Side Information Generation in Distributed Video Coding

Akshay kumar

(Roll No: 213CS1144)

Department of Computer Science and Engineering
National Institute of Technology, Rourkela
Rourkela-769 008, Odisha, India
May, 2015.

Side Information Generation in Distributed Video Coding

Thesis submitted in partial fulfillment

of the requirements for the degree of

Master of Technology

in

Computer Science and Engineering

by

Akshay kumar

(Roll No: 213CS1144)

under the guidance of

Prof. B. Majhi

Department of Computer Science and Engineering
National Institute of Technology, Rourkela
Rourkela-769 008, Odisha, India

May, 2015.

Department of Computer Science and Engineering
National Institute of Technology Rourkela
Rourkela - 769008, Odisha, India

CERTIFICATE

This is to certify that the work in the thesis entitled ***Side information generation in Distributed Video Coding*** by ***Akshay kumar***, having roll number **213CS1144**, is a record of an original research work carried out by him under my supervision and guidance in partial fulfillment of the requirements for the award of the degree of ***Master of Technology*** in ***Computer Science and Engineering Department***. Neither this thesis nor any part of it has been submitted for any degree or academic award elsewhere.

Place: NIT Rourkela
Date:

Prof. B. Majhi
Department of Computer Science and Engineering
National Institute of Technology Rourkela
Rourkela-769008, Odisha, INDIA

Acknowledgment

Preeminent, I want to express my true appreciation towards my supervisor Prof. B. Majhi, who has been the managing compel behind this work. I am most obligated to him for acquainting me with the field of Distributed video coding and giving me the chance to work under him. His undivided faith in this topic and ability to bring out the best of analytical and practical skills in people has been invaluable in tough periods. Without his priceless guidance and support it would not have been workable for me to complete this theory. His invaluable suggestions, constant encouragement and assistance were always encouraging for me in every aspect of my academic life. I am enormously obligated to him for his consistent consolation and priceless counsel in every part of my scholastic life. I think of it as my favorable luck to have got a chance to work with such a magnificent individual.

I wish to thank all faculty members and secretarial staff of the CSE Department for their sympathetic cooperation.

During my studies at N.I.T. Rourkela, I made many friends. I would like to thank them all, for all the great moments I had with them.

When I look back at my accomplishments in life, I can see a clear trace of my family's concerns and devotion everywhere. My dearest mother, whom I owe everything I have achieved and whatever I have become; my beloved father, for always believing in me and inspiring me to dream big even at the toughest moments of my life; and my brother and sister; who were always my silent support during all the hardships of this endeavor and beyond.

Akshay kumar

Abstract

Distributed Video Coding (DVC) coding paradigm is based largely on two theorems of Information Theory and Coding, which are Slepian-wolf theorem and Wyner-Ziv theorem that were introduced in 1973 and 1976 respectively. DVC bypasses the need of performing Motion Compensation (MC) and Motion Estimation (ME) which are largely responsible for the complex encoder in devices. DVC instead relies on exploiting the source statistics, totally/partially, at only the decoder. Wyner-Ziv coding, a partivular case of DVC, which is explored in detail in this thesis. In this scenario, two correlated sources are independently encoded, while the encoded streams are decoded jointly at the single decoder exploiting the correlation between them.

Although the distributed coding study dates back to 1970's, but the practical efforts and developments in the field began only last decade. Upcoming applications (like those of video surveillance, mobile camera, wireless sensor networks) can rely on DVC, as they don't have high computational capabilities and/or high storage capacity. Current coding paradigms, MPEG-x and H.26x standards, predicts the frame by means of Motion Compensation and Motion Estimation which leads to highly complex encoder. Whilst in WZ coding, the correlation between temporally adjacent frames is performed only at the decoder, which results in fairly low complex encoder.

The main objective of the current thesis is to investigate for an improved scheme for Side Information (SI) generation in DVC framework. SI frames, available at the decoder are generated through the means of Radial Basis Function Network (RBFN) neural network. Frames are estimated from decoded key frames block-by-block. RBFN network is trained offline using training patterns from different frames collected from standard video sequences.

Keywords : Distributed video coding, radial basis function network, side information, wyner-ziv coding, low-complexity encoding.

Contents

List of Figures

List of Acronyms

Acronym	Description
DVC	Distributed Video Coding
SW	Slepian-Wolf
WZ	Wyner-Ziv
SI	Side Information
TDWZ	Transform Domain Wyner-Ziv
STD	Stanford
MCFI	Motion Compensated Frame Interpolation
MCI	Motion Compensated Interpolation
PSNR	Peak Signal to Noise Ratio

Chapter 1

Introduction

With the advent of high resolution images and high definition videos, they are very popular and can be easily found in daily use by several people. Relying on quality data for processing led to the development of the multimedia products such as Mobile phone video capture, Wireless camera, Sensor Networks etc. The increase in crime and elevated Terrorist threats has also been a reason for the increase in video surveillance system. More often than not, these applications and/or devices requires storing and/or transmitting of the recorded media. Compression becomes important in such cases, where the video is need to be of minimal space possible but not degrading the visual quality too much. Due to the scarcity of storage space and computational capabilities in the handheld and monitoring devices, we need an algorithm with good compression rate. For some applications/devices it is imperative that they consume low power at both the ends of the codec, as in mobile phone camera.

Modern digital video coding schemes are governed by the ITU-T (International Telecommunication Unit-Telecommunication) and ISO/IEC MPEG (Moving Picture Experts Group) (2) standards, which relies on combination of transformations, block-based, and inter-frame prediction to exploit spatial and temporal correlations within encoded video. This results in high complexity encoders because of the motion estimation (ME) process run at the encoder side. On the other hand, the resulting decoders are simple and around 5 to 10 times less complex than the corresponding encoders (26). However this types of architecture is more suited for the applications where the media is once encoded and might be decoded multiple times. Few such areas include on-demand-video, broadcasting etc. It presents a challenge for the traditional video coding paradigms to fulfill the requirements posed by these applications. So, there is a need for the low cost and power encoding device possibly at the expense of slightly complex decoder. Additional challenge arises while trying to achieve the efficiency as of those achieved by the traditional coding techniques, like those of MPEG-x or H.26x when the complexity shifts from encoder to decoder.

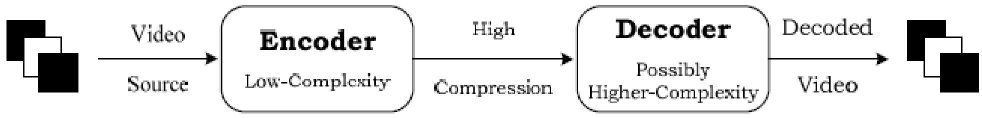

Figure 1.1: Ideal coding architecture for upcoming video applications

1.1 Information Theory Background

Distributed source coding (DSC) mainly depends on the principle of independent encoding and joint decoding. 'Distributed' in DSC points to the distributed nature of encoding operation, not the location as in distributed computing. DSC regard the compression of correlated information resources that do not communicate with each other (1). DSC models the correlation between multiple sources together with channel code and hence able to shift complexity from encoder to decoder. Hence DSC, DVC in current context, can be used to develop the devices having complexity-constrained encoder.

