OPTIMIZATION OF ASYMMETRIC STEREO VIDEO CODING AND STREAMING

by

Görkem Saygılı

A Thesis Submitted to the

Graduate School of Engineering

In Partial Fulfillment of the Requirements for

the Degree of

Master of Science

in

Electrical and Computer Engineering

Koc University

February 2010

Koc University

Graduate School of Sciences and Engineering

This is to certify that I have examined this copy of a master's thesis by

Görkem Saygılı

and have found that it is complete and satisfactory in all respects,

and that any and all revisions required by the final

examining committee have been made.

Committee Members:

_____

Prof. A. Murat Tekalp (Advisor)

_____

Assist. Prof. Serdar Kozat

_____

Assoc. Prof. Çağatay Başdoğan

Date: _____

to my family

OUTLINE

ACKNOWLEDGEMENTS

First, I would like to thank my advisor Prof. A. Murat Tekalp for his understanding, guidance. I would also thank to Prof. Serdar Kozat for always helping me with great interest and guiding me about my future plans. I will never forget their teachings on technical and non-technical issues.

It will be impossible for me without the help and support of my family to write this thesis. I would especially thank to my future wife for her understandings about my shortage of time and for my extra ordinary stress.

Finally my office friends, C. Göktuğ Gürler, Burak Görkemli and Emrah Cem have always helped me, they never excused about not helping me on any subject. I appreciate their support in my research and also their support as being valuable friends of mine. I will never forget how I laugh to the jokes of Mehmet Ali Yatbaz (a.k.a MALİ).

LIST OF TABLES

LIST OF FIGURES

NOMENCLATURE

3D	Three Dimensional
3DTV	Three Dimensional Television
SVC	Scalable Video Coding
QP	Quantization Parameter
MD	Multiple Description
MDC	Multiple Description Coding
JVT	Joint Video Team
JSVM	Joint Scalable Video Model
DCCP	Datagram Congestion Control Protocol
TFRC	TCP-Friendly Rate Control
MGS	Medium Grain Scalability
FEC	Forward Error Correction
NAL	Network Abstraction Layer
NALU	Network Abstraction Layer Unit

ABSTRACT

It is well known that the human visual system can perceive high frequencies in 3D, even if that information is present in only one of the views. Then, the best 3D stereo quality may be achieved by asymmetric rate allocation to the reference (left) and auxiliary (ri1ght) views. However, the question of what should be the level of asymmetry between reference and auxiliary views without disturbing the perceived quality and whether the rate reduction for the auxiliary view should be achieved by spatial resolution reduction (coding a downsampled version of the video followed by upsampling after decoding) or quality (QP) reduction is an open issue. Subjective tests indicate that when the reference view is fixed at sufficiently high quality (i.e. 38 dB), the auxiliary view can be encoded around 31 dB without losing much on the perceived stereo video quality depending on the type of utilized 3D display technology. The threshold is ~31 dB for a parallax barrier display and ~33 dB for a polarized projection display. Beyond this threshold, symmetric coding starts to perform better than asymmetric coding in terms of perceived stereoscopic video quality. In addition to those, the result of subjective tests shows that users prefer SNR scalability over spatial scalability on both parallax barrier and polarized projection displays above the threshold..

Although it is more desirable to have more than 1 MGS NAL unit(NALU) since number of extraction points increase as the number of MGS NAL units increases, as number of NALUs increases, the coding efficiency decreases. We have tested the R-D performance of streams including only 1 NALU with streams including 2 NALUs. We used our own Bitstream Extractor program to extract the streams at various bitrates and compare their R-D performances.

Additionally we present the performance of different multi description generation algorithms. The streams of each descriptions are generated by MD generator program that is coded by us. Moreover we coded a channel simulator program to investigate the R-D performance of these descriptions under different network conditions (channel loss probabilities).

ÖZET

İnsan görme sisteminin bir görüntüdeki yüksek frekans eksikliğini diğer görüntüde bu bilgi mevcut olduğu sürece fark etmemesi bugün bilindik bir gerçektir. Öyleyse en iyi 3B görüntü sıkıştırma tekniğinin referans (sol) ve yardımcı görüntünün (sağ) asimetrik kodlanmasıyla elde edilmesi beklenir. Fakat bu iki görüntü arasındaki asimetri derecesinin ne kadar olması gerektiği ve bu asimetriyi sağlarken kullanılması gereken sıkıştırma yönteminin kalite (QP) yada boyut ölçeklenmesi mi olması gerektiği henüz yanıtlanmamış sorulardır. Görsel testlerimizin sonuçları gösteriyor ki, referans görüntü yüksek kalitede gösterildiği sürece (~38 dB) yardımcı görüntü 31 dB kalite seviyesinde kodlanabilir ve böylelikle kullanılan 3B görüntüleme teknolojisine de bağlı olarak görüntü algılamada bozukluk yaşanmaz. Bu kalite derecesi paralaks bariyerli görüntüleyicide yaklaşık 33 dB, polarize projektör görüntüleyicide ise yaklaşık 31 dB'dir. Bu derecenin ötesine gidildiğinde, simetrik kodlamanın, simetrik kodlamanın asimetrik kodlamaya kıyasla daha iyi görsel kalite sunduğu görülmüştür. Dahası izleyicilerin bu derece yukarısında kalite(QP) sıkıştırma yöntemini boyut sıkıştırma yöntemine tercih ettikleri gözlemlenmiştir.

Birden çok MGS NAL ünitesi ölçeklenebilir betimlemelerin özütleme noktalarının eriminin arttırılmasını sağlar ancak bu sayının artması kodlama performansının azalmasına sebep olur. Yaptığımız hız-bozunum testleriyle 1 ve 2 adet MGS NAL ünitesine sahip görüntü katarlarını karşılaştırdık. Bunu yaparken de görüntü katarlarının boyutlarının ayarlanmasında kendi yazdığımız katar özütleyici programını kullandık.

Bunlara ek olarak, değişik betimleme oluşturma yöntemlerinin performanslarını inceledik. Betimlemelerin oluşturulmasında kendimizin yazdığı betimleme oluşturma programını kullandık. Dahası yine kendimize ait kanal simülatör programı ile oluşturduğumuz betimlemeleri değişken ağ koşullarında (değişken paket kayıp ihtimallerinde) test edip hız-bozunum performanslarını inceledik.

INTRODUCTION

1.1 Background and Motivation

3DTV technology has been developing rapidly with the contributions of research projects such as EC founded 3DTV project, SARACEN and DIOMEDES. 3D video comes with a price of double bandwidth requirement in comparison to its 2D counterpart which means rate allocation of views has a crucial importance for stereoscopic video coding. Scalable Video Coding (SVC) enables the adaptation of video stream to the network environment nonetheless the optimum adaptation strategy to be followed is still an open issue. Asymmetric rate scaling of views that relies on Human Visual System (HVS) is a serious alternative to symmetric rate scaling. However there is a limitation of HVS which means degrees of asymmetry should be carefully determined.

This thesis presents threshold values for the levels of asymmetry such that above this threshold asymmetric coding performs better than symmetric coding in terms of perceived 3D video quality.

1.2 Optimization of Stereoscopic Video Transmission

In order to stream 3D video content through a network, the source video needs to be adapted to the available network throughput, such as to the DCCP rate or TCP send rate. The methods that are aiming to optimize rate distortion performance in case of packet loss scenario such as retransmission and buffering of packets [1] and utilizing forward error correction (FEC) [2] are out of scope of this work. Our interest is finding the optimum multi-view video compression vs. perceived video quality by comparing the performances of different scalability options of SVC at different bitrates. Researchers addressing the best multi-view video compression at a given total bitrate (e.g., 2 Mbps) mainly focus on three issues: The first is better inter-view prediction methods, e.g., illumination compensation, for improving compression efficiency [3-4]. The second one is encoding and sending the reference view and a low resolution depth-map, hence rendering the auxiliary view at the receiver. There are many approaches addressing scalability performance increase such as

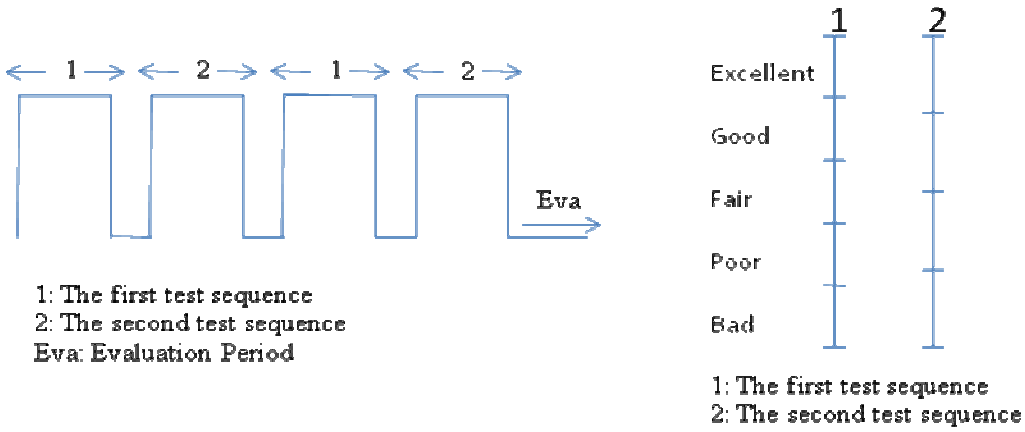
region of interest(ROI) based coding [5], changing the depth quantization parameter index within a threshold level [6] and using algorithms that rely on exploiting statistical dependencies of inter-view reference pictures in order to predict the depth information of the views [7]. Third, asymmetric (unequal) rate allocation between the reference and auxiliary views in stereo video coding by exploiting human visual system (HVS) which compensates for the lack of high frequency components in one view if the other view is at sufficiently high quality [8-11]. It is a well known fact that better rate-distortion performances can be achieved by using spatial, temporal or SNR scaling on one of the views while holding the other view at sufficiently high quality [12].

In Section 2, we present the best asymmetric coding strategy. In other words we demonstrate which scalability option and coding method (asymmetric or symmetric) performs the best in terms of perceived 3D quality. In Section 3 we discuss different models for MD generation and observe their behavior under varying channel packet loss probability rates. Section 4 presents the results of subjective tests and comments on them. Finally we draw our conclusions in Section 5 and talk about the possible future work.

1.3 Subjective Evaluation Procedures and Qualitative Metrics

Peak Signal to Noise Ratio (PSNR) is frequently used and widely accepted as a quantitative metric for assessing the quality of impaired monoscopic video. However, in the stereoscopic case, averaging over the individual PSNR values of the views can yield misleading results. The aim of this work is to find a metric or threshold values to explain the perceived quality in terms of each view's PSNR.

The perceived quality is measured by subjective tests. We have used Double-Stimulus Continuous Quality-Scale (DSCQS) method for the subjective tests and evaluation of the results. According to the standard, the viewers are shown one test sequence and original video consecutively each time and repeat once more in the same order as depicted in fig.1.1.a. At the end of each repetition, the viewers are expected to grade the perceived quality of the test sequences on a continues scale shown in fig.1.1.b[13].



(a) (b)

Fig. 1.1.(a) Test sequence presentation procedure (b) Grading scale

1.4 Contributions of this Thesis

The HVS has altering behavior on different types of displays. The type of scalability to be chosen for the best perceived quality with lowest bitrate depends on the utilized 3D display technology. This result is presented at IEEE International Conference on Image Processing (ICIP) 2009.

In addition to that, the type of scalability to be used for asymmetric coding depends on the maximum quality of the views that can be achieved under the bitrate constraint. Although HVS works well above the threshold levels that we have found, when the degree of asymmetry is increased beyond the threshold, symmetric coding starts to perform better than asymmetric coding. This result is prepared and ready to submit to a journal paper and also ready to submit conference paper to ICIP 2010.

Chapter 2

ASYMMETRIC RATE SCALING FOR ADAPTIVE VIDEO STREAMING

2.1 Stereo Video Coding with Asymmetric Rate Allocation Among Views

Asymmetric stereo video coding relies on properties of HVS. According to the HVS theory, as long as one of the views is at high quality, human eye can perceive the overall 3D video quality close to the high quality view even the other view is shown at lower quality levels.

Scalability can be done by using spatial, temporal and SNR layer scalability options of SVC [14]. We used SNR and spatial scalability, since frame rate reduction has been observed to be more noticeable by users [11].

Spatial scalability can be implemented by using one of the two following ways;

I- Half resolution video can be encoded at the desired QP first and then interpolated to have full spatial resolution

II- Half resolution video can be interpolated to full spatial resolution (low pass filtering) and then encoded at the desired QP.

Table 2.1 shows RD performance of both methods. The PSNR difference at the same bitrate is around 0.3 dB for both videos hence we conclude that both methods can be used for performing spatial scalability

.

