3.4 Successive Approaching Algorithm

After the establishment of the functions $R_p(q_p)$, $D_p(q_p)$, $R_r(q_r)$ and $D_r(q_r)$, the computation of the condition of optimality as derived in (3.8) can be obtained via (3.9) and (3.10), respectively. Note that the derived function (3.8) is a necessary condition to achieve *optimal* R-D performance, given the total bitrate R (subjected to (3.5)) and the probability of channel failure, ρ . This is equivalent to find an *optimal* partition, $2R_p^*$ and $2R_r^*$ such that $D'_p(R_p^*) = \rho D'_r(R_r^*)$ and $R = 2R_p^* + 2R_r^*$. For that, we propose the following successive approaching algorithm based on the assumption that an R-D function is convex, and thus a global minimum solution that satisfies the condition (3.8) is guaranteed. The algorithm works out as follows.

Initially, the amount of R_p is set to $R/2$ for one description (i.e., all the bits are allocated to the primary part, and consequently the amount of redundancy $R_r = 0$). The value of R_p is then gradually decreased, while increasing R_r for maintaining (3.5), until (3.8) is completely or nearly satisfied. Note that we only need to find q_p^* and q_r^* for one description. For the other description, we will encode it by simply using the same parameters, since both descriptions are considered equal or ‘balanced’, thus having the same R-D behavior.

3.3 Proposed Post-processing: Switching Gaussian Filter (SGF)

Due to the operation of polyphase downsampling process, the consecutive samples from the source image are allocated into different descriptions and then independently encoded. As a consequence, consecutive samples are quite likely subject to different quantizations and yield different amount of quantization errors. This will inevitably cause disturbing artifacts (coined as the *granular noise* in this thesis)

Figure 3.12. A portion of the reconstructed “Lena” image coded at 0.25 bpp for two description together with or without exploiting the proposed *switching Gaussian filter* (SGF).

appeared on the reconstructed images especially encoded at low bitrates [see Fig. 3.12 (a) for a demonstration]. In order to detect and attenuate this kind of granular noise, a simple and effective filter, called *switching Gaussian filter* (SGF), is proposed in this thesis and applied to the reconstructed image when both descriptions are received, as follows.

Denote $\tilde{x}(n, m)$ as a decoded image sample based on both descriptions received. The proposed SGF method examines the absolute-valued differences between the current sample $\tilde{x}(n, m)$ and its eight neighboring samples of the decoded image. Note that these differences could be caused either by edges presented in the image or by the quantization errors, or by both of them. We further observe that a difference is usually quite small if it is simply due to the quantization errors. To filter granular noise while preserving edge, it would be intuitive to assume that only non-edge pixels with small differences are subject to be filtered. To identify these non-edge pixels, an effective approach is employed as follows. If all the absolute-valued differences, $|\tilde{x}(n, m) - \tilde{x}(i, j)|$, are smaller than a pre-determined threshold μ for all $(i, j) \in D$, where the region of support D is a 3×3 window centered on the current sample $\tilde{x}(n, m)$, then $\tilde{x}(n, m)$ will be filtered by a 2-D Gaussian

Table 3.3
PSNR PERFORMANCE (IN dB) WITH OR WITHOUT USING THE SGF FOR THE TEST
IMAGE “LENA” ENCODED AT DIFFERENT BITRATES

Total bitrate (bpp)	PCP-MDC		
	without SGF	with SGF	Improvement
1.03	38.70	39.00	0.30
0.50	34.56	35.08	0.52
0.25	31.20	31.77	0.57

filter; otherwise, $\tilde{x}(n, m)$ remains unchanged without being filtered. For the 2-D Gaussian filtering, a 3×3 coefficient mask is applied, and the coefficient values are determined by the 2-D Gaussian function $G(x, y) = \frac{1}{2\pi\sigma^2} \exp(-\frac{x^2+y^2}{2\sigma^2})$, where $\sigma = 0.35$ is empirically determined. The threshold μ is set as $\mu = 1.5\Delta_p$, where Δ_p is the quantization step size for the primary part, which reflects the magnitude of quantization errors.

The resulted PSNR with and without using our proposed SGF post-filtering operation is shown in Table 3.3 for the test image “Lena” encoded at different low bitrates, in which the SGF is individually applied to the reconstructed image of the PCP-MDC method decoded from the two descriptions. One can see that the lower the bitrate, the higher the post-filtering improvement delivered by using our SGF. To appreciate the visual quality improvement, a portion of the reconstructed “Lena” image is shown in Figure 3.12 (b).

3.4 Simulation Results

The performance evaluation of the proposed MDC system, consisting of our two thrust contributions PCP-MDC and ARC, has been conducted by using the JM 11.0 H.264/AVC intra encoder [51]. The H.264/AVC encoder is configured as follows: 1) the *rate-distortion optimization* (RDO) is enable; 2) CABAC entropy coding is

used. Fig. 3.13 displays eight test images (with a size of 512×512 each) used in our simulation.

Our proposed MDC system is compared with the works proposed in [37], [38] and [22]. To introduce controllable redundancy in [37], the original image is oversampled over 2-D DCT domain. Specifically, the original image is first DCT transformed as a whole, followed by inserting zeros in the DCT domain and then by performing IDCT transformed. The oversampled image is then subject to interlaced downsampling to generate two descriptions, and each is independently encoded by the H.264/AVC intra coding. On the other hand, the method proposed in [38] oversamples the source image either in the vertical direction or in the horizontal direction over the 1-D DCT domain. In order to select a better oversampling direction, the yielded reconstruction errors obtained from both directions are individually computed at the encoder side. Then, the source image is oversampled along the direction with smaller reconstruction errors. The oversampled image is subject to *interlaced* downsampling to generate two descriptions, which are independently encoded by the H.264/AVC intra coding. The method in [22] is a state-of-the-art frequency-domain-based method, which exploited the JPEG2000 image coding [29]. In this thesis, the side R-D performance (in PSNR) is plotted versus the central R-D performance (in PSNR) for each method under evaluation and comparison. To conduct a fair comparison, the total bitrate for each method is fixed, and two bitrate values are considered: 2 bpp (i.e., 1 bpp for each description as shown in Fig. 3.14) and 1 bpp (i.e., 0.5 bpp for each description as shown in Fig. 3.15). Note that the feedback channel is not considered in these experiments, since we focus on the performance evaluation of DRG (Fig. 3.9) for the moment. The advantages of the proposed MDC system can be appreciated by noticing that better side PSNR values are mostly obtained under the same central PSNR. As an example, for the “Splash” image in Fig. 3.14 (a), if we fix the central PSNR, the side PSNR of the

(a) Splash

(b) Peppers

(c) Lena

(d) Goldhill

(e) Woman

(f) Plane

(g) Baboon

(h) Building

Figure 3.13. Eight test images used for simulation experiments.

proposed MDC system is able to achieve about 7 dB improvement over that in [37], 3 dB over that in [38], and 2 dB over that in [22], respectively. Note that, for the test image “Lena,” the method of [38] and [22] provide comparable or slightly better results than our proposed method. This is because the vertical correlation of the image “Lena” is much stronger than that in horizontal direction. Therefore, in this case, utilizing the strong correlation from only one direction (as done in the method [38]) is enough to recover the missing description at the decoder, and the proposed method can not make full use of the advantage on exploiting the correlation from both directions for recovering the missing description.

Another advantage of the proposed MDC system is the consideration of dynamic network conditions for optimally determining the amount of redundancy; that is, our MDC system guarantees that the expected distortions yielded at the receiver under various probabilities of channel failure ρ are *minimum*. To justify this through simulations, the test image “Lena” is used, and the target total coding bitrate R for two descriptions altogether is set at 2 bpp (i.e., 1 bpp for each description). The coding performance measurement in terms of average PSNR versus redundancy R_r (thus $R_p = R/2 - R_r$) are shown in Fig. 3.16 for three different probabilities of channel failure: (a) $\rho = 5\%$, (b) $\rho = 30\%$ and (c) $\rho = 50\%$. Optimal R_r^* computed by exploiting the proposed system for each case is marked in the respective figure (using symbol Δ). The average PSNR is calculated by averaging all the PSNR results independently obtained from 1000 simulations for each case of ρ . Note that, for computing the average PSNR, we do not count the cases that both channels are in failure. Based on the simulation results, one can see that our proposed system (those marked by points R_r^*) always yields the highest average PSNR results.

In Table 3.4-Table 3.7, the average PSNR results of the proposed MDC system and that in [38] are compared under various channel conditions (i.e., probabilities of channel failure) and the total bitrates. Note that there is no optimal redundancy

(a) Splash

(b) Peppers

(c) Lena

(d) Goldhill

(e) Woman

(f) Plane

Figure 3.14. Objective comparisons among the proposed MDC system and the methods in [37], [38] and [22] in terms of the *side distortion* versus *central distortion*, both measured in PSNR (dB). For a fair comparison, the total bitrate for each method is set as 2 bpp (i.e., 1 bpp for each description).

Figure 3.15. Objective comparisons among the proposed MDC system and the methods in [37], [38] and [22] in terms of the *side distortion* versus *central distortion*, both measured in PSNR (dB). For a fair comparison, the total bitrate for each method is set as 1 bpp (i.e., 0.5 bpp for each description).

Figure 3.16. The average PSNR generated by allocating different amounts of redundancy R_r for the test image “Lena” under three probabilities of channel failure: (a) $\rho = 5\%$, (b) $\rho = 30\%$, and (c) $\rho = 50\%$. The target total bitrate R is 2 bpp. For each case, the amount of redundancy allocated by the proposed NORA method is indicated in the figure by the position of the dashed line; that is, $R_r^* = 0.005$ bpp in (a), $R_r^* = 0.050$ bpp in (b), and $R_r^* = 0.245$ bpp in (c). One can see that they consistently achieve the highest average PSNR result in each case.

Table 3.4

OBJECTIVE COMPARISONS BETWEEN THE PROPOSED MDC SYSTEM AND THE METHOD IN [38] IN TERMS OF THE AVERAGE PSNR (dB) VERSUS THE TOTAL BITRATE (BPP) UNDER PROBABILITY OF CHANNEL FAILURE, $\rho = 1\%$. THE OVERSAMPLING FACTOR OF THE METHOD [38] IS FIXED AT $v_p = 0.5$.