WZ coding scheme has its advantage as it shifts the computational complexity from encoder to decoder. However, WZ coding do not put restrictions on the decoder complexity. Complexity of the decoder also can't be too high either as it has straightforward affect the efficiency of decoding and hence the delayed output. In stanford-based Wyner-Ziv video coding schemes, some feeble operations results in a complex decoder. Some impractical assumptions such as arbitrary size for input block also leaves scope for further optimizations of decoding efficiency.

Moreover, current traditional coding techniques like H.264 and MPEG-4 are capable of achieving high compression by utilizing Motion Estimation (ME) and Motion Compensation (MC). Due to the complexity-constrained encoder, ME and MC cannot be applied in the WZ video coding. However, apart from predictive coding, discrete cosine transformation (DCT) also helps in realizing compression temporally while keeping complexity to a low at the encoder.

From information theoretic perspective, consider X and Y be two statistically dependent sources. According to Shannon's theorem, for reconstruction of the encoded stream X to be lossless, the rate, R(X), should at least be equal to the entropy, H(X), of the same stream.

Fig 1.2 shows independent encoding of two statistically dependent sequences X and Y. Both the sequences can be reconstructed losslessly only if

$$R(X) \geq H(X)$$

and

$$R(Y) \geq H(Y)$$

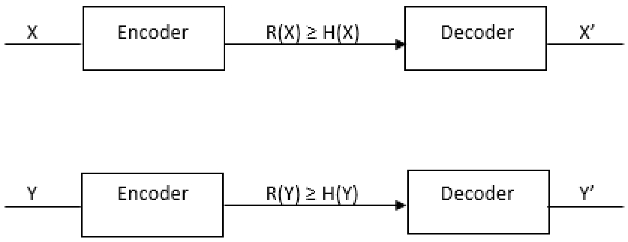

Figure 1.2: Independent encoding and independent decoding

Fig 1.3 shows independent encoding of two statistically dependent sequences X and Y which are jointly decoded. For reconstructing X and Y perfectly, Rate combination R would be

$$R \geq H(X) + H(Y) \geq H(X, Y)$$

The information theoretic background of DVC takes its basis on Wyner-Ziv theorem. Which are discussed in detail.

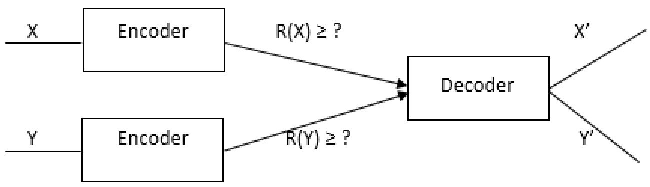

Figure 1.3: Distributed compression of two statistically dependent sequences

1.1.1 Slepian Wolf Theorem

In information Theory and communication, Slepian-wolf theorem provides a method for theoretical coding of two correlated and losslessly compressed sources. DVC is practical realization of the work of Slepian and Wolf, introduced in 1973.
Given two or more, in this case two, dependent sources X and Y which are encoded with separate encoders and are decoded together. Slepian-wolf theorem provides the theoretical bounds for the lossless coding rate as

$$R_X \geq H(X|Y), R_Y \geq H(Y|X), R_X + R_Y \geq H(X,Y)$$

Where $H(X|Y)$ is the conditional entropy of X with given Y and $H(Y|X)$ is conditional entropy of Y with given X and H(X,Y) is joint entropy.

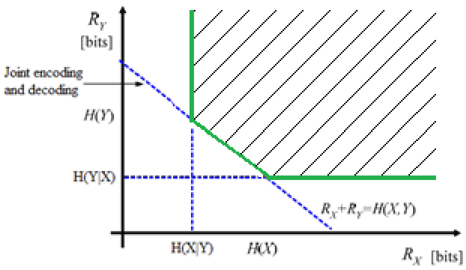

Figure 1.4: Region for Achievable rate as per Slepian-Wolf theorem

Fig. 1.4 illustrates the region for achievable rate for which the distributed compression of two dependent streams X and Y, allows reconstruction with small error probability. The shaded region represents the bounds for achievable rate combination of R_X and R_Y.
Slepian-wolf coding is the term used to depict the architecture followed in the scenario described above in Fig. 1.3. It is also referred as lossless distributed source coding that allows a small error probability at joint decoder. It is to be noted that "lossless" is not exactly lossless as defined in mathematics, since Slepian-wolf allows a controlled margin of error in the sequences.

1.1.2 Wyner-Ziv Theorem

In 1976, A. Wyner and J. Ziv extended the Slepian-wolf theorem for lossy case of distributed video coding. They exploited the source coding of sequence X, when other sequence Y, known as side information (SI), is available at the decoder. The lossy compression of the sequence is due to the fact that an acceptable distortion d is allowed.

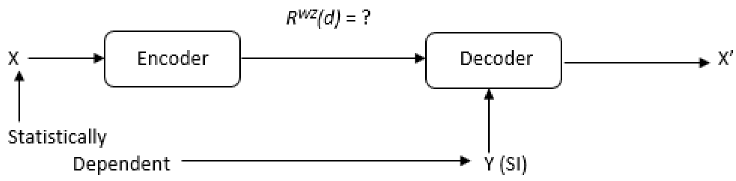

Figure 1.5: Lossy compression scenario with SI available at decoder

Fig. 1.5 depicts the approach taken by Wyner-Ziv (WZ) coding for the lossy compression of the sequence X, allowing an acceptable distortion d, with another sequence available at the decoder as the side information, Y.
WZ theorem can be mathematically summarized as

$$R^{WZ}(d) \geq R_{X|Y}(d), d \geq 0$$

Where R^{WZ}(d) is the minimum bit rate for transmission, given a finite distortion d. $R_{X|Y}$(d) is the encoding rate of X, with Y available both ends of the codec simultaneously.

For d=0, no distortion, WZ theorem behaves similar to that of Slepian-wolf theorem. Also the reconstructed information X' consist of small error probability even when the correlation is exploited only at the decoder.

1.2 Possible applications targeting DVC

1.2.1 Video Surveillance and Monitoring

With the increase in crime and elevated terrorist threats, public safety has become critical and hence the need for Video surveillance. Multiple cameras sense same event from different locations. Video streams from multiple cameras are generally correlated due to the fact that neighboring cameras cover partially overlapping areas. Since the system is centralized and most likely the video stream is needed to be decoded only once, the system can do with encoder of low complexity, to accommodate the low computational capability, and a slightly complex decoder. Hence cost reduction is possible in the system if encoder of low-complexity are used. WZ coding is well suited for the situations where correlation between streams can be explored only at the decoder.

1.2.2 Video based Sensor networks

Sensor networks of smart cameras distributed spatially is capable of fusing and processing (3) scenic images various viewpoints into some that is of more use than individual images. Since sensor network has a scarce of computational power and storage capability, DVC will be much suited for the need of video coding performed by nodes of sensor network.