I. Method	**1(Mb/s)**	**0.8(Mb/s)**	**0.6(Mb/s)**
Adile	31.754	31.674	31.498
Flower	30.362	30.311	29.946
Train	30.311	30.051	29.302
II. Method	**1(Mb/s)**	**0.8(Mb/s)**	**0.6(Mb/s)**
Adile	31.516	31.375	31.067
Flower	30.194	29.997	29.627
Train	29.989	29.666	29.05

Table 2.1: RD Performance of Spatial Scalability Methods

SNR scalability can be achieved by discarding mgs units from SVC bit stream. This operation causes degradation on the resulting PSNR and bitrate. Accordingly it corresponds to encoding the video at higher QP with single layer SVC.

Spatial scaling causes blurring artifact on the image and SNR scaling brings blockiness on to the image as a side effect.

2.2 Determination of Perceptually Unnoticeable Level of Asymmetry

In asymmetric coding/rate scaling, potential visual artifacts in the auxiliary view are concealed by the human visual system if the PSNR of the auxiliary view is higher than a threshold value, provided that the main view is coded at a high enough PSNR. Furthermore, this threshold PSNR is display-dependent; namely, it is ~31 dB for a parallax barrier display and ~33 dB for a polarized projection display. We believe this difference is induced by the technology of parallax barrier. These barriers perform filtering operation by blocking light from certain subpixels to certain directions. As a result perceived spatial resolution is halved as a side effect. This effect causes sharp display to conceal visual defects caused by utilized

scalability option. Accordingly test sequences on parallax barrier display take better grades than on projector display from subjective test results that are discussed in the results section.

In order to test this hypothesis, we have performed subjective tests by viewing four different types of stereo video using both a polarized projection display and a parallax barrier autostereoscopic display, where subjects view stereo videos using a player interface, which allows them to change the quality of the auxiliary view up or down as the video continues to play, while the quality of the main view remains fixed. The test starts displaying both left and right views at 40 dB, and the subjects are allowed to decrease the PSNR value of the auxiliary view down to 25 dB by incrementing encoder QP value by steps of 1. Besides varying the quality one step at a time, the player interface allows subjects to switch between the current and highest quality level for better comparison. Table 2.2 presents the PSNR values at which the subjects start observing quality degradation in stereo viewing when compared to symmetric coding at 40 dB. PSNR values reported in Table 2.2 are the average over all subjects taking the test.

Content			Projector	Parallax Barrier
Name	Resolution	Type	PSNR	PSNR
Adile	640x480	Computer Generated	33.07	31.90
Train	704x576	Captured, fixed background	32.88	31.74
Flower	704x448	Captured, camera motion	33.20	31.19
Iceberg	640x384	2D-3D Conversion	33.05	31.64

Table 2.2: Visibility threshold values for asymmetric coding

2.3 Performance Analysis of Asymmetric and Symmetric Coding

The asymmetry for the auxiliary view can be achieved either by increasing the quantization parameter (QP) to reduce the SNR quality or by decreasing the spatial resolution of the input sequence. We have performed subjective tests in which the bitrate for the auxiliary view is fixed to four different values and both methods are evaluated by the users to select the best method. The bitrate are PSNR values for test sequences are presented in Table 2.3. In almost all cases above the threshold levels that we have found in the previous section, the viewers favored SNR scaling therefore, we have chosen SNR scaling method to achieve asymmetry for encoding the auxiliary view above the threshold level.

Additionally, SNR scalability constitutes our main concern for scalability since when an SNR enhancement NAL unit is discarded the resultant bitstream is still considered as a complaint bitstream. However, when a spatial enhancement NAL unit is discarded the stream becomes non-conforming unless it is signaled explicitly by conforming switching points. Therefore SNR scalability can be used for adaptive streaming purposes in IP network that has varying bitrate over time and spatial scalability is more suitable for terminal adaptation way without signaling decoder. Therefore, when compared to spatial scalability, SNR scalability is more suitable for rate adaptation purposes.

We have observed that asymmetric coding with SNR scaling performs better than asymmetric coding with spatial scaling around the threshold levels. However the effect of blockiness starts to become more disturbing than the effect of blur caused by spatial scaling. Hence, we expect to notice that asymmetric coding with spatial scalability performs better than asymmetric coding with SNR scaling and symmetric coding with SNR scaling at very low rates. This statement also answers why Stelmach [15] had favored spatial scaling over SNR and points a second threshold level.

In order to prove this argument, we performed subjective tests and compare the performances of asymmetric coding with spatial scaling and with symmetric coding. The results prove this hypothesis and they are shown in the Results section. Between these two threshold levels symmetric coding performs better than asymmetric coding.

Adile				Flower				Train			
SNR		Spatial		SNR		Spatial		SNR		Spatial	
38.03	699.07	31.41	729.15	34.25	730.66	29.90	716.66	33.79	956.62	29.97	914.65
35.57	496.96	31.04	513.77	33.84	678.33	29.79	643.18	32.48	748.63	29.59	729.06
33.07	352.96	30.26	342.27	32.06	481.06	29.49	496.82	29.27	510.04	28.71	533.64
29.499	262.14	29.43	262.49	29.47	343.45	28.66	340.24	28.48	495.62	28.31	492.57

Table 2.3: RD performance of test sequences for determining the type of asymmetry

As long as the auxiliary view of asymmetrically coded stereo video is fixed around the threshold level, asymmetric coding performs better than symmetric coding in terms of perceived video quality. However if the quality of the auxiliary view is further decreased, human eye starts to be inadequate for compensating the quality difference between two views and hence symmetric coding becomes more favorable than asymmetric coding below the threshold levels.

In order to test this hypothesis, we have conducted two subjective tests at different PSNR and bitrates. In the first test, we have constrained the total bitrate for left and right views to about 1 Mbps, and hence the auxiliary view is encoded below the PSNR threshold of 31 dB determined in hypothesis 1. In the second test, we have encoded the auxiliary view above 31 dB.

The results clearly show that below the threshold level, symmetric coding achieves better performance than asymmetric coding as expected. Furthermore, the results demonstrates that when asymmetric coding is performed at the threshold level, the results of asymmetric coding overwhelms the results of the symmetric coding. The results are presented and argued extensively in the results section.

2.4 Extension to Scalable Multi-View Coding

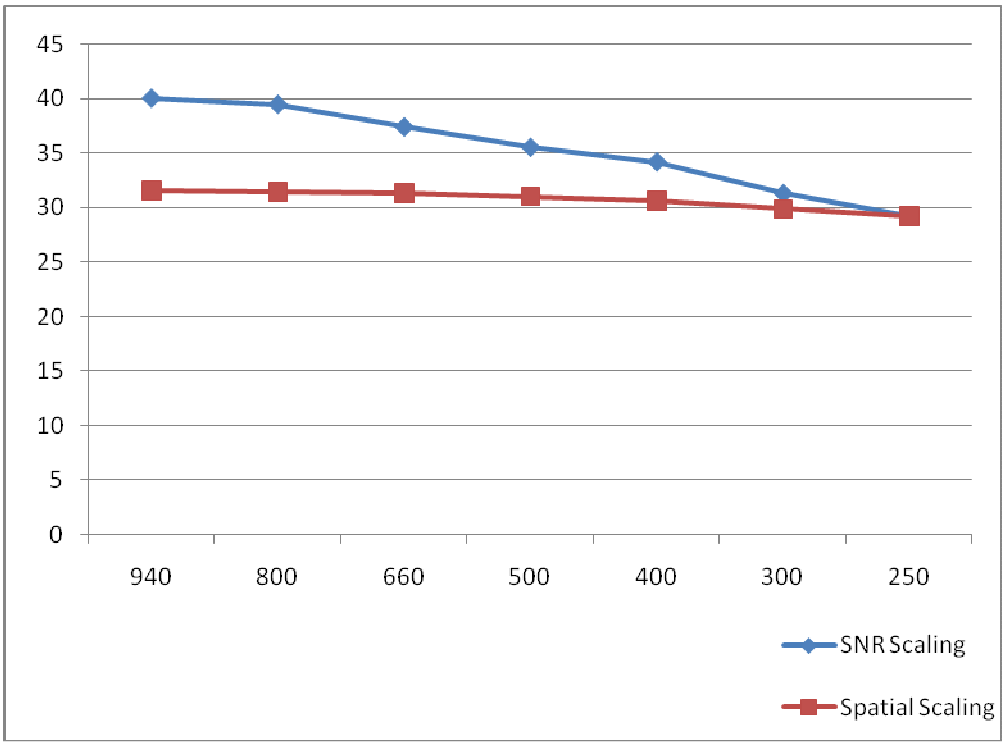
Scalable extension of H.264/AVC is used as encoding codec in this work [16]. Scalability has a vital importance in network adaptation of the video stream that is to be transmitted, however gain of scalability comes with a cost of bitrate overhead around %10 with respect to H.264/AVC.

We have used two types of scalability, spatial and SNR in this work as discussed in Section 2.A. Table 2.4 shows the bitrate PSNR values for both SNR scaling (SNR) and spatial scaling (SPA). From this table, we can observe that spatial scalability causes PSNR to drop drastically at the first step however the speed of PSNR drop with spatial scaling is slower the speed of PSNR drop of SNR scaling. Hence even SNR scaling supports more extraction points and high PSNR at high levels of bitrate, as the bitrate gets smaller, spatial scaling starts to perform better than SNR scaling. It is important to note that this behavior explains why SNR achieves better perceived quality scores than spatial scaling above the threshold levels and why spatial scaling starts to perform better than SNR at lower PSNR levels below the threshold in our subjective test results. For better understanding, Figure 2.1 shows R-D performance of spatial and SNR scaling applied on three of test sequences at some bitrate intervals.

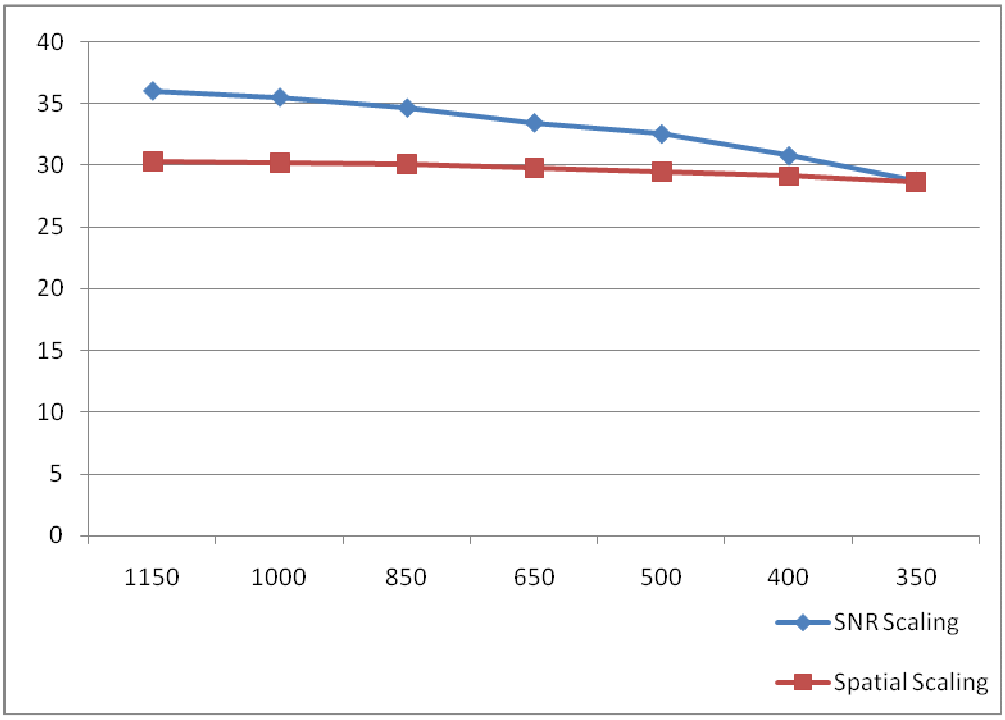
Multi-View Coding (MVC) uses high correlation between views to eliminate inter-view redundancies. Table 2.5 represents compression performance of MVC encoded at different rates. From this table, we can observe that MVC can compress the other view from %12 up to %72, different prediction strategies can be implemented for better performance [17]. The compression efficiency also depends on the type of the content, i.e. MVC performs better on Adile than Flower and Train because Adile has lower noise and it is a computer generated content. Researchers also uses scalable implementations of MVC [18] and asymmetric coding with MVC[19]. Hence, it is a fact that MVC has better R-D performance compared to Scalable extension of H.264/AVC. However, our results can also be extended to MVC with higher PSNR rates.