Test images (Grayscale)	$\rho = 1\%$					
	Proposed		Method in [38]		Improvement	
	2 bpp	1 bpp	2 bpp	1 bpp	2 bpp	1 bpp
Splash	48.45	43.20	46.50	42.22	1.95	0.98
Peppers	43.27	39.02	41.31	38.25	1.95	0.77
Lena	43.56	38.89	41.85	38.36	1.71	0.53
Goldhill	39.22	33.51	37.85	33.12	1.37	0.39
Building	40.67	35.74	38.64	34.79	2.03	0.95
Baboon	33.36	27.82	31.29	27.10	2.07	0.72
Woman	43.42	38.86	41.93	37.37	1.49	1.49
Plane	43.62	37.69	41.97	37.22	1.65	0.48
Overall	42.13	37.07	40.35	36.28	1.78	0.79

control mechanism used in [38]. To control the redundancy, the original image (with size of $n \times n$) is oversampled by a fixed factor at $v_p = 0.5$, resulting in an image of size $(1 + v_p)n^2$. From the simulation results, one can see that our proposed method outperforms the method [38] in different network conditions. As an example, for the case of $\rho = 1\%$ and 2 bpp for total bitrate, the proposed method is able to achieve 1.78 dB improvement over the method [38] in terms of the overall average PSNR (i.e., averaging the PSNR results over 8 test images).

Table 3.8 indicates the encoding time required in each method. One can see that the total encoding time required by our system (PCP-MDC and ARC together) is much less than that in [37] and [38]. The incurred computational complexity in [37] and [38] is quite high, since they oversample the input image to introduce redundancy, and thus a larger oversampled image needs to be encoded by the H.264/AVC coder.

Table 3.5

OBJECTIVE COMPARISONS BETWEEN THE PROPOSED MDC SYSTEM AND THE METHOD IN [38] IN TERMS OF THE AVERAGE PSNR (dB) VERSUS THE TOTAL BITRATE (BPP) UNDER FOUR PROBABILITY OF CHANNEL FAILURE, $\rho = 10\%$. THE OVERSAMPLING FACTOR OF THE METHOD [38] IS FIXED AT $v_p = 0.5$.

Test images (Grayscale)	$\rho = 10\%$					
	Proposed		Method in [38]		Improvement	
	2 bpp	1 bpp	2 bpp	1 bpp	2 bpp	1 bpp
Splash	47.27	42.69	45.79	41.88	1.48	0.81
Peppers	42.17	38.56	40.79	37.86	1.38	0.70
Lena	42.49	38.40	41.45	38.14	1.05	0.26
Goldhill	38.28	33.16	37.23	32.85	1.05	0.31
Building	39.10	34.83	37.47	34.09	1.62	0.74
Baboon	32.14	27.32	30.87	26.89	1.28	0.42
Woman	42.32	38.32	41.35	37.11	0.98	1.21
Plane	42.33	37.13	40.99	36.70	1.34	0.43
Overall	40.94	36.51	39.67	35.92	1.26	0.59

Table 3.6

OBJECTIVE COMPARISONS BETWEEN THE PROPOSED MDC SYSTEM AND THE METHOD IN [38] IN TERMS OF THE AVERAGE PSNR (dB) VERSUS THE TOTAL BITRATE (BPP) UNDER PROBABILITY OF CHANNEL FAILURE, $\rho = 20\%$. THE OVERSAMPLING FACTOR OF THE METHOD [38] IS FIXED AT $v_p = 0.5$.

Test images (Grayscale)	$\rho = 20\%$					
	Proposed		Method in [38]		Improvement	
	2 bpp	1 bpp	2 bpp	1 bpp	2 bpp	1 bpp
Splash	46.49	42.33	45.13	41.56	1.36	0.77
Peppers	41.14	38.13	40.30	37.50	0.84	0.63
Lena	41.50	37.95	41.07	37.93	0.44	0.02
Goldhill	37.40	32.83	36.65	32.60	0.75	0.23
Building	38.00	33.97	36.39	33.44	1.61	0.54
Baboon	31.01	26.84	30.47	26.70	0.54	0.14
Woman	41.40	37.82	40.80	36.87	0.60	0.95
Plane	41.29	36.60	40.07	36.22	1.22	0.39
Overall	39.95	36.01	39.04	35.58	0.91	0.43

Table 3.7

OBJECTIVE COMPARISONS BETWEEN THE PROPOSED MDC SYSTEM AND THE METHOD IN [38] IN TERMS OF THE AVERAGE PSNR (dB) VERSUS THE TOTAL BITRATE (BPP) UNDER FOUR PROBABILITY OF CHANNEL FAILURE, $\rho = 50\%$. THE OVERSAMPLING FACTOR OF THE METHOD [38] IS FIXED AT $v_p = 0.5$.

Test images (Grayscale)	$\rho = 50\%$					
	Proposed		Method in [38]		Improvement	
	2 bpp	1 bpp	2 bpp	1 bpp	2 bpp	1 bpp
Splash	45.12	41.83	43.67	40.86	1.45	0.97
Peppers	40.18	37.33	39.23	36.70	0.95	0.63
Lena	40.10	37.19	40.23	37.46	-0.13	-0.27
Goldhill	35.70	32.15	35.38	32.05	0.32	0.10
Building	36.60	32.88	34.00	32.00	2.60	0.88
Baboon	29.55	25.87	29.60	26.27	-0.05	-0.40
Woman	40.04	36.96	39.60	36.33	0.44	0.62
Plane	39.91	35.71	38.05	35.15	1.85	0.56
Overall	38.55	35.18	37.65	34.83	0.90	0.35

Table 3.8

COMPARISONS OF ACTUAL ENCODING TIME (PER FRAME). THE RESULTS WERE OBTAINED BASED ON THE INTEL CORE 2 DUO 1.86 GHz CPU WITH 1 GB DDR2 533 MHz RAM MEMORY. PARAMETER v_p IS THE OVERSAMPLING FACTOR.

Test images	Encoding time (seconds)				
	Method [37]	Method [37]	Method [38]	Method [38]	Proposed
	$v_p = 0.5$	$v_p = 1$	$v_p = 0.5$	$v_p = 1$	
Lena	5.27	7.01	3.25	4.13	1.69
Splash	3.53	6.77	2.82	3.73	1.41
Peppers	5.49	7.23	3.41	4.39	1.70
Goldhill	5.90	7.51	3.79	4.88	1.91

3.5 Summary

In this chapter, we have proposed a novel MDC system. This system consists of two parts which are two major contributions in this thesis: One is the *prediction-compensated polyphase* MDC (PCP-MDC), which significantly outperforms the traditional polyphase-based MDC method. The other is the proposed *adaptive redundancy control* (ARC) framework, which consists of two schemes: 1) *description-compensated redundancy generation* (DRG) for addressing quality control issue on redundancy control, and 2) *network-aware optimal redundancy control* (NORA) for addressing quantity control issue. Furthermore, we have proposed a post-processing method for the MDC system, which is able to effectively remove the granular noise in the reconstructed image when both descriptions are correctly received and encoded at very low bitrate. Finally, through extensive simulations, we have shown that the proposed MDC system (PCP-MDC and ARC together) significantly outperforms the state-of-the-art MDC methods, while providing optimal balance between coding efficiency and error resilience. Moreover, the computational complexity of the entire MDC system is much lower than that of those compared methods.

In the next chapter, we will propose a novel and generic image interpolation method, called the *contrast-guided image interpolation* method to enlarge the concerned image for better visualization.

Chapter 4

Contrast-guided Image Interpolation

4.1 Review of Interpolation Methods

There are many image interpolation methods that have been proposed in the literature. The most well-known and widely-exploited image interpolation techniques are the linear interpolation techniques, such as the bilinear interpolation [2], the bicubic spline interpolation [3] and the bicubic convolution interpolation [4]. These methods have advantages in algorithmic simplicity for fast implementation. However, they were developed mainly from the standpoint of fulfilling mathematical approximation without considering or incorporating any information that is important to the human visual system (HVS), such as edge, contrast, and so on. Consequently, the interpolated image often suffers from blurring and other annoying artifacts especially introduced around edges.

More sophisticated adaptive image interpolation methods [52–65] have been proposed in order to maintain edge sharpness and improve overall image quality. Ramponi [52] introduced the concept of *warped* distance and incorporated it for

determining the bicubic interpolation filter's coefficients. The distances between the pixel under interpolation and its nearest neighboring pixels are individually warped according to the degree of similarity measured based on the intensities of those involved pixels. Experimental results documented in [52] have shown that smaller interpolation errors are yielded compared with that of the bicubic method. Jesen and Anastassiou [53] proposed a scheme that first detects the edges of the LR image at the sub-pixel accuracy. The interpolation of the HR image is then conducted with a guidance from the detected edges. The resulted HR image delivers sharper edges than that of the interpolated image by using the conventional linear interpolation methods. In [54], Li and Orchard exploited an 8×8 window to compute the covariance matrix based on the LR image for each pixel in the HR image under interpolation, since the covariance matrix reveals the direction of local edge. With the covariance matrix, the Wiener interpolation filter's coefficients can be determined by exploiting classic Wiener filtering theory. Alternatively, Zhang and Wu [55] proposed to interpolate a pixel based on multiple directions by using the *linear minimum mean square-error* (LMMSE) estimation method. This approach is able to achieve a comparable quality on the reconstructed HR image when compared with that of the method proposed in [54], while enjoying a much reduced computational complexity. Recently, Zhang and Wu [56] further proposed a *soft-decision interpolation* (SAI) technique which jointly estimates and simultaneously interpolates a block of pixels, rather than one pixel at a time. By combining block estimation and piecewise autoregressive modeling in [56], the SAI technique achieves superior image interpolation results to the methods proposed in [54] and [56]. However, the computational complexity is fairly high and, in fact, even higher than that of [54]. This is due to the fact that besides the need to compute a covariance matrix for each unknown pixel as done in [54], the SAI needs to invert a 12×12 matrix for a block of 12 pixels under simultaneous interpolation.

It has been observed that, for the pixels situated along a sharp edge, their interpolation should be conducted more carefully for maintaining edge sharpness. This goal can be effectively accomplished by conducting their interpolation only along their respective edge direction. Such *edge-guided interpolation* approaches can be found in [53, 62], and [66], for example. On the other hand, the *contrast-guided* approach as proposed in this thesis is intimately related to, but fundamentally different from, this class of methods, since the contrast information is not considered at all in their generated edge maps.