1.3 Related Work

DVC is realizing DSC principles in practical applications. Theoretic foundations of Slepian-Wolf and Wyner-Ziv theorems were put forward in early 1970s, but the practical implementations of DVC are fairly recent.

Popular areas of research in DVC has been intra-frmae coding and better quality Side Information generation.

In 2002, by using the turbo codes, Aaron, Zhang and Girod (7) have shown results on video coding using an intra-frame encoding and inter-frame decoding scheme where individual frames are independently encoded and jointly decoded.

In 2003, Zhu, Aaron and Girod proposed an approach to Wyner-Ziv based low-complexity coding under the name of "distributed compression for large camera arrays" (33). In this approach, multiple correlated views of a scene are independently encoded with a pixel domain Wyner-Ziv coder but are jointly decoded at a central node. Zhu et al. performed in (33) a comparison between pixel domain Wyner-Ziv coder and an independent encoding and decoding of each view employing JPEG-2000 wavelet image coding standard. The result demonstrate that at lower bitrates the solution presented by Zhu et al. achieves higher PSNR than JPEG-2000 with a lower encoder complexity.

In 2004, Aaron, Rane, Setton and Girod (5) proposed an architecture similar to the one in (7). The key difference being the use of Transform Coding (DCT transformation) at the encoder. The results obtained show the new coding solution leads to a better coding efficiency when compared to the solutions provided by (7).

In the same year, Aaron, Rane and Girod proposed another solution based on intraframe encoding-interframe decoding (4) and beside the resulting bitstream from the current frame-encoding process, supplementary information regarding current frame is also transmitted from encoder to help decoder in motion estimation task.

Also in 2004, Aaron, Rane, Setton and Girod (5) showed that different WZ coding performances were resulted in by using SI of different quality. It was observed that the PSNR gap can exceed 6 dB, between the simplest scheme as average interpolation and complicated ones like MC interpolation.

Later that year, Aaron, Rane and Girod proposed a model with hash-based motion compensation at the receiver (4). The encoder sent from the current frame a hash codeword to help decoder to precisely predict the motion, wherein only the previously reconstructed frame was used to generate SI. It resulted in a system with low-delay, since there was no complicated MCI operation at the decoder end.

In 2006, W. J. Chien, G. P. Abousleman and L. J. Karam explored the case of lossy SI (9).

In 2006, D. Kubasov and C. Guillemot proposed a Mesh-Based MCI for SI Extraction in DVC (14).

In 2008, X. Zhang and J. Zhangs presented SI generation using optimal filtering techniques (31).

Keeping in mind of the related work, we can see that most of the research is done on improving the intra-frame coding scheme and improving the quality of side information generated. There still exists the scope for improving the different modules towards the betterment of whole codec.

1.4 Motivation

It has been seen that due the scarcity of computational power and storage space in the hand-held devices, such as sensor network, surveillance cameras, there is a need for the less complex encoder. DVC can be a significant area for research as it focuses on reducing the complexity of the encoders at the cost of slightly high complex decoders. From the investigations of the literature, it can be seen that DVC is still in its early stages and is not sufficiently mature. It is essential to improve and to create tools for DVC scenario with better rate-distortion performances.

Also, we know that more often than not, the movement of the objects in a video sequence is non-linear. So, a simple linear interpolation can not be relied upon, since it would create blocking artifacts, such as blurred object, jerky motion of objects etc.

Therefore, the ANN techniques are explored for generating the improved SI frame.

1.5 Objective

Hence the motivation to investigate the techniques for better side information generation. The main objective of this thesis is to investigate the quality of SI generated, from the decoded input stream, using the RBF neural network technique.

1.6 Thesis Organization

In this thesis, investigations have been made to propose a scheme for improved SI generation. The thesis is being organized into five chapters. In this chapter, foundations of Distributed Video Coding and related theorems, related work done, research motivation along with objective etc. are discussed. Organization of rest of the thesis is as: In Chapter 2, detailed study of stanford based DVC architecture is done, alongside the advances in field of prominent module. In Chapter 3, importance of side information, its recent advances and the proposed techniques have been discussed. In Chapter 4, results and discussions have been done. In Chapter 5, the conclusions are drawn from thesis and scope for future work is given.

Chapter 2

Distributed Video Coding

The foundations of DVC dates back to 1970's when Slepian and Wolf (SW) (22) established the rates achievable by lossless coding of two sources that are correlated. Wyner and Ziv then later extended the SW theorem for lossy case. It was not until last decade that the practical implementations of DVC were introduced in (18) (10).
Unlike the conventional encoders, H26x/AVC (26), where the source statistics are exploited at the encoder side, DVC can shift this complexity towards the decoder. On the other hand, DVC decoder would be fairly more complex than the traditional decoder. Therefore, DVC is suitable for the applications where the computational power is scarce at the encoding end of the devices, such as wireless video surveillance, mobile phone camera and multimedia sensor network. DVC can be used to design codec independent scalable codes as in (17). In other words, enhancement layer is independent of base layer codec.
Target scenario is the lossy coding of main information with SI available at receiver (WZ coding).

2.1 Practical Wyner-Ziv codec – Stanford

2.1.1 Overall architecture

Fig. 2.1 shows the Transform Domain WZ coding (TDWZ), proposed by Stanford group (6). The operation is similar to that of the Pixel domain WZ coding (PDWZ). TDWZ introduces DCT transformations and bit-plane transmissions, which improves the compression performance despite PDWZ being less complex comparing to that of TDWZ.
System starts by separating the series of frames into WZ frames and Key frames.

