MULTIVIEW VIDEO COMPRESSION WITH 1-D TRANSFORMS

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

BURCU KARASOY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
ELECTRICAL AND ELECTRONICS ENGINEERING

JUNE 2013

Approval of the thesis:

MULTIVIEW VIDEO COMPRESSION WITH 1-D TRANSFORMS

submitted by **BURCU KARASOY** in partial fulfillment of the requirements for the degree of **Master of Science in Electrical and Electronics Engineering Department, Middle East Technical University** by,

Prof. Dr. Canan Özgen
Dean, Graduate School of **Natural and Applied Sciences**

Prof. Dr. Gönül Turhan Sayan
Head of Department, **Electrical and Electronics Engineering**

Assist. Prof. Dr. Fatih Kamışlı
Supervisor, **Electrical and Electronics Eng. Dept.**

Examining Committee Members:

Prof. Dr. Gözde Akar
Electrical and Electronics Engineering Dept., METU

Assist. Prof. Dr. Fatih Kamışlı
Electrical and Electronics Engineering Dept., METU

Assoc. Prof. Dr. İlkay Ulusoy
Electrical and Electronics Engineering Dept., METU

Assoc. Prof. Dr. Şenan Ece Schmidt
Electrical and Electronics Engineering Dept., METU

Dr. Cevahir Çığla
ASELSAN Inc.

Date: 03.06.2013

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last Name: BURCU KARASOY

Signature   :

ABSTRACT

MULTIVIEW VIDEO COMPRESSION WITH 1-D TRANSFORMS

Karasoy, Burcu

M.Sc., Department of Electrical and Electronics Engineering

Supervisor   : Assist. Prof. Dr. Fatih Kamışlı

June 2013, 57 pages

In previous research, it has been shown that motion-compensated prediction residuals can have 1-D structures in many regions and that 1-D directional DCTs can compress such regions more efficiently than the conventional 2-D DCT. In this thesis, we analyze the spatial characteristics of the disparity compensated prediction residuals and the analysis results show that, similar to motion compensated prediction residuals, many regions of disparity compensated prediction residuals also have 1-D structures. Thus, we apply 1-D directional DCTs for compressing disparity compensated prediction residuals in this thesis. To show the achievable compression efficiency gains from using these transforms, we modify the reference software of the Multiview Video Coding (MVC) extension of H.264 so that each residual block can be transformed either with one of the 1-D DCTs or the conventional 2-D DCT. Experimental results indicate that the overall compression efficiency increases.

Keywords: Discrete cosine transforms, Disparity compensation, Video coding

ÖZ

1B DÖNÜŞÜMLER İLE ÇOK BAKIŞLI VIDEO SIKIŞTIRMA

Karasoy, Burcu

Yüksek Lisans, Elektrik ve Elektronik Mühendisliği Bölümü

Tez Yöneticisi   : Yrd. Doç Dr. Fatih Kamışlı

Haziran 2013, 57 sayfa

Son yıllarda yapılan çalışmalar hareket-telafili kestirim artıklarının birçok bölgesinde pikseller arası ilintilerin tek boyutlu olduğunu ve 1-B yönsel ayrık kosinüs dönüşümlerinin (AKD) bu kestirim artıklarını 2-B AKD'den daha iyi sıkıştırabildiğini göstermiştir. Bu çalışmada disparite-telafili kestirim artıklarındaki mekansal ilintiler incelenmiş ve hareket-telafili kestirim artıkları gibi disparite-telafili kestirim artıklarının da birçok bölgesinde tek boyutlu ilintiler olduğu tespit edilmiştir. Bu analizden yola çıkılarak 1-B yönsel AKD'ler disparite-telafili kestirim artıklarının sıkıştırılmasında kullanılmıştır. Bu kullanımın sıkıştırma performansını arttırdığını gösterebilmek için H.264'ün çok bakışlı videoların sıkıştırılmasını içeren genişletilmiş referans yazılımı kullanılmıştır. Bu yazılım her bir artık bloğunun transformunun 1-B AKD'lerin biri ya da 2-B AKD kullanılarak alınabildiği hale getirilmiştir. Elde edilen deney sonuçları bu kullanımın genel sıkıştırma performansını arttırdığını göstermektedir.

Anahtar Kelimeler: Ayrık kosinüs dönüşümü, Disparite telafisi, Video kodlama

To Bora and Burak...

ACKNOWLEDGMENTS

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS AND ACRONYMS

1D	One Dimensional
2D	Two Dimensional
AVC	Advanced Video Coding
BD	Bjontegaard-Delta
CABAC	Context Adaptive Binary Arithmetic Coding
CAVLC	Context Adaptive Variable Length Coding
DC	Disparity Compensation
DCT	Discrete Cosine Transform
HDTV	High Definition Television
ITU	International Telecommunication Union
JMVC	Joint Multiview Video Coding
JVT	Joint Video Team
MC	Motion Compensation
MPEG	Moving Picture Experts Group
MV	Motion Vector
MVC	Multiview Video Coding
PSNR	Peak Signal-to-Noise Ratio
QP	Quantization Parameter
VCEG	Video Coding Experts Group
VLC	Variable Length Coding

CHAPTER 1

INTRODUCTION

The goal of video compression is to represent video data with a minimum number of bits while keeping picture quality at a sufficient level for a given application. Raw video data contains a significant amount of redundant information, and video compression is achieved by removing the redundant information. The major sources of redundancies are temporal redundancy and spatial redundancy. Temporal redundancy results from correlation between two adjacent frames. Similarly, neighboring pixel values within the same frame are correlated, and this correlation is the source of spatial redundancy.

This thesis is mainly concerned with multiview video compression. In multiview video compression, a scene is captured with multiple closely-placed cameras to generate a feeling of depth when the video is played back as a 3-D video. The frames captured at the same time instants by each of these cameras contain redundant information, as these frames contain the same scene at the same time instant from slightly different perspectives. This redundancy which presents only in multiview videos is called inter-view redundancy.

The redundant information is reduced using different methods. Spatial redundancy can be reduced using transforms such as discrete cosine transforms and wavelet transforms. Another way of reducing spatial redundancy is intra prediction. In this method, each block is predicted from previously coded neighboring blocks within the same frame. Furthermore, temporal redundancy is exploited by predicting the current frame from previously encoded frames; this method is called motion compensation prediction. Similarly, inter-view redundancy is also reduced with the prediction technique, namely, disparity compensation prediction. In disparity compensation prediction, the current frame is predicted from a previously coded frame in a neighboring view.

The difference between predicted and original values is called prediction residuals. In many cases, the prediction is not accurate enough, and the prediction residuals are also coded. Prediction residuals also have some spatial redundancy, and in order to exploit this kind of redundancy, the prediction residuals are transformed. The main focus of this thesis is the transformation of the prediction residuals.

2-D Discrete Cosine Transform (DCT) is commonly used in video compression for the transformation of residuals. Many years ago, when digital images and video compression began to be studied, 2-D Discrete Cosine Transform (DCT) was in fact used in image compression [38]. It was directly applied to the image intensities without prediction. In many regions of typical images, pixel intensities vary smoothly. In these regions, high spatial 2-D correlation exists between pixel intensities. 2-D DCT can remove this correlation well since its smooth basis functions may provide a good representation of smoothly varying regions, therefore, 2-D DCT is effective at compressing image intensities.

In recent video compression standards such as H.264/AVC, 2-D DCT is also used to transform resid-

uals. However, images and prediction residuals may have different spatial characteristics. Smoothly varying regions in images can be predicted well, and these regions may constitute a small fraction of residuals. A significant part of the residuals comes from high contrast regions, such as object boundaries or edges, which cannot be predicted well. Therefore, 2-D correlation between residual pixels is smaller than images, and most of the residual energy is concentrated on these specific regions. Because the statistical characteristics of the residuals in such regions may be different from the smooth image regions, 2-D DCT may not perform well in such regions of residuals, and more efficient transforms can be used for the compression of such regions of the prediction residuals.

Spatial characteristics of the motion compensation residual have been previously investigated in [1]. By referring to these analysis results, 1-D directional transforms have been designed for compressing the motion compensation residuals, and it has been demonstrated that using 1-D directional transforms for compression of the motion compensation residuals increases overall compression efficiency. In this thesis, the spatial characteristics of the disparity compensation residuals transformed in multiview video compression are analyzed and compared with the characteristics of the images and motion compensation residuals. Based on the analysis results, 1-D directional transforms are proposed for compression of disparity compensation residuals besides the compression of motion compensation residuals.

1.1   Overview of Thesis

This thesis is composed of six chapters, including the introduction chapter. The second chapter is focused on previous work related to this thesis. Firstly, brief information about video compression is given, and the H264/MPEG-4 AVC video compression standard is discussed with its Multiview Video Coding (MVC) extension. Then, the characteristics of motion compensation (MC) residuals are analyzed based on previous research. In the final section of this chapter, 1-D directional transforms proposed for compressing of MC residuals in [1] are introduced. In the third chapter, the characteristics of disparity compensation (DC) residuals are investigated in detail. At the end of this chapter, based on the analysis results, 1-D transforms are also proposed for compression of DC residuals. To evaluate the performance of the proposed transforms, 1D directional transforms were implemented in reference to the multiview video codec of H.264/AVC. Chapter four gives details of this implementation. Chapter five presents the experimental results obtained with a comparison of both the conventional multiview coding system (which uses 2-D DCT) and modified multiview coding system (which uses 2-D DCT and 1-D directional transforms). Finally, chapter six summarizes the thesis and suggests ideas to increase the performance of the proposed transforms.

CHAPTER 2

PREVIOUS RESEARCH

This chapter provides a brief background on video compression and discusses previous research related to this thesis. Section 2.1 presents the basics of the video compression and the H264/MPEG-4 AVC video compression standard with its Multiview Video Coding (MVC) extension. Section 2.2 provides analysis results of the MC residual and introduces the 1D transforms used for the compression of the MC residual.

2.1 Video Compression

2.1.1 Overview Of Video Compression

In today's world, there is growing need for transmission and storage of videos. Before transmission and storage processes, it is required to compress the video since raw video has an excessive amount of data. For example, 24 fps progressive raw video in size of 1280x720 with 24 bits color depth requires 1.59 Gbps transmission speed, and 5 minutes of this video requires 59.72 GB memory [18].

Video compression can be defined as the representation of the video data using fewer numbers of bits. Video compression techniques are based on exploiting redundancies in the video data such as spectral, coding, spatial, temporal and inter-view redundancies. Spectral redundancy is due to correlation between different color planes, and transformations between color planes can be used to reduce this type of redundancy. Coding redundancy is the extra amount of bits used to represent a given sequence of data. Entropy coding techniques such as Huffman coding and arithmetic coding are used to reduce coding redundancy [24].

The correlation between neighboring pixel values within the same frame is the source of spatial redundancy. This type of redundancy can be reduced using discrete cosine transform or wavelet transforms. These transforms are also used in image compression to exploit spatial redundancy. Intra prediction is another way of reducing spatial redundancy and this method is used in recent video compression standards. In this approach, each block is predicted from previously coded neighboring blocks within the same frame, and the prediction residual is transformed.

Temporal redundancy arises from the similarities between two adjacent frames since in typical sequences, adjacent frames differ slightly. The correlation between two consecutive frames can easily be seen in Figure 2.1. In order to reduce this type of redundancy, most video coders predict the local region of the current frame from previously encoded frames, and this technique is known as motion compensation. In this approach, generally translational motion is assumed. Especially in stationary

(a) View 0 - Frame 15 of the Exit video sequence

(b) View 0 - Frame 14 of the Exit video sequence

Figure 2.1: Frames which show the correlation between two adjacent frames [6].

(a) View 0-Frame 15 of the Exit video sequence

(b) View 1-Frame 15 of the Exit video sequence

Figure 2.2: Frames which show the correlation between neighboring views [6].

and slowly moving smooth regions, motion compensation produces effective results. On the other hand, in unsmooth regions such as edges and texture regions, this technique can produce large prediction errors.

In multiview video coding, a scene is recorded from different perspectives by closely placed cameras. The images captured from these cameras at the same time are highly correlated. Figure 2.2 shows the correlation between two views. This type of redundancy is called inter-view redundancy, and it is similar to temporal redundancy. The correlation between views is reduced using disparity compensated prediction, which is also similar to motion compensated prediction. In disparity compensated prediction, the current frame is predicted from a previously coded frame in a neighboring view, whereas in motion compensated prediction, the current frame is predicted from a previously coded frame in the same view [2]. Like motion compensated prediction, disparity compensated prediction works well in smooth regions and can produce a significant amount of prediction errors in unsmooth regions.

In this section, inter-view, inter-picture and intra-picture redundancies and predictions are explained. However, this study is mainly focused on inter-view and inter-frame predictions. In video compression, besides prediction information, prediction errors also need to be coded. Especially in the regions where inter-view and inter-picture prediction produce a significant amount of errors, coding of residuals becomes important.

2.1.2 Overview of H264/MPEG-4 AVC

In the implementation step of this thesis, H.264/AVC codec is used. H.264/AVC is the newest video coding standard of the ITU-T Video Coding Experts Group and the ISO/IEC Moving Picture Experts Group [9] which provides better compression performance compared with previous standards [9]. In this part, steps of the video compression with H.264/AVC are explained shortly.

In general, a video compression system consists of an encoder which converts video into a compressed format and a decoder which converts video back into an uncompressed format. H264 syntax does not include encoder steps, but in practice, encoder process is the mirror of the decoding process. Figure 2.3 represents the encoding and decoding steps [5].
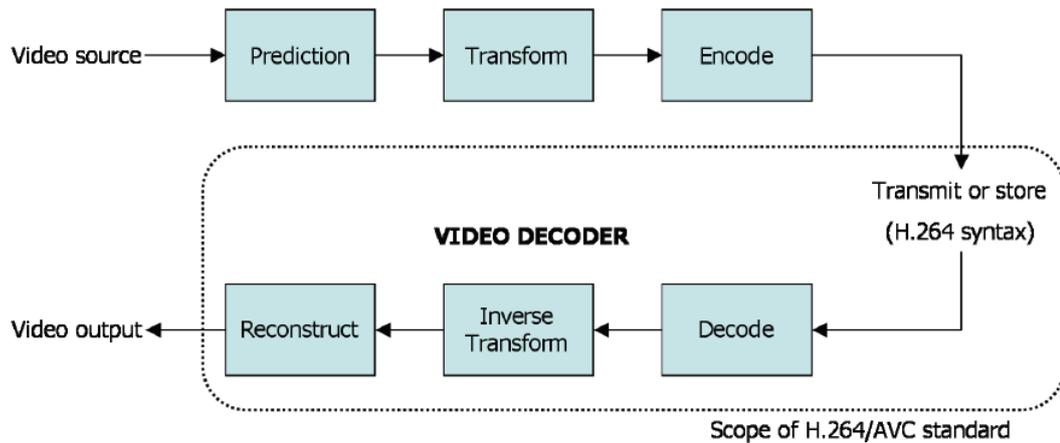
Figure 2.3: The H264/MPEG-4 AVC encoding and decoding processes [5].

In the sampling structure of H.264, which is called 4:2:0 sampling, number of chroma components is one forth of the number of luma components. Codec steps of luma and chroma components have some differences, and for simplicity, only codec steps of luma components are discussed in this section.

Before the prediction step, a frame is split into macroblocks of 16x16 pixels. Macroblocks are the basic blocks of a codec. In the prediction process, each macroblock is predicted from previously coded parts of the current frame (intra prediction) or previously coded frames (inter prediction). Prediction process does not have to be performed on the full macroblock. Smaller blocks can be used in the prediction step. H.264 is more flexible than previous standards in prediction block sizes. Intra prediction can be made on 16x16 or 4x4 blocks, and inter prediction can be made on 16x16, 16x8, 8x16, 8x8 blocks where 8x8 blocks also can be divided into 4x4 blocks. While frames coded using intra prediction are named as I Frame, frames coded using intra and inter prediction are named as P and B frames. In B frames, a predicted block can be taken as the weighted average of two predicted blocks which are predicted from two different reference blocks, and this type of blocks are named as bi-predicted blocks.

After the prediction step, residuals, which are the differences between the prediction of the macroblock and the original macroblock, are transformed using 8x8 or 4x4 2D DCT. Next, coefficients, which are the output of the transform, are quantized using quantization parameter (QP), i.e. each coefficient is divided by QP and rounded to the nearest integer. Increase in QP increases compression ratio and decreases picture quality. Transformation and quantization steps are combined as the integer transform

[9]. As the last step of the coding process, quantized coefficients are encoded using variable length coding or arithmetic coding. Quantized coefficients are not only data needed to be encoded; all other data that decoder needs to reconstruct the block, such as the reference and block size of the estimation, is also encoded.

At the decoder side, information sent as an H.264 stream from the encoder is decoded. Quantized transform coefficients are re-scaled and with the inverse transformation of these re-scaled coefficients, residual macroblock is obtained. Using intra prediction or inter prediction, decoder predicts the macroblock and adds residual data to reconstruct the decoded macroblock [5].

Block-based approach in H.264 can cause visually disturbing discontinuities at the block boundaries. These artifacts are known as blocking effects. To reduce these effects, adaptive in-loop deblocking filter is performed in H.264. The filter is applied after the inverse transform in the encoder and decoder. [9]. Since the filtered frames are used as reference frames, encoder and decoder have to perform identical filtering to stay in synchronization.

2.1.3 Extending H264/ MPEG-4 AVC For Multiview

Due to the increasing interest in multiview video, ITU-T Video Coding Experts Group and the ISO/IEC Moving Picture Experts Group has published multiview video coding (MVC) technology as an extension of H264/ MPEG-4 AVC that presents the technique for the compression of multiview videos. Because MVC is an extension of H264/ MPEG-4 AVC, in this section, all encoder and decoder processes will not be repeated, only basic differences of MVC from H264/ MPEG-4 AVC will be explained. In this thesis, the term MVC is used for the multiview extension of H264/ MPEG-4 AVC.

The main aim of MVC is to increase multiview compression efficiency using redundancy between views. For this purpose, it enables inter-view prediction. Figure 2.4 shows a sample prediction structure.

Figure 2.4: Sample prediction structure for MVC. [2]

MVC design includes one "base view" coded independently of other views, and this view can be reconstructed using decoders which do not have multiview support. As expected, while coding base view, inter-view prediction is not used. Frames are encoded using intra-frame prediction and inter-frame prediction. The left view in Figure 2.4 is an example of base view. Inter-view prediction can be used for coding views other than base view. In prediction step, the best reference frame is selected from the candidate list composed of inter-frame and inter-view references. Consider the right view in Figure 2.4. P frame in right view is predicted from I frame in left view using inter-view prediction. First B frame in right view is predicted from first B frame in the left view and P frame in right view. In MVC, each block in a B frame of the right view can be predicted either from a previous frame in the right view and/or from the adjacent frame in the left view.

6

2.2 Coding of MC Residual

2.2.1 Motion Compensation

Motion compensation (MC) is used to reduce bitrate by using the correlation along the temporal dimension in the video signal. This technique is based on predicting a local region of the current frame from previously encoded frames. Encoder estimates the motion between previously encoded frames and current frame. Although there are many motion compensation algorithms, in this section, block matching method will be discussed because it is used in H264/ MPEG-4 AVC.

In the block matching algorithm, frame is divided into blocks, and for each block, the best matching block from previously encoded blocks is determined. Estimated motion between current block and best matched block is mostly referred as motion vector (MV).

Since all pixels in the block are predicted using the same motion vector, size of the block is critical in block matching method. In H264/AVC, 16x16, 16x8, 8x16 and 8x8 blocks are available, and 8x8 blocks can be further divided into 8x4, 4x8 or 4x4 blocks.