Adile				Train				Flower			
SNR		SPA		SNR		SPA		SNR		SPA	
PSNR	bitrate	PSNR	bitrate	PSNR	bitrate	PSNR	bitrate	PSNR	bitrate	PSNR	bitrate
42.88	1324.49	42.88	1324.49	38.39	2773.54	38.39	2773.54	38.28	2059.26	38.28	2059.26
42.89	1324.49	31.58	946.18	37.91	2773.54	30.61	1831.87	38.28	2059.26	30.29	1162.04
42.18	1215.90	31.53	872.46	37.52	2466.01	30.61	1653.51	37.81	1818.89	30.22	1051.81
41.49	1113.80	31.48	801.43	37.15	2197.69	30.54	1494.42	37.35	1616.63	30.15	953.94
40.75	1009.15	31.41	729.15	36.72	1938.88	30.43	1336.95	36.87	1418.58	30.06	856.74
40.06	921.09	31.33	667.97	36.34	1735.78	30.34	1214.73	36.42	1266.18	29.98	779.48
39.50	854.89	31.26	620.75	35.97	1576.11	30.25	1115.25	36.05	1145.26	29.90	716.66
38.72	764.01	31.16	558.00	35.49	1389.13	30.11	999.25	35.53	1006.53	29.79	643.18
38.03	699.07	31.04	513.77	35.04	1253.18	29.97	914.65	35.07	902.74	29.68	589.49
37.43	641.33	30.92	472.89	34.64	1141.51	29.85	844.84	34.67	817.12	29.58	542.98
36.76	576.73	30.80	429.25	34.22	1027.50	29.73	775.24	34.25	730.66	29.49	496.82
36.12	537.87	30.64	401.88	33.79	956.62	29.59	729.06	33.84	678.33	29.38	467.40
35.57	496.96	30.48	374.30	33.40	889.76	29.44	688.15	33.46	626.42	29.26	439.11
34.83	449.88	30.26	342.27	32.88	811.75	29.27	637.94	32.95	566.02	29.12	405.68
34.29	410.35	30.13	316.96	32.48	748.63	29.11	597.47	32.55	516.50	29.00	378.98
33.65	383.47	29.88	298.72	31.98	700.46	28.91	565.34	32.06	481.06	28.82	360.54
33.07	352.96	29.64	278.49	31.50	648.58	28.71	533.64	31.60	442.23	28.66	340.24
32.49	326.37	29.43	262.49	31.03	606.82	28.50	506.46	31.15	411.19	28.49	324.00
32.03	315.10	29.22	254.61	30.63	582.27	28.31	492.57	30.77	394.90	28.33	315.54
31.34	292.13	28.91	240.16	30.09	546.78	28.02	469.79	30.25	368.30	28.11	302.90
30.91	283.08	28.70	236.36	29.69	528.17	27.81	458.64	29.89	356.79	27.94	296.94
30.421	272.59	28.42	229.33	29.27	510.04	27.59	448.82	29.47	343.45	27.73	292.07
29.897	268.83	28.14	228.12	28.84	498.23	27.35	445.89	29.07	336.67	27.53	290.87
29.499	262.14	27.9	227.55	28.48	495.62	27.13	443.98	28.73	334.94	27.36	291.05

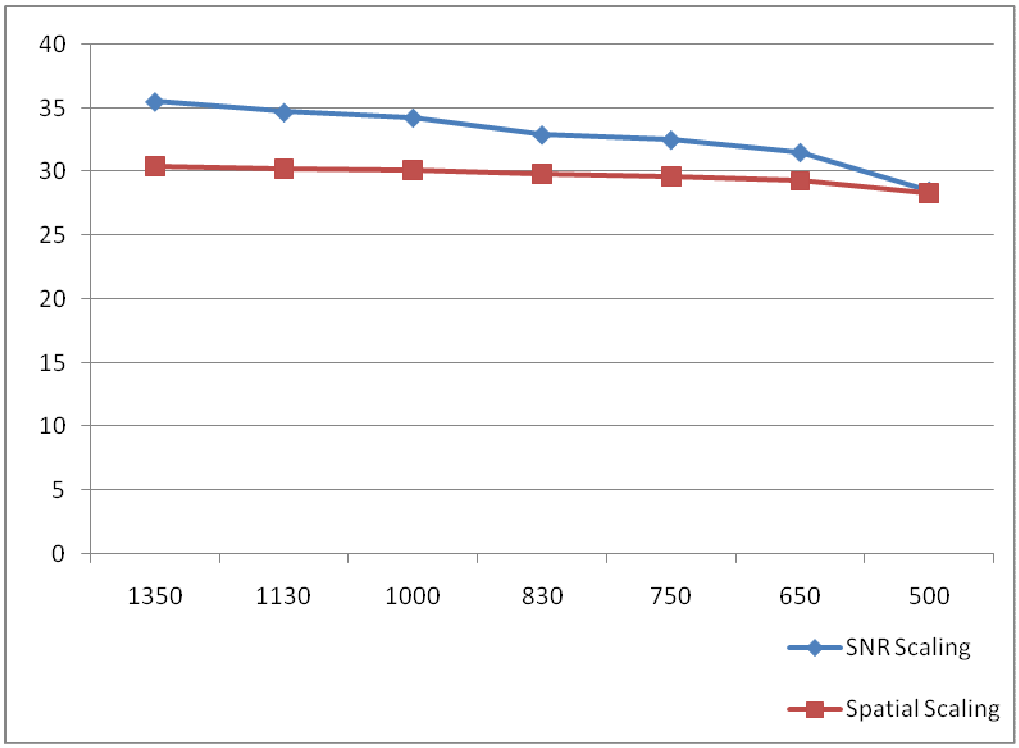
Table 2.4: R-D performance of SNR and spatial scalability of SVC


45
40
35
30
25
20
15
10
5
0
940
800
660
500
400
300
250
SNR Scaling
Spatial Scaling

(a)


40
35
30
25
20
15
10
5
0
1150
1000
850
650
500
400
350
SNR Scaling
Spatial Scaling

(b)



(c)

Fig.2.1: R-D performances of SNR and Spatial Scaling. (a)Adile,(b)Flower,(c)Train

Adile				Flower				Train			
PSNR	Bitrate L	Bitrate R	Gain (%)	PSNR	Bitrate L	Bitrate R	Gain (%)	PSNR	Bitrate L	Bitrate R	Gain (%)
42,03	1206,06	488,00	0,60	38,25	2107,20	1801,65	0,14	38,31	2890,80	2557,12	0,12
36,51	828,12	286,73	0,65	36,51	1250,91	1002,82	0,20	36,50	1776,70	1451,91	0,18
34,32	569,64	171,33	0,70	34,32	778,09	581,79	0,25	34,32	1134,22	852,19	0,25
33,47	382,33	106,40	0,72	31,93	483,43	335,74	0,31	31,86	717,05	508,78	0,29

Table 2.5: Compression performance of MVC

Chapter 3

MULTIPLE DESCRIPTION ASYMMETRIC STEREO VIDEO CODING

Joint Video Team (JVT) has developed Scalable Video Coding Extension of H.264/AVC (SVC) and it is widely used for rate adaptive streaming of video content over congested networks. In order to generate scalable video streams, Joint Scalable Video Model (JSVM) is being used as a reference software[20]. The necessary bandwidth information for adaptive streaming is obtained by using DCCP featuring TCP-friendly rate control (TFRC) protocol [21-22]. Although scaling the video stream according to the feedback of TFRC offers efficient rate adaptation and error resilience, burst packet losses caused by the network can still perturb the received video quality.

Multiple description coding addresses the problem of increasing the robustness of the stream to the distortions caused by the network environment [23]. In the most general case, two descriptions of a single stream is generated such that each of them can be decoded separately and provides acceptable quality furthermore receiving both of them produces higher quality. There are many ways of generating descriptions for both single view and multiview video sequences [24-27]. In the case of stereoscopic video, descriptions can be generated either coding the monoscopic videos directly or coding each view by exploiting the inter-view correlations. The second approach is similar to Multi-View Coding (MVC). The redundancy between views is minimized which means significant gain is obtained for the bandwidth of the network depending on the features of the stereoscopic video such as distance of the cameras, density of the motion inside views, etc [28]. The disadvantage of this approach is if the reference view gets distorted by the packet loss, the other view also gets

perturbed. In this work, we have generated our descriptions by coding the monoscopic videos directly.

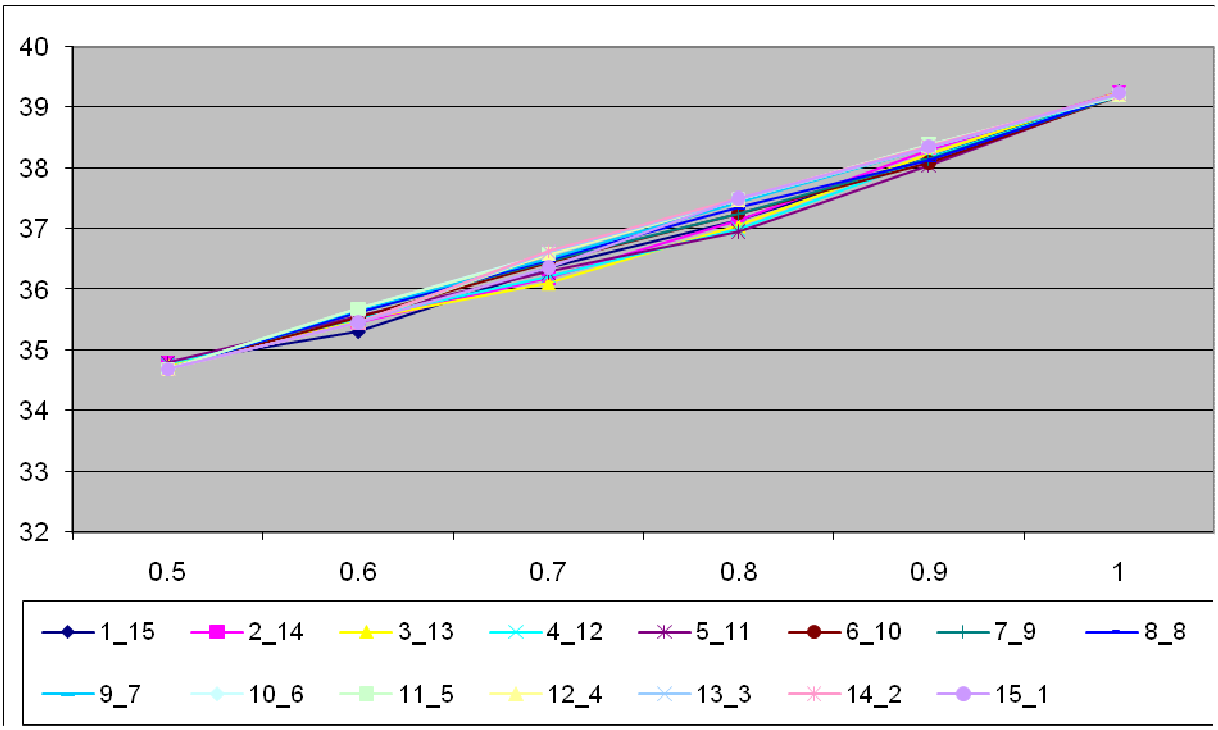
There are two different theories for rate allocation between stereoscopic views which are fusion and suppression theories. Fusion theory, similar to symmetric coding , suggest equal weighting between views in terms of bitrate. Suppression theory, analogous to asymmetric coding, defends unequal rate allocation between views. According to this theory, the resulting perceptual video quality would be dominated by the high quality view of stereoscopic pair. There are many researches related to asymmetric coding of stereoscopic video[29]. In our work, we used both techniques to extract our descriptions for stereoscopic test sequences.

SVC supports three scalability types, spatial, temporal and SNR scalability. SNR scalability is utilized for generating the descriptions in this work. The number of enhancement units in a scalable bitstream effects the total number of extraction points directly. Incrementing their number increases the number of extraction points however it also decreases compression efficiency at those points[30]. Section 3.1 discusses the effect of the number of enhancement units to the compression efficiency and determines the adequate number for extracting the descriptions. In section 3.2 we present the structure of the generated descriptions for channel simulation. Channel simulation is a stochastic process which means it has to be done iteratively similar to Monte Carlo Simulation to obtain meaningful results. Section 3. 3. presents channel simulation results that obtained from 100 iterations. Finally in Section 3.4. we comment on results and draw our conclusions.