4.2 Fundamental Issues and Ideas

4.2.1 Using A Decision Map for Image Interpolation—The Turnkey Solution

As mentioned earlier, the to-be-interpolated (or ‘missing’) pixels should be interpolated along their respective edge direction (or, alternatively speaking, orthogonal to their respective gradient direction, which is 90° apart from the edge direction). Such *directional* filtering is able to maintain edge sharpness which avoiding possible artifacts. Obviously, if we know the ground truth of the edge direction of each to-be-interpolated pixel somehow, it would be highly advantageous to conduct the interpolation process based on those existing pixels situated along the given edge direction, rather than using a *directionless* (i.e., *isotropic*) filtering that will inevitably blur the image. On the other hand, if the ground truth says that there is no clear gradient direction in that to-be-interpolated pixel location, then it is better off using a directionless filtering method to avoid unwanted artifacts that might be possibly generated by using a directional filtering approach, otherwise. Unfortunately, we do not know the aforementioned ground truth in advance! Therefore, the

whole issue boils down to how to determine whether a directional or directionless filtering should be applied at each to-be-interpolated pixel. Equivalently to say, we need to generate a *decision map* to guide the interpolation process.

The ‘decision map’ generated in the existing edge-guided methods (e.g., [53, 62, 66]) is the so-called *edge map*, which is used for guiding the image interpolation. Consequently, the missing ‘edge’ pixels are interpolated by using the directional filtering (e.g., the 1-D cubic spline interpolation [3]) along the estimated edge direction, while the ‘non-edge’ pixels are conducted by exploiting the directionless filtering (e.g., the 2-D bicubic interpolation). For the ease of referring, we denote the ‘edge map’ used in the edge-guided methods as the binary *edge-guided decision map* (EDM), while coining ours as the binary *contrast-guided decision map* (CDM)¹, since the *contrast* information is incorporated on the generation of CDM. Both maps will be used to guide the interpolation process—1-D cubic interpolation for the directional filtering and 2-D bicubic interpolation for the directionless (or, isotropic) filtering. Extensive simulation results presented later on clearly show that the interpolation results of using the CDM are superior to that of exploiting the EDM. This indicates that more accurate decisions regarding directional versus directionless were made due to the success of incorporating contrast information on the generation of the CDM. At this juncture, let us discuss what is edge and what is contrast, as follows.

4.2.2 Edge Versus Contrast

From the standpoint of the HVS, any edge presented in a typical natural image is resulted by two adjacent uniform regions having sufficiently different levels of pixel-intensity values. That is, the so-called ‘edge’ is perceived by the HVS through

¹Hereafter, we shall drop the term ‘binary’ for both the *binary* EDM and the *binary* CDM throughout the thesis for convenience.

Figure 4.1. (a) The presence of an edge can only be *perceived* by the human visual system once a sufficiently-large contrast is formed across two adjacent uniform regions; the larger the pixel-intensity difference across two regions, the stronger their contrast, which leads to a sharper edge being perceived. (b) The resulted binary edge map (or, edge-guided decision map) by exploiting any edge detection method. Note that the detected edge must be drawn on the boundary of either side of region's borderline. (c) An identical image content as that in (a), except with much less contrast being perceived. However, the same edge-guided decision map as shown in (b) will be obtained by exploiting the same edge detection method.

such pixel-intensity ‘*contrast*’ formation across two adjacent regions. An illustration example is shown in Fig. 4.1 (a), in which a distinct edge is clearly being perceived by our HVS due to a sufficiently-large contrast being created by two uniform regions that are adjacent to each other. Applying any edge detection algorithm to Fig. 4.1 (a) will result in an edge as shown in Fig. 4.1 (b), which is an edge map itself that contains one edge. However, the indicated edge line in Fig. 4.1 (b) must be consistently drawn on the region boundary of either side of the regions. If the Fig 4.1 (a) is subject to a 2×2 image interpolation, then Fig. 4.1 (b) can be used as the EDM to guide the interpolation’s filtering process. Therefore, a typical binary edge map (or, the binary EDM) resulted by exploiting any edge detector can only reveal the approximated locations of contrast boundaries. This observation leads to our view that a *true* edge can only be perceived by the HVS, once a sufficiently large contrast is formed by two adjacent regions. Further note that, the greater the contrast, the stronger the edge will be perceived by the HVS; however, this is not able to be reflected in the EDM. For example, applying the same edge detection algorithm to Fig. 4.1 (c), which has the same image content as that in Fig. 4.1 (a) but with much less contrast, could result in the same EDM as shown in Fig.

Figure 4.2. (a) Illustration of *confidence band* and its *width* through an image consisting of three uniform regions with three different gray levels. (b) The resulted binary *contrast-guided decision map* of (a), reflecting the local contrast; the higher the contrast, the wider the confidence band.

4.1 (b). That is to say the EDM completely ignores the information regarding how strong the edge is. Based on the insights as discussed above, a more accurate term to describe what the so-called ‘edge’ should be coined as *contrast boundary* in our view. Consequently, how to interpolate the LR image without incurring disturbing artifacts (especially around the contrast boundaries) becomes the main goal on the development of the CDM for guiding the follow-up interpolation’s filtering process. This is main point of our image interpolation algorithm.

One should note that the information regarding contrast might not be an issue or could be even irrelevant for certain image processing and computer vision tasks (e.g., object recognition and tracking). But, it is of tremendous importance to image interpolation. Instead of utilizing the EDM as practiced in the existing edge-guided image interpolation methods (e.g., [62] and [66]), we shall generate the CDM for guiding the image interpolation process. The CDM not only indicates the locations of contrast boundaries but also adaptively reflects the amount of contrast for each detected contrast boundary in terms of *confidence band*—symmetrical with respect to the contrast boundary and with a width that reflects the amount of contrast associated with it. To illustrate, Fig. 4.2 (a) shows three adjacent uniform regions with three different levels of pixel intensities as reflected by their gray scales. Fig.

4.2 (b) is a demonstration of a possible CDM which consists of three confidence bands with different band widths; each one reflects the degree of contrast formed in each boundary. That is, for a contrast boundary formed with less contrast, the width of its confidence band will be narrower; otherwise, a wider band will be yielded for the case when the contrast is of more distinct. As a result, the CDM not only conveys the same information as the EDM does, but additionally, conveying the amount of contrast presented in each contrast boundary. The follow-up image interpolation operation will then be guided by the generated CDM (which has the same resolution as that of the LR image) in the following manner: For those LR-image pixels with the same locations as the ones locating on the confidence bands, they will be treated as *contrast-boundary* (or, ‘*edge*’) pixels such that the interpolation process will be only conducted along the indicated contrast-boundary direction. Extensive simulation results clearly demonstrated that this approach is simple and yet effective on maintaining the edge sharpness in the interpolated image. It also greatly mitigates unwanted image artifacts that might be incurred in the interpolated HR image otherwise. Note that edge sharpness and free of artifacts are intimately related from the view point of HVS.

At this juncture, one might be curious how the proposed CDM map could effectively achieve the above-mentioned goals and improves the interpolation. Recall that for edge-guided image interpolation, the pixels will be conducted by directional filtering (for ‘edge’ pixel) or directionless filtering (for ‘non-edge’ pixel) guided by EDM. Now, imagine that there is a strong edge presented, and the interpolation window size is 5×5 . This means that a non-edge pixel (say, “Y”) that is immediately adjacent to an edge pixel (say, “X”) will be inevitably involved on the interpolation process and lead to undesired filtering result with distinct artifacts. That is, the pixels from one side of a strong edge will be viewed as unwanted ‘outlier’ pixels in the eye of the pixel from the other side of the edge. Therefore, if we can implement

a mechanism to re-classify those adjacent non-edge pixels “Y’s”, then each pixel “Y” will be handled in the same way as that of “X” as if it were an edge pixel. Consequently, both pixels will be processed along the same direction as indicated by our CDM. In our work, we consider 0° , 90° , 45° , and 135° only; therefore, four CDMs will be generated according to these four directions respectively. Hence, the generation of CDM is based on the EDM and further decides which surrounding non-edge pixels of the edge pixels are re-classified as edge pixels. Furthermore, the stronger the edge (or the larger the contrast), the more such non-edge pixels will be re-classified as edge pixels. This is the main reason that our contrast-guided approach outperforms those edge-guided methods. In the following, we shall discuss how to generate the CDMs based on the given LR image in detail.

4.3 Framework of Contrast-guided Image Interpolation

Let $I(i, j)$ denote the final interpolated HR image with a size of $2M \times 2N$. With respect to the HR image, the given original LR image can be thus represented as $I_{LR}(i, j) = I(2i - 1, 2j - 1)$, where $1 \leq i \leq M$ and $1 \leq j \leq N$. As depicted in Fig. 4.4, there are two sequential stages of pixel interpolations involved in our approach: In Stage 1, the pixels $I(2i, 2j)$ —which can be viewed as *diagonal* pixels—are interpolated based on the original LR image pixels. In Stage 2, the remaining pixels $I(2i - 1, 2j)$ and $I(2i, 2j - 1)$ —which can be thus viewed as *row* pixels and *column* pixels, respectively—are interpolated based on the original LR image pixels as well as the interpolated diagonal pixels $I(2i, 2j)$ obtained in Stage 1. With that, the framework of the proposed contrast-guided interpolation is shown in Figs. 4.3 and 4.4, which is to be detailed as follows.

Figure 4.3. Framework of the proposed contrast-guided image interpolation, which consists of two sequential stages of pixel interpolation with a lattice arrangement as depicted in Fig 4.4.

Figure 4.4. The proposed two-stage image interpolation, where the hollow circles denote the pixels under interpolation for the HR image and the black dots denote available pixels for conducting the interpolation of those pixels represented in the hollow circles: (a) interpolating the 'diagonal' pixels $I(2i, 2j)$ based on the LR image, and (b) interpolating the 'row' pixels $I(2i-1, 2j)$ and 'column' pixels $I(2i, 2j-1)$ based on the LR image and the interpolated diagonal pixels $I(2i, 2j)$ obtained in Stage 1.

4.3.1 Generation of Directional Variation Fields via Directional Derivative

As depicted in Fig. 4.3, *two* CDMs will be generated in Stage 1 and in Stage 2, respectively, for guiding the interpolation process. A key observation is that in the vicinity of contrast boundary, the variation of pixel-intensity values along the contrast-boundary (or ‘edge’) direction (i.e., in the *tangent* direction along the boundary) is much smaller than that of crossing over the boundary (i.e., in the *normal* direction). Therefore, we need to compute two *directional variation* (DV) values along these two orthogonal directions for each pixel under interpolation to determine whether it is situated on a contrast boundary: If the two DV values have a large difference, the pixel is classified as a contrast-boundary pixel. The specific two orthogonal directions of the above-mentioned are 45° and 135° for the diagonal pixels in Stage 1, and 0° and 90° for the row and column pixels in Stage 2, respectively. The computation of two DV fields is described in detail, as follows.