It is possible to make integer pixel and fractional pixel accurate prediction with the block matching algorithm. Searching best matching block in original resolution gives integer pixel accuracy. Using interpolation while searching the best matching block gives fractional pixel accuracy [32]. In H264/MPEG-4 AVC, quarter-pixel accurate motion compensation is used [9]. Typically, fractional pixel accurate motion compensation gives better results compared to integer pixel accurate motion compensation because motion vector between two video frames is not generally integer multiples. The difference between the two methods can also be seen in Figure 2.5. This figure includes current frame, reference frame, integer-pixel accurate MC residual and quarter-pixel accurate MC residual. It can be observed the quarter-pixel accurate MC residual has smaller components compared to integer-pixel accurate one.

Motion compensation can also be achieved using more than one previously encoded frame. In this approach, predicted block is the weighted average of two predicted blocks which are predicted from best matching blocks. This method is named as multihypothesis motion compensation [25, 29, 31]. In H264/MPEG-4 AVC, multihypothesis motion compensation is available for B frames.

(a) Current frame

(b) Reference Frame

(c) Integer-pixel accurate MC residual

(d) Quarter-pixel accurate MC residual
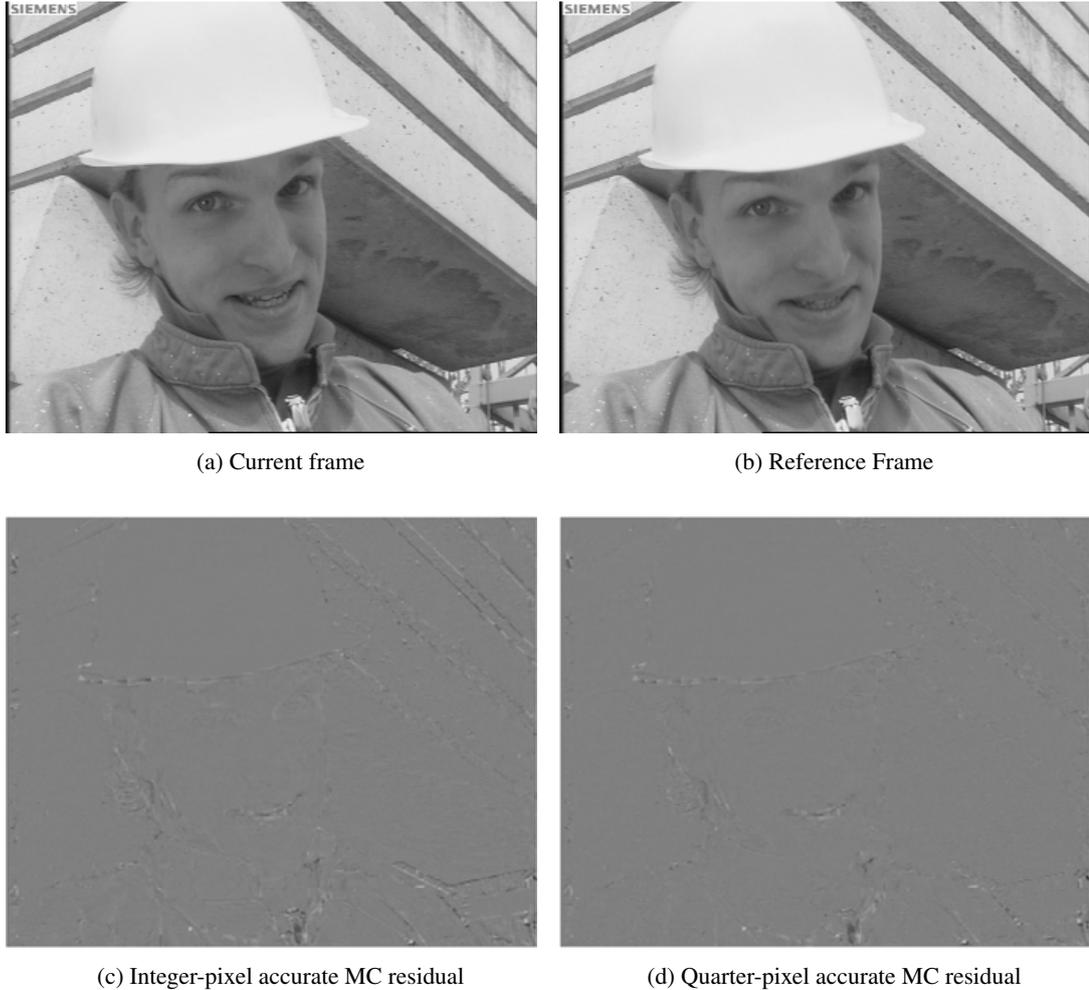
Figure 2.5: Frame 151 of Foreman sequence, its reference frame which is frame 150 of Foreman sequence, integer and quarter-pixel accurate MC residuals with 8x8 blocks [19].

2.2.2 Empirical Analysis of the MC Residual

In many cases, motion compensated prediction is not accurate enough, and the motion compensated prediction residual is also coded. To code the motion compensation residual efficiently, the characteristics of the residual have to be investigated. This section includes the empirical analysis of the MC residual based on the study in [1]. Motion estimation residual has different characteristics in smooth regions, texture regions and edges or object boundaries. In the remaining part of this section, each characteristic is examined one by one. In this investigation, a sample original frame, reference frame, predicted frame and prediction residual frame in Figure 2.6 are utilized. Original frame and reference frame are frame 11 and 10 of Stefan sequence at CIF (352x288) resolution [19].

Generally, motion estimation algorithms are successful in predicting smooth regions such as smooth backgrounds. Motion compensation residual of these regions is smaller than texture regions, edges and object boundaries. This is because even if the motion between current and reference frames is not translational, high correlation between pixels of the block to be predicted enables successful motion estimation. Typically, residuals of smooth regions are not coded. Court floor in Figure 2.6 is a good example of the smooth region. Prediction residual in the court floor is almost zero.

In Figure 2.6d, details of the texture regions of the original frame can be seen easily, and this means that prediction is not quite successful in these regions. On the other hand, unlike original frame, only high contrast components of texture regions are visible in residual frame; mean value of the texture regions is predicted well.

Like texture regions, edges or boundary of the objects cannot be predicted well using motion compensated prediction. Since generally motion between video frames is not translational, and motion prediction assumes translational motion, high magnitude motion compensation residuals present in edges or boundary of the objects. These residuals have 1D characteristic. For example; humans and letters in 2.6a transform into 1D structures after motion estimation.

In summary, characteristics of image and motion compensation residual have some differences. In image, most of the local region has 2D structures. On the other hand, motion compensation residual has three different characteristics. In motion compensation residual, pixel values of many local regions are almost zero. Residuals originated from object boundaries and edges form the significant portion of non-zero prediction residual, and these residuals have 1D characteristics. Residuals of the detailed regions are similar to images except the mean.

(a) Current frame

(b) Reference Frame

(c) Reconstructed Frame

(d) Motion Compensation Residual
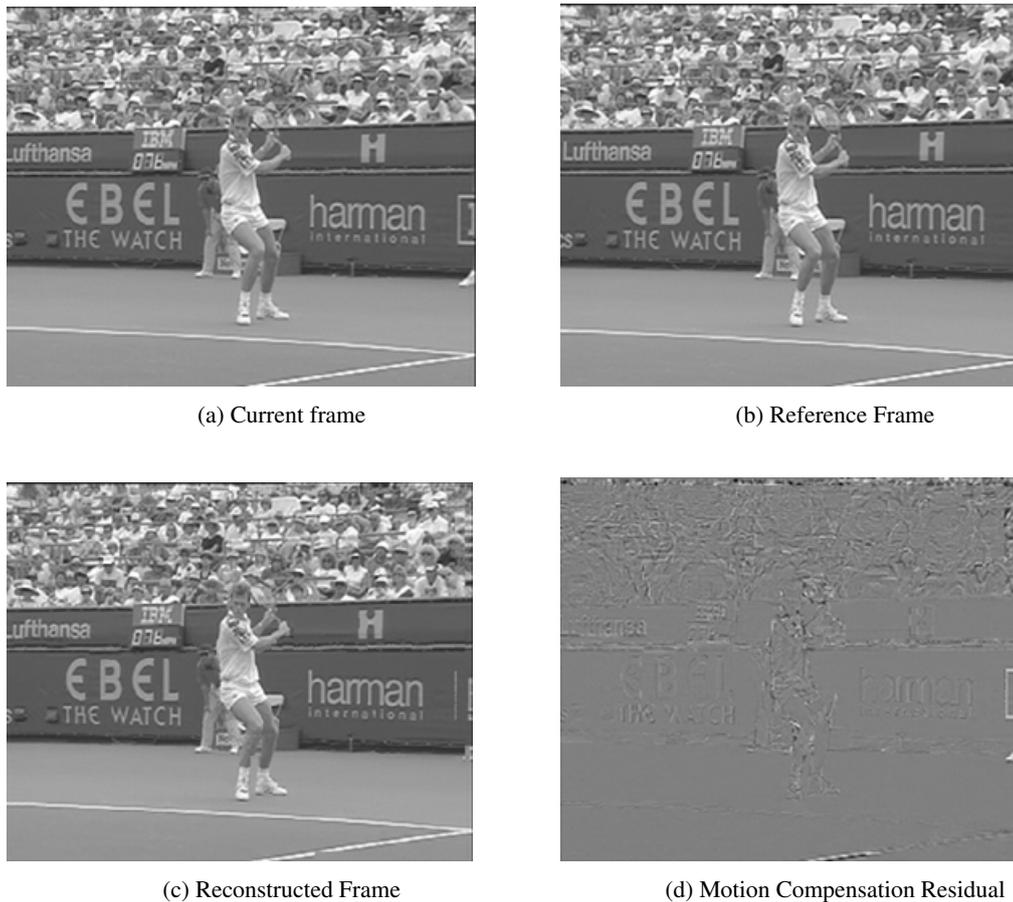
Figure 2.6: Frame 11 of Stefan sequence, its reference frame which is frame 10, reconstructed frame obtained by 8x8 motion compensated prediction of frame 11 from frame 10 and the prediction residual [19].

2.2.3 Auto-covariance Analysis of MC Residual

In the previous section, characterizations of MC residual are analyzed empirically. It has been discussed that while local regions of the image have 2D anisotropic structures, local region characteristic of the MC residual is different. This chapter reviews the statistical characteristics of the MC residual based on the studies in [1] and [38].

To analyze the characteristics of the images statistically, many models have been used. One of these models is Markov-1 model. Stochastic processes whose conditional distribution depends only finite number of past values have Markov property. If conditional distribution of the signal depends on only a single past value, the signal can be named as Markov-1 signal. Auto-covariance equation of stationary Markov-1 signal is shown in equation 2.1 [38].

$$C(I) = \rho^{|I|} \tag{2.1}$$

In the equation, I is the distance between the two points of which the auto covariance is computed, and ρ is the correlation parameter. ρ can take the values between 0 and 1. In [38], decorrelation transform is calculated for this auto-covariance equation and this transform becomes DCT when correlation approaches the maximum ($\rho \rightarrow 1$). Using separable construction, auto-covariance equation of 2D stationary Markov-1 signal can be obtained as equation 2.2 [1].

$$C_s(I, J) = \rho_1^{|I|} \rho_2^{|J|} \tag{2.2}$$

In this separable model, I and J represent horizontal and vertical pixel distances, ρ_1 and ρ_2 represent the horizontal and vertical correlation parameters. Decorrelation transform of this auto-covariance equation is also 2D DCT when correlation reaches its maximum ($\rho_1 \rightarrow 1$ and $\rho_2 \rightarrow 1$) [1].

Correlation parameters for images are expected to be high because of the high spatial correlation between pixels. Taking correlation parameters as $\rho_1 = \rho_2 = 0.95$ is an acceptable approximation for typical images [38], and these analyses explain the success of the 2D DCT in image compression.

MC residual has been also modeled with Markov-1 model and smaller correlation coefficients than 0.95 have been found [40]. This shows that decorrelating MC residual with 2D DCT is not effective as decorrelating image with 2D DCT.

In [1], MC residuals are modeled using a generalized auto-covariance model. This model is directionally adaptive and provides an additional degree of freedom around axes with parameter θ. The generalized model is given by equation 2.3 [1].

$$C_g(\theta, I, J) = \rho_1^{|Icos(\theta)+Jsin(\theta)|} \rho_2^{|-Isin(\theta)+Jcos(\theta)|} \tag{2.3}$$

The aim of the generalized model is to capture local anisotropic features by rotation around axes. When θ is taken as 0, the separable model is obtained.

To compare separable model and generalized model, image and MC residual are modeled by estimating the model parameters for each 8x8 block as in [1]. Firstly, a non-parametric auto-covariance of each 8x8 block is estimated. For this purpose, the mean is removed, the auto correlation is found, and each

correlation element is divided by the number of overlapping points. Secondly, the parameters θ, ρ_1 and ρ_2, which make the mean square error between the non-parametric auto covariance estimate and models minimum, are calculated. In this calculation ρ_1 is taken as the larger correlation parameter, θ takes the values between 0 and 180.

In this thesis, original image and residual in Figure 2.6 are modeled with the separable model and the generalized model as described in [1]. Figure 2.7 shows modeling plots. In these plots, each point represents a ρ_1 ρ_2 pair estimated from one 8x8 block. Color of the point depends on estimated θ of the point. If θ of the point is 0, the point is represented by black. When θ goes from 0 to 180, color of the point becomes lighter and goes from black to blue.

Figure 2.7a and Figure 2.7b are obtained by modeling the image in Figure 2.6a with the separable model and the generalized model. In Figure 2.7a, positions of points are more scattered and fewer points have a ρ_1 larger than 0.5 in compared to Figure 2.7b. These two observations mean that the generalized model is more successful than the separable model at modeling images. This is because modeling with less variety of the parameters is an indicator of more efficient image compression and having more points which have ρ_1 larger than 0.5 is an indicator of capturing high correlation more successfully. In generalized model plot, the points in the region where ρ_1 is larger than 0.5 have many different colors. This variety shows that the generalized model can capture high correlations in the images in many directions.

Figures 2.7c and 2.7d are obtained by modeling the MC residual in Figure 2.6d with the separable model and the generalized model. Comparison of Figures 2.7c and 2.7d gives the MC residual modeling capability of the separable model and the generalized model. In the plot of separable model, points tend to have ρ_1 and ρ_2 smaller than 0.5. On the other hand, points tend to have ρ_1 larger than 0.5 and ρ_2 smaller than 0.5 in the plot of generalized model. Contrary to the separable model, the generalized model catches the correlation on the MC residual along the direction of ρ_1. This means that the MC residual has 1D structures, and the generalized model can capture these 1D structures. Also, difference in the colors of the points shows that these 1D structures can have different directions.

Comparison of Figures 2.7b and 2.7d gives statistical differences between image and MC residual. In the region where both ρ_1 and ρ_2 are smaller than 0.5, Figure 2.7d has more number of (ρ_1,ρ_2) pairs than Figure 2.7b. The smaller correlation parameters are indicators of the smaller correlation between neighboring pixels in MC residuals compared to images. Furthermore, in the region where ρ_1 is larger than 0.5, MC residual plot has smaller ρ_2 values. It means that MC residual pixels have less 2D correlation compared to image pixels, MC residual pixels mostly have 1D correlation.

In summary, the generalized model is generalized form of the separable model, this model provides an additional degree of freedom around axes. Due to an additional degree of freedom, the generalized model is more successful than the separable model on capturing the correlation between pixels in both images and MC residuals. Generalized model plots indicate that image and MC residual have different characteristics. Generally, image tends to have 2D structures, and MC residual tends to have significant amount of 1D structures.
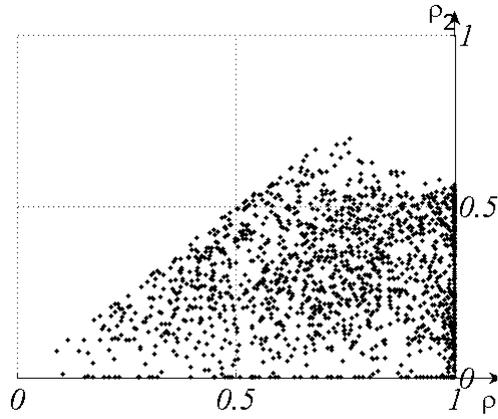
(a) Separable Model, Image
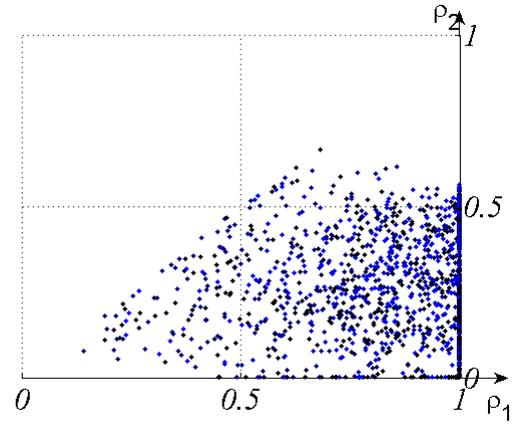
(b) Generalized Model, Image
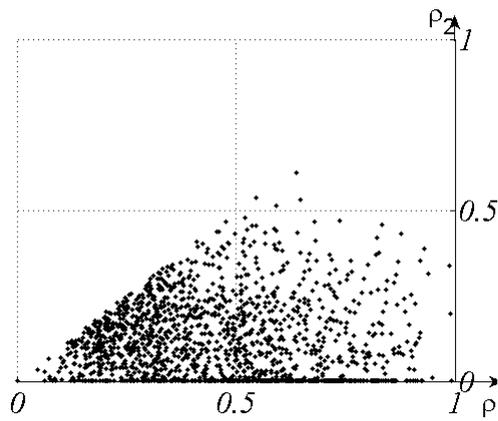
(c) Separable Model, MC residual
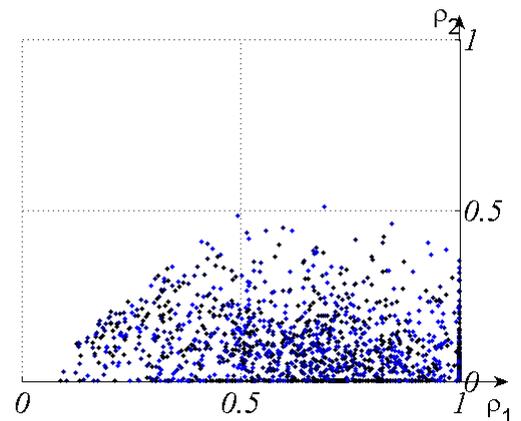
(d) Generalized model, MC residual

Figure 2.7: Plots of correlation parameters (ρ_1 and ρ_2) estimated using the separable and the generalized models. Figures 2.7a and 2.7b are obtained from using previous image in Figure 2.6a. Figures 2.7c and 2.7d are obtained from using previous motion compensation residual in Figure 2.6d. Colors of the points in generalized model depend on estimated θ of the point. If θ of the point is 0, the point is represented by black. When θ goes from 0 to 180, color of the point becomes lighter and goes from black to blue.

2.2.4 Direction-Adaptive 1D-Transforms

Analyses in Section 2.2.2 demonstrate that residuals originated from object boundaries or edges form a significant amount of the MC residual. In this type of MC residual regions, pixels are correlated in one directional way. Statistical analysis in Section 2.2.3 supports this observation with the generalized model plot of MC residuals which shows that MC residuals consist a significant amount of 1D structures. Hence, 2D DCT may not be the optimum choice for decorrelating such regions of the MC residual.