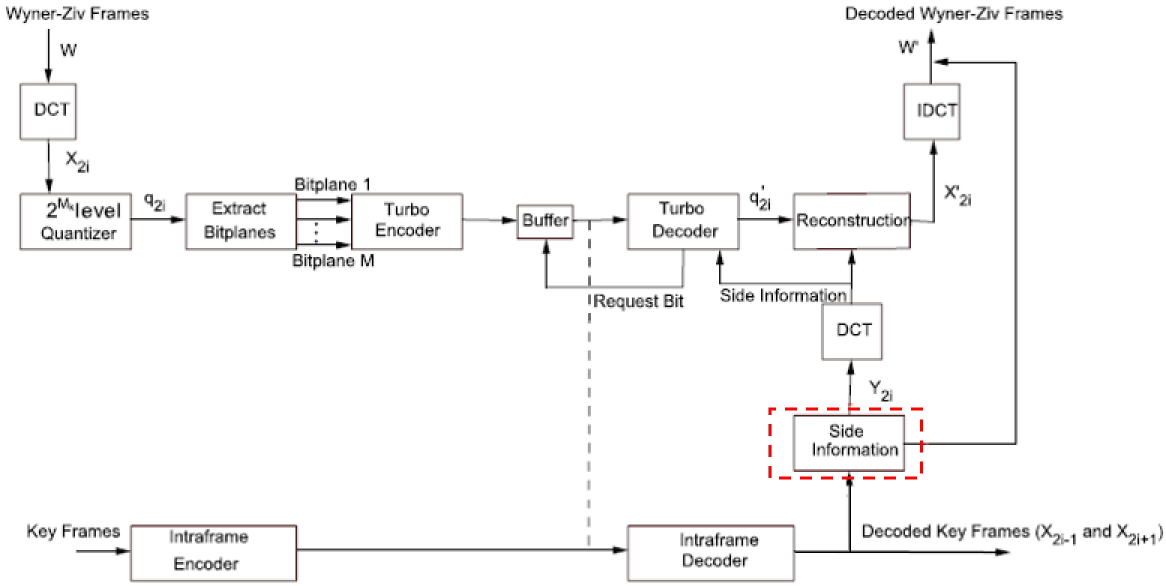

Figure 2.1: Stanford based transform domain Wyner-Ziv codec architecture

- The WZ frame,even numbered X_{2i} frames, are transformed by applying DCT block by block, block size being 4x4.
- The transformed coefficients having corresponding positions are then grouped together to form the DC/AC coefficient bands.
- The bit-plane is extracted and the resulting bit-planes are sent to Turbo encoder in sequence.
- Buffer saves the parity bits generated and transmission takes place after receiving request from decoder. In the meantime, the key frames are sent to decoder via traditional intraframe coding.
- SI, X_{2i}, is reconstructed from two key frames X_{2i-1} and X_{2i+1}. Then same as in earlier, blockwise DCT is applied on SI, Y_{2i}, and the coefficient bands are grouped in same manner as earlier.
- After decoding all the bit-planes, quantized symbol stream q' can be reconstructed, post which the reconstruction of coefficient band will take place.
- After the availability of all the coefficient bands, inverse DCT is applied to reconstruct the WZ frame.

In (6) terms of performance, Stanford's TDWZ architecture yeilds better performance to that of the Stanford's PDWZ architecture, due to the exploitation of spatial redundancy

DCT.

Few of major modules of STD-TDWZ codec are:

2.1.2 Transformation

In WZ video coding, transformations are a means to achieve high compression at the expense of shifting a part of the complexity to the encoder. Entropy coding is not advised in WZ video coding due to the limitations on energy absorbed. In TDWZ presented in (6), to encode the coefficients, a coefficient band grouping method is devised.

After applying DCT blockwise to the WZ frame of size 4x4, corresponding coefficients from every DCT blocks at same position are put together forming a single band of coefficients.

2.1.3 Quantization

In Information theory and coding, Quantization is a way to compress into a single value, the whole range of values. The signal is broken down into quantization bins, which can be termed as quantization symbol. Compression is achieved by representing the signal using the smaller bit stream since the number of bins is fewer than the total number of the values assumed by a signal.

Different Rate Distortion points are associated with, 2^M=$\{$2 4 8 16 32 64 128 256$\}$ levels of quantization. In TDWZ scenario, different quantization tables, as shown in Fir. 2.2, are utilized to quantise the DCT block. It defines quantization levels for the coefficient bands in the DCT block.

2.1.4 Slepian-Wolf Encoder

Systematic channel codec is one of possible ways to realize Slepian-Wolf codec, just same as TDWZ architecture using the turbo codes. Identical recursive systematic convolutional (RSC) codes with parallel concatenation, forms the basic building block of Turbo code encoder as shown in Fig.2.3.

Interleaver is used to separate two component encoders, as discussed above. One each of systematic and parity output stream is produced for each one of the component encoder. Systematic outputs used are only from the first component encoders, since other

16	8	0	0
8	0	0	0
0	0	0	0
0	0	0	0

(a)

32	8	0	0
8	0	0	0
0	0	0	0
0	0	0	0

(b)

32	8	4	0
8	4	0	0
4	0	0	0
0	0	0	0

(c)

32	16	8	4
16	8	4	0
8	4	0	0
4	0	0	0

(d)

32	16	8	4
16	8	4	4
8	4	4	0
4	4	0	0

(e)

64	16	8	8
16	8	8	4
8	8	4	4
8	4	4	0

(f)

64	32	16	8
32	16	8	4
16	8	4	4
8	4	4	0

(g)

128	64	32	16
64	32	16	8
32	16	8	4
16	8	4	0

(h)

Figure 2.2: Eight quantization matrices relating to different rate-distortion performances

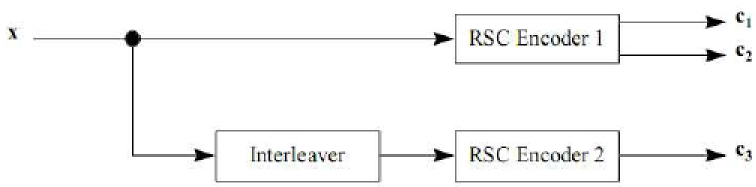

Figure 2.3: The encoder architecture of turbo encoder

component encoder is just an interleaved version of the output yielded by first encoder.

2.1.5 Side Information

In WZ coding, exploiting correlation happens only at the decoder between the SI and the WZ frame. Accuracy of the generated SI has critical impact on overall performance of compression of WZ coding, due to the fact that encoder is unaware of the SI frame at the time of encoding. Relatively very few parity bits will be required to be transmitted which would result in efficient compression, if the generated SI is similar, accurately reconstructed, to WZ frame. Or else, to correct the 'errors' between WZ frame and generated SI, more parity bits would be needed to sent to the encoder. This would have an expected effect on efficiency of compression.

Hence, it is of concern as to how to generate accurate SI. Only two frames, adjacent to the frame to be generated, are available in STD-WZ coding scheme. These two given key frames are used to generate SI frame. Of the several ways to generate the SI in TDWZ architecture.

- Generating SI by directly using one of the key frames in its place.
- Generating SI by using the two frames and taking out the average of the corresponding pixel location intensities to construct a new frame.
- Generating SI by the method of Motion Compensated Interpolation (MCI), where motion vectors are calculated between two key frames at time t-1 and t+1

SI generation can be summarized as the frame interpolation from two frames. It already has been an area of research, to generate SI with better frame rate up conversion. Wherein, the frame is reconstructed at the decoder after being skipped to be transmitted from the encoder.

2.1.6 Frame Reconstruction

Quantized symbol stream, q', is reconstructed at the decoder. Metric function used for evaluating the performance of reconstructed coefficient band, or the whole WZ frame in TDWZ video coding, is Mean Square Error (MSE).
Reconstructed pixels will take same value as SI value, if it lies within the reconstructed bins. In case, if SI values overflows the quantization bin, the values are forced to fall inside of bin by the reconstruction function and boundary value of bin is assigned.

2.2 Other advances in WZ coding

This section covers, a review of other advancements from the research community, done in the field of WZ video coding. In last few years, quite a good number of approaches have been made have been researched regarding DVC. Discussing each and every one of then would not be feasible, so a handful of the advancements are taken up for discussion that had significant progress. According to the emphasis on applications, references are categorized into various categories.