Given a gray-scaled image $I(x, y)$, which is viewed as a continuous function of two-dimensional spatial variables (x, y) , two DV fields can be computed as

$$\begin{cases} U_\theta(x, y) = |\nabla_\theta I(x, y)|; \\ U_\phi(x, y) = |\nabla_\phi I(x, y)|, \end{cases} \quad (4.1)$$

where $\theta \perp \phi$; ∇_θ and ∇_ϕ perform the directional derivatives along two *orthogonal* directions θ and ϕ , respectively; $|\cdot|$ denotes the absolute value. For the discrete image $I_{LR}(i, j)$, the directional derivative can be implemented by the convolution process [53]. In Stage 1, the 45° -direction and 135° -direction DV fields (denoted as $U_{45^\circ}(i, j)$ and $U_{135^\circ}(i, j)$, respectively) will be generated, since available neighboring pixels around a diagonal pixel $I(2i, 2j)$ are all situated along these two directions only as illustrated in Fig. 4.4 (a). The two DV fields can be therefore obtained by convolving the LR image with a 45° mask and a 135° mask, respectively. For that,

two 3×3 convolution masks, one for each above-mentioned diagonal direction, are proposed in (4.2) and (4.3):

$$U_{45^\circ}(i, j) = \left[\begin{bmatrix} 0 & 1 & 1 \\ -1 & 0 & 1 \\ -1 & -1 & 0 \end{bmatrix} * I_{LR}(i, j) \right]; \quad (4.2)$$

$$U_{135^\circ}(i, j) = \left[\begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & -1 \\ 0 & -1 & -1 \end{bmatrix} * I_{LR}(i, j) \right], \quad (4.3)$$

where $*$ denotes convolution.

Likewise, two DV fields, $U_{0^\circ}(i, j)$ and $U_{90^\circ}(i, j)$, are thus generated in Stage 2 by convolving the LR image with the 0° and 90° masks, respectively, which basically perform the Prewitt edge detection [2], as follows:

$$U_{0^\circ}(i, j) = \left[\begin{bmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{bmatrix} * I_{LR}(i, j) \right]; \quad (4.4)$$

$$U_{90^\circ}(i, j) = \left[\begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{bmatrix} * I_{LR}(i, j) \right]. \quad (4.5)$$

In the next section, we shall exploit the variational approach [67] to diffuse these four DV fields to fulfill the goal on reflecting the degree of contrast for each detected contrast boundary through a confidence band and the width of the band is proportional to the amount of contrast by exploiting the diffusion process. Consequently, more non-edge pixels in the vicinity of contrast boundaries will be re-classified as contrast-boundary pixels; the larger the contrast, the more such non-edge pixels will be re-classified.

4.3.2 Generation of Contrast-guided Decision Maps via Diffusion Process

4.3.2.1 Formulation of the Diffusion Process

Diffusing the DV fields is to realize the objective on proportionally reflecting the degree of contrast for each contrast boundary presented in these DV fields (generated in Section 4.3.1). Let u_θ and u_ϕ respectively denote the *diffused* DV fields of two DV fields U_θ and U_ϕ generated previously². We shall begin with the discussion on the generation of the diffused DV field u_θ as follows. First, an energy functional of u_θ , denoted as $E(u_\theta)$, is formulated as [67]

$$E(u_\theta) = E_d(u_\theta) + \lambda E_s(u_\theta). \quad (4.6)$$

In (4.6), $E(u_\theta)$ consists of two energy terms: the *data* term $E_d(u_\theta)$ and the *smoothness* term $E_s(u_\theta)$. The variable λ is a regularization parameter that provides a tradeoff between $E_d(u_\theta)$ and $E_s(u_\theta)$. Furthermore, since we prefer both $E_d(u_\theta)$ and $E_s(u_\theta)$ to be as small as possible, the energy functional $E(u_\theta)$ is subject to be minimized.

The data term $E_d(u_\theta)$ is used to reflect how well the diffused DV field u_θ has maintained the ‘property’ of the original DV field U_θ after diffusion. The $E_d(u_\theta)$ can be quickly obtained by simply computing the energy of field-intensity difference between u_θ and U_θ . In this work, the data term is given by

$$E_d(u_\theta) = \iint [U_\theta^2(u_\theta - U_\theta)^2] dx dy. \quad (4.7)$$

Note that unlike the conventional least-squares approximation method, i.e., $E_d(u_\theta) = \iint (u_\theta - U_\theta)^2 dx dy$, we further multiply the square of the difference term, $(u_\theta - U_\theta)^2$,

²Hereafter, the function argument (x, y) of two dimensional functions $u_\theta(x, y)$, $u_\phi(x, y)$, $U_\theta(x, y)$, and $U_\phi(x, y)$ shall be dropped for the convenience of mathematics notation and development.

by a factor U_θ^2 in (4.7) to prevent any large variation that originally presented in U_θ from being over-smoothed out. This is desirable since the large-magnituded variations in U_θ indicate that this region contains multiple contrast boundaries. For those locations in U_θ containing large variations, the multiplication factor U_θ^2 will effectively magnify the data term $E_d(u_\theta)$ and thus dominate the energy functional (4.6). In this case, $E(u_\theta)$ can only be minimized by setting $u_\theta = U_\theta$ (i.e., forcing u_θ being ‘no change’) at these locations. By doing so, the properties of U_θ are well maintained. Similar approach has been observed in [68], but the process is performed over the gradient vector field rather than on the DV field. Furthermore, the application goal in [68] is to find the object boundaries for conducting image object segmentation.

On the other hand, the smoothness term $E_s(u_\theta)$ is used to penalize any *unsmoothness* that might be incurred in the diffused DV field u_θ . This is because that in order to diffuse or spread the large variations into their vicinity, we prefer the diffused DV field u_θ being as smooth as possible. The smoothness term, $E_s(u_\theta)$, is given by

$$E_s(u_\theta) = \iint [(u_\theta)_x^2 + (u_\theta)_y^2] dx dy, \quad (4.8)$$

where the $(u_\theta)_x$ and $(u_\theta)_y$ denote the partial derivatives of u_θ with respect to the directional variables x and y , respectively. Note that such incorporation of a smoothness term can be found in the computation of optical flow by Horn and Schunck [69]. By substituting (4.7) and (4.8) into (4.6), we can get the energy functional for u_θ as follows,

$$E(u_\theta) = \iint \{U_\theta^2(u_\theta - U_\theta)^2 + \lambda[(u_\theta)_x^2 + (u_\theta)_y^2]\} dx dy, \quad (4.9)$$

where the parameter $\lambda = 0.2$ is empirically determined in our work. For convenience, we define the integrand function in (4.9) as

$$F[x, y, u_\theta, (u_\theta)_x, (u_\theta)_y] = U_\theta^2(u_\theta - U_\theta)^2 + \lambda [(u_\theta)_x^2 + (u_\theta)_y^2]. \quad (4.10)$$

Using the *calculus of variations* [70], the functional $E(u_\theta)$ is minimized if the integrand function $F[x, y, u_\theta, (u_\theta)_x, (u_\theta)_y]$ (simply denoted as F , in symbols, for mathematical convenience) satisfies the following *Euler-Lagrange differential equation*:

$$\frac{\partial F}{\partial u_\theta} - \frac{d}{dx} \frac{\partial F}{\partial (u_\theta)_x} - \frac{d}{dy} \frac{\partial F}{\partial (u_\theta)_y} = 0. \quad (4.11)$$

By solving (4.11), we get

$$U_\theta^2(u_\theta - U_\theta) - \lambda \nabla^2 u_\theta = 0, \quad (4.12)$$

where ∇^2 is the Laplacian operator [i.e., $\nabla^2 u_\theta = (u_\theta)_{xx} + (u_\theta)_{yy}$]. That is, the diffused DV field u_θ can be optimally obtained from the DV field U_θ by satisfying (4.12).

Similarly, the diffused DV field u_ϕ can be generated from the DV field U_ϕ likewise by satisfying

$$U_\phi^2(u_\phi - U_\phi) - \lambda \nabla^2 u_\phi = 0. \quad (4.13)$$

Note that the closed-form analytical solutions of u_θ and u_ϕ in (4.12) and (4.13), respectively, are difficult to obtain. Therefore, we shall resort to a numerical approach that can iteratively find the solution until a stopping criterion is met, as described in the next sub-section.

4.3.2.2 Numerical Approach on Finding the Diffused DV Fields

Besides the existing two continuous spatial variables x and y , a new continuous-time variable t is introduced into the continuous functions u_θ and u_ϕ in (4.12) and (4.13) for conducting the iterative process. The partial derivatives of functions u_θ and u_ϕ with respect to variable t are then set and equal to the left-hand side of the equal sign in (4.12) and (4.13), respectively, as

$$\begin{cases} \frac{\partial u_\theta(x, y, t)}{\partial t} = U_\theta^2[u_\theta(x, y, t) - U_\theta] - \lambda \nabla^2 u_\theta(x, y, t); \\ \frac{\partial u_\phi(x, y, t)}{\partial t} = U_\phi^2[u_\phi(x, y, t) - U_\phi] - \lambda \nabla^2 u_\phi(x, y, t). \end{cases} \quad (4.14)$$

Finding a solution of $u_\theta(x, y)$ and $u_\phi(x, y)$ in (4.12) and (4.13) is now equivalent to find a *steady-state* solution of $u_\theta(x, y, t)$ and $u_\phi(x, y, t)$ in (4.14), which means that the solutions of $u_\theta(x, y, t)$ and $u_\phi(x, y, t)$ can be obtained when no further noticeable changes of $u_\theta(x, y, t)$ and $u_\phi(x, y, t)$ incurred after some period of time; that is, $\frac{\partial u_\theta(x, y, t)}{\partial t} = 0$ and $\frac{\partial u_\phi(x, y, t)}{\partial t} = 0$. Note that the equations in (4.14) are in the form of *generalized diffusion equation* (or the so-called *generalized heat equation*) [71], which is commonly encountered in modeling the temperature of heat dissipated in a uniform material or the spread of a drop of ink in water [71]. Here, we exploit (4.14) to conduct the diffusion process on the DV fields U_θ and U_ϕ to spread out the directional derivatives.