By considering the characteristics of the MC residual, directional 1D transforms for 4x4 and 8x8 block sizes were designed in [1] to exploit 1D correlations in MC residuals, and code the MC residuals more efficiently. Sets of directional transforms can be seen in Figures 2.8 and 2.9. Moreover, some experiments are conducted by modifying H264/AVC codec to find out the effect of proposed transforms on compression efficiency.

In conventional H264/AVC codec, each macroblock is coded using one of the motion compensation block sizes which are 16x16, 16x8, 8x16, 8x8 [9]. Decision on the block size is made using Lagrangian-based rate distortion optimization. Only 2D DCT is used for the transformation of the motion compensation residual. In the modified H264/AVC, MC residuals can be transformed using one of the direction adaptive 1D transforms or 2D DCT. 2D DCT is still kept as an option because the 2D DCT is a globally good transform, and in many regions that do not have strong 1D anisotropic features, 2D DCT gives better results.

Figure 2.10 is used to explain usage of transforms more explicitly. Residual image in Figure 2.10 is the same with residual image in Figure 2.6. On the right of the residual image, four zoomed 8x8 residual blocks can be seen. Each block has different spatial characteristics. Sample block one has high magnitude textured components, and it is expected that in this kind of textured regions, 2D DCT gives better results compared to proposed 1D transforms. Sample block two has horizontally correlated pixels, and ninth 8x8 transform pattern is designed for horizontally correlated blocks. Pixels in the sample block three are correlated diagonally from the left, and thirteenth 8x8 transform pattern is designed for this type of regions. Lastly, sample block four is the example of quite smooth regions. In quite smooth regions, 2D DCT gives better results compared to proposed transforms. On the other hand, since sample block four has highly small components, this block can be ignored. In this case, coefficients of this region are not transmitted to the decoder.

Experiments are conducted for three different transform configurations in [1]. For each transform configuration, results are taken for four quantization parameters: 24, 28, 32 and 36. In the first configuration, the encoder uses 4x4 2D DCT and 1D directional transforms, in the second configuration, the encoder uses 4x4-8x8 2D DCT and 1D directional transforms, in the third configuration, the encoder uses 8x8 2D DCT and 1D directional transforms. The output bitrate and PSNR values are compared with 4x4, 4x4-8x8 and 8x8 2D DCT cases respectively. Bjontegaard-Delta (BD) bitrate metric is used to calculate the average bitrate savings. On average 4.1%, 11.4% and 4.8% bitrate savings are achieved for 4x4, 4x4-8x8 and 8x8 cases.

Usage of directional 1D transforms on compression of MC residuals is also investigated in [4]. In this research, instead of using 1-D directional transforms in all directions, 1D directional transforms in horizontal and vertical directions are used. It is reported that by using only 1D directional transforms in horizontal and vertical directions, almost 50 percent of the bitrate savings obtained by adding all of the 1-D directional transforms can be achieved.

Additionally, the effect of 1D directional transforms on energy compaction capability of the video compression system has been investigated. Energy compaction can be defined as the ability to pack a large fraction of signal energy into just a few transform coefficients. The research in [3] presents the increase in energy compaction capability of the video compression system when 1D directional transforms are added for compression of MC residuals. It is reported that using 1D directional transforms and 2D DCT instead of using only 2D DCT significantly increases preserved energy.
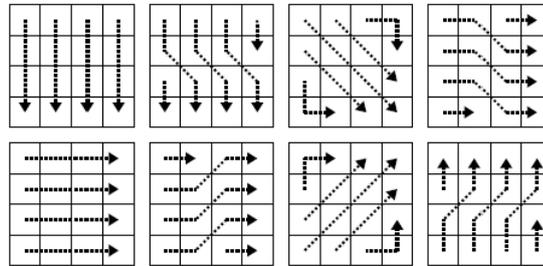
Figure 2.8: Eight 4x4 1D transforms proposed in [1].
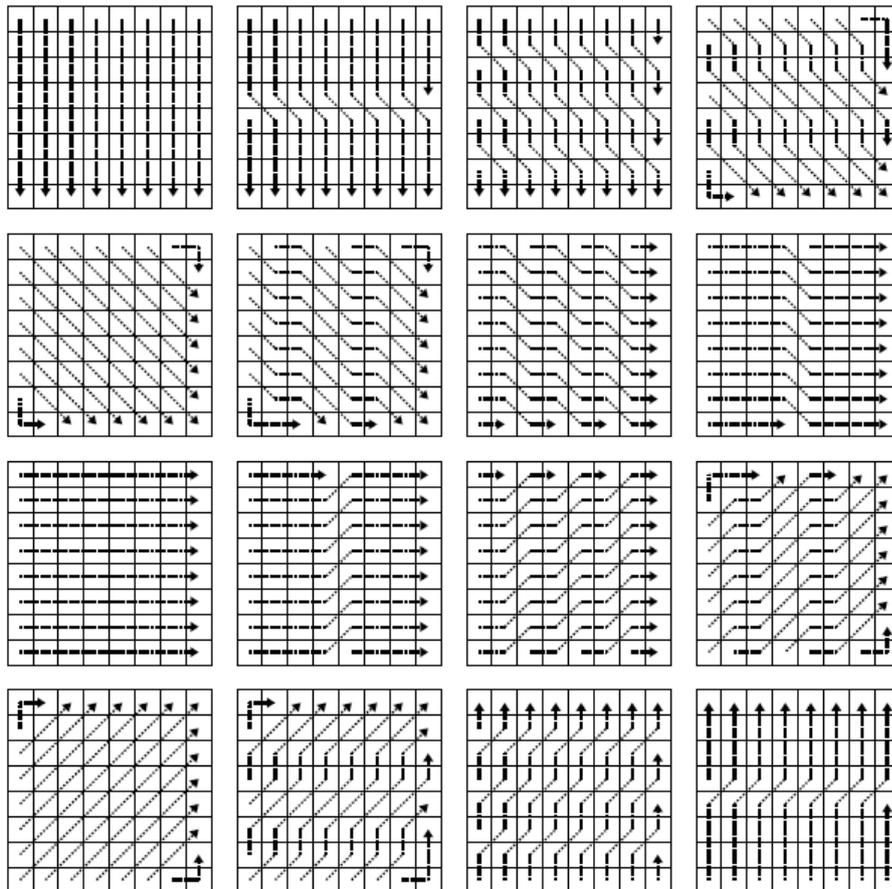
Figure 2.9: Sixteen 8x8 1D transforms proposed in [1].

14
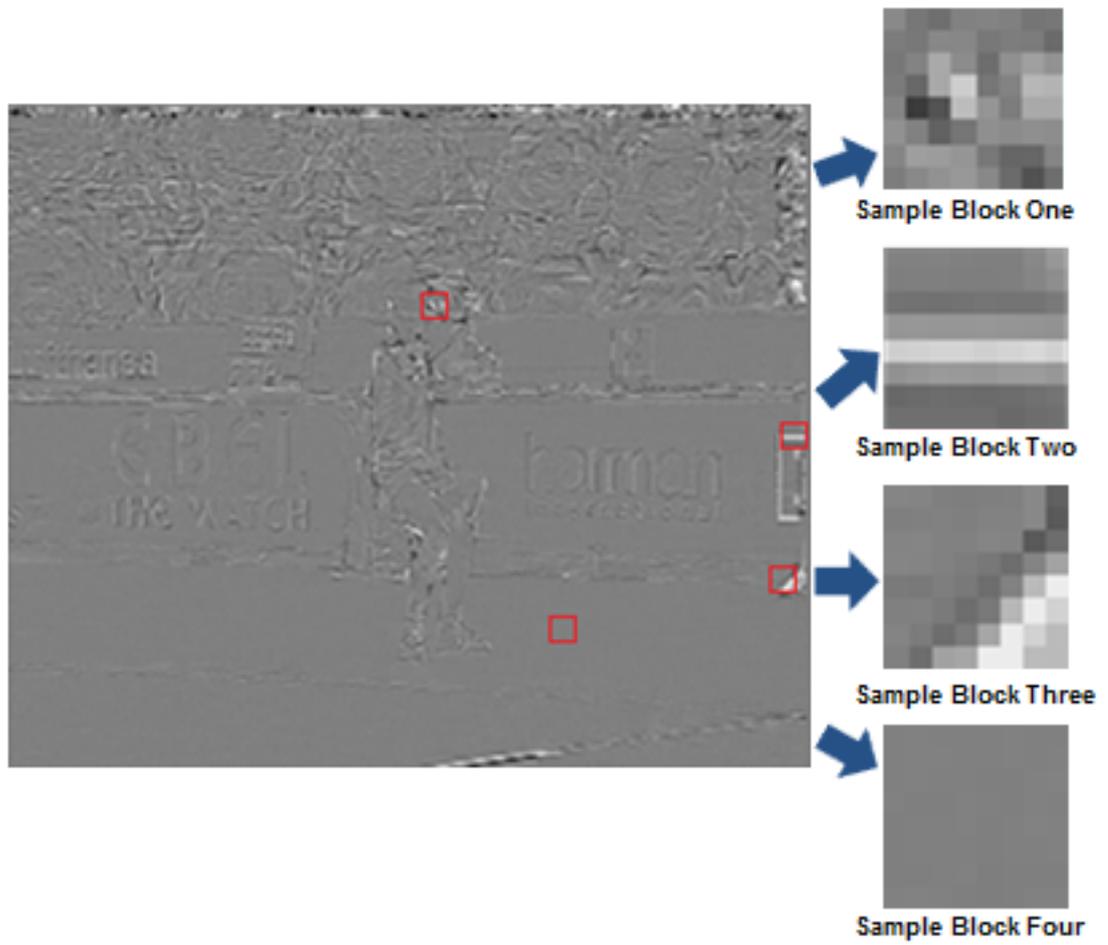
Figure 2.10: Sample 8x8 MC residual blocks.

CHAPTER 3

ANALYSIS OF DISPARITY COMPENSATION RESIDUALS

Investigation of disparity compensation (DC) residual characteristics is important to design better transforms for this residual. This chapter mainly focuses on analysing the statistical characteristics of disparity compensation residual. Section 3.1 gives brief information about disparity compensation for completeness. Section 3.2 includes empirical analyses of disparity compensation residual. In Section 3.3, disparity compensation residual is analyzed statistically. Both sections 3.2 and 3.3 also include differences and similarities between image and disparity compensation residual. In Section 3.4, disparity and motion compensation residuals are compared.

3.1 Disparity Compensation

In multiview video, same scene is captured with closely placed cameras. Multiview video coding (MVC) aims to represent these videos captured by different cameras with as few bits as possible.

The same point in an object is mapped to different coordinates in videos captured from different perspectives. The difference between the coordinates is named as disparity. Disparity compensation (DC) focuses on the correlation between the views to reduce the bitrate, and tries to estimate disparity for each pixel or block. In this approach, the current frame is predicted from the previously encoded frames of neighboring views. There exists many disparity compensation methods in the literature such as horizontal scaling and shearing-based disparity compensated prediction [11], multiview image coding based on geometric prediction [13] and overlapped block disparity compensation [14].

MVC extension of H264/ MPEG-4 AVC is the multiview coding standard published by ITU-T Video Coding Experts Group and the ISO/IEC Moving Picture Experts Group [2]. In this thesis, sample codec for MVC extension of H264/MPEG-4 AVC is used. The block matching technique that is used for inter-frame prediction is also used for inter-view prediction in MVC extension of H264/MPEG-4 AVC [2]. As in inter-frame prediction, in inter-view prediction, frames are divided into blocks and the best matching block is determined from previously encoded blocks. These two techniques differ in terms of the reference frame. In inter-frame prediction, reference frame is selected from the same view within the current frame. On the other hand, in inter-view prediction reference frame belongs to neighboring views.

Motion compensation and disparity compensation have the same block size options in MVC extension of H264/ MPEG-4 AVC. 16x16, 16x8, 8x16 and 8x8 blocks are available, and 8x8 blocks can be further divided into 8x4, 4x8 or 4x4 blocks. As mentioned in Section 2.2.1, the block size is important because all pixels in the block are represented by the same motion/disparity vector.

As explained in Section 2.2.1, searching best matching block in original resolution gives integer pixel accuracy and using interpolation while searching the best matching block gives fractional pixel accuracy [32]. As H264/MPEG-4 AVC, MVC extension uses the block matching algorithm with quarter-pixel accuracy to make an estimation between current and reference frames [2]. Compared to integer pixel accurate block matching algorithm, quarter pixel accurate one gives better results because the disparity between two frames of different views is not generally integer multiples. Figure 3.1 shows current frame, reference frame, integer-pixel accurate DC residual and quarter-pixel accurate DC residual. It can be observed the quarter-pixel accurate DC residual has smaller energy compared to integer-pixel accurate one.

(a) Current Frame

(b) Reference Frame

(c) Integer-pixel accurate DC residual

(d) Quarter-pixel accurate DC residual
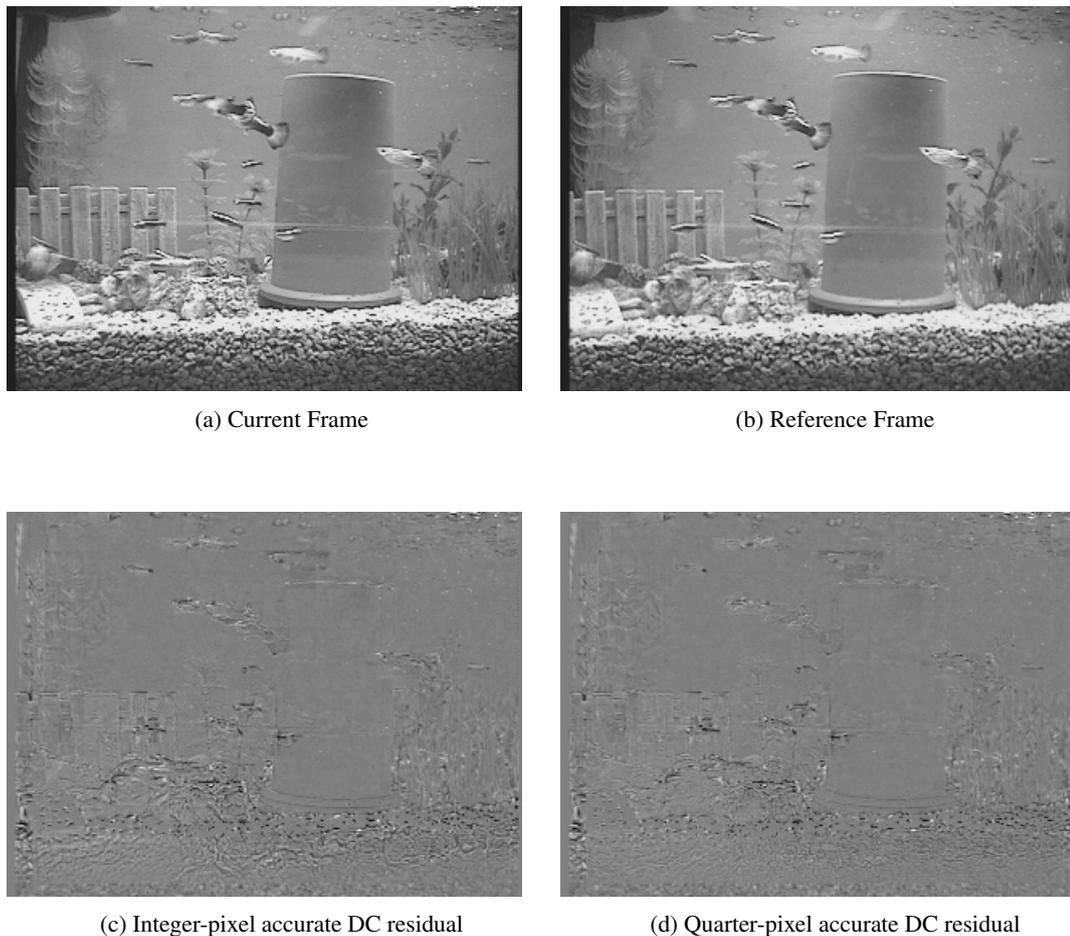
Figure 3.1: Frame 20 of aquarium view 1 sequence, its reference frame which is frame 20 of aquarium view 0 sequence, integer and quarter-pixel accurate DC residuals with 8x8 blocks [6].

3.2 Empirical Analysis of the Disparity Compensation Residual

Disparity compensation (DC) residual is the difference of original and disparity compensated frame. Typically, disparity compensation residual is too large to ignore, and residual is also transmitted to the decoder. For efficient compression, coding of disparity compensation residual is important. In order to code the disparity compensation residual efficiently, the characteristic of the residual has to be investigated. In this section, DC residual is analyzed empirically using Figures 3.2, 3.3 and 3.4.

Figures 3.2, 3.3 and 3.4 show current frame, reference frame, disparity compensated frame and disparity compensation residual frames from several test sequences. Current frame of Figure 3.2 is frame 15 of Exit view 1 sequence at 640x480 resolution [6]. Current frame of Figure 3.3 is frame 25 of Akko&Kayo view 1 sequence at 640x480 resolution [6]. Current frame of Figure 3.4 is frame 30 of Uli view 1 sequence at 1024x768 resolution [6]. Reference frames are adjacent frames of view 1 frames in view 0. Disparity compensated frame is the outcome of inter-view prediction of the current frame from the reference frame. Disparity compensation residual is the difference between the current frame and the compensated frame.

In occluded regions, smooth regions, texture regions and object boundaries or edges, disparity compensation residuals can have different spatial characteristics. In this section, characteristic of DC residual in these regions is discussed.

The regions captured by only one of the views are referred as occluded regions. Mainly, the difference of viewing area and overlapping objects cause occluded regions. In occluded regions, since the region exists on either one of the views, prediction fails. Therefore, prediction residual has high magnitude components in these regions. In occluded regions, images and DC residuals show similar characteristics. Papers on the right side of the Figure 3.2a are occluded region examples.

Smooth regions in all three example figures (Figure 3.2, Figure 3.3, Figure 3.4) have near zero prediction errors. It has been mentioned that block matching algorithm in MVC estimates translational disparity between current and reference frames. However, in smooth regions, even if the disparity of block between two frames is not translational, high spatial correlation enables successful prediction. Therefore, in smooth regions, images and DC residual have similar characteristics. However unlike in images, the mean of such regions is zero in DC residuals.

Bookshelf seen in Figure 3.2a can be given as an example of texture regions, and this kind of regions have large prediction errors so that details of bookshelf are visible. The characteristic of the texture regions in the residual frame is similar to the original image, but in the residual frame, mean of the texture regions is almost zero.

In object boundaries or edges, disparity compensation produces large errors. Translational estimation causes mismatch along the edges and boundaries. For example, the boundary of the peoples' body in Figures 3.2, 3.3, 3.4 and edges of the figure on the box in Figure 3.3 have high magnitude components in the prediction residual frame. Object boundaries or edges in DC residual have different characteristics than in image. In the DC residual, boundary or edges shows 1D characteristic. On the other hand, in the image, these regions unite with smooth regions and show 2D characteristics. Edges or object boundaries in DC residuals are mostly vertical because the disparity is mainly in horizontal direction.