2.2.1 Advancements in Quantisation

Most of the WZ coding solutions popularly uses Uniform scalar quantization. A further improvement can be done on the account of engaging more sophisticated quantization techniques. Introduction of the sophisticated quantization technique would be a problem due to the higher complexity WZ encoder. The Lattice Vector Quantisation(LVQ)(24) was introduced on the basis of WZ coding solutions with LDPC. In comparison to scalar quantization, the introduced LVQ provides better coding performance with low complexity. Lloyd maximum quantisation, was modified and two of its variants (21; 27) were used as a replacement to uniform quantization. In the reconstruction process, both the approaches claims its advantages and superiority over the uniform quantization. The framework mentioned in (30) uses a nested scalar quantisation.

2.2.2 Advancements in Transformation

DWT inherits advantages by the fact that the block artifacts are reduced and over DCT are made scalable. Of the several approaches for DVC based on wavelet transformations, EZW (20) and SPIHT (19) are two main alogorithms used for the encoding of wavelet coefficients. Zerotree entropy coding(ZTE) is used in (13) as a method for encoding the wavelet coefficients for DVC. To distinguish between the significant and insignificant coefficients, the wavelet coefficients after quantization are rearranged in terms of Zero tree structure. Slepian-wolf coding takes place for the significant coefficients, while intra-coding takes place for the significance-map and then transmitted. SPIHT method based DVC, proposed in (15), WZ encoding is done for low-frequency coefficients and for that of high frequency coefficients, SPIT is used for coding. This has shown to perform better comparing to intra-coding algorithm that purely relies on SPIT.

2.2.3 Advancements in Slepian-Wolf codec

Low Density Parity Check (LDPC) codec is one of the substitute choices to the turbo codec for being used in current WZ coding. LDPC, a new systematic channel codec, has shown a similar or even superior performance over the turbo codes that has been proved in many references. Literature shown in (24; 30; 11; 23; 28; 12; 16) etc can be seen using the Slepian-wolf codec based on LDPC. Just as the turbo codes is in use currently, most of them are used in identical manner. From the quantized symbol, bit-plane extraction is done and is then sent to the LDPC encoder. Now only syndrome bits are generated and sent to the decoder, after being operated by syndrome coding. Decoding is then performed over the SI frame and syndrome bits. LDPC is used as basic slepian-wolf codec in most of the reviewed references, for the purpose of proposing other advanced techniques, it becomes a tough decision to be certain if LDPC offers better performanceover the turbo codes for WZ coding solution. (25) has shown in his work the comparison between bit-plane and symbol based approach against the LDPC codes in WZ coding. The conclusion reached that both performs similarly. Even so, computations are reduced significantly in the bit-plane based approach which is thus advantageous and preferred in the practical applications. Digital fountain code presented in (29), in addition to LDPC codes.

2.2.4 Advancements in Side Information generation

The factors that can significantly impact the coding performance of WZ coding also includes simple SI generation. It was noticed that the PSNR difference is noticeable between even the simplest of the technique, average interpolation, and that of the complicated MC interpolation. For generating high quality SI frame, many approaches are proposed in the research community. One of the prominent research area is refinement of the SI frame by using decoded WZ frame, then whole process of decoding is repeated to obtain better output with the refined SI frame. The SI frame can be repeatedly refined by the mentioned process upto the point till where output of a fixed quality is obtained. (4) employs a WZ architecture where hash codeword of the current frame is sent from the encoder to help decoder in estimating the motion of objects. In the process mentioned SI generation is based only on the previously reconstructed frame, which results in system of low-delay due to the absence of MC interpolation performed at the decoder. In (8), a solution is proposed by the authors where the use of multiple SI is generated from multiple reference frames. In (9) the case of lossy SI generation is explored by the authors. In (31) a method to generate SI is introduced where optimal filtering technique is presented, where Motion vectors are predicted ,using the optimal filter, between the SI and WZ frame, which will be corrected using a traditional motion search in WZ frame after decoding. These refined Motion vectors interpolates SI frame before passing it to the decoder to generate high quality WZ frame.

2.3 Chapter Summary

An overall followup regarding the current video coding based on WZ is presented in this chapter, outlining the progress starting from the basic theoretic background to various mature WZ coding schemes presented by different researching groups. Specifically, detailed working of each component is presented in the chapter and various experiments and results have been outlined undertaken by various researching groups. Pros and cons, possible applications with limitations of presented solutions are also taken up for discussion. Apart form the discussed schemes from the leading researching groups, several other schemes are also categorized and reviewed with the inclusion of Transformations, Quantizations, Channel coding and SI generation and others. Furthermore, potential directions for research based on the current WZ coding review and motivations responsible for the commencement of this thesis is discussed in subsequent chapters.

Chapter 3

Side Information generation

Most popular distributed video coding (DVC) solutions use the correlation between original frame with a frame predicted at the decoder. This predicted frame is known as side information (SI), which is a key function in the DVC decoder. The more accurate is the predicted SI, fewer number of bits need to be sent to decode the Wyner-Ziv (WZ) frame. So, SI generation is one of the most focused area of research that directly influence the DVC performance.

This chapter presents a SI generation scheme for distributed video coding based on Motion Compensated Frame Interpolation (MCFI). The suggested scheme predicts a WZ frame from two decoded key frames. MCFI processes the video frames block by block to calculate the motion vectors between two frames. The proposed scheme is simulated along with other standard video coding schemes. Performance comparisons have been made with respect to peak signal to noise ratio (PSNR). In general, it is observed that the proposed scheme has a superior SI frame generation capability as compared to its competent schemes.

As discussed in the previous chapter that SI generation is mostly performed from the decoded key frames information. In this chapter, we investigate a similar but efficient scheme based on blockwise motion estimation to generate a quality SI frame. The proposed scheme utilizes Block matching algorithms for the purpose.

3.1 MCFI based SI generation

Motion Compensated Frame Interpolation methods takes the assumption of a video being smooth in its flow and there can be only continuous and translational motion of the objects. Mentioned assumption might be true for a small number of videos where there is only a little to no motion at all of the objects in the video. Traditional approach works by considering the previous frame and calculating the motion vectors of macro-blocks with regard to the current frame, and then interpolating frame by calculating the average value of the pixels by taking down half of the motion vectors obtained. As a consequence, most of the efforts in research are aimed at improving motion vector predictions. As the residual information from the frame that was skipped is unavailable, accuracy of the motion vectors becomes more important since it directly affects the result via the generated frame. In proposed scheme, Block matching algorithms are utilized to obtain motion vectors between two odd numbered key frames.

To generate the intermediate even key frame, two odd numbered key frames are given as input. The frames are then divided into macro-blocks of size 4x4. Of the two corresponding blocks from two input frames, earlier is used as the source block whose vector is to be calculated, and the latter frame is then searched for best available match of the input block. To have a generalized approach, 4x4 blocks from key frames and WZ frames are collected from video sequence with various motion patterns. The matching algorithm, for the blocks, used is Diamond Search (DS). DS returns the motion vectors as the difference in coordinates of previous block and newly calculated position of the matched block, which is discussed ahead.

3.1.1 Diamond Search

DS (32) algorithm has the search point pattern of a diamond rather than a square, with no limitations regarding the number of steps undertaken by the algorithm.