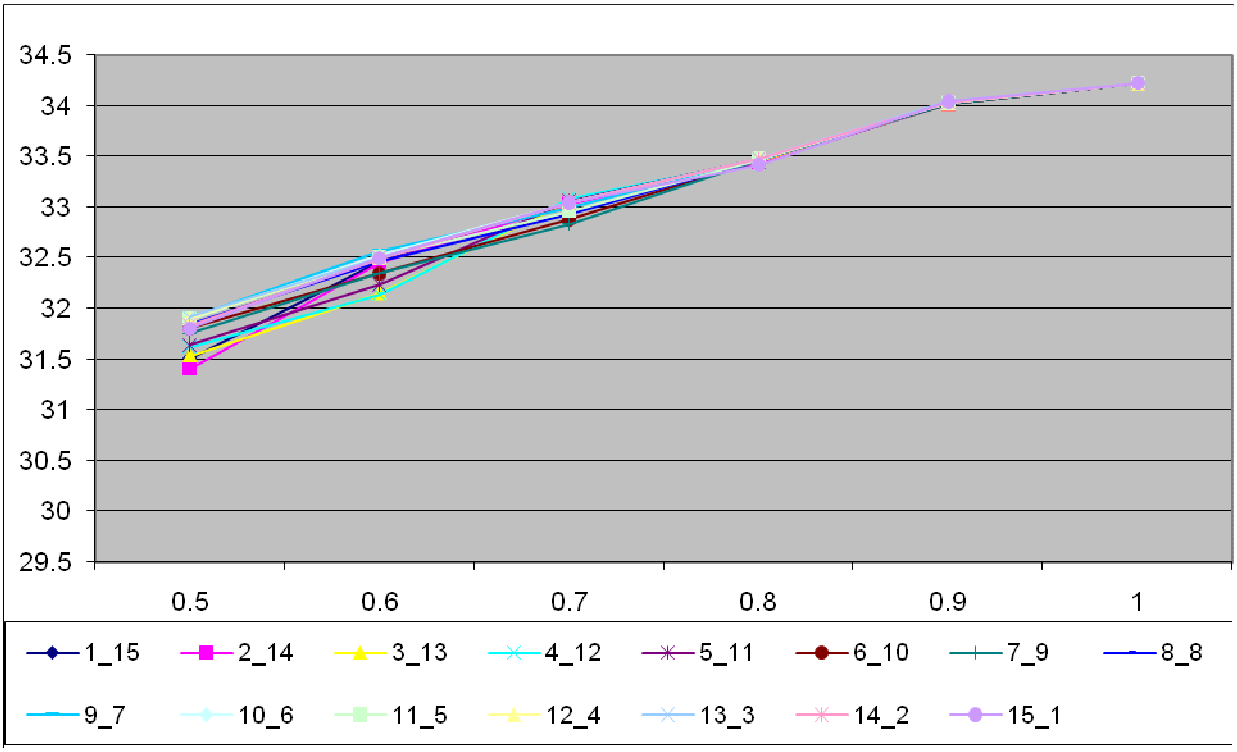
3.1. Determination of The Number of Enhancement Units

It is a well known fact that, the number of enhancement units has a direct influence on the compression efficiency as well as on the number of extraction points. Therefore as long as the number of extractions points is sufficient for rate allocation, choosing the smaller number of enhancement units increases the compression efficiency. SNR NAL units and spatial scalability NAL units are examples of enhancement layer units. There are decoders that cannot decode the streams with more than one SNR NAL units. Hence at first we try to find the optimum MGS vector configurations in terms of R-D performance for scalable video streams with two SNR NAL units and afterwards we compare the resulting R-D performance with the R-D performance of the scalable video stream with only one SNR NAL unit.

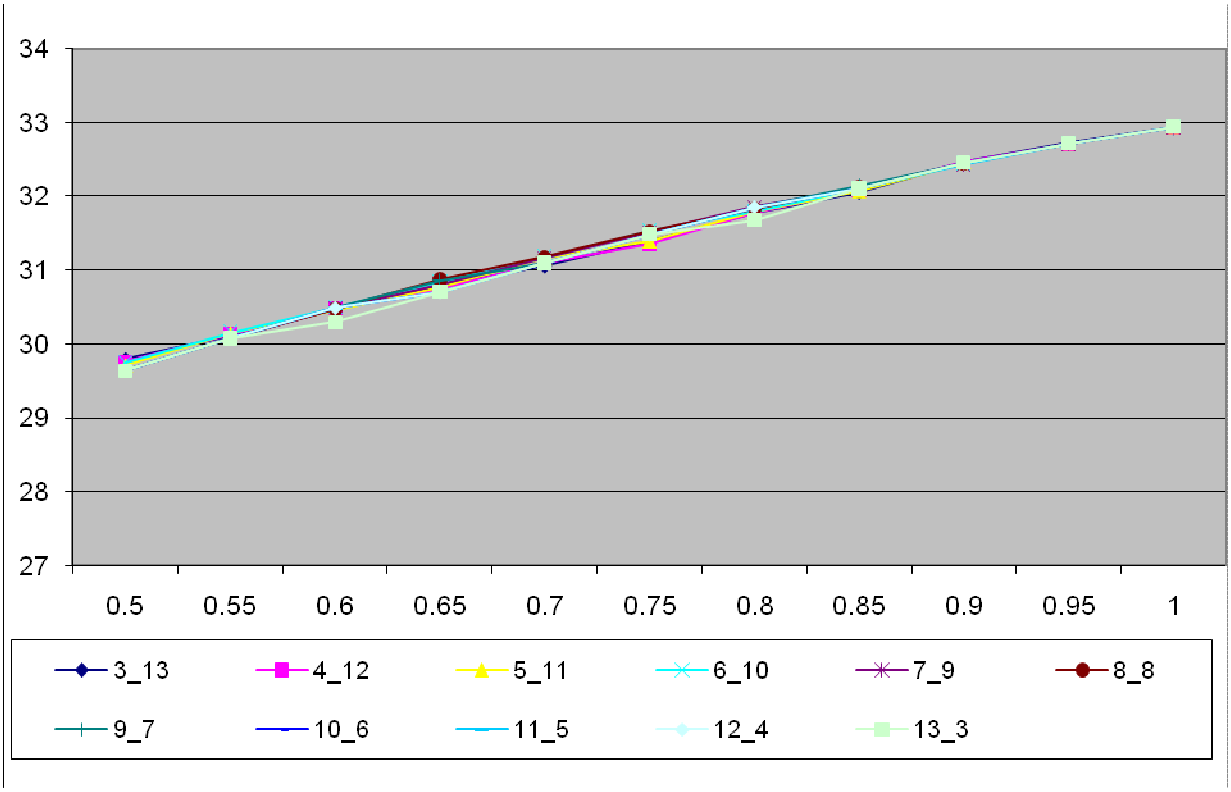
Figure 3.1 shows the R-D performance of video streams with two SNR NAL units and with altering MGS vector configurations also extracted between 0-1 Mbits with 0/5 Mbit step sizes. The plots show that, even the performances are too close to each other, However, vector configuration with 10-6 achieved the best R-D performance for all of the contents. At the second step, compare R-D performance of video stream with 2 SNR NAL units that has 10-6 mgs vector configuration with video stream including only one SNR NAL unit. Figure 3.2 presents the results. The final results show that in fact there isn't any significant difference between 2 and 1 SNR NAL unit configurations. Hence we prefer to use 1 SNR NAL unit configuration for channel simulation scenario.


40
39
38
37
36
35
34
33
32
0.5
0.6
0.7
0.8
0.9
1
1_15
2_14
3_13
4_12
5_11
6_10
7_9
8_8
9_7
10_6
11_5
12_4
13_3
14_2
15_1

(a)

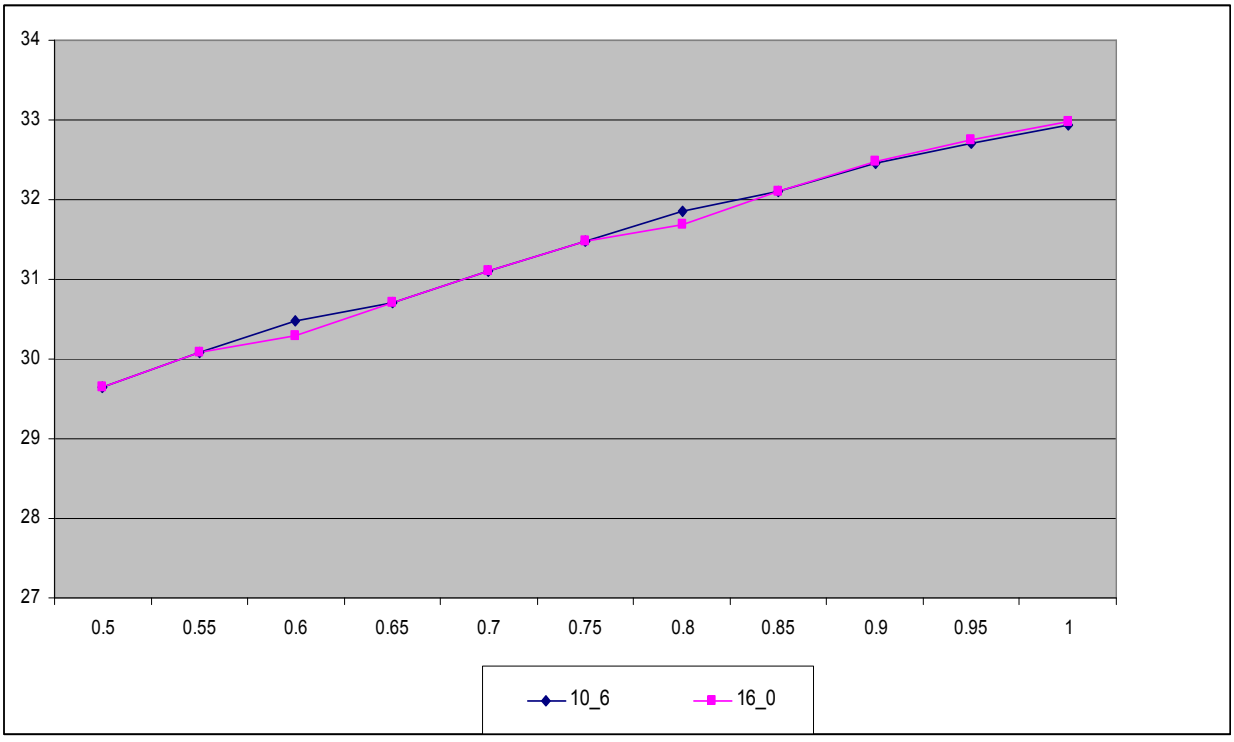

34.5
34
33.5
33
32.5
32
31.5
31
30.5
30
29.5
0.5
0.6
0.7
0.8
0.9
1
1_15
2_14
3_13
4_12
5_11
6_10
7_9
8_8
9_7
10_6
11_5
12_4
13_3
14_2
15_1

(b)

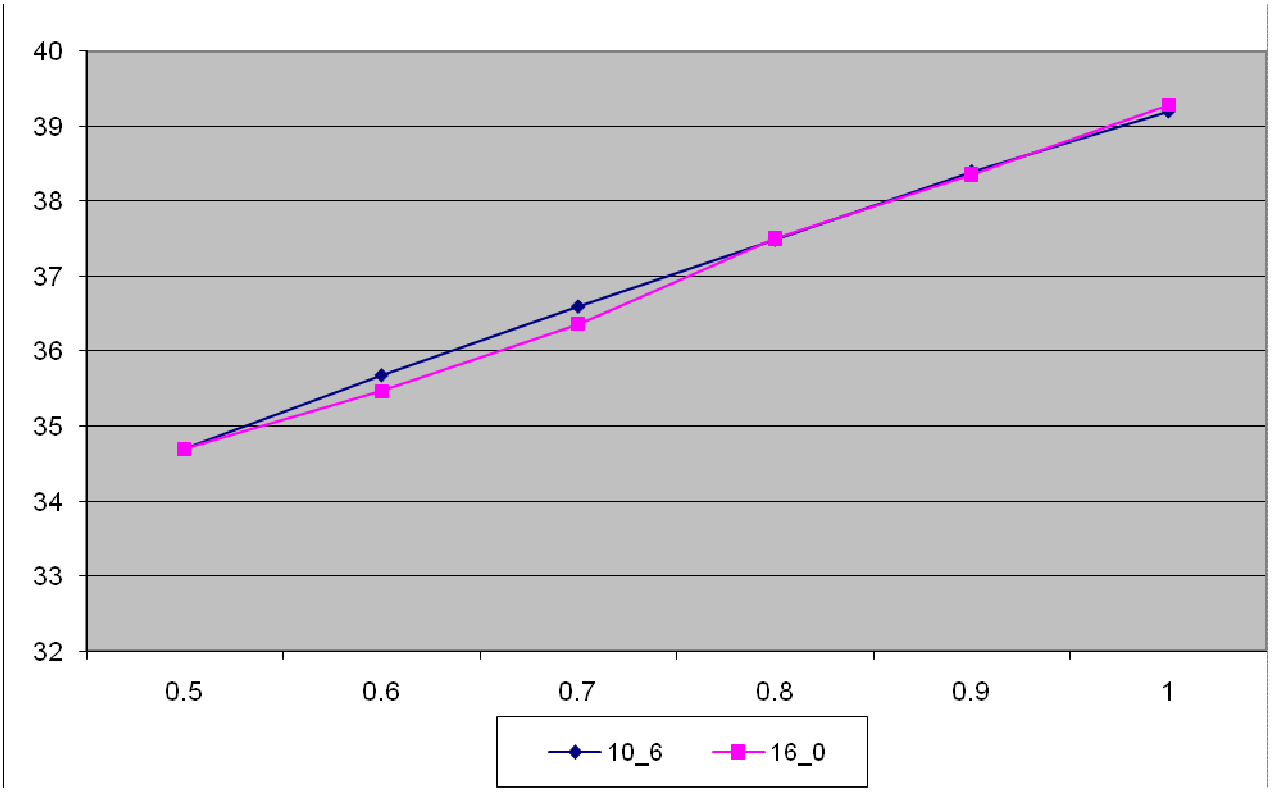


(c)