(a) Current Frame

(b) Reference Frame

(c) Reconstructed Frame
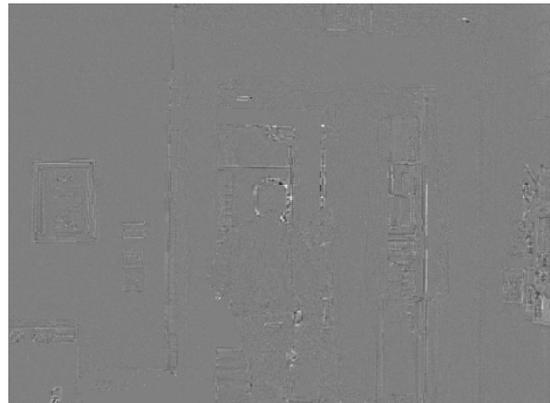
(d) Disparity Compensation Residual

Figure 3.2: Frame 15 of Exit view 1 sequence, its reference frame which is frame 15 of view 0 sequence, the reconstructed frame obtained by 8x8 disparity compensated prediction of the current frame from the reference frame and the prediction residual [6].

(a) Current Frame

(b) Reference Frame

(c) Reconstructed Frame

(d) Disparity Compensation Residual

Figure 3.3: Frame 25 of Akko&Kayo view 1 sequence, its reference frame which is frame 25 of view 0 sequence, the reconstructed frame obtained by 8x8 disparity compensated prediction of the current frame from the reference frame and the prediction residual [6].

(a) Current Frame

(b) Reference Frame

(c) Reconstructed Frame
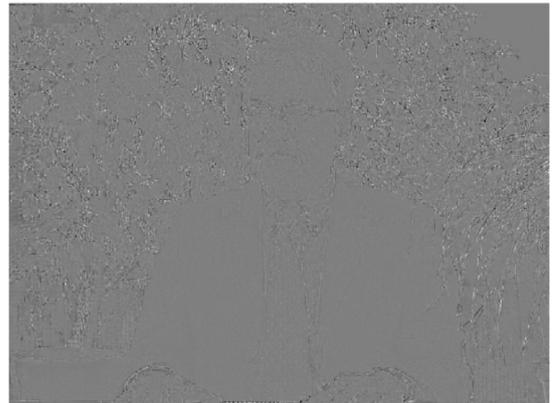
(d) Disparity Compensation Residual

Figure 3.4: Frame 30 of Uli view 1 sequence, its reference frame which is frame 30 of view 0 sequence, reconstructed frame obtained by 8x8 disparity compensated prediction of the current frame from the reference frame and the prediction residual [6].

In summary, spatial statistical characteristics of images and disparity compensation residuals are different in some regions. This difference appears especially in object boundaries or edges. In these regions, 2D structures in images transform into 1D structures in disparity compensation residuals. Edges or object boundaries constitute a significant portion of DC residual and compression efficiency of these regions can have an important impact on the overall compression efficiency of multiview video.

3.3   Auto-covariance Analysis Of Disparity Compensation Residual

In the preceding section, characteristics of disparity compensation residual are analyzed visually/empirically. This section focuses on the statistical characteristics of the disparity compensation (DC) residual based on the studies in [38] and [1].

In this section, images and DC residuals in Figures 3.2, 3.3 and 3.4 are modeled using the separable model (equation 3.1) and the generalized model (equation 3.2). Previously, images, MC residuals and DC residuals have been modeled with the separable model or its derivatives [38, 35, 34, 21, 30]. Images have also been modeled using similar generalized models [20]. Moreover, MC residuals have been modeled with a direction-adaptive model [1]. However, DC residuals have not been modeled with a direction adaptive model.

The modelling process is the same with Section 2.2.3. Firstly, image or DC residual is divided into 8x8 blocks, and a non-parametric auto-covariance of each block is estimated. To estimate auto-covariance, mean is removed, the auto correlation of zero-mean block is found, and each correlation element is divided by the number of overlapping points. Then, the parameters θ, ρ_1 and ρ_2 which minimize the mean square error between the auto covariance estimate and models are found. In this calculation, ρ_1 is taken as the larger correlation parameter, θ is in the range of 0 to 180.

$$C_s(I, J) = \rho_1^{|I|} \rho_2^{|j|} \tag{3.1}$$

$$C_g(\theta, I, J) = \rho_1^{|Icos(\theta)+Jsin(\theta)|} \rho_2^{|-Isin(\theta)+Jcos(\theta)|} \tag{3.2}$$

The model plots are shown in Figures 3.5, 3.6 and 3.7. In these plots, each point represents one 8x8 block. Color of the point depends on estimated θ of the point. If θ of the point is 0, the point is represented by black. When θ goes from 0 to 180, the color of the point becomes lighter and goes from black to blue. Because analysis results of figures are similar to each other, only results in Figure 3.5 are investigated in this section.

Figures 3.5a and 3.5b are obtained by modeling image in Figure 3.2a with the separable model and the generalized model. In the separable model plot (Figure 3.5a), points are located more dispersedly compared to the generalized model plot (Figure 3.5b). More compact positions are indicator of better modeling and higher compression efficiency. In the generalized model plot, more points have a ρ_1 larger than 0.5. Therefore, the generalized model is more successful at capturing higher correlation in images. Higher correlation parameter gives higher compression efficiency. Moreover, colors of the points which have ρ_1 higher than 0.5 are indicator of the directions of the high correlations. Difference in the colors of these points shows that the image has high correlation in different directions.

Figure 3.5c and 3.5d are obtained by modeling disparity residual in Figure 3.2d with the separable model and the generalized model. In the separable model plot (Figure 3.5c), most of the blocks have ρ_1

and ρ_2 smaller than 0.5. In the generalized model plot (Figure 3.5d), blocks tend to have ρ_1 higher than 0.5 and ρ_2 smaller than 0.5. High values of correlation parameter ρ_1 and small values of correlation parameter ρ_2 point out 1D correlation between DC residual pixels. In other words, DC residual has 1D structures and these structures can be captured by generalized model but not by the separable model. Moreover, since the points in the generalized model plot which have ρ_1 higher than 0.5 and ρ_2 smaller than 0.5 have different colors, it can be concluded that 1D structures in DC residual can have different directions.

(a) Separable Model, Image

(b) Generalized Model, Image

(c) Separable Model, DC residual

(d) Generalized model, DC residual
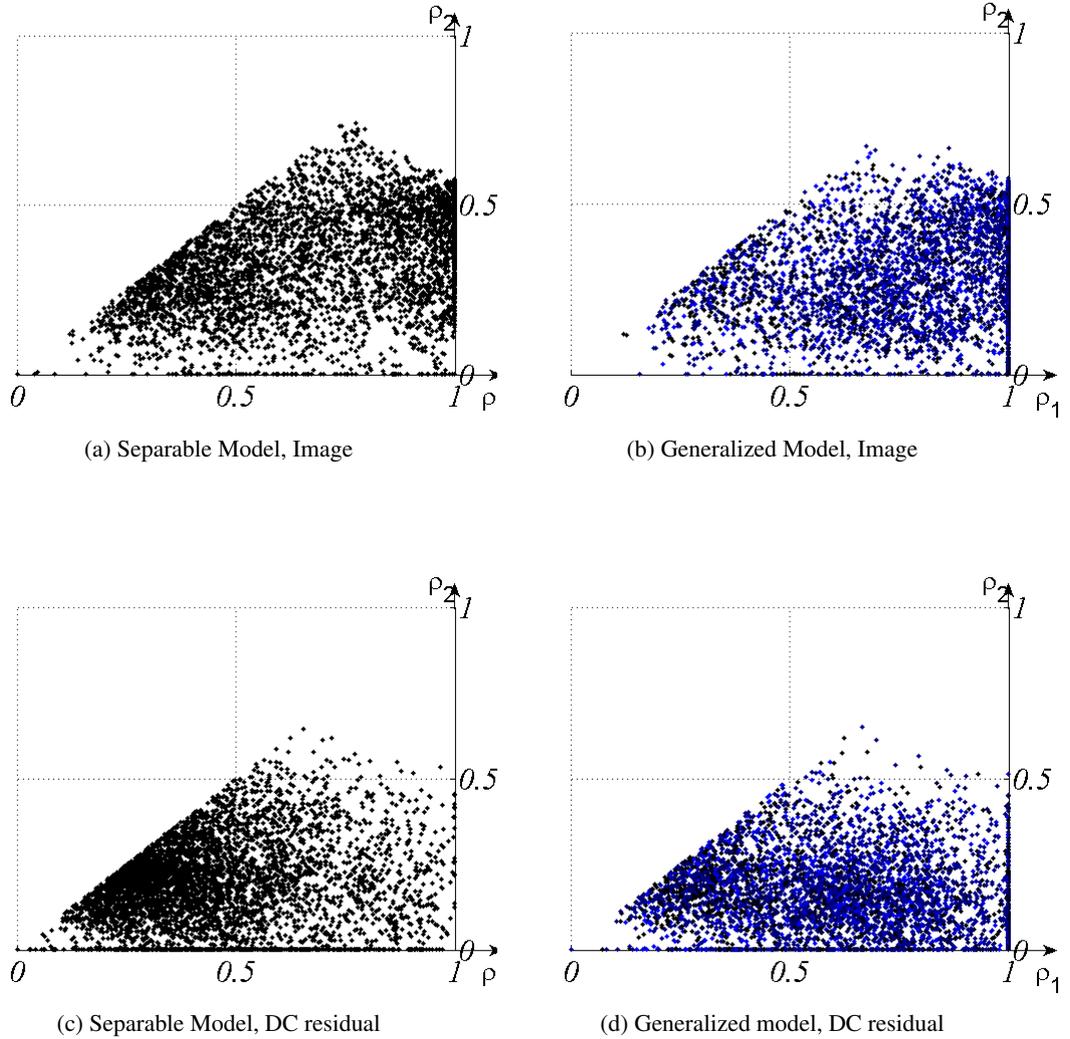
Figure 3.5: Plots of correlation parameters (ρ_1 and ρ_2) estimated using separable and generalized models. Figures 3.5a and 3.5b are obtained from using previous image in Figure 3.2a. Figures 3.5c and 3.5d are obtained from using previous disparity compensation residual in Figure 3.2d. Colors of the points in the generalized model plot depend on estimated θ of the point. If θ of the point is 0, the point is represented by black. When θ goes from 0 to 180, the color of the point becomes lighter and goes from black to blue.

(a) Separable Model, Image

(b) Generalized Model, Image

(c) Separable Model, DC residual

(d) Generalized model, DC residual
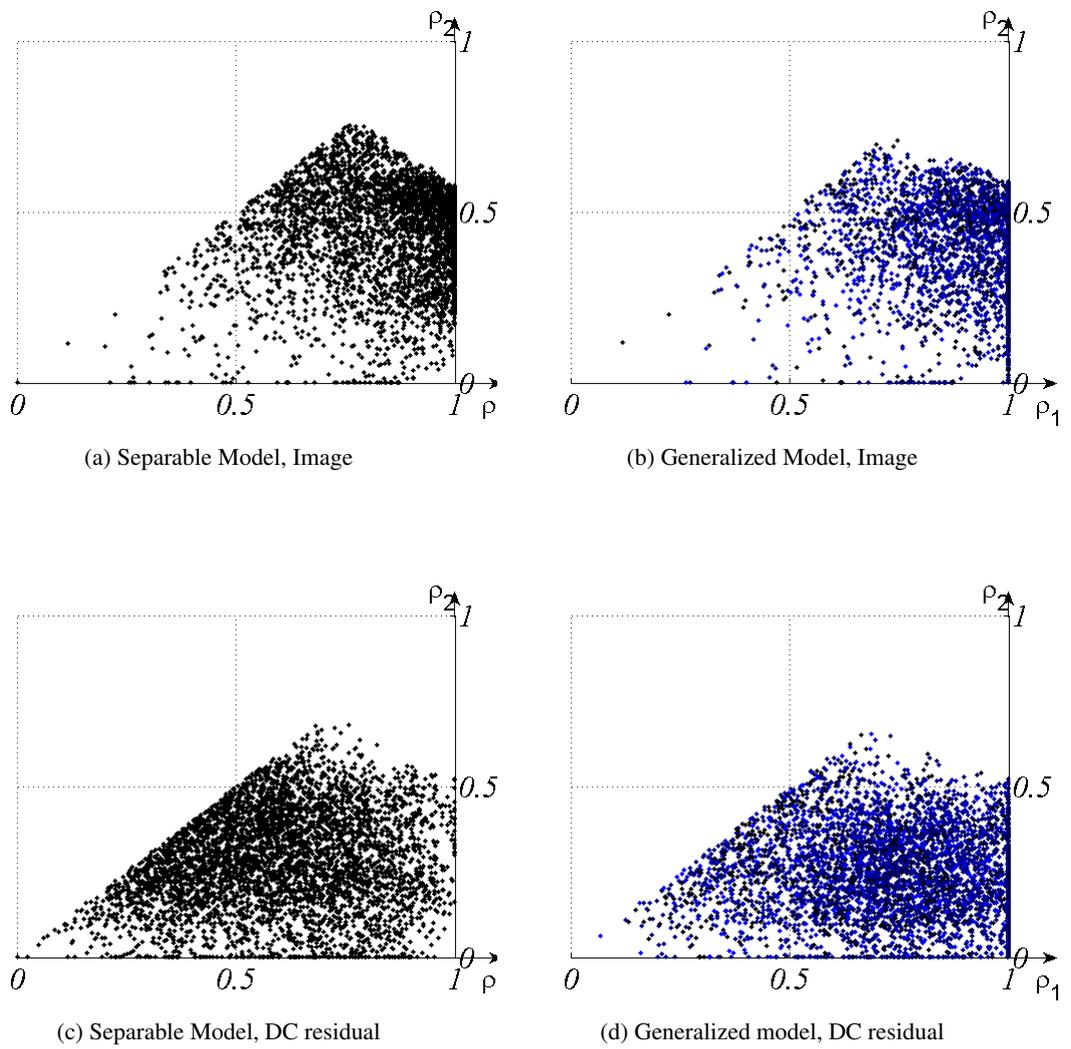
Figure 3.6: Plots of correlation parameters (ρ_1 and ρ_2) estimated using separable and generalized models. Figures 3.6a and 3.6b are obtained from using previous image in Figure 3.3a. Figures 3.6c and 3.6d are obtained from using previous disparity compensation residual in Figure 3.3d. Colors of the points in the generalized model plot depend on estimated θ of the point. If θ of the point is 0, the point is represented by black. When θ goes from 0 to 180, color of the point becomes lighter and goes from black to blue.

(a) Separable Model, Image

(b) Generalized Model, Image

(c) Separable Model, DC residual

(d) Generalized model, DC residual
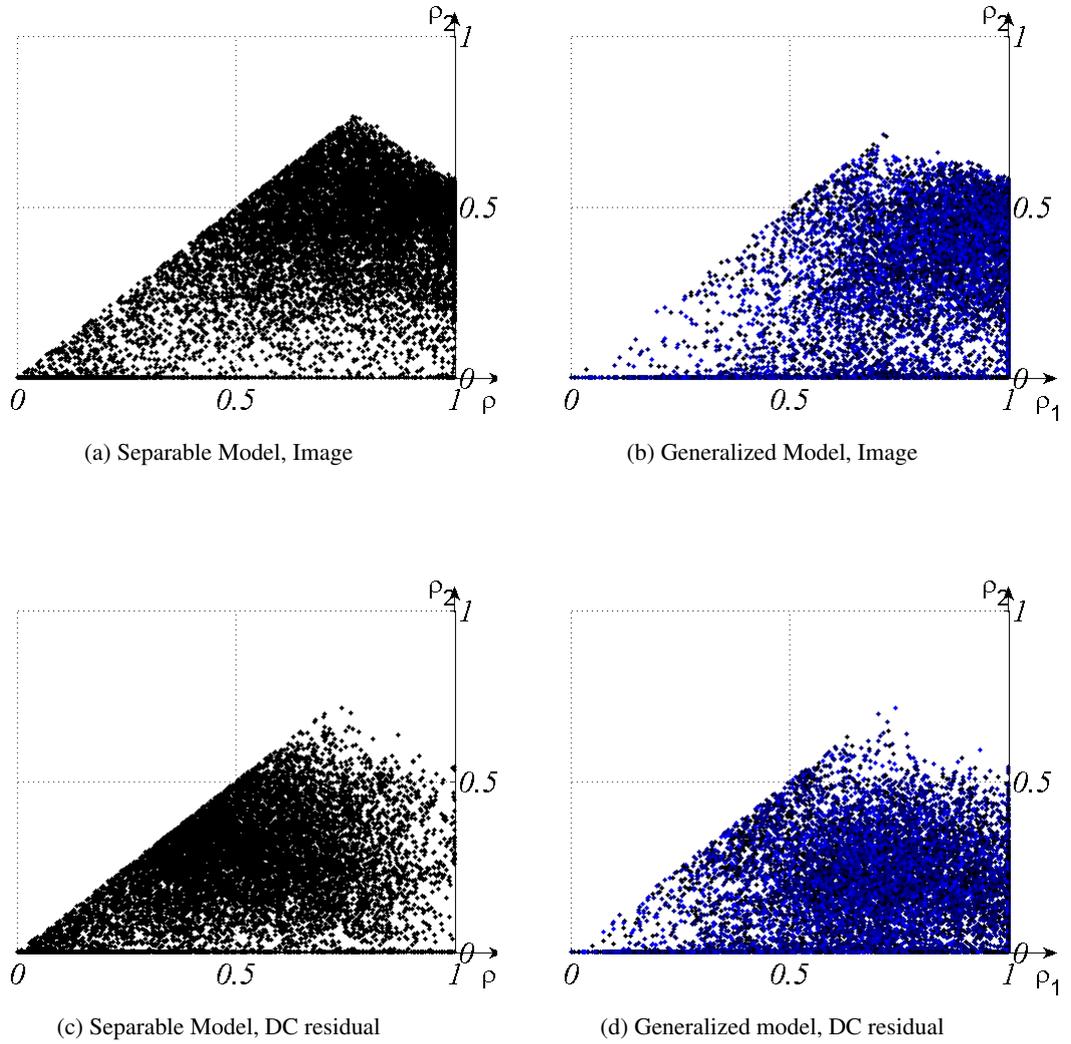
Figure 3.7: Plots of correlation parameters (ρ_1 and ρ_2) estimated using separable and generalized models. Figures 3.7a and 3.7b are obtained from using previous image in Figure 3.4a. Figures 3.7c and 3.7d are obtained from using previous disparity compensation residual in Figure 3.4d. Colors of the points in the generalized model plot depend on estimated θ of the point. If θ of the point is 0, the point is represented by black. When θ goes from 0 to 180, the color of the point becomes lighter and goes from black to blue.

Comparison of the generalized model plot of image (Figure 3.5b) and DC residual (Figure 3.5d) gives statistical differences between image and DC residual. The points with ρ_1 larger than 0.5 represent the blocks where high correlation captured in the direction of the ρ_1. ρ_2 values of these points are smaller in DC residual plot. It means that the probability of having two directional correlation is smaller in DC residual. Actually, the number of points which have both ρ_1 and ρ_2 larger than 0.5 is very small in DC residual plot. These observations point out that contrary to image, DC residual tends to have 1D structures instead of 2D structures.

In summary, by modeling images and DC residuals with the generalized model, higher correlation parameters can be achieved compared to separable model. The generalized model is better at capturing higher correlations in images and DC residuals because it has more parameters. The generalized model provides an additional degree of freedom around axes. By setting angle parameter to zero, the separable model can be obtained. Moreover, generalized model plots reveal the difference between statistical characteristics of image and DC residual. Correlation between neighboring pixels is less in DC residuals. Images tend to have exclusively 2D structures while DC residuals tend to have a significant amount of 1D structures.