DS utilizes two separate patterns for marching of he blocks. First one of them is known as Large Diamond Search Pattern (LDSP) and the other one is called as Small Diamond Search Pattern (SDSP). Mentioned patterns of search alongwith the DS mechanism is illustrated in Fig 3.1. First step makes use of LDSP and looks for the minimum weight depending on the cost function used and if minimum weight is found to be at the center, algorithm jumps directly to the Small Diamond Search pattern. The subsequent steps, but the last, are repeatedly checked by LDSP and the working procedure is illustrated in Fig 3.1. Last of the steps uses SDSP with a new search origin as obtained from the lowest cost function pointed by LDSP. Step size is reduced to half and again the cost is calculated for minimum value.

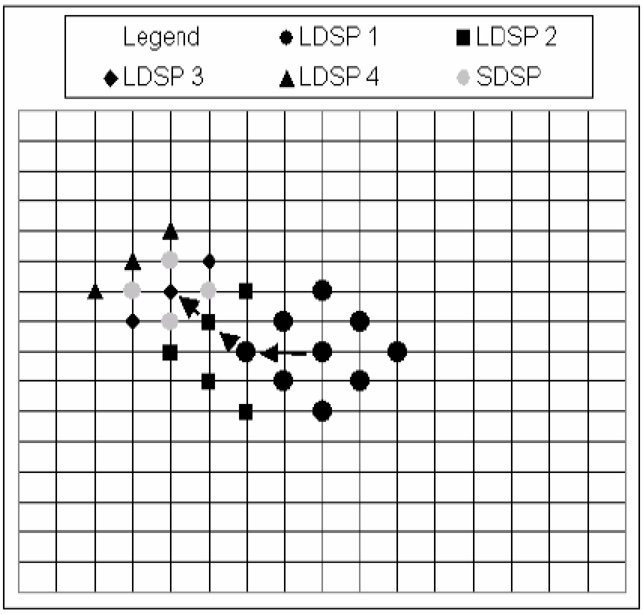

Figure 3.1: Diamond Search Pattern

The fact that there is no limitations on the steps taken by the algorithm and also that the search pattern size is justifiable, that is its not too large neither too short. Therefore, DS is able to find the global minimum value for the cost very soon and very accurately. The PSNR level of the end result shouldbe close to that of Exhaustive Search, Brute force method, while significantly decreasing the computational expense.

Large Diamond Search Problem

Algorithm: LDSP

Step 1: Start with search location at center

Step 2: Set step size 'S' = 2

Step 3: Search 8 locations pixels (X,Y) such that ($|X| + |Y|$=S) around location (0,0) using a diamond search point pattern

Step 4: Pick among the 9 locations searched, the one with minimum cost function

Step 5: If the minimum weight is found at center for search window, go to SDSP step

Step 6: If the minimum weight is found at one of the 8 locations other than the center, set the new origin to this location

Step 7: Repeat LDSP

Small Diamond Search Problem

Algorithm: SDSP

Step 1: Set the new search origin

Step 2: Set the new step size as S = S/2 = 1

Step 3: Repeat the search procedure to find location with least weight

Step 4: Select location with the least weight as motion vector

Cost function used for matching a macro-block with another block is is Mean Absolute Difference (MAD), which can be calculated as

$$MAD = \frac{1}{N^2} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} |C_{ij} - R_{ij}|$$

where C_{ij} and R_{ij} are the current and Referenced frame respectively.

3.1.2 Intermediate Frame generated by MCFI

Fig 3.2 shows motion vectors between frame number 49^{th} and 51^{st} using DS algorithm with parameters as, mbSize = 4 and p = 4, are

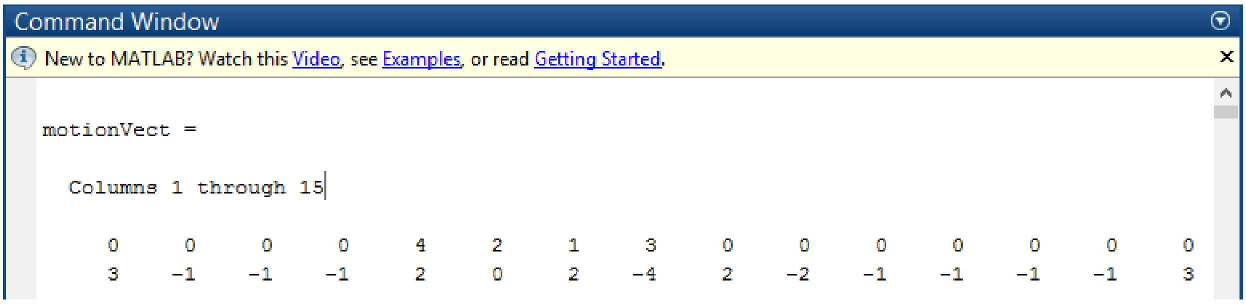

Figure 3.2: Motion vectors between frame number 49^{th} and 51^{st}

Figure 3.3: Original and MCFI generated 50^{th} frame

3.2 RBF based SI generation

A radial basis function (RBF) is a function based on a scalar radius.

$$\phi(x) = \phi(|x - x_i|)$$

The scalar distance is usually calculated using the Euclidean norm, although other distance metrics are also possible. Radial basis functions are means to approximate multivariate functions by linear combinations of terms based on a single univariate function (the radial basis function). Radial basis functions (RBF) can be used for interpolation and approximation of scattered data in any dimension.
Radial Basis Function Networks (RBFN) consists of 3 layers

- An Input layer
- A Hidden layer
- An Output layer

The hidden units provide a set of functions that constitute an arbitrary basis for the input patterns and are known as radial centers. Each hidden unit has its own receptive field in input space. An input vector x_i which lies in the receptive field for center c_j , would activate c_j and by proper choice of weights the target output is obtained.
The output is given as

$$y = \sum_{j=1}^{h} \phi_j w_j$$

$$\phi_j = \phi(||x - c_j||)$$

where w_j : weight of j^{th} center, ϕ: some radial function
Although there a number of options that can be used, but for our purpose, we have used the Gaussian Radial function, which can be written as

$$\phi(z) = e^{-z^2/2\sigma^2}$$

where $z = ||x - c_j||$ (euclidean distance)

3.2.1 Need of RBF

In the previous section, we have seen that MCFI does not completely remove the unwanted artefacts, which do not ensure the smooth playback of the resultant output video.