Now, we investigate how to find the solution of (4.14) for discrete image, where the spatial coordinates of image are now denoted as (i, j) and the time variable t becomes the index of iteration n . For convenience, we further move n to the superscript position with a round bracket; i.e., $u_\theta(i, j, n) \equiv u_\theta^{(n)}(i, j)$. The required first partial derivative and the second partial derivative (i.e., Laplacian) in (4.14) can be calculated via the finite-difference approximation [72]; that is,

$$\begin{cases} \frac{\partial u_\theta(x, y, t)}{\partial t} \approx u_\theta^{(n+1)}(i, j) - u_\theta^{(n)}(i, j); \\ \frac{\partial u_\phi(x, y, t)}{\partial t} \approx u_\phi^{(n+1)}(i, j) - u_\phi^{(n)}(i, j); \\ \nabla^2 u_\theta(x, y, t) \approx u_\theta^{(n)}(i+1, j) + u_\theta^{(n)}(i, j+1) + u_\theta^{(n)}(i-1, j) + u_\theta^{(n)}(i, j-1) - 4u_\theta^{(n)}(i, j); \\ \nabla^2 u_\phi(x, y, t) \approx u_\phi^{(n)}(i+1, j) + u_\phi^{(n)}(i, j+1) + u_\phi^{(n)}(i-1, j) + u_\phi^{(n)}(i, j-1) - 4u_\phi^{(n)}(i, j). \end{cases} \quad (4.15)$$

The iterative solution is obtained by substituting (4.15) into (4.14) and arrives at,

$$\begin{cases} u_\theta^{(n+1)}(i, j) - u_\theta^{(n)}(i, j) = U_\theta^2(i, j) \left[u_\theta^{(n)}(i, j) - U_\theta(i, j) \right] - \lambda \left[u_\theta^{(n)}(i+1, j) \right. \\ \quad \left. + u_\theta^{(n)}(i, j+1) + u_\theta^{(n)}(i-1, j) + u_\theta^{(n)}(i, j-1) - 4u_\theta^{(n)}(i, j) \right]; \quad (4.16) \\ u_\phi^{(n+1)}(i, j) - u_\phi^{(n)}(i, j) = U_\phi^2(i, j) \left[u_\phi^{(n)}(i, j) - U_\phi(i, j) \right] - \lambda \left[u_\phi^{(n)}(i+1, j) \right. \\ \quad \left. + u_\phi^{(n)}(i, j+1) + u_\phi^{(n)}(i-1, j) + u_\phi^{(n)}(i, j-1) - 4u_\phi^{(n)}(i, j) \right]. \quad (4.17) \end{cases}$$

Figure 4.5. The *mean absolute difference* $MAD(n)$ incurred in the n -th iteration for three test images, “Lena,” “Station,” and “Cameraman,” where the $MAD(n)$ is the difference measured between the diffused 45° -direction DV field of the current iteration and that of the previous iteration. One can see that $MAD(n)$ converges when the number of iterations n increases. In addition, $\epsilon = 0.006$ is set as an empirically-determined threshold to flag that the numerical solution on obtaining the diffused DV field is converged, once $MAD(n) < 0.006$.

The initial DV fields used for numerical iterative computation are set to be $u_\theta^{(0)}(i, j) = U_\theta(i, j)$ and $u_\phi^{(0)}(i, j) = U_\phi(i, j)$. Obviously, a stopping criterion is required for reaching the final numerical solution as follows.

The stopping criterion can be set by checking whether the diffused DV fields computed in the current iteration (n) is similar enough to that of the previous iteration ($n-1$). For that, we compute the *mean absolute difference* (MAD) between them; that is,

$$MAD(n) = \frac{1}{MN} \sum_i^M \sum_j^N \left| u_\theta^{(n)}(i, j) - u_\theta^{(n-1)}(i, j) \right|. \quad (4.18)$$

It has been observed that indeed the value of $MAD(n)$ monotonically decreases when the number of iterations, n , increases; this indicates that the diffusion process converges. To demonstrate, three test images “Lena,” “Station,” and “Cameraman” are experimented, and their simulation results are shown in Fig. 4.5. Therefore, once the $MAD(n)$ is smaller than a pre-determined threshold ϵ , the iterative diffusion process can be terminated. Note that all the values of the diffused DV fields

are normalized to the range $[0, 1]$ in our algorithm, and the threshold $\epsilon = 0.006$ is empirically determined.

As expected, the diffused DV field $u_\phi(i, j)$ has similar converging behavior as described previously in the diffused DV field $u_\theta(i, j)$ and can be determined in the same way and applies the same threshold value. Our extensive experiments have shown that, for most of the test images, the computation of the diffused DV fields will be stopped within 10 iterations.

4.3.2.3 Generation of CDMs via Thresholding

The diffused DV fields generated in the previous step are further thresholded in order to produce CDMs as follows. In Stage 1, if $u_{45^\circ}(i, j) - u_{135^\circ}(i, j) > T$, the diagonal pixel $I(2i, 2j)$ will be regarded as a contrast-boundary pixel located in the 135° -direction, since the 135° direction yields a smaller variation. On the 135° -direction CDM, the location of these pixels will be marked in ‘white.’ The parameter $T = 0.02$ is empirically determined and applied in all diffused DV-field thresholding process. On the other hand, if $u_{135^\circ}(i, j) - u_{45^\circ}(i, j) > T$, pixel $I(2i, 2j)$ will be regarded as a contrast-boundary pixel located in the 45° -direction CDM. If $|u_{135^\circ}(i, j) - u_{45^\circ}(i, j)| \leq T$, this indicates that there is no distinct contrast boundary in the vicinity of the pixel $I(2i, 2j)$; thus, it is *not* classified as a contrast-boundary pixel and marked as a ‘black’ pixel in both 45° -direction and 135° -direction CDMs. In Stage 2, the 0° -direction and 90° -direction CDMs can be generated in the same way by using the diffused DV fields $u_{0^\circ}(i, j)$ and $u_{90^\circ}(i, j)$, and the same parameter $T = 0.02$ is applied.

Fig. 4.6 shows the DV fields (generated in Section 4.3.1) and the diffused DV fields (generated in Section 4.3.2) using the test image “Wheel,” and Fig. 4.7 shows the EDMs and the CDMs after applying the aforementioned thresholding process on the DV fields and the diffused DV fields in Fig. 4.6, respectively, where the ‘white’

Figure 4.6. The DV fields (left column) and the *diffused* DV fields (right column) generated from the test image “Wheel.” (a)-(b) 45° direction and (c)-(d) 135° direction obtained in Stage 1; (e)-(f) 0° direction and (g)-(h) 90° direction obtained in Stage 2.

Figure 4.7. EDMs (left column) and CDMs (right column) generated from the test image “Wheel” by exploiting a thresholding process on the DV fields and the diffused DV fields as shown in Fig. 4.6, respectively: (a)-(b) 45° direction and (c)-(d) 135° direction obtained in Stage 1; (e)-(f) 0° direction and (g)-(h) 90° direction obtained in Stage 2.

pixels denote that the LR image pixels located in these positions are considered as contrast-boundary (or ‘edge’) pixel and subject to directional filtering along the indicated direction according to the map generated. As one can observe from Fig. 4.7, each EDM (on the left-hand column) only identifies the pixel locations falling on the contrast boundaries (or ‘edge’) for conducting 1-D directional filtering, while the proposed CDMs (on the right-hand column) further identify more non-edge pixel locations in the vicinity of the contrast boundary for conducting 1-D directional filtering. Obviously, the number of such re-classified non-edge pixels (as edge ones) is proportional to the amount of contrast measured at each detected contrast boundary. Further note that the direction indicated by the individual band in each CDM in Fig. 4.7 is orthogonal to that of the counterpart DV field in Fig. 4.6, respectively. The filtering operation of the process will be guided by the generated CDMs, which is a decision map—indicating whether a directional or directionless filtering should be conducted at each to-be-interpolated pixel. This will be discussed in detail in the next subsection.

4.3.3 Image Interpolation Guided by the Contrast-guided Decision Maps

The image interpolation of the HR image will be conducted with a guidance from the generated CDMs on each stage as depicted in Fig. 4; that is, exploiting the 45°-direction and 135°-direction CDMs for the interpolation of the diagonal pixels in Stage 1, and the 0°-direction and 90°-direction CDMs for the row pixels and the column pixels, respectively, in Stage 2. With reference to the CDMs, for those LR-image pixels with the same locations as the ones falling on the confidence band, the one-dimension (1-D) cubic spline interpolation [3] involving four nearest neighboring pixels along the indicated direction of the CDM will be used to conduct the

interpolation process. For example, if a diagonal pixel is locating on a confidence band in the 45° -direction CDM, its interpolation will be conducted based on four nearest neighboring pixels from the LR image along the 45° direction. For those pixels that are outside of the confidence band, the 2-D bicubic interpolation will be exploited.

4.4 Simulation Results

Extensive simulation experiments are conducted to evaluate the proposed *contrast-guided* image interpolation technique and compare its performance with that of four comparable methods as shown in Table 4.1, including the traditional bicubic interpolation [3]. In particular, we compare our method with the edge-guided method which basically exploits binary EDMs to conduct interpolation adaptively but without considering contrast information. In both cases, either 1-D cubic spline (along the indicated edge direction) or bicubic (when there is no distinct edge direction) will be chosen to apply at each pixel position, depending on whether the pixel on the guiding maps (EDMs for the edge-guided or CDMs for the contrast-guided) is ‘white’ pixel or ‘black’ pixel. Our method is also compared with several state-of-the-art methods—i.e., the *covariance-guided interpolation* method proposed by Li and Orchard [54], and a *linear minimum mean square-error* (LMMSE)-based method proposed by Zhang and Wu [55]. Fig. 4.8 displays eight test images used in our simulation experiments. Each test image is directly downsampled by a factor of 2×2 without conducting any low-pass filtering to get their LR images as the input images to begin with, so that the *peak signal-to-noise ratio* (PSNR) is measurable after the 2×2 interpolation process is completed by using the above-mentioned methods. As shown in the PSNR results (in dB) that reported in Table 4.1, the proposed method achieves the highest PSNR results in all eight test images. For

(a)

(b)

(c)

(d)

(e)

(f)

(g)

(h)

Figure 4.8. Eight test images used for simulation experiments: (a) “Lena” (256×256), (b) “Station” (256×256), (c) “Cameraman” (256×256), (d) “House” (256×256), (e) “Wheel” (300×160), (f) “Parrot” (768×512), (g) “Butterfly” (480×288), and (h) “Bike” (768×512).

Table 4.1

PSNR RESULTS (IN dB) OF THE INTERPOLATED HR IMAGES USING FIVE DIFFERENT INTERPOLATION METHODS, WHERE THE LR IMAGES ARE OBTAINED BY DIRECTLY DOWNSAMPLING THE HR IMAGES. THE HIGHEST PSNR RESULT ACHIEVED IN EACH TEST CASE IS HIGHLIGHTED IN BOLDFACE. FOR THE PROPOSED METHOD, ITS GAIN (IN dB) OVER THE SECOND-BEST METHOD IS ALSO PROVIDED IN THE PARENTHESIS.