3.4 Comparison of Disparity Compensation Residual and Motion Compensation Residual

Section 2.2.2 and 2.2.3 provide analysis of MC residual. Section 3.2 and 3.3 provide analysis of DC residual. In this section, similarities and differences between MC and DC residuals are presented by based on the analyses in these four sections.

Firstly, MC and DC residuals are compared empirically based on the analyses in Section 2.2.2 and 3.2. In smooth regions, both MC and DC residuals show similar characteristics. Smooth regions are easy to predict, and residuals in these regions are almost zero. Hence, in most of the cases, residuals in these regions are not coded.

MC and DC residuals in texture regions are also similar to each other. Texture regions are difficult to predict, and residuals in these regions show similar characteristics to the image except the mean. Mean of texture regions in these regions is almost zero.

As distinct from MC residual, DC residual has occluded regions. In these regions, DC residual has 2D structures.

In object boundaries or edges, as MC residual, DC residual has a significant amount of high magnitude components. In images, these regions unite with 2D regions and have 2D characteristics. Due to high contrast in these regions, 1D structures are formed in the residuals after the prediction step.

Lastly, when MC and DC residuals are compared statistically based on the analyses in Section 2.2.3 and 3.3, it is realized that statistical characteristics of both residuals are also similar to each other. In both residuals, correlation between pixels is weaker than the images. More importantly, both residuals have 1D anisotropic characteristics in many regions rather than 2D characteristics.

3.5 Summary and Outcomes

In [1], 1D directional transforms are proposed for compression of the MC residual. Results in [1] indicate that using these 1D transforms increases compression efficiency of single view video compression.

In this chapter, we analyzed DC residual. According to analysis results, DC residuals have 1D anisotropic characteristics as MC residuals.

In this thesis, referring to the characteristics of DC residual and similarities between MC and DC residual, we propose the same 1D transforms for the compression of the DC residual.

CHAPTER 4

SYSTEM IMPLEMENTATION WITH 1-D DIRECTIONAL TRANSFORMS

This chapter includes some of codec steps designed to add 1D transforms to the codec. Section 4.1 discusses implementation of transforms. Section 4.2 includes entropy coding of transform coefficients. In original codec, residuals are always transformed using integer transform which combines 2D DCT and quantization steps. In our implementation, residuals are transformed using one of 1D transforms or 2D DCT, and quantization step is implemented separately. To choose the optimum transform, transform selection algorithm is needed. Section 4.4 is about the transform selection. Information representing the selected transform is named as side information. Side information is sent to the decoder for each block in the residual. Coding process of the side information is described in Section 4.3.

4.1   Transformation of Residuals

H.264/ MPEG-4 AVC uses integer transform in transformation step [10, 27]. In integer transform, 2D DCT and quantization steps are merged. In this thesis, floating-point implementation of DCT algorithms [38, 39] is used. Using floating point arithmetic instead of integer arithmetic increases overall computational complexity but it does not affect the results. Our goal in this thesis is proof of concept, computational complexity is not our major concern.

4.2   Entropy Coding of Transform Coefficients

According to H.264/ MPEG-4 AVC standard, transform coefficients can be coded using one of two methods; context adaptive variable length coding (CAVLC) and context adaptive binary arithmetic coding (CABAC) [9, 28, 22]. In this thesis, CAVLC is preferred, and only the details of CAVLC are discussed.

In CAVLC method, coefficients are ordered using predefined scans. The aim of the scans is to order coefficients from biggest to smallest in magnitude. In regular H.264/ MPEG-4 AVC codec, scans are specified considering characteristics of the 2D DCT. In this thesis, we add new scan patterns for 1D transforms. These scan patterns are the same with the patterns in [1]. Except scan patterns, CAVLC implementation is not modified.

The scan patterns for 4x4 and 8x8 1D transforms are shown in Figure 4.1 and 4.2. H.264/MPEG-4 AVC uses four length-16 scans instead of using one length-64 scan for 8x8 2D DCT. As seen in Figure

4.2, we also use four length-16 scans for 8x8 1D transforms. With scans for 1D transforms, we try to order coefficients from largest to smallest and to keep neighboring coefficients close to each other.
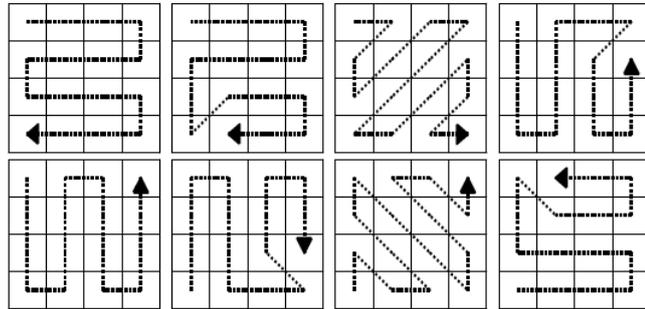
Figure 4.1: Scan patterns used in the coefficient coding of proposed 4x4 1D transforms [1].
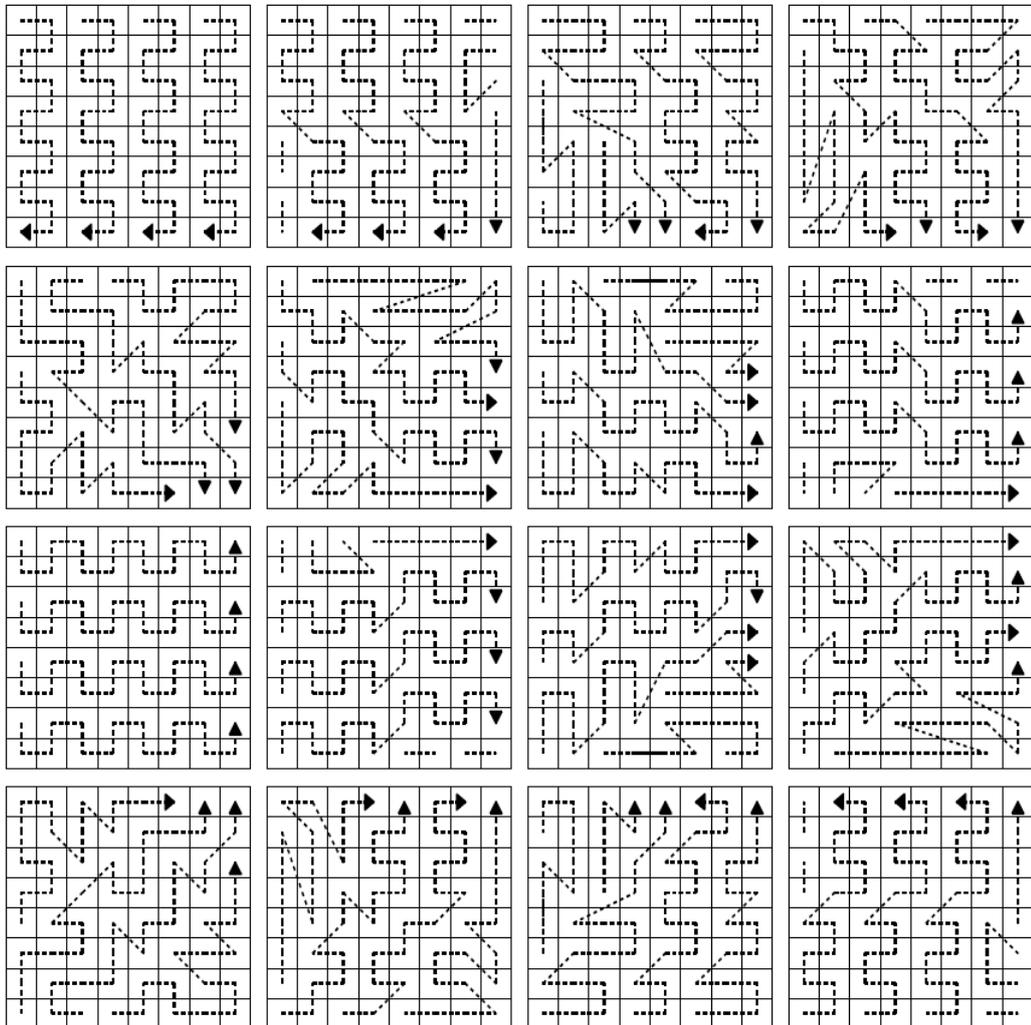
Figure 4.2: Scan patterns used in the coefficient coding of proposed 8x8 1D transforms [1].

30

4.3 Entropy Coding of Side Information

The decoder needs the selected transform information to perform appropriate inverse transform. As in [1], the selected transform information is named as the side information in this thesis. In our experiments, we run H.264/MPEG-4 AVC in CAVLC mode, and side information is also coded using variable length coding (VLC) method.

The codewords used in the experiments are seen in table 4.1. For transform size 4x4 and 8x8, 4-bit and 5-bit codewords are used as side information of 1-D DCTs. On the other hand, for both transform sizes, 2D DCT is represented by 1-bit codeword. In most of the region, 2D DCT is still optimum transform or 1-D DCTs give slightly better results compared to 2D DCT. That is why shorter codeword is given to 2D DCT. This case helps to increase overall efficiency, but decreases the probability of 1D DCTs being selected. 1D DCT is selected over 2D DCT if the decrease in distortion is enough to compensate for the higher side information bits.

Table 4.1: Side information codewords

Transform :	Codeword
2-D DCT :	1
1-D Transform:	0XXX

(a) 4x4 Block Transforms

Transform :	Codeword
2-D DCT :	1
1-D Transform:	0XXXX

(b) 8x8 Block Transforms

4.4 Transform Selection

For each residual block to be transformed, we present different transform options. Each transform has an effect on distortion and bitrate of the block. Comparing the transform results and choosing the best transform are critical for compression efficiency. We use Lagrangian based rate-distortion optimization for selecting the best transform option [33, 36].

Simple formulation of Lagrangian based rate-distortion optimization is shown in equation 4.1. In this equation, D represents distortion cost, R represents bitrate cost. Lagrange multiplier λ controls the trade-off between distortion and rate. Thus, joint cost J is calculated.

$$J = D + \lambda R \qquad (4.1)$$

Lagrangian based Rate-Distortion optimization is also used by H.264/MPEG-4 AVC to chose best coding mode of each macroblock [37, 26]. Each macroblock is coded with all available coding options (such as 16x16, 16x8, 8x16 prediction) and coding mode with minimum joint cost is chosen.

In our transform selection process, total number of bits used to code the residual block is taken as bitrate, mean square error of the residual block is taken as distortion. The value of the Lagrange multiplier is the same value used for selecting macroblock coding mode. Best transform of the each 8x8 block is decided independently of other blocks. One side information codeword is coded for each 8x8 block. If transform size is 8x8, all available 8x8 transforms are applied to the 8x8 block, and transform with minimum cost is selected as best transform. If transform size is 4x4, transform selection is a little different. In this case, one common transform is selected for four 4x4 blocks in the

8x8 block. All available 4x4 transforms are applied to the 4x4 blocks in 8x8 block, and the transform with minimum cost is selected as best transform option for the 8x8 block. Using the same transform for all 4x4 blocks in the 8x8 block reduces flexibility but increases overall efficiency. Diagram in Figure 4.3 summarizes the transform selection algorithm.

4.5   Complexity Increase

To add 1D transforms, encoder and decoder are needed to be modified, and this modification increases codec complexity depending on the implementation. Moreover, due to the transform selection process, encoding time increases while decoding time remains almost the same. Increase in the encoding time highly depends on implementation, and our discussion is based on JMVC software (JMVC 8.5). The JMVC (Joint Multiview Video Coding) software is the reference software for the Multiview Video Coding (MVC) project of the Joint Video Team (JVT) of the ISO/IEC Moving Pictures Experts Group (MPEG) and the ITU-T Video Coding Experts Group (VCEG) [23]. This software is designed for general purpose processors, each command is executed successively with no parallel processing support.

As mentioned earlier, to find the optimum one, each macroblock is coded with all available coding options. In our implementation, for each available coding option, all available transforms are applied to choose the optimum transform option. In other words, transformation process composed of transform, quantization, entropy coding of quantized coefficients, inverse quantization, inverse transform processes are repeated for each transform. Therefore, if sixteen 8x8 1D directional transforms are added as an option, encoding time increase is about sixteen times of one transformation process time. It should be also noted that transformation process time of 1D transforms takes less time than 2D DCT.

Encoder time increase ratio depends on the percentage of the transformation process time in overall encoding time. This ratio changes depending on encoder configuration. If 8x8 floating point transform and inter-view prediction with one view are used for encoding of the P frame, transformation process time takes about 20% of the total encoding time. Hence, encoding time increase factor originated from the addition of sixteen 1D directional transforms can be approximated as 4.2 (=17x20%+80%) if it is assumed that transformation process time of 1D transforms and 2D DCT are equal. When 8x8 integer transform is used instead of floating one, transformation process time takes about 4% of the total encoding time. In this case, if 1D transforms are also implemented with integer arithmetic, encoding time increase would be approximately 1.64 (=17x4%+96%). For 4x4-block transforms, transformation processes with integer arithmetic time take about 8% of the entire encoding time, and the factor of increase in the overall encoding time would be 1.6 (=9x8% + 92%).

Actually, video compression standards are mostly implemented with parallel processing support. Many encoder steps are suitable for parallel processing. For example, all available coding options can be executed in parallel. Similarly, 1D transforms and 2D DCT options can be executed in parallel. In this manner, addition of 1D transforms does not change encoder time.

Figure 4.3: Transform Selection Algorithm Diagram.
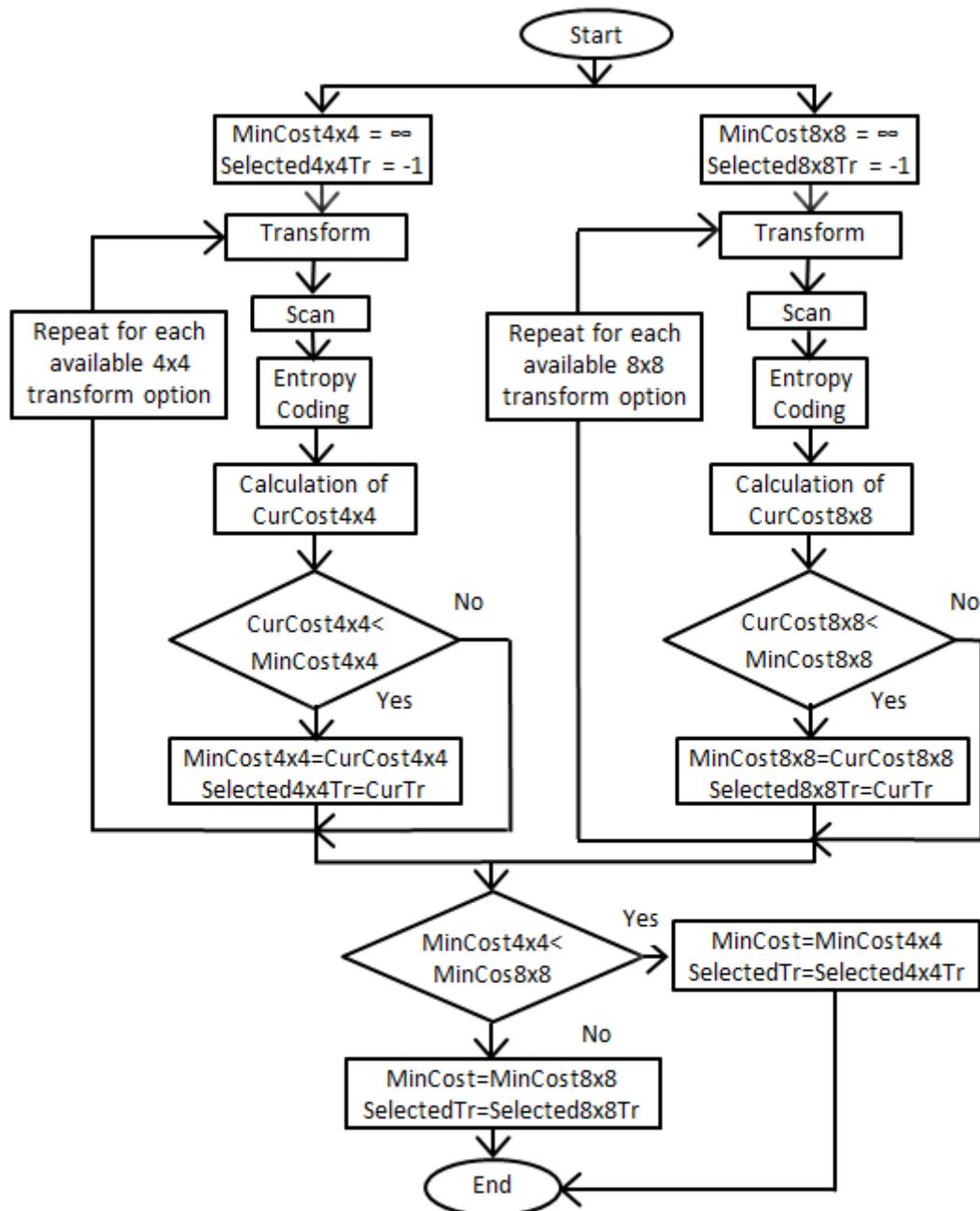
Term	Explanation
Transform	Explained in Sec 4.1
Scan, Entropy Coding	Explained in Sec 4.2
MinCost4x4, MinCost8x8	Minimum cost obtained with 4x4 or 8x8 transforms
CurCost4x4, CurCost8x8	Cost of 4x4 or 8x8 transform candidate
Selected4x4Tr, Selected8x8Tr	Selected 4x4 or 8x8 transform
MinCost	Minimum cost of the block
SelectedTr	Selected transform for the block

CHAPTER 5

EXPERIMENTAL RESULTS AND ANALYSIS

In this chapter, compression efficiency of using directional 1D transforms in addition to 2D DCT in multiview coding is analyzed. The compression efficiency of directional 1D transforms in multiview and single view coding is also compared.

In Section 5.1, common environment of these experiments are explained. In ection 5.2, experimental results related to single view compression are presented. In single view experiments, 1D transforms and 2D DCT are used to transform MC residuals. Section 5.3 provides multiview compression experiment results. In multiview experiments, 1D transforms and 2D DCT are used to transform MC and DC residuals. Section 5.4 compares the effect of adding 1D transforms in single view and multiview coding. Section 5.5 investigates the multiview efficiency of 1D transforms using different frame types (P and B). Finally, Section 5.6 examines the effect of the luminance compensation on the multiview efficiency of 1D transforms.

5.1   Common Properties of Experiments

In all experiments, the compression efficiency of using 1D directional transforms in addition to 2D DCT for coding MC and/or DC prediction residuals is analyzed. In particular, each block of a MC or DC residual is coded either with one of the 1D transforms or the 2D DCT. The encoder chooses the best transform for each block and conveys its choice by transmitting a variable length code. Results of proposed coding method are compared with the conventional coding method, which uses always the 2D DCT. Experiments are conducted by modifying JMVC software (JMVC 8.5). The JMVC (Joint Multiview Video Coding) software is the reference software for the Multiview Video Coding (MVC) project of the Joint Video Team (JVT) of the ISO/IEC Moving Pictures Experts Group (MPEG) and the ITU-T Video Coding Experts Group (VCEG).

To conduct the experiments, the JMVC software is modified as described in Section 4. Directional 1D transforms are provided as an option to transform MC and/or DC residual blocks. Only the transforms of luminance component residuals are modified. Chrominance components use only 2D DCT.

Sequences used in the experiments are shown in Table 5.1 and Figure 5.1. We use first 180 frames of 640x480 resolution videos and first 120 frames of 1024x768 resolution video. More information about the test sequences can be found in [6] and [7].