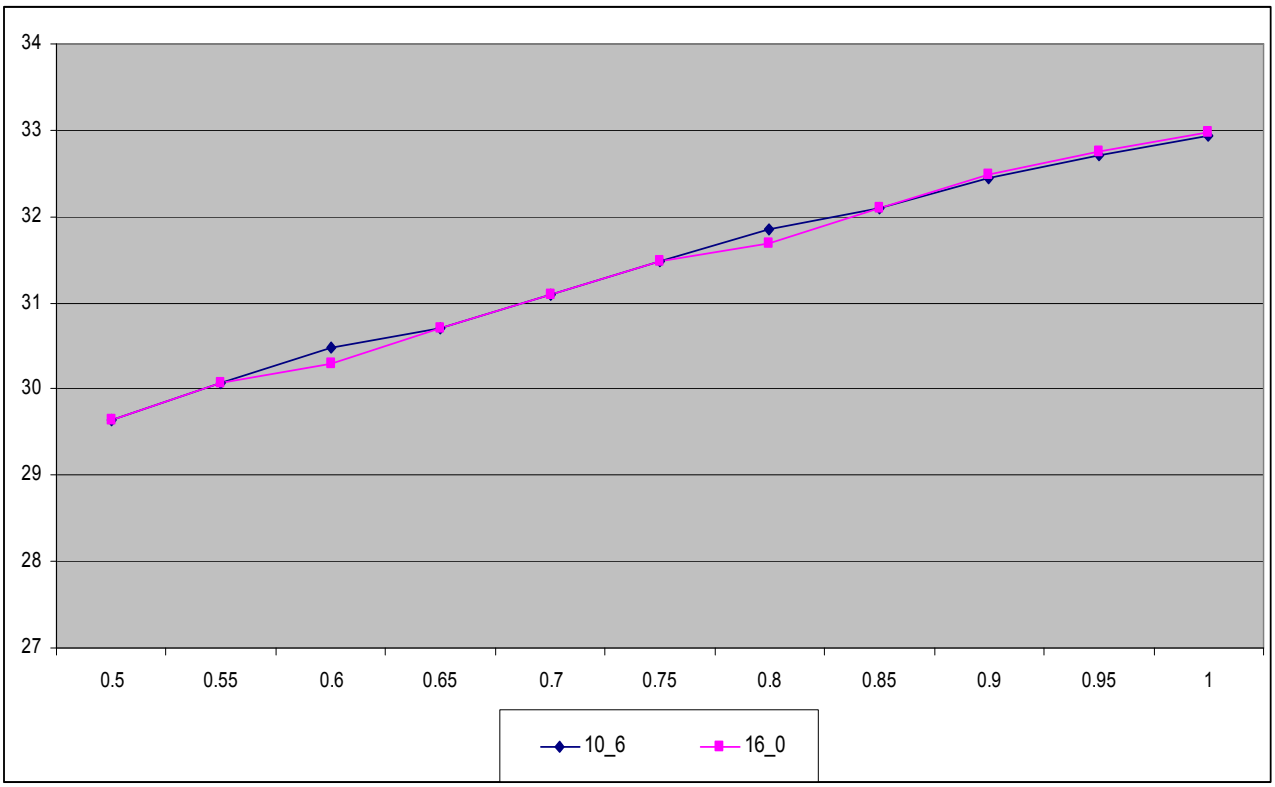
Fig. 3.1: R-D performance of video streams with 2 SNR enhancement units (a) Adile (b) Flower (c)Train



(a)



(b)



(c)

Fig. 3.2: R-D performance of scalable stream with 1 SNR enhancement unit & 2SNR enhancement unit (a)Adile, (b) Flower, (c)Train

3.2. Scalable MD Generation Method Candidates

In this section, we describe 4 different MD generation methods. In order to generate those descriptions, we have implemented a C++ program that has the ability of channel simulating the resulting description as an optional work. The program also generates the symmetric of those descriptions. For stereoscopic multiple description coding, we need to generate at least 2 independently decodable descriptions for two different channels such that each description has one left view bitstream and one right view bitstream.

In common, all descriptions has the base layer of views, therefore each description guarantees base layer quality when they are successfully received by the recipient. However MGS NAL units of each view are periodically distributed into even and odd frames of each of these descriptions. This period corresponds to two group of pictures (GOP) for each description. The description generation models are explained blow. For all models, "B" indicates that the frame that only contains the base layer and "B+E" denotes that the frame contains base layer as well as all the MGS NAL units. In our case, there is only one MGS NAL unit as discussed in the previous section.

Model 1:

The left view contains only the base layer inside every GOP and right view includes base layer and MGS NAL units. Therefore left view is encoded at base quality and right view is encoded at full quality which means this generation method relies on suppression theory.

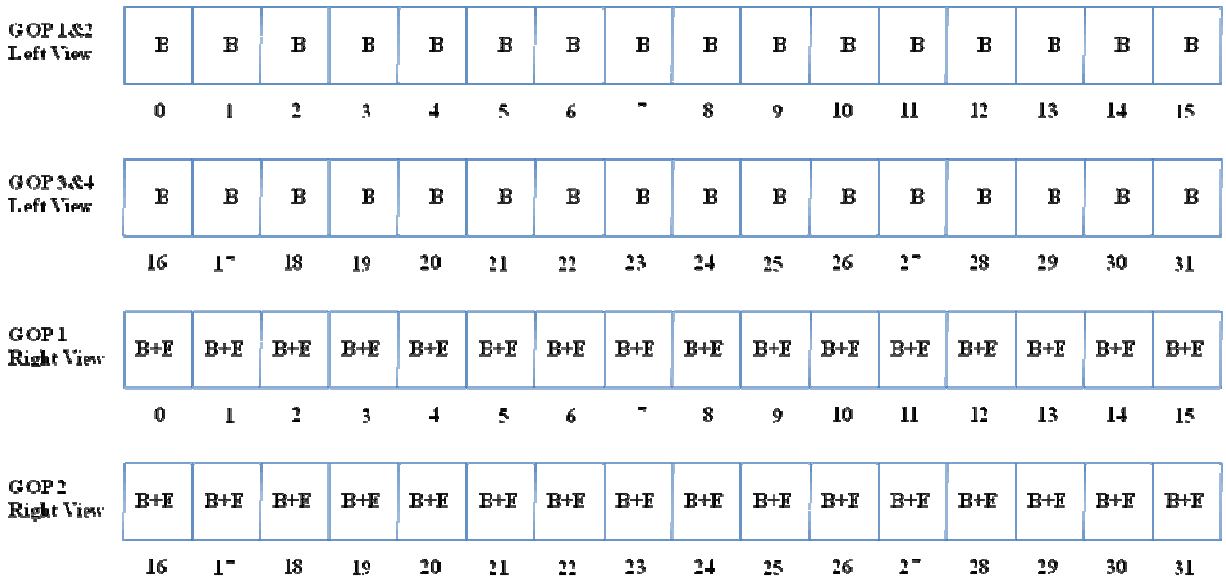


Fig. 3.3: Scalable MDC Description 1 of Model 1(Description 2 is the symmetric of Description 1)

Model 2:

In this model, first and the second GOP are encoded at base quality, third and fourth GOP are encoded at full quality and the remaining GOPs are also designed to be in this order. This strategy is built upon fusion theory since it supports equal weighting of bitrate for left and right views.

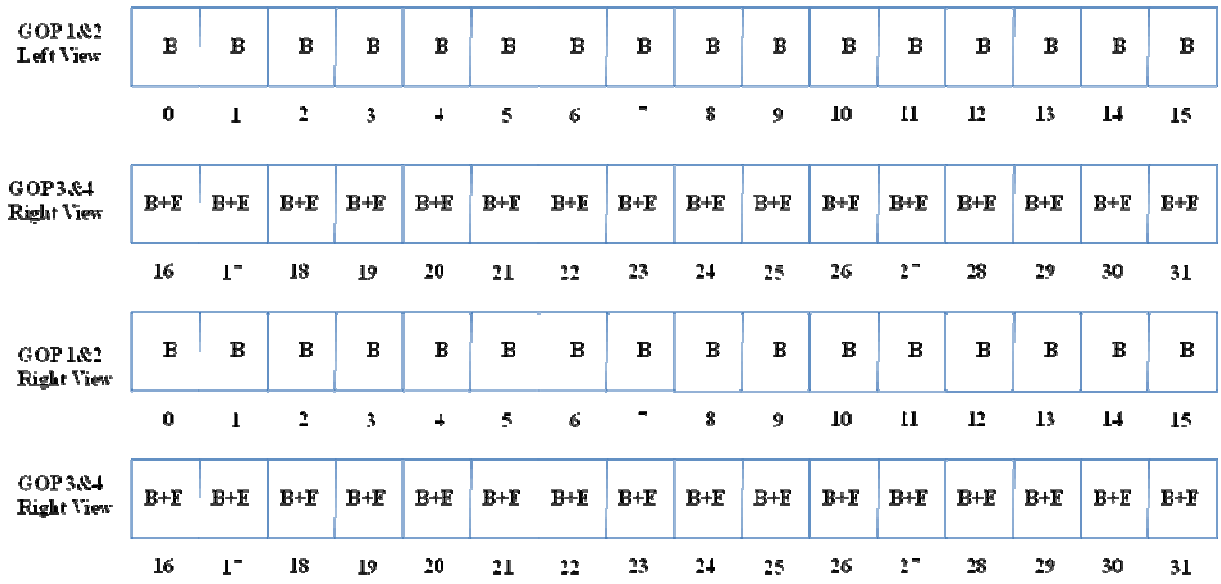


Fig. 3.4: Scalable MDC Description 1 of Model 2(Description 2 is the symmetric of Description 1)

Model 3:

As represented in the figure below, the MGS NAL units are assigned to even and odd according to their GOP number. Since this assignment follows the same routine for all GOPs, the resulting stream is still built according to fusion theory.

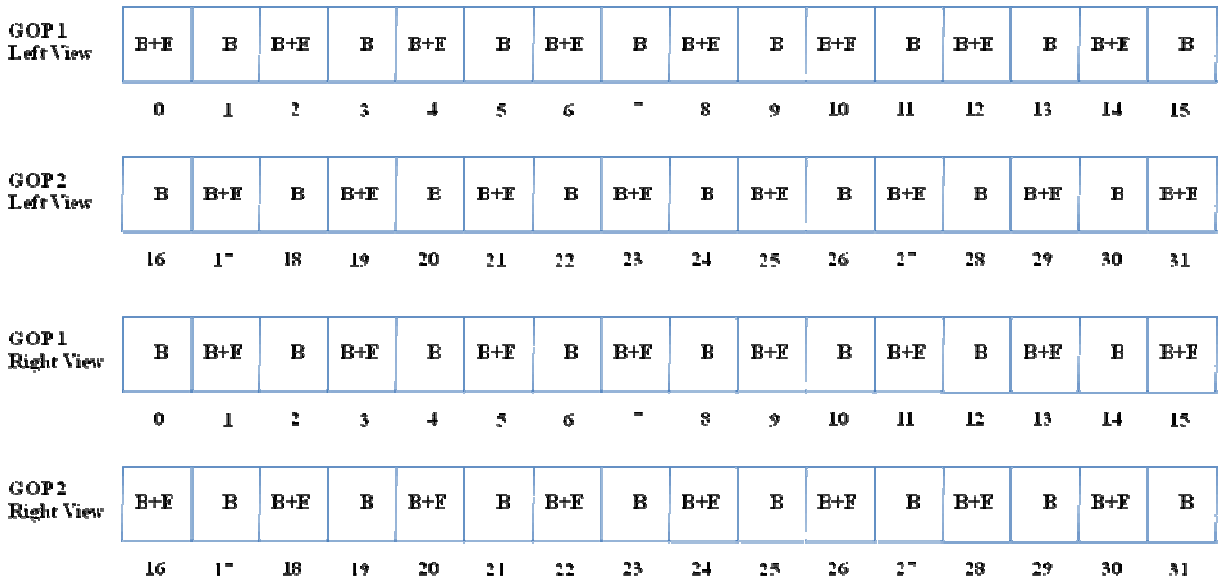


Fig. 3.5: Scalable MDC Description 1 of Model 3(Description 2 is the symmetric of Description 1)

Model 4:

This model is the exactly the same as model 3 except all I frames are encoded at full quality. Although more redundancy is added, the bitrate is still equally distributed to both of views which means this model also relies on fusion theory.