$(i-1)^{th}$ frame pixel coordinate	New pixel coordinate in $(i)^{th}$ frame	New pixel coordinate in $(i+1)^{th}$ frame	Euclidean distance of pixel coordinate in $(i-1)^{th}$ frame and $(i)^{th}$ frame	Euclidean distance of pixel coordinate in $(i)^{th}$ frame and $(i+1)^{th}$ frame
(1,2)	(3,1)	(4,2)	2.23	1.414
(1,3)	(1,5)	(2,4)	2	1.414
(1,4)	(3,2)	(2,6)	2.82	4.123
(1,5)	(1,7)	(2,3)	2	7.071
(1,6)	(1,8)	(7,6)	2	6.324
(1,8)	(2,8)	(3,4)	1	4.123
(2,3)	(5,3)	(3,2)	3	2.23
(2,4)	(7,5)	(6,3)	5.099	2.23
(2,6)	(6,7)	(1,7)	4.123	5
(2,8)	(7,6)	(1,4)	5.385	6.324
(3,2)	(5,2)	(4,3)	2	1.414
(3,3)	(3,6)	(3,7)	3	1
(5,1)	(8,1)	(6,8)	3	7.280
(5,3)	(4,4)	(4,5)	1.414	1
(5,4)	(6,3)	(6,4)	1.414	1
(5,6)	(7,3)	(8,3)	3.605	1
(6,1)	(8,2)	(8,5)	2.23	1.73
(6,2)	(7,2)	(1,2)	1	2.449
(6,8)	(8,5)	(1,6)	3.605	7.071
(7,3)	(6,4)	(4,8)	1.414	4.472
(7,7)	(5,8)	(1,8)	2.23	2

Figure 3.4: Non-linear pixel movement in three consecutive frames of Foreman video sequence

From the Fig 3.4, we can say that inter-frame pixel movement is non-linear due to 3-D motions of an object moving back and forth in horizontal, vertical, and diagonal direction or may be due to some directional orientation, camera zoom, and panning. The linear motion and non-linear motion of a pixel in a video between frames is shown in Fig 3.5.

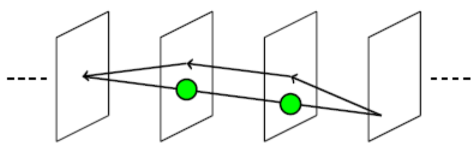

Figure 3.5: Linear vs. non-linear motion between adjacent frames

The Euclidean distance measures reflect the non-linear motion property of a pixel. Similar observations are found in pixels with other frames of Foreman, Coastguard, and other video sequences.
Therefore, Artificial neural network (ANN), RBFN in current proposal, being a potential tool for non-linear prediction is utilized.

Figure 3.6: Architecture of the Neural Network Predictor

3.2.2 Intermediate Frame generated by RBF Interpolation

Figure 3.7: Original and RBF generated 50^{th} frame

Chapter 4

Result and Discussion

In previous chapter, we saw that there are some anomalies and artefacts occurring in the SI generated through the MCFI methodology. Since linear interpolation is not enough to account all the motion inside a video, which is due to the fact that there exist non-linear motion of objects inside almost all videos.

So, by means of ANN, RBFN in this case, is used to account for the non-linear motion of objects. By the use of RBFN, the artefacts were removed and a smooth playback of the video is ensured.

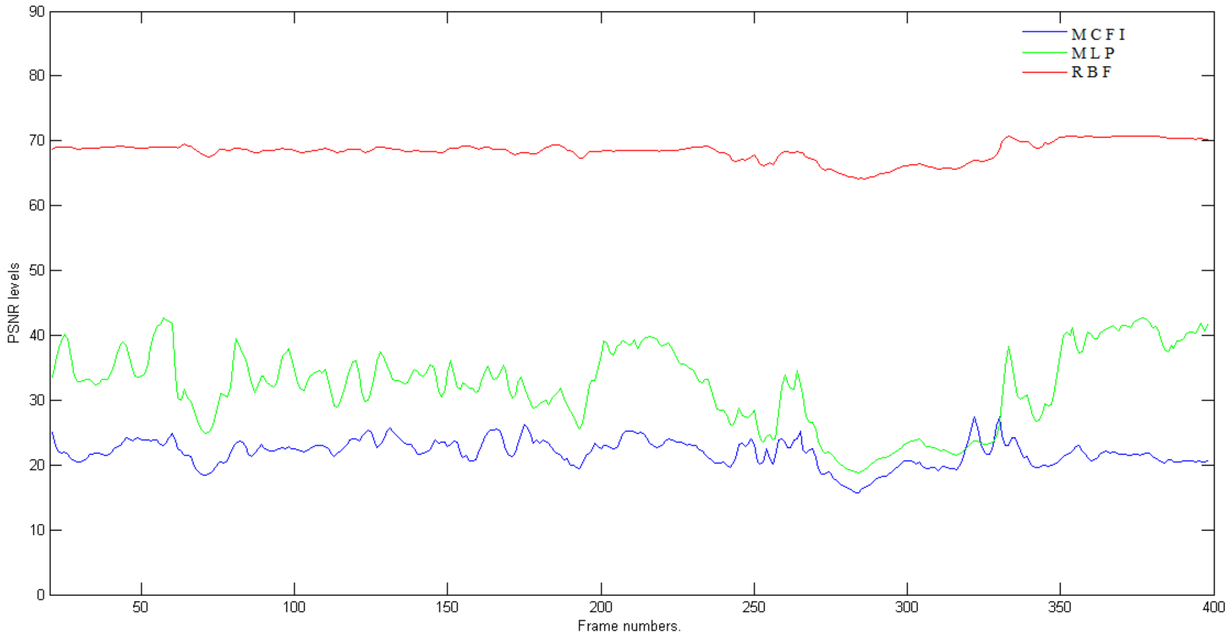

Figure 4.1: PSNR comparison among MCFI, RBF-based and MLP-based SI frames

In the Fig 4.1, PSNR levels of the applied techniques, i.e. MCFI and RBF based SI generation are compared with the existing MLP-based SI generation scheme.

PSNR values for the RBF-based SI generation were found to be much higher than counterpart schemes. Visually, the frames is found to be slightly darkened but the output video has smooth playback.

In MCFI, the PSNR level is slightly lower and few of the frames have the blocking artefacts. Visually, output video is not so smooth when compared to the one generated by RBFN.

Chapter 5

Counclusion and Future Work

In this thesis, two schemes have been investigated for side information (SI) generation in a distributed video coding (DVC) framework. In DVC, intra-frame coding and side information are dependent on each other as SI uses decoded key frames. Hence superior quality key frames generated through intra-key frame coding in turn help in generating high quality SI frames. As a result, DVC needs less number of parity bits to reconstruct the WZ frames at the decoder. MLP-based SI generation scheme is compared with the suggested schemes, which are

- Motion Compensated Frame Interpolation (MCFI)
- Interpolation via Radial Basis Function (RBF)

MLP-SI scheme utilizes a multilayer perceptron to estimate SI frames from the decoded key frames block-by-block.

MCFI scheme utilizes Block matching algorithm to calculate motion vectors and hence interpolate the intermediate SI frame.

RBF-SI scheme utilizes a Radial based function to estimate the weights of the input vectors and generates the intermediate SI frame with the help of calculated weights.

Research directions are open to explore several other ANN techniques and optimizations for improved performance. Apart from SI module, DVC is a vast architecture where the initial modules can be very well explored for improvements, such as Intra-frame coder, Quantizer etc. Different modules are still not fast enough to be implemented in real-time, which can still be improved further. This shows that there is a lot of scope for improvement of overall architecture.

Bibliography

[1] Distributed source coding. http://en.wikipedia.org/wiki/Distributed_source_coding.

[2] Moving picture experts group. http://mpeg.chiariglione.org/.

[3] Visual sensor network. http://en.wikipedia.org/wiki/Visual_sensor_network.