Table 5.1: Properties of the test sequences

Name	Resolution	Fps	Scene Content	Disparity Range (min-max)(in pixel)
Exit	640x480	25	Low Motion-Medium Texture	(10, 70)
Ballroom	640x480	25	High Motion-High Texture	(5, 40)
Race1	640x480	30	High Motion-Low Texture	(20, 50)
Vassar	640x480	25	Low Motion-Medium Texture	(5, 30)
Uli	1024x768	25	Low Motion-High Texture	(10, 125)

Results are obtained from four different experiments. For each experiment, after configuration is determined, encoding and decoding are repeated for four different quantization parameters (QP); 24, 28, 32, 36. Lower QP provides higher picture quality, and thus higher bitrate.

In order to compare the coding efficiency of the conventional coding system (which uses only 2D DCT) and the modified coding system (which uses also 1D transforms), their compression outputs are analyzed as follows; average bitrate (in kbit/s) of the compressed video stream and PSNR (in dB) between compressed and original frames are recorded for each QP. PSNR is obtained from compressed and original luminance component of the frames. However, bitrate includes compressed bitstream of luminance and chrominance components. With Bitrate and PSNR values for different QPs, rate-distortion plots are plotted to compare the two codecs' compression efficiency for a range of picture qualities. In order to provide average bitrate and PSNR differences between rate distortion curves of the two systems, Bjontegaard-Delta (BD) bitrate metric [8] is used.

In summary, experiments are conducted following these four steps;

- Configure the codecs.

- Compress the video using original codec for four different QP, record bitrate and PSNR values.

- Compress the video using modified codec for four different QP, record bitrate and PSNR values.

- Compare the results.

(a) Exit

(b) Ballroom

(c) Race1

(d) Vassar

(e) Uli

Figure 5.1: First frames of the test sequences used in the experiments [6].

5.2 Single View Compression Experiment

5.2.1 Experiment Properties

The goal of this experiment is to analyze the efficiency increase obtainable with 1D directional transforms in single view compression. For this purpose, the coding efficiency of the conventional coding system (which uses only 2D DCT) and the modified coding system (which uses also 1D transforms) are compared. Although similar results for single view experiments are provided in [1], we provide single view compression results for completeness here. We will use these results also to compare efficiency increase obtainable with 1D directional transforms in single view and multiview compression.

Conventional coding system and the modified coding system are compared using different transform sizes which are 4x4, 8x8 and both 4x4 and 8x8. Thus, this experiment includes three comparisons:

- 4x4 DCT vs. 4x4 1D (includes 4x4 DCT)

- 8x8 DCT vs. 8x8 1D (includes 8x8 DCT)

- 4x4-and-8x8 DCT vs. 4x4-and-8x8 1D (includes 4x4-and-8x8 DCT)

Only view 1 sequences are used in the experiments. Prediction structure of compression can be seen in Figure 5.2. As seen in the Figure, the first frame is coded as I frame, remaining frames are coded as P frames.

Figure 5.2: Prediction structure of the single view compression experiment

5.2.2 Bjontegaard-Delta Bitrate Results

Figure 5.3 shows Bjontegaard-Delta bitrate savings of the modified encoder, which uses 1D directional transforms and 2D DCT together, with respect to the original encoder, which uses only 2D DCT. Each plot shows the bitrate saving with different transform sizes. Comparisons with 4x4 DCT and 4x4 1D, 8x8 DCT and 8x8 1D, 4x4-8x8 DCT and 4x4-8x8 1D are given in Figure 5.3.

As seen in Figure 5.3, the encoder which can also use 1D directional transforms provides bitrate savings for all the test sequences and the transform sizes. It means that instead of using always 2D DCT, using one of the 1D directional transforms or 2D DCT for the compression of the MC residual increases compression efficiency. This efficiency increase depends on the characteristics of the video sequence and transform size. For all test sequences, maximum bitrate saving is reached with 8x8 transform size and minimum bitrate saving is reached with 4x4 transform size.
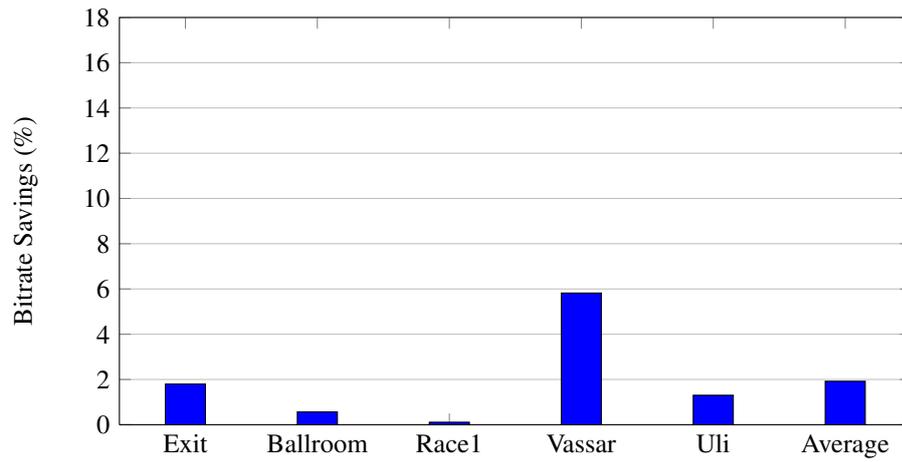
(a) 4x4 1D vs 4x4 DCT
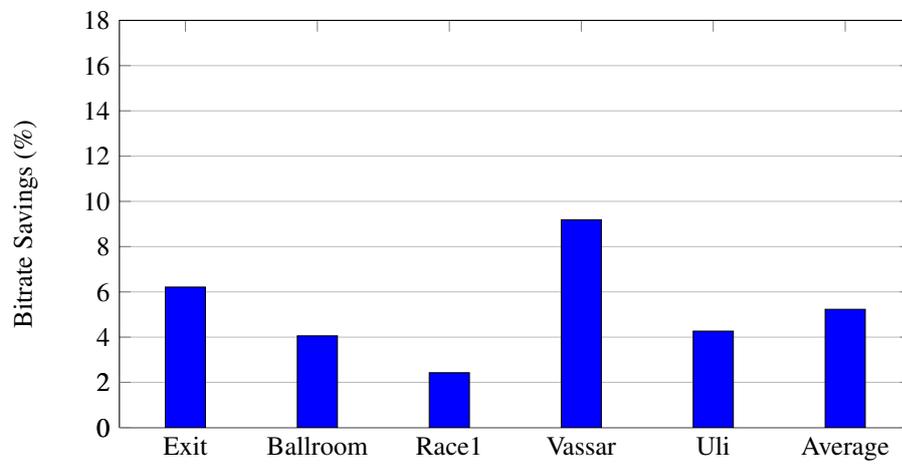
(b) 8x8 1D vs 8x8 DCT
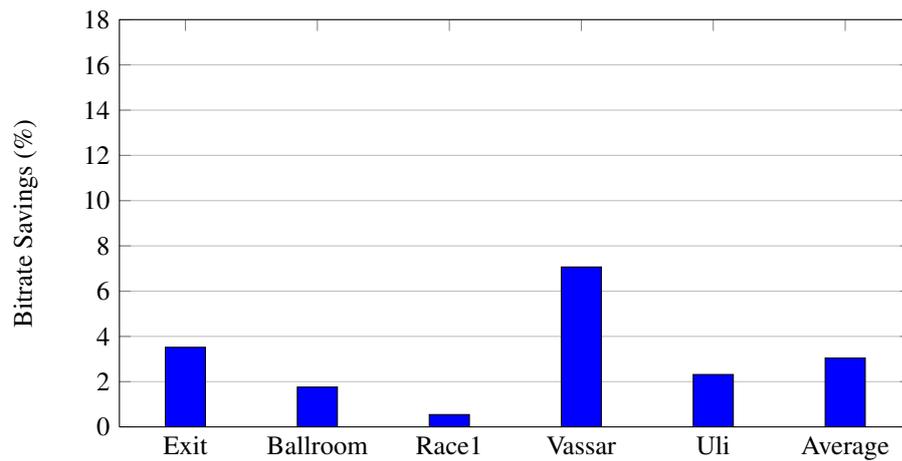
(c) 4x4-8x8 1D vs 4x4-8x8 DCT

Figure 5.3: Average bitrate savings gained by adding 1D directional transform to the encoder. These results are single view compression results of test sequences.

5.3 Multiview Compression Experiment

5.3.1 Experiment Properties

Here, we want to investigate compression efficiency of 1D directional transforms in multiview compression. For this purpose, compression efficiency of the conventional multivew coding system and the modified multiview coding system are compared. Conventional multivew coding system always applies 2D DCT to transform residuals. Modified multiview coding system uses 1D directional transforms or 2D DCT for transforming each block of MC and DC residuals. The efficiency of 1D transforms on multiview compression is analyzed using different transform sizes; 4x4, 8x8 and both 4x4 and 8x8. This experiment includes three comparisons:

- 4x4 DCT vs. 4x4 1D (includes 4x4-DCT)

- 8x8 DCT vs. 8x8 1D (includes 8x8-DCT)

- 4x4-and-8x8 DCT vs. 4x4-and-8x8 1D transform (includes 4x4-and-8x8-DCT)

Two views, view 0 and view 1, are used in the experiments. View 0 is taken as the base view, so view 0 is compressed independently from view 1. View 1 is compressed by using view 0. In compression of view 0, the first frame is coded as I frame, remaining frames are coded as P frames. In compression of view 1, all of the frames are coded as P frames. Prediction structure can be seen in Figure 5.4.
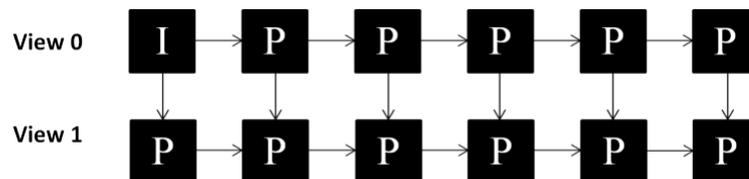
Figure 5.4: Prediction structure of the multiview compression experiment

5.3.2 Rate-Distortion Plots

Figure 5.5 includes the rate-distortion plots of 4x4 1D vs 4x4 DCT, 8x8 1D vs 8x8 DCT and 4x4-8x8 1D vs 4x4-8x8 DCT comparisons. These plots indicate multiview coding results of the Uli view 1 sequence. Since the rate distortion plots of all test sequences are similar to each other, this section presents only plots of one test sequence.

The plots show that for all transform sizes, in lower bitrate coding, modified multiview coding system provides slightly better results than conventional multivew coding system. When bitrate increases, bitrate saving of modified multiview coding system also increases. One of the reason for better efficiency in higher quality is that fraction of bitrate used in the coding of transform coefficients increases as bitrate increases. In lower bitrate coding, more transform coefficients are zero compared to higher bitrate coding, so the effect of 1D directional transforms becomes less visible. The other reason is bitrate cost of the side information. Bitrate of the side information requires higher percentage of the total bitrate in lower bitrate coding compared to higher bitrate coding.

The plots also show that increase in the transform size also increases the difference between the two curves. The difference between two curves is most in 8x8 case and least in 4x4 case. This originates from the fact that difference between 1D and 2D transforms becomes smaller with smaller transform size. In the extreme case, taking 1 point 1D DCT and 2D DCT give the same results.

5.3.3 Bjontegaard-Delta Bitrate Results

Figure 5.6 shows Bjontegaard-Delta bitrate savings of the modified codec with 1D transforms and 2D DCT with respect to the conventional codec with 2D DCT. Plots of the experiments where encoder can use 4x4, 8x8 or 4x4-8x8 transforms are given separately in Figure 5.6. 1D transform improves coding performance of all test sequences. However, this improvement changes depending on the transform size and the test sequence.

The increase in transform size affects bitrate saving of the modified codec over conventional codec positively. When three charts are examined, it is seen that bitrate saving of the modified coded is largest in 8x8 case and smallest in 4x4 case. As explained in Section 5.3.2, this is because the difference between 8x8 1D and 2D transforms is higher than the difference between 4x4 1D and 2D transforms.

Characteristic of the video sequence highly affects the bitrate savings obtainable by adding 1D transform option to the codec. In some of the cases, this dependency becomes very explicit. For example, in 8x8 DCT case, bitrate saving of the Exit sequence is almost 2.5 times larger than bitrate saving of Race1 sequence. Typically, video sequences which have more 1D structures tend to use 1D transforms more frequently. Therefore, these sequences can obtain higher bitrate saving by using 1D transforms and 2D DCT together. Furthermore, magnitude of 1D structures affects bitrate savings. When magnitude of a 1D structure in a local region increases, bitrate saving obtained by transforming the region using one of 1D transforms instead of 2D DCT also increases.
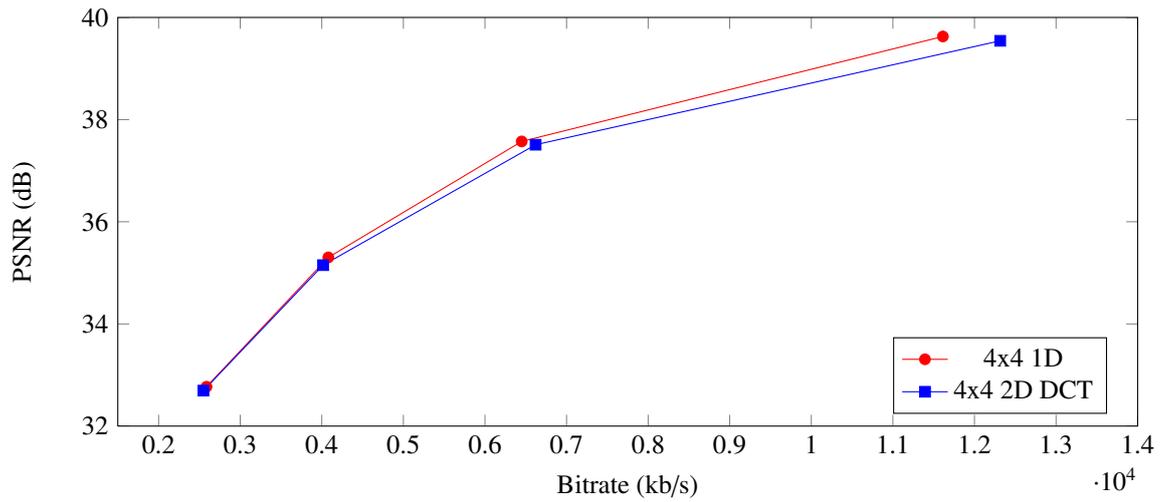
(a) 4x4 1D vs 4x4 2D DCT
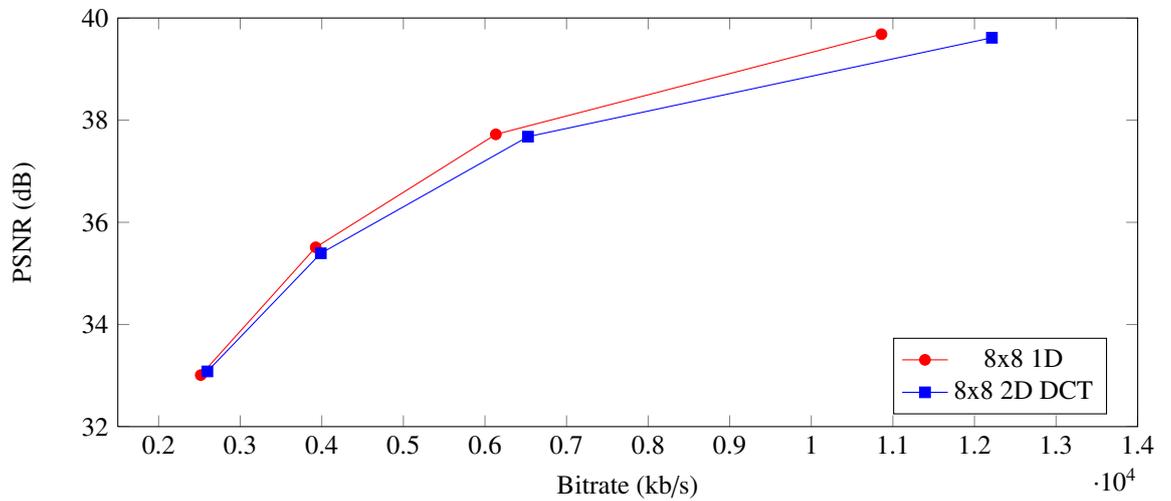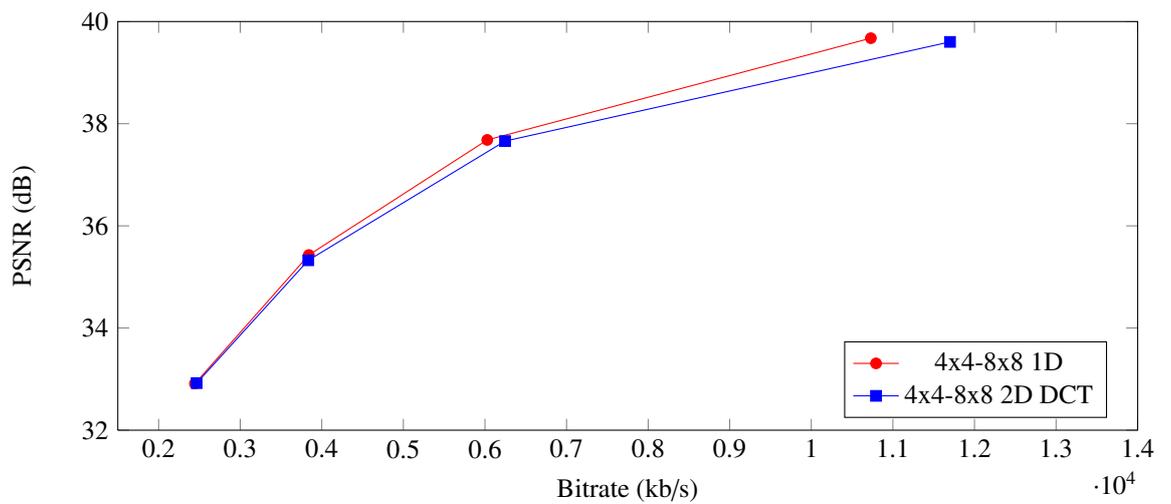
(b) 8x8 1D vs 8x8 2D DCT