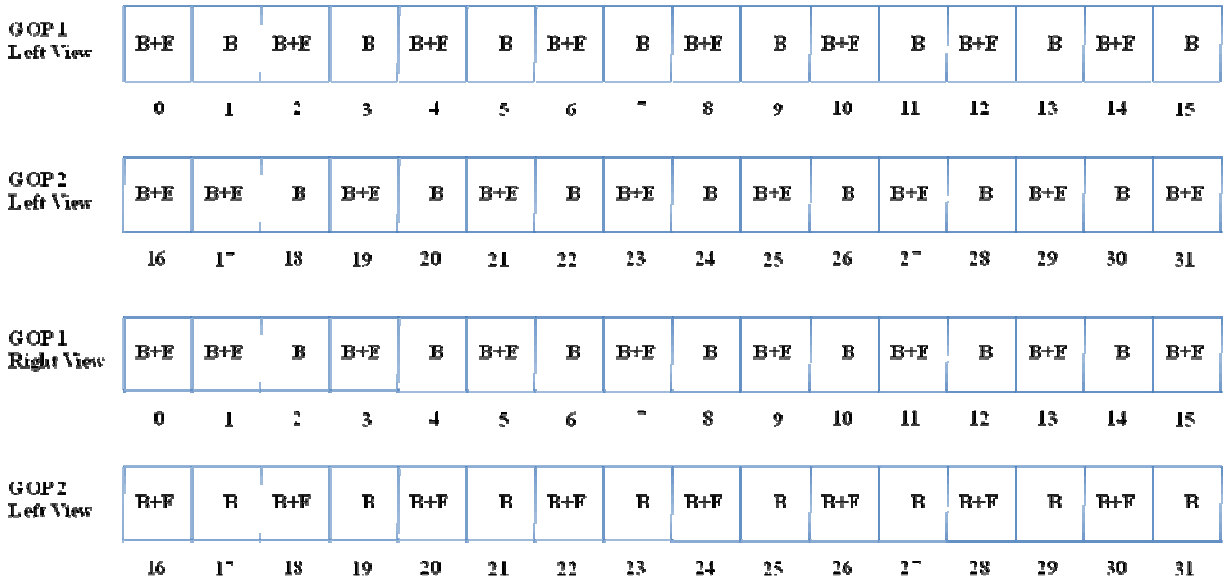


Fig. 3.6: Scalable MDC Description 1 of Model 4(Description 2 is the symmetric of Description 1)

For all of the descriptions, the quantization parameter (QP) of the base layer is the same, however the QP of the enhancement layer is different for model 7 since there is more

redundancy added in this model in comparison to other modes. The bitrates and PSNR values for all descriptions of all models are presented in Table 3.1. Table 3.2 shows the size of original video sequence that all of the models are generated from. It is important to note that, in order to have all the descriptions at the same bitrate, we scaled the source stream of model 7 with higher QP, therefore the PSNR of the resulting descriptions of this model is 1.5-3 dB lower than the output descriptions of other models on the average. In other words, additional redundancy caused around 1.5-3 dB loss on the resulting description's PSNR.

		Desc1L	Desc1R	Desc2L	Desc2R	TotDesc1	TotDesc2	Desc1L	Desc1R	Desc2L	Desc2R
Adile	**Model1**	0.198	0.771	0.770	0.197	0.970	0.968	29.514	37.061	37.029	29.533
	Model2	0.483	0.484	0.484	0.484	0.968	0.969	33.141	33.18	33.158	33.164
	Model3	0.484	0.484	0.483	0.485	0.969	0.969	33	33.051	33.025	33.004
	Model4	0.495	0.496	0.494	0.496	0.991	0.990	34.097	34.098	34.075	34.106
Flower	**Model1**	0.175	0.787	0.814	0.168	0.962	0.982	27.11	32.437	32.119	27.351
	Model2	0.492	0.478	0.497	0.477	0.971	0.974	29.502	29.848	29.623	29.768
	Model3	0.492	0.478	0.497	0.477	0.971	0.974	29.446	29.747	29.449	29.651
	Model4	0.512	0.494	0.511	0.494	1.006	1.005	30.578	30.818	30.561	30.821
Train	**Model1**	0.207	0.776	0.802	0.199	0.983	1.002	25.972	31.189	31.134	26.07
	Model2	0.494	0.500	0.517	0.477	0.994	0.994	28.398	28.612	28.55	28.493
	Model3	0.508	0.486	0.503	0.490	0.994	0.993	28.361	28.463	28.366	28.433
	Model4	0.508	0.489	0.505	0.492	0.997	0.997	29.21	29.287	29.181	29.311

Table 3.1: Bitrate and PSNR rates of description models for all contents.

	Adile		**Flower**		**Train**	
	Left View	**Right View**	**Left View**	**Right View**	**Left View**	**Right View**
Model 1	37.029	37.061	32.119	32.437	31.134	31.189
Model 2	37.029	37.061	32.119	32.437	31.134	31.189
Model 3	37.029	37.061	32.119	32.437	31.134	31.189
Model 4	34.287	34.302	30.855	31.087	29.509	29.597

Table 3.2: Bitrate and PSNR rates of description models for all contents.

3.3 Channel Simulation Scenario and Results

The crucial part of multi description coding simulation is simulating the packet loss over multiple paths (channels). In our case, we prefer sending two independently decodable descriptions over two independent network channels as shown in Figure 3.7. Since the channels are independent from each other, we can assume that their packet loss probability rates do not have a correlation in between.

There are two states in a real network environment in terms of packet loss probability. This model is also called two-state Gilbert model and approximates the burst behavior of packet losses over network channels. In our tests, packet loss ratios are selected as %3, %5 and %10 respectively for both channels.

Channel simulation is a stochastic process. In order to characterize such a process, it must be done several times iteratively similar to Monte Carlo method. In this work, we choose the iteration count as 100 and we calculated the average and standard deviation of each simulation. Table 3.3,4,5 presents the results of simulations.

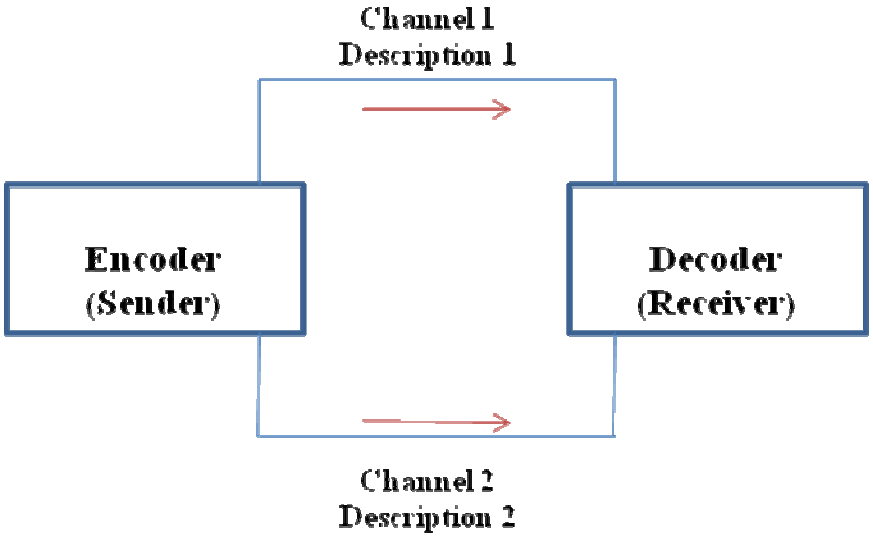


Fig. 3.7: Channel simulation scenario of two independently decodable descriptions

Method	Channel 1	Channel 2	Mean L	Mean R	std L	std R
1	3	3	36.64	36.658	0.259	0.227
1	3	5	36.376	36.722	0.31	0.202
1	3	10	35.7	36.67	0.372	0.241
1	5	3	36.604	36.4	0.238	0.274
1	5	5	36.361	36.358	0.309	0.329
1	5	10	35.799	36.397	0.383	0.284
1	10	3	36.637	35.827	0.217	0.342
1	10	5	36.326	35.7	0.351	0.396
1	10	10	35.777	35.822	0.389	0.404
2	3	3	36.65	36.708	0.212	0.241
2	3	5	36.628	36.458	0.246	0.286
2	3	10	36.574	35.809	0.265	0.466
2	5	3	36.485	36.557	0.309	0.267
2	5	5	36.384	36.369	0.292	0.284
2	5	10	36.044	36.016	0.372	0.398
2	10	3	36.21	36.186	0.341	0.372
2	10	5	36.071	36.04	0.306	0.421
2	10	10	35.667	35.82	0.384	0.385
3	3	3	36.634	36.631	0.234	0.253
3	3	5	36.526	36.503	0.267	0.341
3	3	10	36.194	36.133	0.352	0.352
3	5	3	36.548	36.567	0.224	0.262
3	5	5	36.397	36.41	0.298	0.331
3	5	10	36.046	36.035	0.373	0.338
3	10	3	36.183	36.195	0.355	0.361
3	10	5	35.983	36.04	0.324	0.396
3	10	10	35.649	35.705	0.47	0.533
4	3	3	34.26	34.28	0.034	0.023
4	3	5	34.261	34.276	0.026	0.028
4	3	10	34.229	34.242	0.058	0.061
4	5	3	34.258	34.271	0.03	0.038
4	5	5	34.245	34.267	0.044	0.035
4	5	10	34.219	34.236	0.055	0.054
4	10	3	34.234	34.253	0.051	0.04
4	10	5	34.212	34.247	0.056	0.049
4	10	10	34.172	34.178	0.077	0.088

Table 3.3: Mean and standard deviation results for channel simulation of Adile

Method	Channel 1	Channel 2	Mean L	Mean R	std L	std R
1	3	3	31.967	32.15	0.138	0.197
1	3	5	31.769	32.19	0.195	0.169
1	3	10	31.355	32.19	0.296	0.156
1	5	3	31.952	32.038	0.155	0.219
1	5	5	31.786	32.039	0.203	0.191
1	5	10	31.423	32.006	0.268	0.177
1	10	3	31.931	31.634	0.177	0.265
1	10	5	31.778	31.598	0.192	0.299
1	10	10	31.412	31.582	0.243	0.293
2	3	3	31.967	32.173	0.14	0.174
2	3	5	31.867	32.1	0.197	0.187
2	3	10	31.665	31.926	0.188	0.178
2	5	3	31.867	32.088	0.176	0.173
2	5	5	31.795	32.033	0.206	0.213
2	5	10	31.548	31.809	0.247	0.236
2	10	3	31.631	31.9	0.236	0.209
2	10	5	31.577	31.794	0.262	0.258
2	10	10	31.384	31.632	0.259	0.263
3	3	3	31.965	32.205	0.152	0.157
3	3	5	31.863	32.071	0.203	0.185
3	3	10	31.641	31.874	0.248	0.197
3	5	3	31.864	32.108	0.157	0.197
3	5	5	31.779	32.026	0.196	0.183
3	5	10	31.625	31.834	0.236	0.214
3	10	3	31.684	31.883	0.224	0.212
3	10	5	31.584	31.781	0.268	0.281
3	10	10	31.349	31.603	0.246	0.293
4	3	3	30.83	31.63	0.02	0.022
4	3	5	30.826	31.06	0.018	0.016
4	3	10	30.796	31.028	0.033	0.039
4	5	3	30.822	31.057	0.03	0.023
4	5	5	30.813	31.038	0.028	0.47
4	5	10	30.776	31.013	0.062	0.047
4	10	3	30.8	31.034	0.036	0.036
4	10	5	30.771	31.015	0.051	0.041
4	10	10	30.727	30.796	0.075	0.062

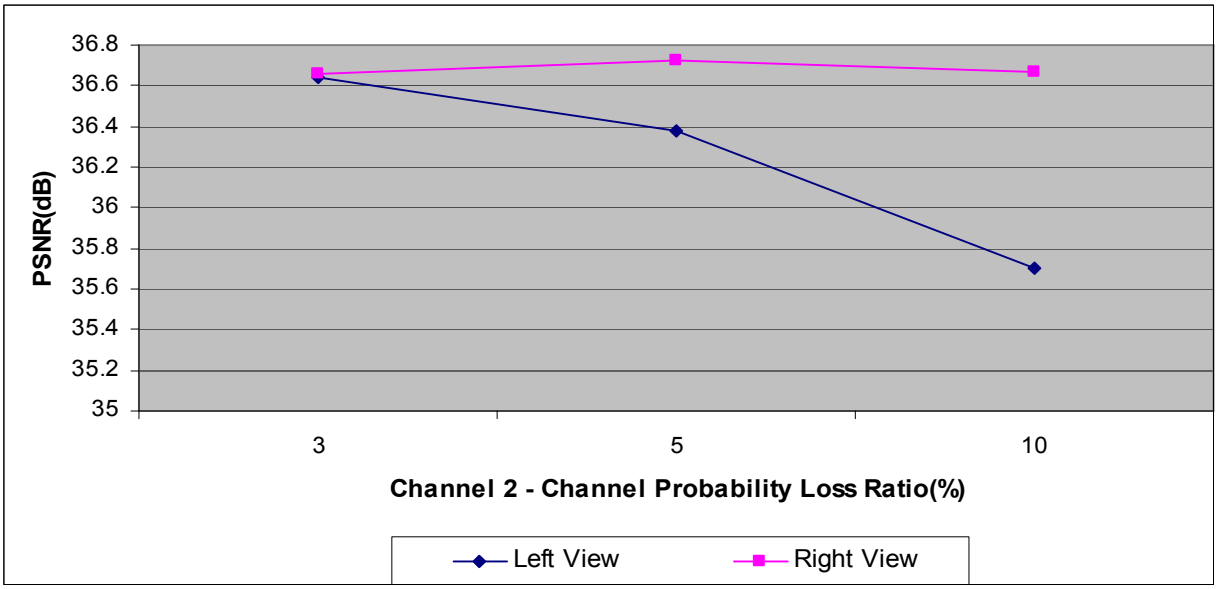
Table 3.4: Mean and standard deviation results for channel simulation of Flower