Test Images (Gray-scaled)	Bicubic [3]	Edge-guided	Covariance- guided [54]	LMMSE -based [55]	Contrast-guided [proposed]
Lena	30.19	30.33	30.57	30.50	31.08 (0.51 dB)
Station	24.64	25.08	25.04	25.06	26.21 (1.13 dB)
Cameraman	25.25	25.44	25.38	25.55	25.86 (0.31 dB)
House	31.80	32.69	31.84	32.47	32.84 (0.15 dB)
Wheel	19.54	20.19	21.06	19.63	22.33 (1.27 dB)
Parrot	34.77	34.41	34.99	34.60	35.69 (0.50 dB)
Butterfly	26.09	26.31	26.88	26.42	27.64 (0.76 dB)
Bike	25.95	25.72	25.99	26.06	26.75 (0.69 dB)

Figure 4.9. A downsampled test image “Wheel” as shown in (a) is interpolated by a factor of 2×2 by using different interpolation methods: (b) bicubic [3], (c) covariance-guided [54], (d) edge-guided, (e) LMMSE-based [55], and (f) the proposed contrast-guided.

those images containing more high-frequency components such as test images “Station” and “Wheel,” the proposed method can even outperform the second best method by more than 1 dB. Figs. 4.9-4.12 show a set of 2×2 interpolated images for conducting subjective evaluation and comparison. Without considering local edge directions, the bicubic interpolation [3] tends to blur more image details than others, and it also generates more distinct jaggy edges [e.g., see Figs. 4.9 (b) and 4.11 (b) along the rim of the hat in test image “Lena”]. Although the covariance-guided method [54] might yield sharper edges than the bicubic method, distinct ringing artifacts and ‘speckle’-like noises are also presented as shown in Figs. 4.9 (c) and 4.10 (c). The LMMSE-based method [55] achieves comparable interpolation

Figure 4.10. A downsampled test image “Station” as shown in (a) is interpolated by a factor of 2×2 by using different interpolation methods: (b) bicubic [3], (c) covariance-guided [54], (d) edge-guided, (e) LMMSE-based [55], and (f) the proposed contrast-guided.

Figure 4.11. A downsampled test image “Lena” as shown in (a) is interpolated by a factor of 2×2 by using different interpolation methods: (b) bicubic [3], (c) covariance-guided [54], (d) edge-guided, (e) LMMSE-based [55], and (f) the proposed contrast-guided.

Figure 4.12. A downsampled test image “Cameraman” as shown in (a) is interpolated by a factor of 2×2 by using different interpolation methods: (b) bicubic [3], (c) covariance-guided [54], (d) edge-guided, (e) LMMSE-based [55], and (f) the proposed contrast-guided.

results compared to the covariance-guided method [54], but this method sometimes creates jaggy edges, as shown in Figs. 4.9 (e) and 4.10 (e). Obviously, the proposed algorithm delivers the best subjective results with sharper edges and least amount of artifacts.

For the color image (recorded in the RGB color space), each previously-mentioned interpolation method is independently applied to R-, G-, and B-planes. The color test images, “Butterfly,” and “Bike,” are used in our simulation experiments. Each RGB test image is separately downsampled by a factor of 2×2 to get an LR image in order to conduct a 2×2 interpolation for each test image by using these methods. The subjective performance evaluation of the interpolated color images are presented in Fig. 4.13-4.15. One can see that the proposed method also consistently outperforms other methods.

Note that the purpose of downsampling of the original image (without low-pass filtering) is to produce an LR image as the input test image for the experiment of interpolation in order to measure the PSNR of the interpolated image for conducting objective evaluation. However, this downsampling process will inevitably introduce some artifacts due to aliasing and these artifacts will be enlarged through interpolation as well. This will interfere subjective evaluation due to the upsampled artifacts. Therefore, we shall conduct another set of interpolation experiments *without* downsampling the original test image and simply focus on the comparison of the subjective quality of the interpolated image. (Certainly, the PSNR is not measurable in this case.) For that, as shown in Fig. 4.16, the five test images are to be enlarged by a factor of 8×8 through three stages of progressive interpolations—i.e., first 2×2 , then 4×4 , and finally 8×8 . Based on the experimental results as shown in Figs. 4.17-4.21, one can see that our contrast-guided interpolation algorithm avoids most artifacts and delivers better overall image quality when compared to other methods.

Figure 4.13. A downsampled color test image “Butterfly” as shown in (a) is interpolated by a factor of 2×2 by using different interpolation methods: (b) bicubic [3], (c) covariance-guided [54], (d) edge-guided, (e) LMMSE-based [55], and (f) the proposed contrast-guided.

Figure 4.14. A downsampled color test image “Bike” as shown in (a) is interpolated by a factor of 2×2 by using different interpolation methods: (b) bicubic [3], (c) covariance-guided [54], (d) edge-guided, (e) LMMSE-based [55], and (f) the proposed contrast-guided.

Figure 4.15. A downsampled color test image “Parrot” as shown in (a) is interpolated by a factor of 2×2 by using different interpolation methods: (b) bicubic [3], (c) covariance-guided [54], (d) edge-guided, (e) LMMSE-based [55], and (f) the proposed contrast-guided.

Figure 4.16. The test images to be enlarged: (a) “Wheel,” (b) “Tripod,” (c) “Butterfly,” (d) “Cloth,” (e) “Roof.”

Table 4.2

THE INTERPOLATION TIMES (IN SECONDS) CONSUMED IN THREE DIFFERENT METHODS. ALL THESE METHODS ARE IMPLEMENTED IN MATLAB 7.8.0. THE COMPUTER CONFIGURATION USED IN OUR SIMULATION IS THE INTEL CORE 2 DUO 1.86 GHZ CPU AND 4 GB 533 MHZ DDR2 RAM.

Test images	Interpolation time (seconds)		
	Covariance-guided [54]	LMMSE-based [55]	Contrast-guided [proposed]
Lena	17.35	11.35	0.92
Station	18.01	11.15	0.98
Cameraman	17.12	11.17	0.81

Another very attractive advantage of the proposed method is that the *computational complexity* is fairly low. In Table 4.2, the actual interpolation times of the covariance-guided interpolation [54], LMMSE-based method [55], and our contrast-guided method are documented and compared. They are the actual times (in seconds) consumed to enlarge three test images “Lena,” “Station,” and “Cameraman” (with a size of 256×256 each) by a factor of 2×2 . All these three methods are implemented in software using MATLAB 7.8.0 [73]. The proposed method only requires about 1 second to complete the interpolation for one image. On the other hand, the computational complexities of the covariance-guided method and the LMMSE-based method are quite high, which require about 18 seconds and 11 seconds, respectively, to complete the interpolation of one image. Furthermore, the *algorithmic complexity* of our proposed method is much lower than that of these

Figure 4.17. The test image (a) “Wheel” is directly enlarged by a factor of 8×8 via three stages of progressive interpolations by using: (b) bicubic [3], (c) edge-guided, (d) covariance-guided [54], (e) LMMSE-based [55], and (f) the proposed contrast-guided.

Figure 4.18. The test image (a) “Tripod” is directly enlarged by a factor of 8×8 via three stages of progressive interpolations by using: (b) bicubic [3], (c) edge-guided, (d) covariance-guided [54], (e) LMMSE-based [55], and (f) the proposed contrast-guided.

Figure 4.19. The test image (a) “Butterfly” is directly enlarged by a factor of 8×8 via three stages of progressive interpolations by using: (b) bicubic [3], (c) edge-guided, (d) covariance-guided [54], (e) LMMSE-based [55], and (f) the proposed contrast-guided.

Figure 4.20. The test image (a) “Cloth” is directly enlarged by a factor of 8×8 via three stages of progressive interpolations by using: (b) bicubic [3], (c) edge-guided, (d) covariance-guided [54], (e) LMMSE-based [55], and (f) the proposed contrast-guided.

Figure 4.21. The test image (a) "Roof" is directly enlarged by a factor of 8×8 via three stages of progressive interpolations by using: (b) bicubic [3], (c) edge-guided, (d) covariance-guided [54], (e) LMMSE-based [55], and (f) the proposed contrast-guided.

methods, which is quite essential and advantageous on hardware/software realization.

4.5 Summary

In this chapter, we have proposed a novel image interpolation method, called the *contrast-guided* image interpolation. Different from the traditional edge-guided method, our method utilizes the *contrast-guided decision map* rather than the edge map (or, edge-guided decision map) to guide the interpolation process. Through that, the contrast information is optimally incorporated into the interpolation process. Finally, extensive simulations obtained from experimenting both monochrome and color images reveal that our approach significantly outperforms multiple state-of-the-art image interpolation methods in both objective performance evaluation and subjective visual quality assessment. Moreover, both computational complexity and algorithmic complexity of the proposed interpolation method are much lower than that of the compared methods.

Chapter 5

Conclusion and Future Works

5.1 Conclusion

The thesis consists of two parts: an efficient *multiple description coding* (MDC) system and a novel image interpolation method. In the first part of this thesis, we have proposed a novel polyphase-based MDC system and evaluated the developed system from various aspects—namely, coding efficiency (i.e., rate-distortion performance), error resilience, network adaptability and computational complexity. Extensive simulation results have been provided along the way at each milestone stage of development to justify the purpose and efficiency achieved at respective stages. The works done in this part is concluded as follows.

An overview of existing MDC methods was first succinctly presented; in particular, we focused on the classic *polyphase downsampling* MDC (PD-MDC) methods as this approach is our starting foundation with a mathematical analysis of pixel correlation. For that, we thoroughly studied and analyzed two downsampling methods of the PD-MDC—that is, the *interlaced* downsampling versus the *quincunx* downsampling. By computing the average Euclidean distance measured between all adjacent sample pairs in the interlaced sampling lattice and the quincunx sampling lattice,

respectively, we observed that the quincunx lattice yields a smaller average distance. This directly reflects the fact that the sample correlation in the quincunx lattice is stronger, since the shorter the distance, the larger the correlation. This observation led us to adopt the quincunx downsampling as our starting point.

The first thrust contribution of the MDC system is on the development of the *prediction-compensated polyphase* MDC (PCP-MDC) scheme. The proposed PCP-MDC further partitions each decomposed quincunx descriptions into two sub-descriptions, denoted as *primary* sub-description (PS) and *dual* sub-description (DS), such that all the sub-descriptions are in rectangular shape and ready to be further encoded. Our partition approach avoids the problem of dislocating relative positions of the image content presented in the source image that is encountered in the traditional method (e.g., [36]), and thus avoids the unwanted high-frequency artifacts that could be often created otherwise. After the partition, the PS is directly encoded by the H.264/AVC intra coding, while the DS will be encoded by employing a prediction approach. That is, the prediction of DS is conducted based on the H.264/AVC intra decoded PS, and only the prediction errors (or residuals) are encoded. This is a more efficient approach than directly encoding the DS. Furthermore, to obtain a fast and accurate prediction for the DS, the *mode-guided directional prediction* (MDP) is developed, which exploits the prediction mode information generated from the previously encoded PS so that the prediction can be conducted along the majority direction of edges within the block.