(c) 4x4-8x8 1D vs 4x4-8x8 2D DCT

Figure 5.5: MVC results of the Uli view 1 sequence. This view is predicted from view 0 different transform sizes and different transforms.
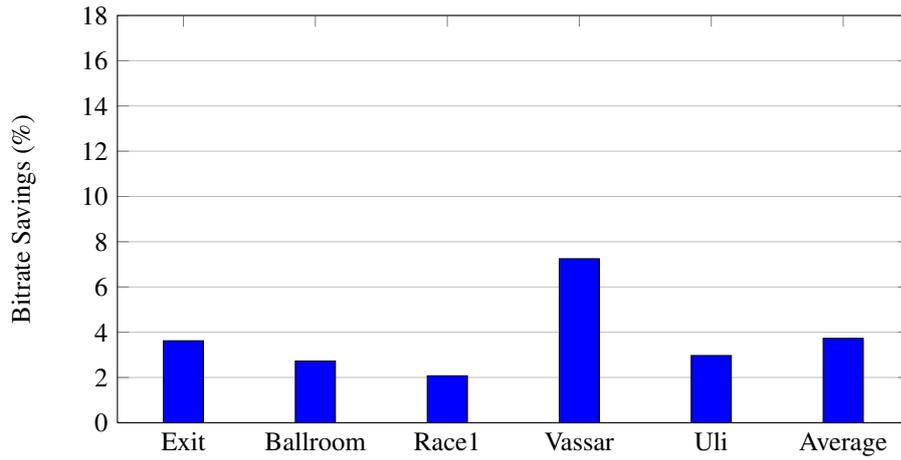
(a) 4x4 1D vs 4x4 2D DCT
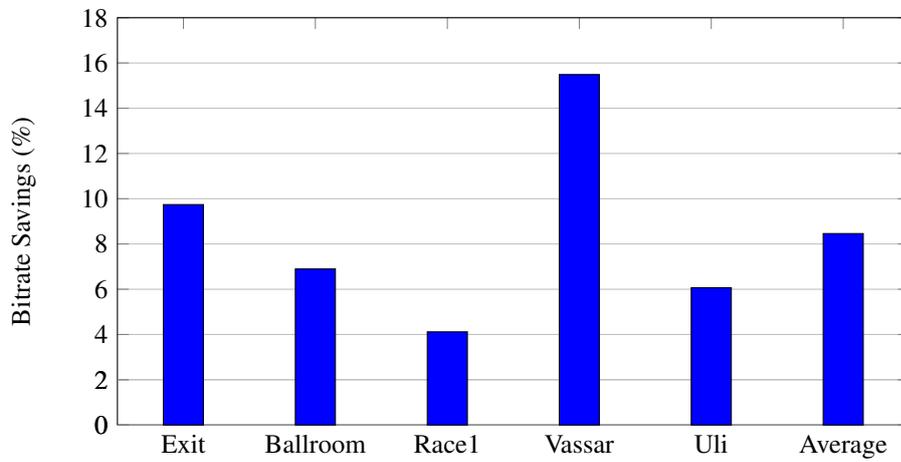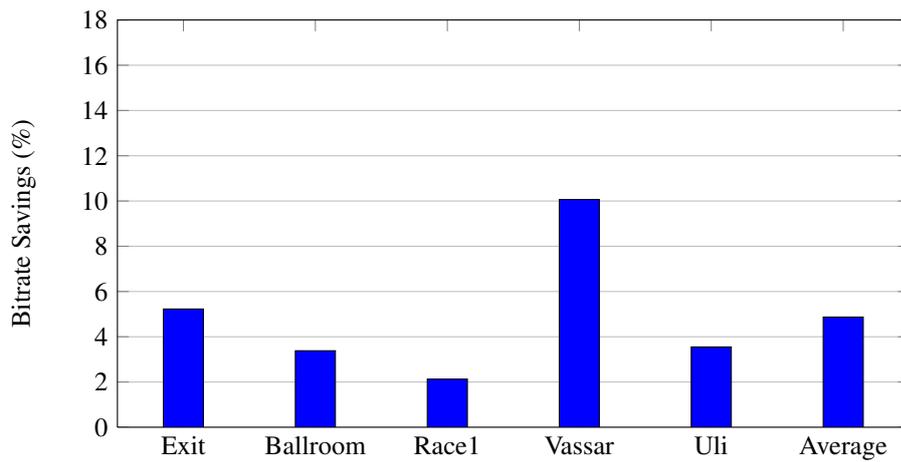
(b) 8x8 1D vs 8x8 2D DCT

(c) 4x4-8x8 1D vs 4x4-8x8 2D DCT

Figure 5.6: Average bitrate savings gained by adding 1D directional transform to the encoder. These results are MVC results of view 1 sequences. This view is predicted from view 0 by using different transform sizes and different transforms.

5.3.4 Bitrate for Coding Side Information

In Section 4.3, it is mentioned that the chosen transform for each block (either one of several directional 1D transforms or the 2D DCT) is transmitted by the encoder to the decoder as side information. In this section, the bitrate cost of the side information is examined.

Figure 5.7 shows the percentage of bits sent as side information for the Exit sequence. Bar charts are obtained using encoders with different transform sizes. In the bar chart, each column represents side information bitrate percentage with different quantization parameters. In all of three charts, when QP increases, bitrate percentage of the side info decreases. In Section 5.3.2, it is explained that decreasing QP affects the performance of 1D transform positively. In general, if QP is reduced, usage of 1D increases, increase in the usage of 1D transform increases both bitrate saving and bitrate percentage of the side information.

Figure 5.8 indicates the average side information bitrate percentage of all test sequences when encoder uses both 4x4 and 8x8 1D transforms. There is a correlation between average bitrate saving and average side info bitrate percentage. Sequence with higher bitrate saving tends to have higher average side info bitrate percentage. However, this correlation is strong only if the difference between bitrate savings is high. Vassar sequence, which has by far the largest average bitrate saving according to Figure 5.6c, also has the largest average side information bitrate percentage. On the other hand, although Exit sequence has higher bitrate saving than Ballroom sequence in Figure 5.6c, it has slightly smaller side info bitrate percentage.

Both Figure 5.7 and Figure 5.8 reveal that side information costs significant amount of bitrate. In our implementation, we send side information bits by using a simple algorithm. More efficient algorithms may be developed.

(a) 4x4 Encoder   (b) 8x8 Encoder   (c) 4x4-8x8 Encoder
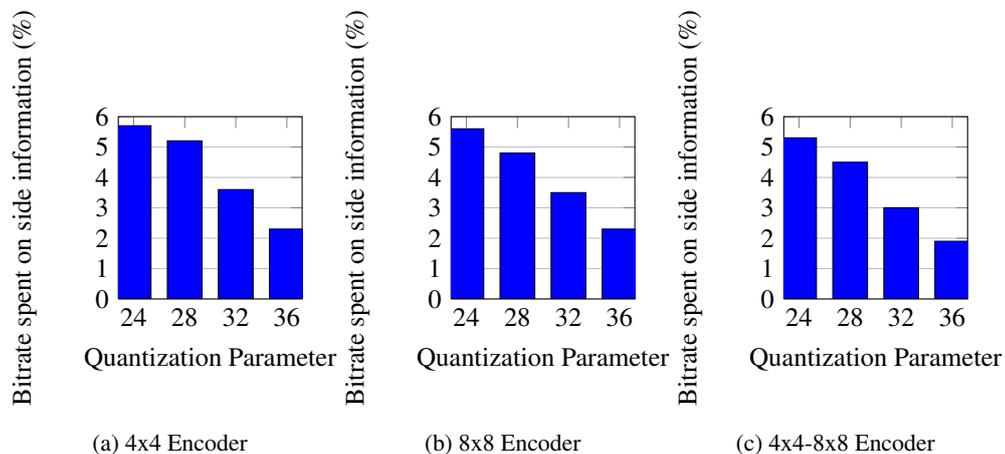
Figure 5.7: Percentage of bits sent as side information. These graphs are obtained from MVC coding of Exit view 1 sequence using different transform sizes. As a multiview reference, view 0 is used.
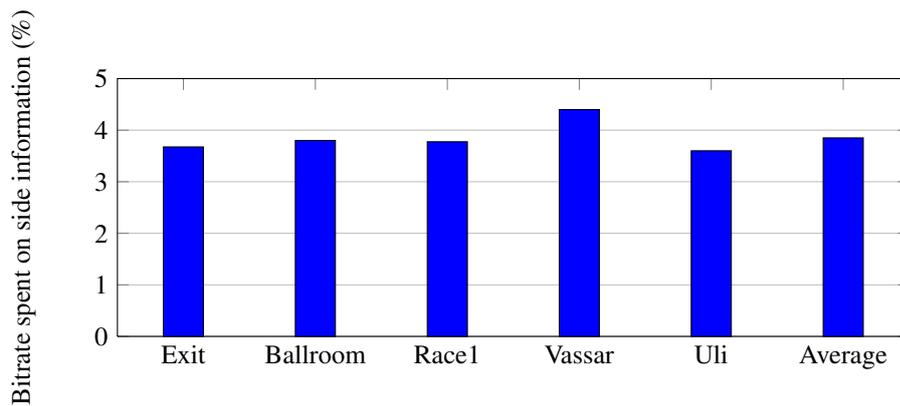
Figure 5.8: Average side information bitrate percentage of all test sequences when encoder uses both 4x4 and 8x8 1D transforms.The average probabilities are obtained with quantization parameters 24, 28, 32 and 36.

5.3.5 Probabilities for Selection of Transforms

Probabilities for selection of transforms depend on quantization parameter (video quality), characteristics of the sequence and transform size. Figure 5.9 shows average transform selection probabilities of all test sequences when 4x4 and 8x8 transforms are used with quantization parameter 24 and 36. Quantization parameter 24 represents high picture quality, quantization parameter 36 represents low picture quality.

2D DCT performs well in many regions. On the other hand, each one of 1D directional transforms performs well in specific regions. Therefore, as seen in Figure 5.9, 2D DCT is selected more often than any 1D transform. However, total usage probability of 1D directional transforms is considerable. The other reason for high selection probability of 2D DCT is the length of the side information bits. It has been mentioned in Section 4.3 that side information codeword of 2D DCTs, any 4x4 1D transform and any 8x8 1D transform are 1, 4 and 5 bits respectively. Since 1D transforms have longer codewords, they have a disadvantage in rate distortion optimization step compared to 2D DCTs. 1D transforms are selected if their distortion score is good enough to compensate for their higher bitrate of the side info.

In high quality videos, 1D transforms are selected more often compared to low quality videos because the effect of the difference between side information bits decreases. In high bitrate coding, cost of side information bits is less important compared to total bitrate of transformed block.
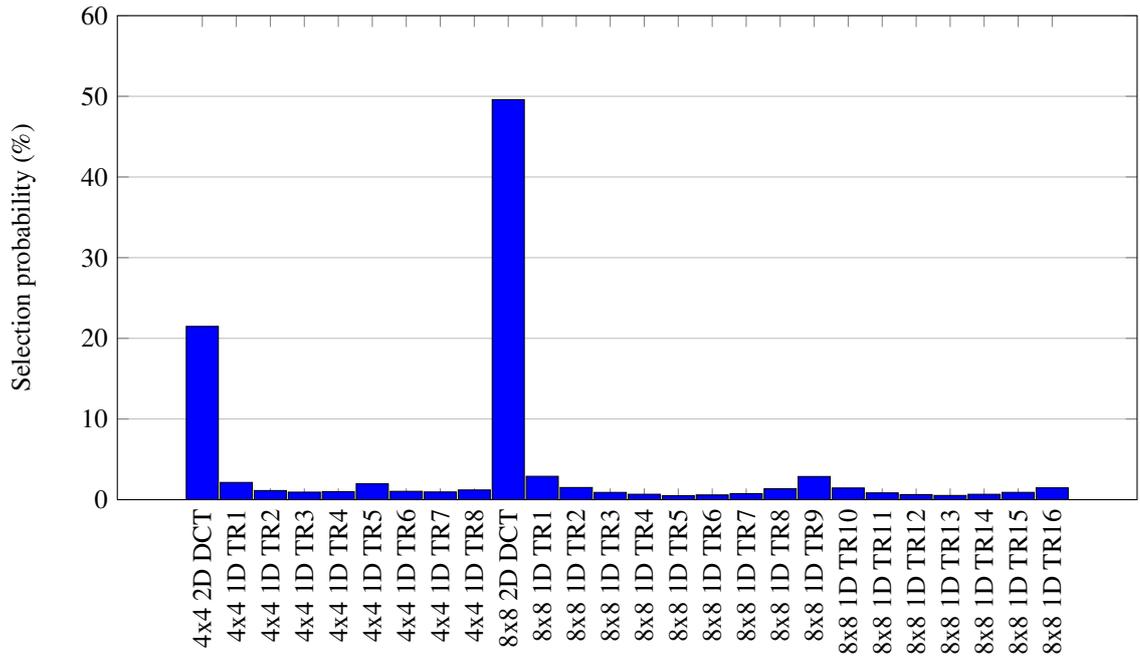
(a) High picture quality (QP=24)
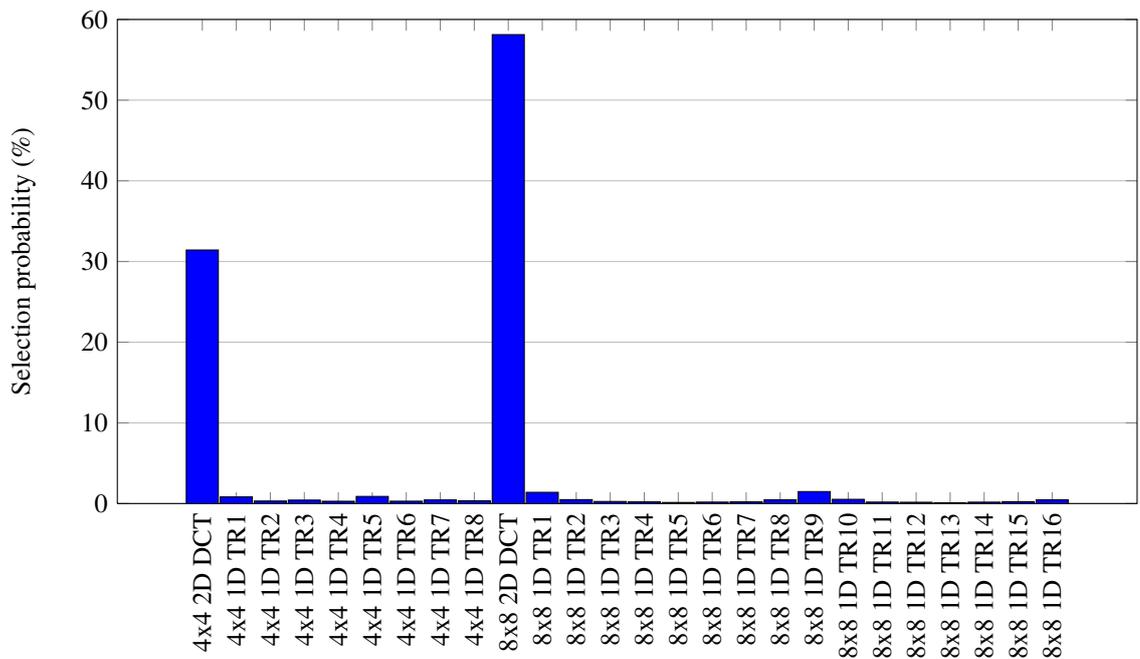

(b) Low picture quality (QP=36)

Figure 5.9: Average transform selection probabilities of all test sequences at different picture quality levels when 4x4 and 8x8 transforms are used.

5.4 Comparison of 1D Directional Transform Efficiency on Single View and Multiview Compression

In this section, bitrate savings obtainable by adding 1-D directional transforms to the codec in single view and multiview compression are compared using findings in previous sections. Section 5.2 provides single view compression experiment results, and Section 5.3 provides multiview compression experiment results. Configurations of experiments only differ in terms of prediction structure. Unlike in single view compression experiments, inter-view prediction between view 1 and view 0 is available in multiview compression experiments.

Figure 5.3 and Figure 5.6 show bitrate saving results of single view and multiview compression experiments. Both figures are very similar to each other. Firstly, correlation between transform size and bitrate saving is the same. For both experiments, increase in transform size affects bitrate savings positively. As explained in 5.3.2, this is because difference between 1D and 2D transforms decreases with decreasing block size. In the extreme case, 1 point 1D and 2D DCT give the same results.

The ranking in bitrate savings among video sequences is also similar for single view and multiview compression experiments.The Vassar and Race1 test sequences have the largest and smallest bitrate savings in Figures 5.3 and 5.6.

From the figures it is apparent that adding 1-D directional transforms to the codec improves compression efficiency in both single view and multiview compression. Bitrate savings in multiview compression are slightly higher compared to single view compression. The reason of higher bitrate saving in multiview compression is characteristics of disparity compensation residual. Since the disparity between two views is in only horizontal direction, many regions in disparity compensation residual have vertical 1D structures.

5.5 Experiments with Different Frame Types

5.5.1 Experiment Properties

The aim of this experiment is to investigate the effect of frame type on directional 1D transform efficiency. In multiview compression experiment (Section 5.3), we have used the prediction structure which has I and P frames. In this experiment, we use the prediction structure which has I, P and B frames. To find out the effect of frame type on directional 1D transform efficiency, the results of this experiment and multiview compression experiment will be compared.

As in multiview compression experiment, conventional multiview coding system and the modified multiview coding system are used to analyze the compression efficiency increase obtained by adding directional 1D transforms to codec. Conventional multiview coding system always uses 2D DCT to transform residuals. Modified multiview coding system uses 1D transforms or 2D DCT to transform each block of MC and DC residuals.

Throughout this experiment, 4x4 and 8x8 transforms are available in the codec. The compression results of view 1 are used. View 0 is taken as base view, inter-view prediction between view 0 and view 1 is available in the encoder. Prediction structure can be seen in Figure 5.10. In compression of view 1, both P and B frames are used.
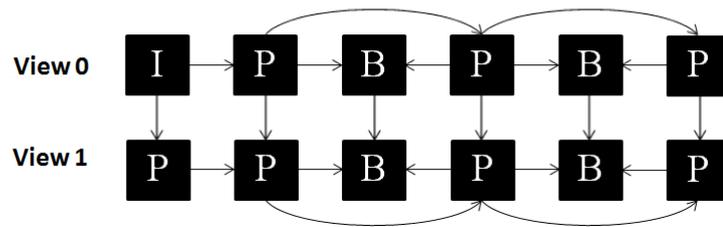
Figure 5.10: Prediction structure with B frames

5.5.2 Rate-Distortion Plots

In this section, plot of Uli sequence is utilized. The rate-distortion plot of 4x4-8x8 1D vs 4x4-8x8 2D comparison is shown in Figure 5.11.

The plot illustrates that in lower bitrate, conventional multiview coding system and the modified multiview coding system give almost the same results. In higher bitrate, modified multiview coding system gives better results than conventional multiview coding system. As in multiview compression experiment, the performance of 1D transforms increases with quality increase.

The Figure 5.5c in multiview coding experiment is obtained by using the same encoder configuration with Figure 5.11 except the prediction structure. As explained in experiment properties sections, in multiview compression experiment, only P frame residuals exist. However, in this experiment, P and B frame residuals exist. In Figure 5.11, difference between 1D transform and 2D DCT curve is smaller compared to Figure 5.5c. It means that adding B frames between P frames decreases performance increase of modified multiview coding system.
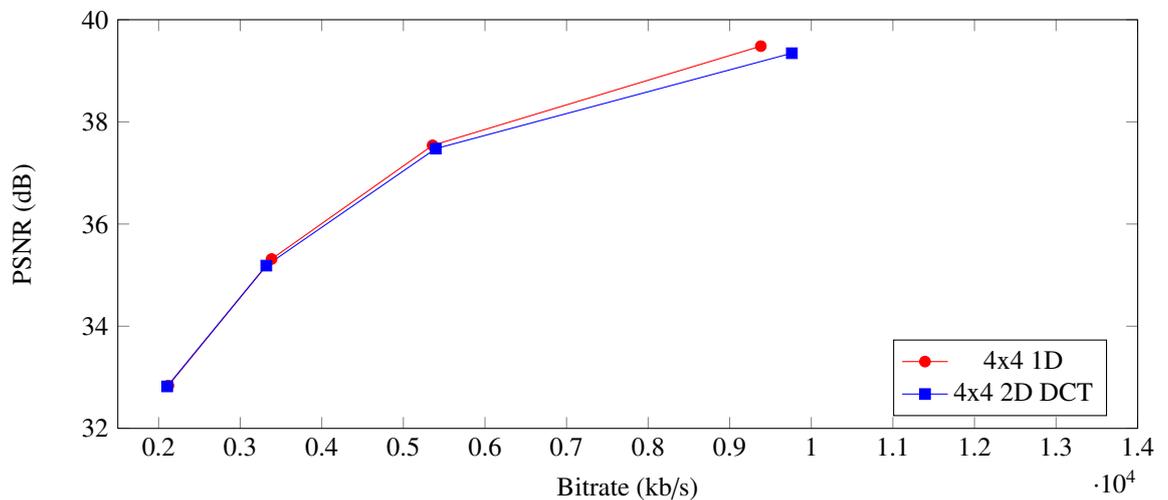

Figure 5.11: MVC results of the Uli view 1 sequence. This view is predicted from view 0 by using 4x4-8x8 1D (2D DCT is also included) and 4x4-8x8 2D DCT transforms.