Method	Channel 1	Channel 2	Mean L	Mean R	std L	std R
1	3	3	30.867	30.932	0.138	0.118
1	3	5	30.732	30.927	0.183	0.146
1	3	10	30.379	30.926	0.208	0.16
1	5	3	30.88	30.761	0.144	0.166
1	5	5	30.693	30.779	0.174	0.166
1	5	10	30.307	30.778	0.253	0.181
1	10	3	30.873	30.414	0.145	0.255
1	10	5	30.691	30.421	0.166	0.203
1	10	10	30.322	30.402	0.216	0.205
2	3	3	30.89	30.95	0.134	0.16
2	3	5	30.804	30.836	0.159	0.171
2	3	10	30.597	30.685	0.193	0.174
2	5	3	30.828	30.861	0.143	0.158
2	5	5	30.71	30.798	0.196	0.168
2	5	10	30.527	30.624	0.219	0.215
2	10	3	30.626	30.652	0.19	0.165
2	10	5	30.511	30.583	0.212	0.183
2	10	10	30.293	30.406	0.262	0.196
3	3	3	30.873	30.943	0.14	0.14
3	3	5	30.811	30.89	0.17	0.15
3	3	10	30.609	30.662	0.188	0.199
3	5	3	30.828	30.864	0.157	0.178
3	5	5	30.754	30.775	0.161	0.181
3	5	10	30.53	30.61	0.221	0.212
3	10	3	30.583	30.662	0.19	0.197
3	10	5	30.541	30.611	0.209	0.223
3	10	10	30.334	30.382	0.24	0.241
4	3	3	29.484	29.573	0.017	0.014
4	3	5	29.472	29.563	0.024	0.022
4	3	10	29.447	29.54	0.033	0.031
4	5	3	29.474	29.562	0.025	0.025
4	5	5	29.467	29.551	0.026	0.028
4	5	10	29.431	29.516	0.039	0.044
4	10	3	29.445	29.54	0.029	0.029
4	10	5	29.431	29.525	0.038	0.033
4	10	10	29.391	29.483	0.053	0.062

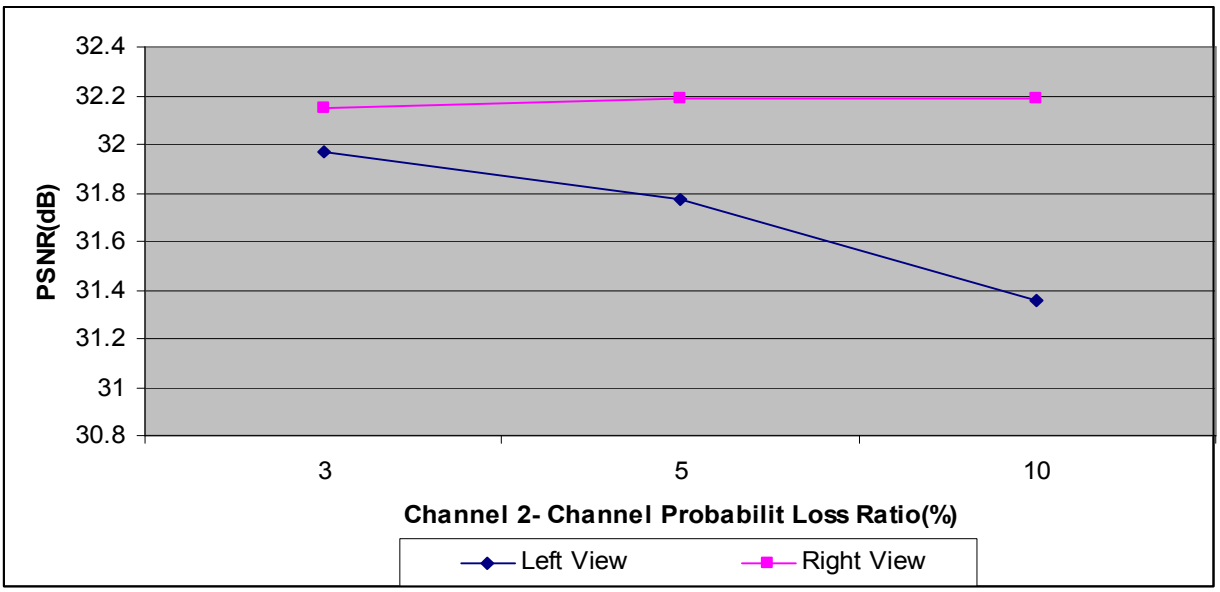
Table 3.5: Mean and standard deviation results for channel simulation of Train

In order to control the correctness of the results, the results of the first algorithm can be used. Base layer of the left view is sent through channel 1 and base and enhancement layer is sent through channel 2. Therefore what we need to observe is, as the channel loss probability ratio of the second channel increases, the resulting PSNR of the left view needs to decrease

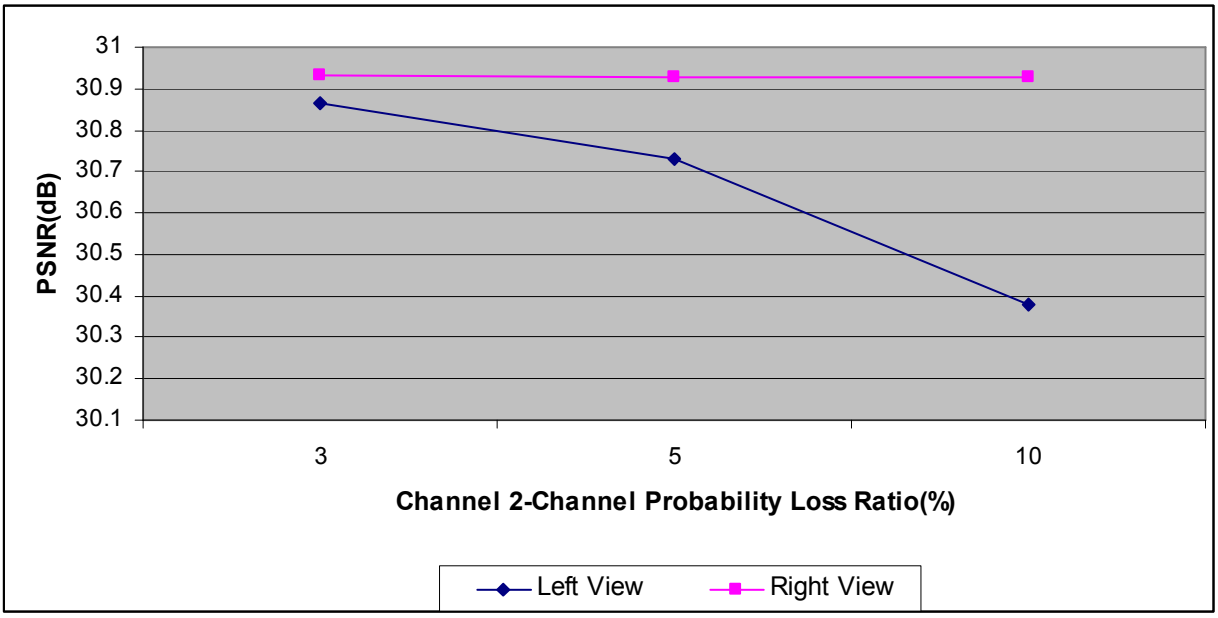
however the PSNR of decoded right view does not need to decrease since MGS NAL units of right view are sent through channel 1 and they are not affected from channel loss probability ratio of channel 2. Figure 3.8 shows that while channel 1 probability loss ratio(PLR) holds constant and channel 2's PLR varies, PSNR of right view stays constant however PSNR of left view decreases for all of the contents.



(a)



(b)



(c)

Fig. 3.8: PSNR of Left and Right View with varying channel 2 probability loss ratio. (a)Adile, (b)Flower, (c)Train

If we consider the performances of the methods from Tables 3.3,4and 5, we observe that model 1, 2and 3 are close to each other in terms of the PSNR values of both views under the highest possible loss rate for both views. However they perform better than model 4. The reason is the extra redundancy in model 4 had caused a PSNR degradation initially because there is a bandwidth limitation of 1Mbit for each channel therefore as the results show, even PLR of each channel is %10, the resulting PSNR of all views for models 1,2 and 3 never falls below the levels of model 4. In other words, the results demonstrate that the additional redundancy is useless even when the PLR of both channels are at the highest possible level. However this extra redundancy lowers the speed of PSNR degradation when the PLR increases and sinks the standard deviation of resulting PSNRs %75 approximately for all PLR.

Even %10 percent PLR is not sufficient to make the design of model 4 feasible. Therefore we increase PLR to %20 to see whether model 4 performs better than the other models in terms of R-D performance. The results are shown in Tables 3.6,7 and 8 respectively. From these results, we conclude that even the probability packet loss rate of the channels are both %20, the first three models should still be preferred.

Mode	Channel 1	Channel 2	Mean L	Mean R	std L	std R
5	3	20	35.684	35.746	0.367	0.288
5	5	20	35.563	35.604	0.383	0.384
5	10	20	35.268	35.302	0.433	0.451
5	20	3	35.708	35.675	0.353	0.385
5	20	5	35.553	35.653	0.397	0.421
5	20	10	35.181	35.275	0.447	0.429
5	20	20	34.794	34.691	0.478	0.453
7	3	20	34.202	34.197	0.059	0.072
7	5	20	34.162	34.197	0.085	0.066
7	10	20	34.101	34.13	0.116	0.089
7	20	3	34.205	34.211	0.057	0.062
7	20	5	34.184	34.197	0.066	0.067
7	20	10	34.103	34.103	0.097	0.112
7	20	20	33.999	34.01	0.129	0.124

Table 3.6: Mean and standard deviation results for channel simulation under %20 channel probability packet loss ratio of Adile

Mode	Channel 1	Channel 2	Mean L	Mean R	std L	std R
5	3	20	31.394	31.595	0.249	0.255
5	5	20	31.276	31.514	0.276	0.278
5	10	20	31.05	31.288	0.269	0.311
5	20	3	31.367	31.609	0.218	0.241
5	20	5	31.261	31.497	0.272	0.285
5	20	10	31.057	31.281	0.3	0.299
5	20	20	30.782	30.99	0.265	0.285
7	3	20	30.761	30.989	0.046	0.062
7	5	20	30.733	30.969	0.066	0.066
7	10	20	30.656	30.893	0.094	0.096
7	20	3	30.761	30.994	0.04	0.044
7	20	5	30.729	30.974	0.067	0.053
7	20	10	30.672	30.908	0.084	0.095
7	20	20	30.557	30.8	0.129	0.111

Table 3.7: Mean and standard deviation results for channel simulation under %20 channel probability packet loss ratio of Flower

Mode	Channel 1	Channel 2	Mean L	Mean R	std L	std R
5	3	20	30.328	30.402	0.213	0.216
5	5	20	30.209	30.328	0.237	0.255
5	10	20	30.022	30.135	0.237	0.255
5	20	3	30.331	30.387	0.198	0.22
5	20	5	30.196	30.338	0.247	0.225
5	20	10	30.013	30.123	0.263	0.246
5	20	20	29.698	29.814	0.234	0.267
7	3	20	29.403	29.506	0.046	0.046
7	5	20	29.387	29.475	0.048	0.06
7	10	20	29.309	29.431	0.075	0.08
7	20	3	29.415	29.506	0.031	0.052
7	20	5	29.383	29.492	0.053	0.047
7	20	10	29.334	29.439	0.075	0.061
7	20	20	29.226	29.33	0.093	0.092

Table 3.8: Mean and standard deviation results for channel simulation under %20 channel probability packet loss ratio of Train

Chapter 4

EXPERIMENTAL RESULTS

4.1 Subjective Test Procedures

We prepared three subjective tests to investigate the effect of the display type, using Double-Stimulus Continuous Quality-Scale (DSCQS) method [13]. 11 male and 4 female assessors are selected for the tests using polarized projection. 9 male and 1 female had attended to the tests with the 3D Display that has the parallax barrier. It is noted that in order to qualify the perceived quality of the test sequences, some rules need to be set such as perceived level of blur or blockiness [31]. In our tests, we also inform the viewers about these rules. The level of blur and blockiness are carefully determined so that they do not corrupt the depth perception.

4.1.A. Test Material

Source sequences are Train, Flowerpot and Adile at 704x576, 704x448, and 640x480 spatial resolutions respectively. The first 241 frames are encoded and streams are displayed at 30 fps.

4.1.B. Test Setup

We perform subjective tests on two different 3D display systems. The first setup consists of a pair of Sharp MB-70X projectors, a silver dielectric screen, polarized filter glasses and a PC to drive the projectors as shown in Fig. 4.1.a. The light emitted from each projector passes through two polarization planes. These planes polarize the light in opposite directions and allow filter glasses to allow only the corresponding view for each eye. The projectors are connected to a PC with a virtual desktop of size 2048x768. Each projector displays one half of the desktop at 1024x768 pixels resolution. The views are displayed on the silver screen on top of each other. The assessors sit approximately 3 meters away from the

screen. Unlike the horizontal or vertical polarization, the circular filtering allows users to tilt their heads without the risk of losing 3D perception.