The second thrust contribution is the *adaptive redundancy control* (ARC), which is developed to address two fundamental issues on the redundancy control of PCP-MDC: 1) how to generate effective redundancy information—*quality* issue; and 2) how to incorporate dynamic network conditions into our design for optimally determining the amount of redundancy to be inserted in each description—*quantity* issue. To our best knowledge, the proposed ARC scheme is the first work that

practically considers network condition and with an aim to achieve an optimal redundancy allocation for the polyphase-based MDC system. For the *quality* issue, we have proposed the *description-compensated redundancy generation* (DRG) scheme. The DRG inserts useful and efficient redundancy information into each description, which is the prediction residual of one quincunx-downsampled description (i.e., the complementary part), resulted by conducting cross prediction based on the other reconstructed (or decoded) description (i.e., the primary part) by exploiting the proposed *mode-guided directional prediction* (MDP) approach.

For the *quantity* issue, the *network-aware optimal redundancy allocation* (NO-RA) scheme has been proposed, in which the amount of redundancy bits allocated to each description are determined according to the network condition; for that, the probability of channel failure is used to represent the network condition and is incorporated into our mathematical formulation to derive the optimal condition for redundancy allocation. Based on our derived results, it is interesting to observe that the network condition, in terms of the probability of channel failure, is equal to the ratio of the first-order derivative of the R-D function of the primary part to that of the complementary part.

To implement the derived optimal condition, a fast algorithm for the estimation of these R-D functions is needed. For that, we proposed a *successive approaching* algorithm to identify the optimal partition of the target bitrate between the primary part of the description and its complementary part (i.e., the redundancy part). We have also proposed a post-processing method for facilitating the proposed MDC system as an image enhancement technique, which is called the *switching Gaussian filter* (SGF). It is specifically designed for the proposed MDC system. The proposed SGF is simple and yet efficient; it is able to effectively detect and attenuate the granular artifacts that tend to present in the reconstructed image when *both* received descriptions were encoded at *low* bitrates at the encoder.

Extensive simulation results have clearly demonstrated that the proposed MDC system (PCP-MDC and ARC together) significantly outperforms the existing PD-MDC methods. As an example, for the “Splash” test image and under a similar central performance for a fair comparison, the side performance of the proposed MDC system is able to achieve more than 3 dB of improvement compared with that of a state-of-the-art PD-MDC method as reported in [38]. Moreover, thanks to the proposed ARC method, our MDC system can provide an optimal balance between coding efficiency and error resilience. We also evaluated the computational complexity of the entire MDC system and found that the computational complexity of our proposed MDC system is much lower than that of those comparable methods. This is quite attractive for the hardware implementation.

The second part of this thesis is on the development of a novel and generic image interpolation algorithm, which is called the *contrast-guided image interpolation*. The proposed interpolation method utilizes the generated binary *contrast-guided decision map* (CDM) rather than the binary *edge map* [or, *edge-guided decision map* (EDM)] which has been commonly utilized in some existing works, to guide the interpolation process. The key success of our method lies at the fact that our binary CDM is able to adaptively reflect the local contrast to facilitate the image interpolation. This is fundamentally different from existing edge-guided methods in which the contrast information is totally ignored in the EDMs. Extensive simulation results have clearly shown that the proposed contrast-guided image interpolation consistently and significantly outperforms other state-of-the-art image interpolation methods on both objective and subjective evaluations, besides its clear advantage on implementation aspect due to its much lower computational and algorithmic complexities.

5.2 Future Works

5.2.1 Multiple Description Video Coding

The PCP-MDC framework developed for image coding as demonstrated in this thesis can be directly exploited for error-resilient video coding. Similar to H.264/AVC, both intra-frame coding and inter-frame coding should be considered. For example each inter-frame can be decomposed into two quincunx descriptions for further encoding. Due to the strong spatial correlation between two quincunx descriptions, the new PCP-MDC system is advantageous on generating better reconstructed results compared to that of the traditional multiple description video coding framework which uses the *interlaced* downsampling to generate descriptions. Besides the spatial correlation exploited in the current system, how to exploit the temporal correlation to help the reconstruction of the received image when errors are encountered during transmission is worth to be further investigated.

5.2.2 Scalable Multiple Description Image Coding

Most of the MDC techniques proposed so far can only be incorporated into the *non-scalable* source coding framework to improve the error resilience of image/video transmission without addressing the adaptation to bandwidth variations and the receiving devices (e.g., HDTV, mobile device). To accommodate the varying network environments and the *quality of service* (QoS) requirements, it is desirable to design a *scalable* MDC coder that could offer both error resilience *and* bitrate scalability. One possible way to achieve that is to create multiple descriptions for each scalable layer [74–78]. However, existing works do not specify how to efficiently allocate the redundancy information among the scalable layers under various network conditions. Therefore, one of the major challenges lies at the investigations of

optimal bit allocation of redundancy information among the layers.

Publications

1. Zhe Wei, Kai-Kuang Ma, and Canhui Cai, “Edge-contrast-guided Image Interpolation Using Directional Variation Field Diffusion,” in *Proc. Int. Conf. on Image Process. (ICIP)*, Hong Kong, 2010.
2. Zhe Wei, Kai-Kuang Ma, and Canhui Cai, “Prediction-compensated Polyphase Multiple Description Image Coding With Adaptive Redundancy Control,” *IEEE Trans. Circuits and Syst. Video Technol.* vol. 22, pp. 465-478, Mar. 2012.
3. Zhe Wei and Kai-Kuang Ma, “Contrast-guided Image Interpolation,” submitted to *IEEE Trans. Image Process.* (Revised).

Bibliography

- [1] V. K. Goyal, “Multiple description coding: compression meets the network,” *IEEE Signal Process. Mag.*, vol. 18, pp. 74–93, Sep. 2001.
- [2] A. Jain, *Fundamentals of digital image processing*. Prentice-Hall, Inc. Upper Saddle River, NJ, USA, 1989.
- [3] H. Hou and H. Andrews, “Cubic splines for image interpolation and digital filtering,” *IEEE Trans. Acoust. Speech Signal Process.*, vol. 26, no. 6, pp. 508–517, 1978.
- [4] R. G. Keys, “Cubic convolution interpolation for digital image processing,” *IEEE Trans. Acoust. Speech Signal Process.*, vol. 29, no. 6, pp. 1153–1160, 1981.
- [5] A. A. E. Gamal and T. M. Cover, “Achievable rates for multiple descriptions,” *IEEE Trans. Inf. Theory*, vol. IT-28, pp. 851–857, Nov. 1982.
- [6] L. Ozarow, “On a source coding problem with two channels and three receivers,” *Bell Syst. Tech. J.*, vol. 59, pp. 1909–1921, Dec. 1980.
- [7] R. Ahlswede, “The rate-distortion region for multiple descriptions without excess rate,” *IEEE Trans. Inf. Theory*, vol. 31, pp. 721–726, Nov. 1985.

- [8] Z. Zhang and T. Berger, “New results in binary multiple descriptions,” *IEEE Trans. Inf. Theory*, vol. 41, pp. 502–521, Jul. 1987.
- [9] —, “Multiple description source coding with no excess marginal rate,” *IEEE Trans. Inf. Theory*, vol. 33, pp. 502–521, 1995.
- [10] V. A. Vaishampayan, “Design of multiple description scalar quantizers,” *IEEE Trans. Inf. Theory*, vol. 39, pp. 821–834, May 1993.
- [11] V. Vaishampayan and J. Domaszewicz, “Design of entropy-constrained multiple-description scalar quantizers,” *IEEE Trans. Inf. Theory*, vol. 40, pp. 245–250, 1994.
- [12] V. A. Vaishampayan, N. J. A. Sloane, and S. D. Servetto, “Multiple description vector quantization with lattice codebooks: Design and analysis,” *IEEE Trans. Inf. Theory*, vol. 47, pp. 1718–1734, Jul. 2001.
- [13] H. Bai, C. Zhu, and Y. Zhao, “Optimal multiple description lattice vector quantization for wavelet image coding,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, pp. 912–917, 2007.
- [14] S. Servetto, K. Ramchandran, V. Vaishampayan, and K. Nahrstedt, “Multiple description wavelet based image coding,” *IEEE Trans. Image Process.*, vol. 9, pp. 813–826, 2000.
- [15] J. Hafarkhani and V. Tarokh, “Multiple description trellis-coded quantization,” *IEEE Trans. Commun.*, vol. 47, pp. 799–803, Jun. 1999.
- [16] J. C. Batllo and V. A. Vaishampayan, “Asymptotic performance of multiple description transform codes,” *IEEE Trans. Inf. Theory*, vol. 43, pp. 703–707, Mar. 1997.

- [17] V. K. Goyal and J. Kovacevic, "Optimal multiple description transform coding of Gaussian vectors," in *Proc. Data Compression Conf. (DCC)*, Snowbird, UT, Apr. 1998, pp. 388–397.
- [18] Y. Wang, M. T. Orchard, V. A. Vaishampayan, and A. R. Reibman, "Multiple description coding using pairwise correlating transforms," *IEEE Trans. Image Process.*, vol. 10, pp. 351–367, Mar. 2001.
- [19] S. Hemami, "Reconstruction-optimized lapped orthogonal transforms for robust image transmission," *IEEE Trans. Circuits and Syst. for Video Technol.*, vol. 6, no. 2, pp. 168–181, 1996.
- [20] V. K. Goyal and J. Kovacevic, "Generalized multiple description coding with correlating transforms," *IEEE Trans. Inf. Theory*, vol. 47, pp. 2199–2224, Sep. 2001.
- [21] Y. Wang, A. R. Reibman, M. T. Orchard, and H. Jafarkhani, "An improvement to multiple description transform coding," *IEEE Trans. Signal Process.*, vol. 50, pp. 2843–2854, Nov. 2002.
- [22] J. Wang and J. Liang, "Filter banks for prediction-compensated multiple description coding," *IEEE Trans. Signal Process.*, vol. 57, pp. 1756–1767, May 2009.
- [23] W. Jiang and A. Ortega, "Multiple description coding via polyphase transform and selective quantization," in *Proc. SPIE Conf. Visual Communication Image Process.*, Feb. 1999, pp. 998–1008.
- [24] A. C. Miguel, A. E. Mohr, and E. A. Riskin, "SPIHT for generalized multiple description coding," in *Proc. Int. Conf. on Image Process. (ICIP)*, Kobe, Japan, Oct. 1999, pp. 842–846.