5.5.3 Bjontegaard-Delta Bitrate Results

In this section, we provide Bjontegaard-Delta comparisons. The comparison results can be seen in Figure 5.12. Adding 1D directional transforms to the coding system provides bitrate saving for all test sequences. On average, %2.9 bitrate saving is obtained. Bitrate saving highly depends on characteristics of the video sequence. For example, the bitrate saving of the Vassar sequence is almost six times bigger than the bitrate saving of the Ballroom and Race1 sequences. Some factors such as weight and shape of residuals affect 1D transform performance and cause bitrate saving difference between different video sequences. Increase in number and magnitude of the 1D structures also increases the bitrate savings obtained by adding 1D directional transforms to the coding system.
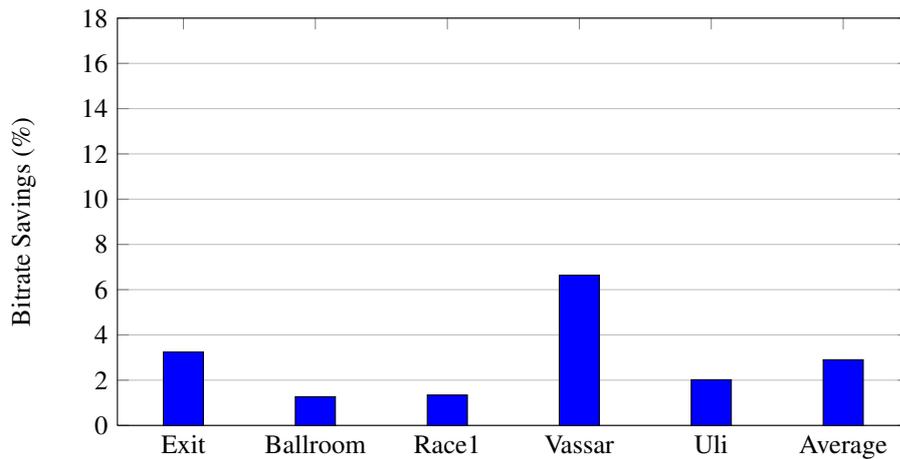

Figure 5.12: Average bitrate savings gained by adding 1D directional transform to the encoder. These results are MVC results of view 1 sequences. This view is predicted from view 0 by using 4x4-8x8 1D (2D DCT is also included) and 4x4-8x8 2D DCT transforms.

Difference between bitrate savings in Figure 5.6c and Figure 5.12 arises from the different prediction structure since those figures are results of the same encoder configuration except prediction structure. As mentioned in experiment properties sections, Figure 5.5c is related to prediction structure in Figure 5.4, and Figure 5.12 is related to prediction structure in Figure 5.10. From the bitrate saving figures, it can be observed that for all video sequences, bitrate savings at Figure 5.5c is higher than bitrate savings at Figure 5.12. Bitrate savings of all sequences are affected by adding B frames between P frames negatively. Also, average bitrate saving decreases from %4.87 to %2.9. Two main reasons that cause this decrease are explained below.

Firstly, adding B frames between P frames in the configuration structure decreases bitrate saving because generally, 1D transforms are more effective in P frames than in B frames. The reason is that B frame prediction residuals have lower magnitude and less 1D structures compared to P frame prediction residuals. In B frames, the magnitude of prediction residuals is lower because the number of reference pictures are higher than P frames, and distinctively from P frames, multihypothesis prediction option is available. In terms of shape, prediction residuals of P frames tend to have more 1D structures compared to B frames because multihypothesis prediction smooths 1D structures by averaging predicted blocks. Hence, P frames tend to use 1D directional transforms more frequently than B frames. For instance, in Vassar sequence with QP 24, average probability of using 1D directional transforms is %32 for P frames and %26 for B frames.

Secondly, as the distance between P frames increases, the interframe correlation and 1D structures in MC residuals decrease. For this reason, increasing the distance between P frames tends to decrease probability of using 1D directional transforms and the bitrate savings of modified coding system. Table 5.2 shows the effect of increasing the distance on probability of using 1D directional transforms. The values in the table are probabilities of the Vassar sequence with QP 24.

Table 5.2: Probabilities of using 1D directional transforms when Vassar sequence is encoded by using different prediction structures with QP 24.

Distance between frames	P Frame #4	P Frame #6	P Frame #8	P Frame #10
1 Frame	%31	%31	%32	%33
2 Frames	%28	%27	%27	%28

5.5.4 Bitrate for Coding Side Information

Bitrates for coding side information of this experiment and multiview compression experiment (Section 5.3) are very similar. Therefore, we do not discuss the bitrate for coding side information results of this experiment to avoid repetition.

5.6 Experiments with Global Luminance Compensation Between Views

Significant variations between global luminance and/or chrominance of the frames in the neighbouring views can be observed due to several reasons such as calibration and view angle differences [16]. Compensating these variations in disparity compensation algorithm may provide better prediction. In [17], block-based illumination compensation method based on predictive coding of the DC coefficient of the integer transform is proposed. Another way of dealing with the intensity variations is to use histogram based prefiltering method in [12]. In this method, histograms of two views are matched before encoding so, encoder and decoder are not modified.

Due to variations between global luminance in views, disparity compensation residual can have DC components. Therefore, compensating these variations can increase bitrate savings of modified coding system (which uses 2D DCT and 1D transforms) compared to the conventional coding system (which uses only 2D DCT).

For experimental investigation of luminance compensation effect, we conduct some experiments. To compensate luminance variations between the views, histogram based prefiltering method in [12] is used. Figure 5.13 shows sample Uli sequences before and after luminance compensation. Original view 1 of the Uli sequence is darker compared to view 0 sequence, and with the luminance compensation process, it becomes lighter. After compensation process, multiview compression experiments in Section 5.3 are repeated for the test sequences with the encoder which uses both 4x4 and 8x8 transforms.

Figure 5.14 shows the bitrate savings with and without global luminance compensation. For the Uli sequence, with luminance compensation, bitrate saving of the modified coding system increases significantly (from %3.5 to %5.8). On the other hand, luminance compensation does not change much the bitrate savings of the other test sequences. The reason of significant bitrate saving increase in Uli sequence is the high luminance variation between the views. Global luminance compensation reduces

this high variation and increases the usage probability of 1D transforms for this sequence. For example, usage probability of 1D transforms increases from %38 to %52 for high picture quality (QP=24). For other sequences, since luminance variation between views is small, the effect of global luminance compensation is not clearly visible.

(a) Uli View 0 - Frame 1

(b) Uli View 1 - Frame 1

(c) Uli View 1 - Frame 1 After Compensation

Figure 5.13: Frame 1 view 0 of the Uli sequence and frame 1 view 1 of the Uli sequence before and after global luminance compensation [6].
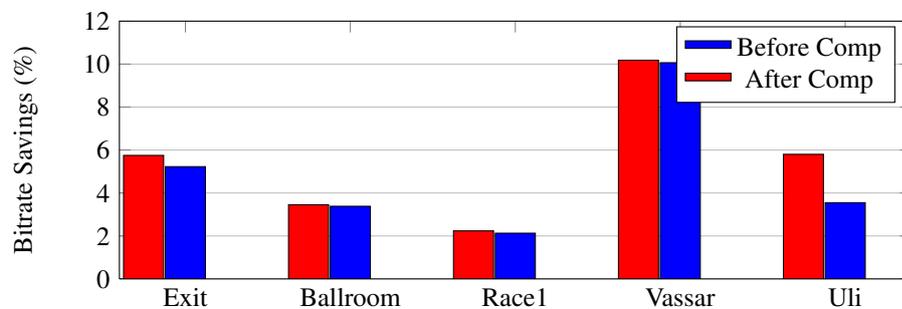

Figure 5.14: Average bitrate savings of modified coding system before and after global luminance compensation with using 4x4-8x8 transforms.

CHAPTER 6

CONCLUSIONS

6.1 Summary

2-D DCT is commonly used to compress prediction residual and image intensities. Considering the characteristics of the images, it can be said that 2-D DCT is a good match to the characteristics of images. When images are modeled with the Markov-1 model, 0.95 can be taken as an acceptable approximation for the correlation parameters [38]. 2-D DCT is an optimum statistical transform for a Markov-1 process if its correlation parameters are close to 1. Images are mostly composed of smoothly varying regions with high 2-D correlation, and these regions can be represented well with the smooth basis functions of 2-D DCT.

On the other hand, the characteristics of the residuals may be different from the characteristics of the images in many regions. Typically, spatial correlation in residuals is smaller than in images. Most of the energy of the residuals is concentrated in the regions which are difficult to predict. Considering the specific statistical characteristics of the residuals, transforms other than 2-D DCT can be used for transformation of residuals.

Previously, the spatial characteristics of images and motion compensation residuals have been analyzed. The analysis results show that images have 2-D anisotropic structures, and many regions of MC residuals have 1-D anisotropic structures. By referring to these analysis results, 1-D directional transforms have been designed. It has been demonstrated that using 1-D directional transforms for the compression of MC residual regions which have 1-D anisotropic structures increases overall compression efficiency.

In this thesis, we analyzed the spatial characteristics of the disparity compensation residual. The analysis shows that, similar to the motion compensation residual, many regions of the disparity compensation residual have 1-D anisotropic structures. Considering these results, we propose the use of 1-D directional transforms for the compression of disparity compensation residuals in addition to the compression of motion compensation residuals. We conducted some experiments to determine the effect of these transforms on the compression efficiency. In these experiments, the reference codec for MVC extension of H.264/MPEG-4 AVC was modified. We compared the coding efficiency of the conventional coding system (which uses only 2D DCT) and the modified coding system (which uses also 1D transforms). Results of the experiments indicate that using 1-D directional transforms in multiview video compression increases overall compression efficiency.

6.2 Future Research

In this section, some future research directions are discussed. In this thesis, CAVLC was chosen as the entropy coding mode of the codec. This also means that side information bits are transmitted using VLC. Although CABAC requires higher processing power to decode than CAVLC, it compresses the data more efficiently. Therefore, using CABAC instead of CAVLC may decrease the percentage of bits sent as side information and increase the performance of the 1-D directional transforms.

Correlation between selected transforms is another topic which may be investigated. In this thesis, the side information of each block was coded independently from each other. On the other hand, there may be a correlation between selected transforms of the neighboring blocks, and side information bits can be coded using this correlation. For example, the side information of one block can be predicted from side information of previously- coded neighboring blocks at the decoder side.

REFERENCES

[1] F. Kamisli and J.S. Lim.Transforms for Prediction Residuals in Video Coding.In *Ph.D, Thesis, Massachusetts Institute of Technology, Dept. of Electrical Engineering and Computer Science,* Sept 2010

[2] A. Vetro, T. Wiegand and G.J. Sullivan.Overview of the Stereo and Multiview Video Coding Extensions of the H.264/MPEG-4 AVC Standard.In *Proceedings of the IEEE, vol.99, no.4*, pages 626-642, Apr 2011

[3] Xun Cai and J.S. Lim.Algorithms for transform selection in multiple-transform video compression.In *Image Processing (ICIP), 2012 19th IEEE International Conference*, pages 2481-2484, Sept. 2012

[4] H.Zhang and J.S Lim. Analysis of one-dimensional transforms in coding motion compensation prediction residuals for video applications.In *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference* , pages 1229-1232, Mar 2012

[5] I.E. Richardson.An Overview of H.264 Advanced Video Coding. *[Online], Available:(http://www.vcodex.com/images/uploaded/337054245340685.pdf)*.(last accessed on 23/05/2013)

[6] Y. Su, A. Vetro and A. Smolic.Common Test Conditions for Multiview Video Coding. *JVT-T207, the 20th Meeting of Video Coding Experts Group (VCEG),*Klagenfurt, Austria, Jul 2006

[7] A. Vetro, M. McGuire, W. Matusik, A. Behrens, J. Lee and H. Pfister.Multiview Video Test Sequences from MERL. *ISO/IEC JTC1/SC29/WG11/MPEG05/m12077,*Busan, Korea, Apr. 2005

[8] G. Bjontegaard. Calculation of average psnr differences between rd-curves. *VCEG-M33, the 13th Meeting of Video Coding Experts Group (VCEG),*Austin, Texas Apr 2001

[9] T. Wiegand, G.J.Sullivan, G.Bjontegaard, A.Luthra.Overview of the H.264/AVC video coding standard. *Circuits and Systems for Video Technology, IEEE Transactions on,* vol.13, no.7, pages 560-576, Jul 2003

[10] I. E. Richardson.*The H.264 Advanced Video Compression Standard.*2nd ed. Upper Saddle River, N.J.: Wiley. pages 185-203,2010

[11] K.-M. Wong, L.-M. Po, K.-W. Cheung, C.-W. Ting, K.-H. Ng, X. Xu.Horizontal Scaling and Shearing-Based Disparity-Compensated Prediction for Stereo Video Coding. *Circuits and Systems for Video Technology, IEEE Transactions on,* vol.22, no.10, pages 1457-1470, Oct 2012

[12] U. Fecker, M. Barkowsky, A.Kaup.Histogram-Based Prefiltering for Luminance and Chrominance Compensation of Multiview Video. *Circuits and Systems for Video Technology, IEEE Transactions on* vol.18, no.9, pages 1258-1267, Sept 2008

[13] X. San, H. Cai, J.-G. Lou, J. Li.Multiview Image Coding Based on Geometric Prediction. *Circuits and Systems for Video Technology, IEEE Transactions on,* vol.17, no.11, pages 1536-1548, Nov 2007

[14] W. Woo and A. Ortega.Overlapped Block Disparity Compensation with Adaptive Windows for Stereo Image Coding. *Circuits and Systems for Video Technology, IEEE Transactions on,* vol.10, no.2, pages 194-200, Mar 2000

[15] A. Kaup and U. Fecker.Analysis of multi-reference block matching for multi-view video coding. *in Proc. 7th Workshop Digital Broadcasting,* pages 33-39, Erlangen, Germany, Sep 2006

[16] J. Lòpez, J. H. Kim, A. Ortega, and G. Chen.Block-based illumination compensation and search techniques for multiview video coding. *in Proc. Picture Coding Symp. (PCS 2004),* pages 15-17, San Francisco, CA, USA, Dec 2004

[17] Y.-L. Lee, J.-H. Hur, Y.-K. Lee, K.-H. Han, S. Cho, N. Hur, J. Kim, J.-H. Kim, P.-L. Lai, A. Ortega, Y. Su, P. Yin, and C. Gomila. Block-based illumination compensation and search techniques for multiview video coding. *CE11:Illumination compensation, in Joint Video Team (JVT) of ISO/IECMPEG ITU-T VCEG, Document JVT-U052r2.*Hangzhou, China, Oct 2006

[18] Video Bitrate Calculator [Online], Available: (http://web.forret.com/tools/video_fps.asp). (last accessed on 23/05/2013)

[19] Sample YUV Video Sequences [Online], Available: (http://trace.eas.asu.edu/yuv/). (last accessed on 23/05/2013)

[20] C.-L. Chang and B. Girod.Direction-adaptive discrete wavelet transform for image compression. *Image Processing, IEEE Transactions on,*16(5): pages 1289-1302, May 2007

[21] K.-C. Hui and W.-C. Siu.Extended analysis of motion-compensated frame difference for block-based motion prediction error. *Image Processing, IEEE Transactions on,*16(5):pages 1232-1245, May 2007

[22] D. Marpe, H. Schwarz, and T. Wiegand.Context-based adaptive binary arithmetic coding in the h.264/avc video compression standard. *Circuits and Systems for Video Technology, IEEE Transactions on* 13(7):pages 620-636, July 2003

[23] The Moving Picture Experts Group website. *[Online], Available:(http://mpeg.chiariglione.org)* (last accessed on 23/05/2013)

[24] R. C. Gonzalez, R. E. Woods.*Digital Image Processing.*3rd ed. Hoboken, N.J.: Prentice Hall, 2008

[25] M. Flierl, B. Girod.Multihypothesis motion estimation for video coding. *Data Compression Conference, 2001. Proceedings. DCC 2001,* pages 341-350, 2001

[26] T. Wiegand and B. Girod.Lagrange multiplier selection in hybrid video coder control.In *Image Processing, 2001. Proceedings. 2001 International Conference on,* volume 3, pages 542-545 vol.3, 2001

[27] H.S. Malvar, A. Hallapuro, M. Karczewicz, and L. Kerofsky. Low-complexity transform and quantization in h.264/avc. emph Circuits and Systems for Video Technology, IEEE Transactions on, 13(7): pages 598-603, July 2003

[28] G.Bjontegaard and K. Lillevold.Context-Adaptive VLC Coding of Coefficients. *JVT-C028, the 3rd Meeting of Video Coding Experts Group (VCEG),*Fairfax, Virginia, USA, May 2002

[29] B. Girod.Why b-pictures work: a theory of multi-hypothesis motion-compensated prediction. *Image Processing, 1998. ICIP 98. Proceedings. 1998 International Conference on,* 2:213,217 vol.2, Oct 1998

[30] M.S. Moellenhoff and M.W. Maier.Characteristics of disparity-compensated stereo image pair residuals. *Signal Processing:Image Communication* vol.14, no.1-2, pages 55-69, Nov 1998

[31] B. Girod.Efficiency analysis of multihypothesis motion-compensated prediction for video coding. *Image Processing, IEEE Transactions on,* 9(2):pages 173-183, Feb 2000

[32] B. Girod.Motion-compensating prediction with fractional-pel accuracy. *Communications, IEEE Transactions on,* vol.41, no.4, pages 604-612, Apr 1993

[33] D. P. Bertsekas. *Nonlinear Programming,* Athena Scientific, 1995

[34] W. Niehsen and M. Brunig. Covariance analysis of motion-compensated frame differences. *Circuits and Systems for Video Technology, IEEE Transactions on,*9(4):pages 536-539, Jun 1999

[35] C.-F. Chen and K.K. Pang. The optimal transform of motion-compensated frame difference images in a hybrid coder. *Circuits and Systems II: Analog and Digital Signal Processing, IEEE Transactions on [see also Circuits and Systems II: Express Briefs, IEEE Transactions on],* 40(6):pages 393-397, Jun 1993

[36] A. Ortega, K. Ramchandran.Rate-distortion methods for image and video compression. *Signal Processing Magazine, IEEE,* 15(6):23,50, Nov 1998

[37] G.J. Sullivan and T. Wiegand.Rate-distortion optimization for video compression. *Signal Processing Magazine, IEEE,* 15(6):74,90, Nov 1998

[38] N. Ahmed, T. Natarajan, and K.R. Rao.On image processing and a discrete cosine transform.In *Computers, IEEE Transactions on,* C-23(1):90-93, Jan 1974

[39] J.S.Lim.*Two-dimensional Signal and Image Processing.*Prentice Hall, 1990

[40] A. Puri, H.-M. Hang, and D. Schilling.An efficient block-matching algorithm for motion-compensated coding.In *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP'87,* volume 12, pages 1063-1066, Apr 1987