[4] Anne Aaron, Shantanu Rane, and Bernd Girod. Wyner-ziv video coding with hash-based motion compensation at the receiver. In *Image Processing, 2004. ICIP'04. 2004 International Conference on*, volume 5, pages 3097–3100. IEEE, 2004.

[5] Anne Aaron, Shantanu D Rane, Eric Setton, and Bernd Girod. Transform-domain wyner-ziv codec for video. In *Electronic Imaging 2004*, pages 520–528. International Society for Optics and Photonics, 2004.

[6] Anne Aaron, Eric Setton, and Bernd Girod. Towards practical wyner-ziv coding of video. In *Image Processing, 2003. ICIP 2003. Proceedings. 2003 International Conference on*, volume 3, pages III–869. IEEE, 2003.

[7] Anne Aaron, Rui Zhang, and Bernd Girod. Wyner-ziv coding of motion video. In *Signals, Systems and Computers, 2002. Conference Record of the Thirty-Sixth Asilomar Conference on*, volume 1, pages 240–244. IEEE, 2002.

[8] ABB Adikari, WAC Fernando, H Kodikara Arachchi, and WARJ Weerakkody. Multiple side information streams for distributed video coding. *Electronics Letters*, 42(25):1447–1449, 2006.

[9] Wei-Jung Chien, Lina J Karam, and Glen P Abousleman. Distributed video coding with lossy side information. In *Acoustics, Speech and Signal Processing, 2006.*

ICASSP 2006 Proceedings. 2006 IEEE International Conference on, volume 2, pages II–II. IEEE, 2006.

[10] Bernd Girod, Anne Margot Aaron, Shantanu Rane, and David Rebollo-Monedero. Distributed video coding. *Proceedings of the IEEE*, 93(1):71–83, 2005.

[11] Marco Grangetto, Enrico Magli, and Gabriella Olmo. Context-based distributed wavelet video coding. In *Multimedia Signal Processing, 2005 IEEE 7th Workshop on*, pages 1–4. IEEE, 2005.

[12] Mei Guo, Yan Lu, Feng Wu, Shipeng Li, and Wen Gao. Scalable wyner-ziv video coding with adaptive bit-plane representation. In *Electronic Imaging 2008*, pages 68221X–68221X. International Society for Optics and Photonics, 2008.

[13] Xun Guo, Yan Lu, Feng Wu, and Wen Gao. Distributed video coding using wavelet. In *Circuits and Systems, 2006. ISCAS 2006. Proceedings. 2006 IEEE International Symposium on*, pages 4–pp. IEEE, 2006.

[14] Denis Kubasov and Christine Guillemot. Mesh-based motion-compensated interpolation for side information extraction in distributed video coding. In *Image Processing, 2006 IEEE International Conference on*, pages 261–264. IEEE, 2006.

[15] Shenyuan Li, Sheng Fang, and Zhe Li. Wyner-ziv video coding for low bitrate using spiht algorithm. In *Signal Processing Systems, 2007 IEEE Workshop on*, pages 341–345. IEEE, 2007.

[16] Limin Liu and Edward J Delp. Wyner-ziv video coding using ldpc codes. In *Signal Processing Symposium, 2006. NORSIG 2006. Proceedings of the 7th Nordic*, pages 258–261. IEEE, 2006.

[17] Mourad Ouaret, Frederic Dufaux, and Touradj Ebrahimi. Codec-independent scalable distributed video coding. In *Image Processing, 2007. ICIP 2007. IEEE International Conference on*, volume 3, pages III–9. IEEE, 2007.

[18] Rohit Puri and Kannan Ramchandran. Prism: A new robust video coding architecture based on distributed compression principles. In *Proceedings of the annual allerton conference on communication control and computing*, volume 40, pages 586–595. Citeseer, 2002.

[19] Amir Said and William A Pearlman. A new, fast, and efficient image codec based on set partitioning in hierarchical trees. *Circuits and Systems for Video Technology, IEEE Transactions on*, 6(3):243–250, 1996.

[20] Jerome M Shapiro. Embedded image coding using zerotrees of wavelet coefficients. *Signal Processing, IEEE Transactions on*, 41(12):3445–3462, 1993.

[21] Fang Sheng, Li Xu-Jian, and Zhang Li-Wei. A lloyd-max-based non-uniform quantization scheme for distributed video coding. In *Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing, 2007. SNPD 2007. Eighth ACIS International Conference on*, volume 1, pages 848–853. IEEE, 2007.

[22] David Slepian and Jack K Wolf. Noiseless coding of correlated information sources. *Information theory, IEEE Transactions on*, 19(4):471–480, 1973.

[23] Yoshihide Tonomura, Takayuki Nakachi, and Tetsuro Fujii. Distributed video coding using jpeg 2000 coding scheme. *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, 90(3):581–589, 2007.

[24] Anhong Wang, Yao Zhao, and Hao Wang. Lvq based distributed video coding with ldpc in pixel domain. In *PRICAI 2006: Trends in Artificial Intelligence*, pages 1248–1252. Springer, 2006.

[25] Ronald P Westerlaken, Stefan Borchert, Rene Klein Gunnewiek, and Reginald L Lagendijk. Analyzing symbol and bit plane-based ldpc in distributed video coding. In *Image Processing, 2007. ICIP 2007. IEEE International Conference on*, volume 2, pages II–17. IEEE, 2007.

[26] Thomas Wiegand, Gary J Sullivan, Gisle Bjontegaard, and Ajay Luthra. Overview of the h. 264/avc video coding standard. *Circuits and Systems for Video Technology, IEEE Transactions on*, 13(7):560–576, 2003.

[27] Bo Wu, Xun Guo, Debin Zhao, Wen Gao, and Feng Wu. An optimal non-uniform scalar quantizer for distributed video coding. In *Multimedia and Expo, 2006 IEEE International Conference on*, pages 165–168. IEEE, 2006.

[28] Geming Wu, Lifeng Sun, and Feng Huang. Consistent-quality distributed video coding framework. In *Advances in Multimedia Information Processing–PCM 2007*, pages 628–637. Springer, 2007.

[29] Qian Xu, Vladimir Stankovic, and Zixiang Xiong. Wyner–ziv video compression and fountain codes for receiver-driven layered multicast. *Circuits and Systems for Video Technology, IEEE Transactions on*, 17(7):901–906, 2007.

[30] Qian Xu and Zixiang Xiong. Layered wyner–ziv video coding. *Image Processing, IEEE Transactions on*, 15(12):3791–3803, 2006.

[31] Xiao Zhang and Jun Zhang. Side information generation for distributed video coding based on optimal filtering. In *Electronic Imaging 2008*, pages 68222D–68222D. International Society for Optics and Photonics, 2008.

[32] Shan Zhu and Kai-Kuang Ma. A new diamond search algorithm for fast block-matching motion estimation. *Image Processing, IEEE Transactions on*, 9(2):287–290, 2000.

[33] Xiaoqing Zhu, Anne Aaron, and Bernd Girod. Distributed compression for large camera arrays. In *Statistical Signal Processing, 2003 IEEE Workshop on*, pages 30–33. IEEE, 2003.