- [25] A. Said and W. Pearlman, "A new, fast, and efficient image codec based on set partitioning in hierarchical trees," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 6, no. 3, pp. 243–250, 1996.
- [26] C. Cai, J. Chen, and S. K. Mitra, "Structure unanimity multiple description coding," *Signal Process.: Image Communication*, vol. 22, pp. 59–68, Jan. 2007.
- [27] C. Cai, C. Chen, K.-K. Ma, and S. K. Mitra, "Multiple description wavelet coding with dual decomposition and cross packetization," *Signal, Image and Video Processing*, vol. 1, pp. 53–61, Apr. 2007.
- [28] T. Tillo, M. Grangetto, and G. Olmo, "Multiple description image coding based on lagrangian rate allocation," *IEEE Trans. Image Process.*, vol. 16, pp. 673–683, Mar. 2007.
- [29] D. Taubman and M. Marcellin, *JPEG2000: Image Compression Fundamentals, Standards and Practice*. Kluwer Academic Publishers, 2001.
- [30] G. Sun, U. Samarawickrama, J. Liang, C. Tian, C. Tu, and T. Tran, "Multiple description coding with prediction compensation," *IEEE Trans. Image Process.*, vol. 18, no. 5, pp. 1037–1047, 2009.
- [31] T. Tran, J. Liang, and C. Tu, "Lapped transform via time-domain pre-and post-filtering," *IEEE Trans. Signal Process.*, vol. 51, pp. 1557–1571, 2003.
- [32] U. Samarawickrama, J. Liang, and C. Tian, "M-channel multiple description coding with two-rate coding and staggered quantization," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, pp. 933–944, 2010.
- [33] D. Chung and Y. Wang, "Robust image coding and transport in wireless networks using non-hierarchical decomposition," in *Proc. Third Int. Workshop on Mobile Multimedia Communications*, New York, 1997.

- [34] M. Caramma, M. Fumagalli, and R. Lancini, "Polyphase downsampling multiple description coding for IP transmission," in *Proc. SPIE, the Int. Society for Optical Engineering*, 2001, pp. 545–552.
- [35] M. Gallant, S. Shirani, and F. Kossentini, "Standard-compliant multiple description video coding," in *Proc. Int. Conf. on Image Process. (ICIP)*, Thessaloniki, Greece, Oct. 2001, pp. 946–949.
- [36] S. Shirani, M. Gallant, and F. Kossentini, "Multiple description image coding using pre- and post-processing," in *Proc. Int. Conf. on Inf. Technol.: Coding and Computing*, Las Vegas, NV, Apr. 2001, pp. 35–39.
- [37] N. Franchi, M. Fumagalli, and R. Lancini, "Flexible redundancy insertion in a polyphase downsampling multiple description image coding," in *Proc. IEEE Int. Conf. Multimedia and Expo*, Lusanne, Switzerland, Aug. 2002, pp. 605–608.
- [38] T. Tillo and G. Olmo, "Data-dependent pre- and postprocessing multiple description coding of images," *IEEE Trans. Image Process.*, vol. 16, pp. 1269–1280, May 2007.
- [39] N. Franchi, M. Fumagalli, R. Lancini, and S. Tubaro, "Multiple description video coding for scalable and robust transmission over IP," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, pp. 321–334, Mar. 2005.
- [40] N. Zhang, Y. Lu, F. Wu, X. Wu, and B. Yin, "Efficient multiple-description image coding using directional lifting-based transform," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, pp. 646–656, May 2008.

- [41] 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.*, vol. 13, pp. 560–576, Jul. 2003.
- [42] A. Al, B. Rao, S. Kudva, S. Babu, D. Sumam, and A. Rao, "Quality and complexity comparison of H.264 intra mode with JPEG2000 and JPEG," in *Proc. Int. Conf. on Image Process. (ICIP)*, Singapore, Oct. 2004, pp. 525–528.
- [43] D. Marpe, V. George, H. L. Cycon, and K. U. Barthel, "Performance evaluation of Motion-JPEG2000 in comparison with H.264/AVC operated in pure intra coding mode," in *Proc. SPIE, the Int. Society for Optical Engineering*, Feb. 2004, pp. 129–137.
- [44] D. Marpe, H. Schwarz, and T. Wiegand, "Context-based adaptive binary arithmetic coding in the H.264/AVC video compression standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, pp. 620–636, 2003.
- [45] H. Malvar, A. Hallapuro, M. Karczewicz, and L. Kerofsky, "Low-complexity transform and quantization in H. 264/AVC," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, pp. 598–603, 2003.
- [46] H. Zeng, C. Cai, and K.-K. Ma, "Fast mode decision for H. 264/AVC based on macroblock motion activity," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 19, pp. 491–499, 2009.
- [47] H. Zeng, K.-K. Ma, and C. Cai, "Hierarchical Intra Mode Decision for H. 264/AVC," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, pp. 907–912, 2010.
- [48] M. Unser, "Splines: A perfect fit for signal and image processing," *IEEE Signal Process. Mag.*, vol. 16, pp. 22–38, Nov. 1999.

- [49] T. Wiegand and B. Girod, “Lagrange multiplier selection in hybrid video coder control,” in *Proc. Int. Conf. on Image Process. (ICIP)*, Thessaloniki, Greece, 2001, pp. 542–545.
- [50] Z. He and S. K. Mitra, “A linear source model and a unified rate control algorithm for DCT video coding,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 6, pp. 970–982, Nov. 2002.
- [51] H.264 JM codec. [Online]. Available: <http://iphone.hhi.de/suehring/tml>.
- [52] G. Ramponi, “Warped distance for space-variant linear image interpolation,” *IEEE Trans. Image Process.*, vol. 8, no. 5, pp. 629–639, 1999.
- [53] K. Jensen and D. Anastassiou, “Subpixel edge localization and the interpolation of still images,” *IEEE Trans. Image Process.*, vol. 4, no. 3, pp. 285–295, 1995.
- [54] X. Li and M. Orchard, “New edge-directed interpolation,” *IEEE Trans. Image Process.*, vol. 10, no. 10, pp. 1521–1527, 2001.
- [55] L. Zhang and X. Wu, “An edge-guided image interpolation algorithm via directional filtering and data fusion,” *IEEE Trans. Image Process.*, vol. 15, no. 8, pp. 2226–2238, 2006.
- [56] X. Zhang and X. Wu, “Image interpolation by adaptive 2-D autoregressive modeling and soft-decision estimation,” *IEEE Trans. Image Process.*, vol. 17, no. 6, pp. 887–896, 2008.
- [57] Y. Takahashi and A. Taguchi, “An arbitrary scale image enlargement method with the prediction of high-frequency components,” *Electronics and Communications in Japan (Part III: Fundamental Electronic Science)*, vol. 86, no. 8, pp. 41–51, 2003.

- [58] W. Carey, D. Chuang, and S. Hemami, “Regularity-preserving image interpolation,” *IEEE Trans. Image Process.*, vol. 8, no. 9, pp. 1293–1297, 1999.
- [59] Y. Zhu, S. C. Schwartz, and M. T. Orchard, “Wavelet domain image interpolation via statistical estimation,” in *Proc. Int. Conf. on Image Process.*, Thessaloniki, 2001.
- [60] D. Muresan and T. Parks, “Prediction of image detail,” in *Proc. Int. Conf. on Image Process.*, Vancouver, Canada, 2000.
- [61] L. Rudin, S. Osher, and E. Fatemi, “Nonlinear total variation based noise removal algorithms,” *Physica D*, vol. 60, no. 1-4, pp. 259–268, 1992.
- [62] J. Allebach and P. Wong, “Edge-directed interpolation,” in *Proc. Int. Conf. on Image Process.*, Lausanne, Switzerland, 1996.
- [63] F. Malgouyres and F. Guichard, “Edge direction preserving image zooming: a mathematical and numerical analysis,” *SIAM J. Numer. Anal.*, vol. 39, no. 1, pp. 1–37, 2002.
- [64] A. Krylov and A. Nasonov, “Adaptive total variation deringing method for image interpolation,” in *Proc. Int. Conf. on Image Process.*, California, USA, 2008, pp. 2608–2611.
- [65] Y. Cha and S. Kim, “Edge-forming methods for color image zooming,” *IEEE Trans. Image Process.*, vol. 15, no. 8, pp. 2315–2323, 2006.
- [66] J. Lee, J. Kim, J. Han, K. Choi, and S. Ko, “Edge-oriented two-step interpolation based on training set,” *IEEE Trans. Consumer Electronics*, vol. 56, no. 3, pp. 1848–1855, 2010.
- [67] T. Chan and J. Shen, *Image processing and analysis: variational, PDE, wavelet, and stochastic methods*. Society for Industrial Mathematics, 2005.

- [68] C. Xu and J. Prince, “Snakes, shapes, and gradient vector flow,” *IEEE Trans. Image Process.*, vol. 7, no. 3, pp. 359–369, 1998.
- [69] B. K. P. Horn and B. G. Schunck, “Determining optical flow,” *Artificial intelligence*, vol. 17, no. 1-3, pp. 185–203, 1981.
- [70] H. Jeffreys and B. Swirles, *Methods of mathematical physics*. Cambridge University Press, 1999.
- [71] M. Holmes, *Introduction to numerical methods in differential equations*. Springer Verlag, 2007.
- [72] G. Smith, *Numerical solution of partial differential equations: finite difference methods*. Oxford University Press, 1985.
- [73] D. Hanselman and B. Littlefield, *Mastering Matlab 7*. Pearson Prentice Hall, 2005.
- [74] S. Cho and W. Pearlman, “A full-featured, error-resilient, scalable wavelet video codecbased on the set partitioning in hierarchical trees (SPIHT) algorithm,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, no. 3, pp. 157–171, 2002.
- [75] H. Wang and A. Ortega, “Robust video communication by combining scalability and multiple description coding techniques,” in *Proc. of the SPIE*, vol. 124, San Jose, 2003.
- [76] L. Kondi, “A rate-distortion optimal hybrid scalable/multiple description video codec,” *IEEE Trans. Circuits and Syst. for Video Technol.*, vol. 15, no. 7, pp. 921–927, 2005.

-
- [77] Y. Chan, P. Cosman, and L. Milstein, “A multiple description coding and delivery scheme for motion-compensated fine granularity scalable video,” *IEEE Trans. Image Process.*, vol. 17, pp. 1353–1367, 2008.
- [78] Z. Zhao, J. Ostermann, and H. Chen, “Low complexity multiple description coding for the scalable extension of H.264/AVC,” in *Proc. Picture Coding Symposium*, Chicago, IL, May 2009.