The second 3D display is a Sharp AL3DU laptop that uses parallax barrier, as shown in Fig. 4.1.b. An operation called interdigitizing is performed to display the content in 3D. This operation reorders color information at subpixel level and allows the light from correct pixels to be emitted at a certain direction. Hence the viewers does not wear polarized filter glasses however they need to fix their head position constant throughout the test.

(a) (b)

Fig. 4.1. (a) Stereoscopic 3D display with polarized projection (b)Autostereoscopic display (Sharp AL3DU)

4.1.C. Test Methodology

Two subjective tests were prepared to observe asymmetric coding and symmetric coding performances on both display types. The PSNR of the auxiliary view for the first test is set to be below the threshold level and the PSNR of the second test's auxiliary view is fixed close to the threshold level. Another subjective test is planned to find out the performance of scalability type that is used for auxiliary view on both displays. The assessors are expected to grade 9 test sequences. Before grading, each test sequence and its original is displayed twice. However, their order is selected randomly so that the assessor cannot know which one is the original. The grading is done after the repetitions on a continuous scale in accordance with the DSCQS standard.

4.2. Subjective Tests

Once all participants grade the test sequences, and overall score is calculated for each sequence. This score is normalized between 0 and 100.

The evaluation procedure defines confidence interval, which is calculated based on the standard deviation of the scores. The confidence interval serves both as a safety margin for the scores and as validity check for the user. During the evaluation period a tester is ignored if scored out of safety margins frequently. The detailed information on the evaluation procedure is given in [32].

The zero point of the autostereoscopic display results is adjusted to the zero of the polarized projection test scores by subtracting the difference between them. Larger scores indicate poorer perceived video quality.

In Section 2, we had 3 hypothesis:

Hypothesis 1: As long as the reference view of asymmetric coding is encoded at sufficiently high quality(~38dB) and the auxiliary view is encoded around 33dB for projector display and 31 dB for parallax barrier, asymmetric coding with SNR scaling performs better than symmetric coding. The PSNR rates for the auxiliary view constitutes our threshold value. We have already proved this statement however in this section we present the result of subjective tests that we performed to support our claim.

The first subjective test aims to see the correctness of hypothesis 1. The test sequences are prepared with both asymmetric coding with SNR scaling and symmetric coding at the rates shown in Table 4.1 and 4.2. The left view is the reference view and it is encoded around 38 dB for all test contents and the right view is encoded at 32 dB which is close to our threshold. Table 4.3 presents normalized scores of the subjective test and Figure 4.2 demonstrates the results for better understanding. The results clearly indicate that asymmetric coding with SNR scaling performs better than symmetric coding at the threshold levels.

	PSNR			Bitrates (Mbps)		
	Adile	Flower	Train	Adile	Flower	Train
Asym Left	38.258	37.145	36.815	0.794	1.486	2.037
Asym Right	32.158	31.818	32.339	0.312	0.469	0.761
Sym Left	34.755	35.012	35.163	0.482	1.023	1.432
Sym Right	34.786	35.251	35.146	0.484	0.98	1.375

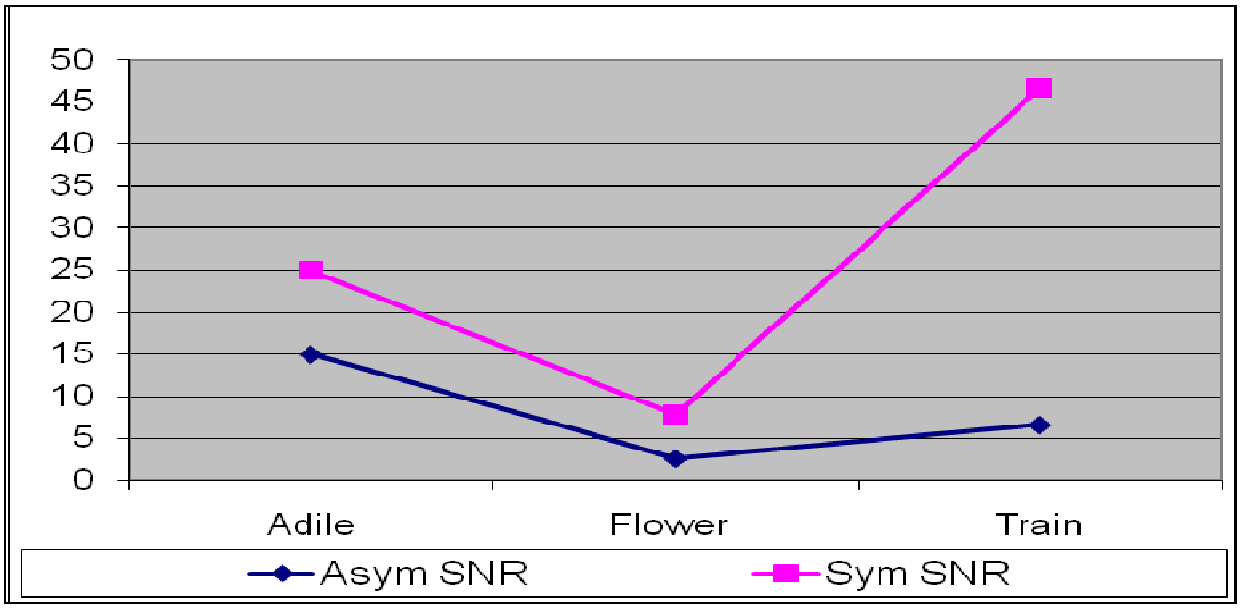
Table 4.1: PSNR and bitrate values for symmetric & asymmetric coding above threshold PSNR (for parallax barrier)

	PSNR			Bitrates (Mbps)		
	Adile	Flower	Train	Adile	Flower	Train
Asym Left	38.258	37.145	36.815	0.794	1.486	2.037
Asym Right	33.394	31.818	32.843	0.312	0.683	0.83
Sym Left	36.174	35.012	35.163	0.595	1.023	1.432
Sym Right	36.176	35.251	35.146	0.598	0.98	1.375

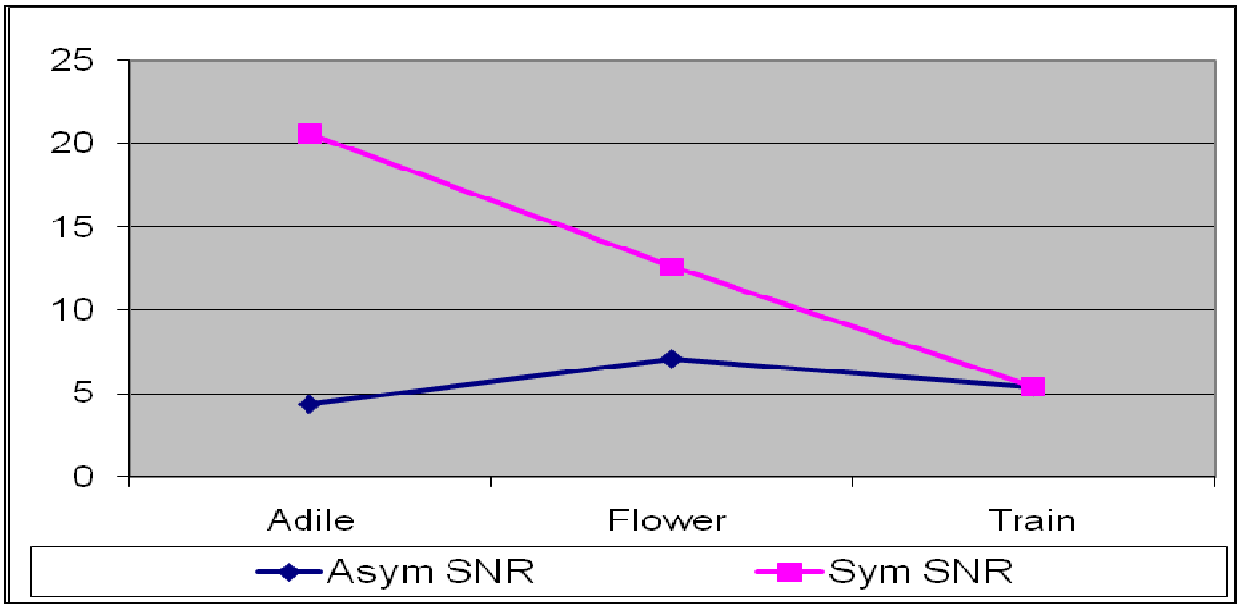
Table 4.2: PSNR and bitrate values for symmetric & asymmetric coding above threshold PSNR (for projector display)

	Parallax			Projector		
	Adile	Flower	Train	Adile	Flower	Train
Asym SNR	15 ± 4	2.6 ± 1	6.6 ± 6	4.4 ± 3	7.1 ± 5	5.4 ± 3
Sym SNR	25 ± 14	7.8 ± 3	47 ± 3	21 ± 10	12.6 ± 7	5.4 ± 6

Table 4.3: Normalized subjective test scores for symmetric & asymmetric coding above threshold PSNR



(a)



(b)

Fig. 4.2: Normalized subjective test scores for symmetric & asymmetric coding above threshold PSNR (a) Parallax, (b) Projector

Hypothesis 2: Below the threshold levels, symmetric coding starts to perform better than asymmetric coding. However as the bitrate constraint becomes tighter which means as the asymmetry between views increase and the total quality decreases, asymmetric coding with spatial scalability achieves better perceived quality degrees than asymmetric coding with SNR scalability. Besides it also starts to perform better than symmetric coding after a second threshold level.

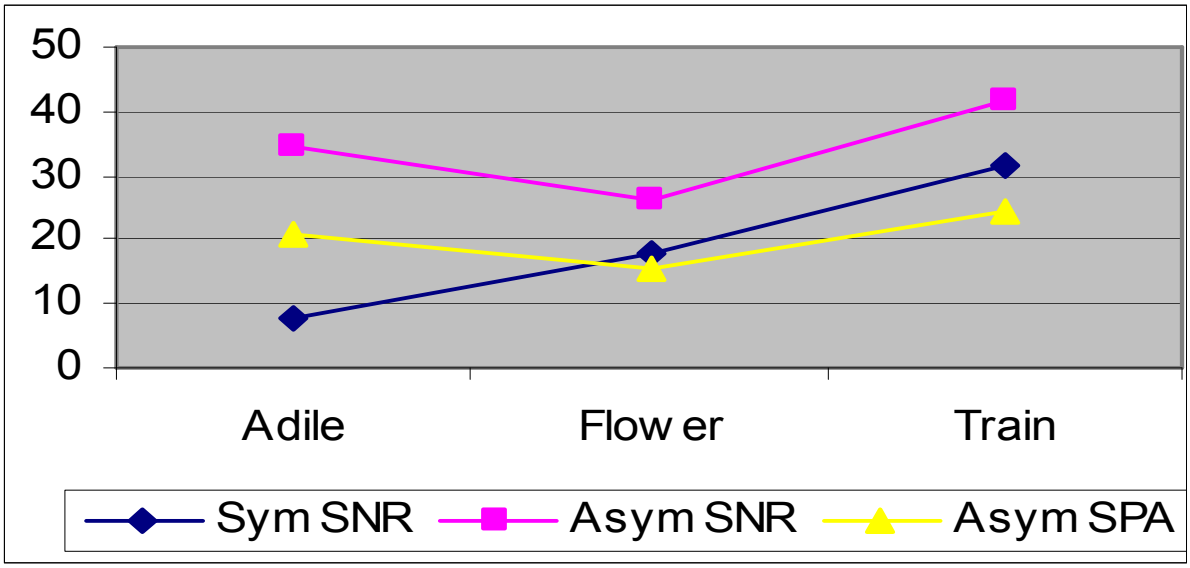
The purpose of the second subjective test is to validate our second hypothesis. This time asymmetric coding with spatial scalability is added to methodologies under investigation and also test sequences have an upper bound bitrate constraint of 1Mbit/sec. The PSNR and bitrate values of the test sequences are shown in Table 4.4. The result of the subjective test is presented in Table 4.5 and they are also depicted in Figure 4.3. The results demonstrates that when the auxiliary view is encoded at 28.6 dB, symmetric coding performs better than all other coding schemes however as the quality of the auxiliary view and symmetric pairs decreases, asymmetric coding with spatial scalability beats symmetric coding for parallax barrier technology. Therefore we expect that there is a second threshold at this point where asymmetric coding performs better than symmetric coding again and projector display's threshold is lower than parallax barrier's threshold. This is again because of the technology that is used in parallax barrier, it tends to conceal the asymmetric distortions.

	PSNR			Bitrate (Mbps)		
	Adile	Flower	Train	Adile	Flower	Train
Asym L	38.258	33.861	32.276	0.794	0.801	0.794
Asym SNR R	29.609	28.485	27.251	0.205	0.221	0.262
Asym SPA R	28.635	27.652	26.587	0.202	0.226	0.254
Sym L	34.755	31.596	29.973	0.482	0.49	0.499
Sym R	34.786	32.426	30.079	0.484	0.524	0.481

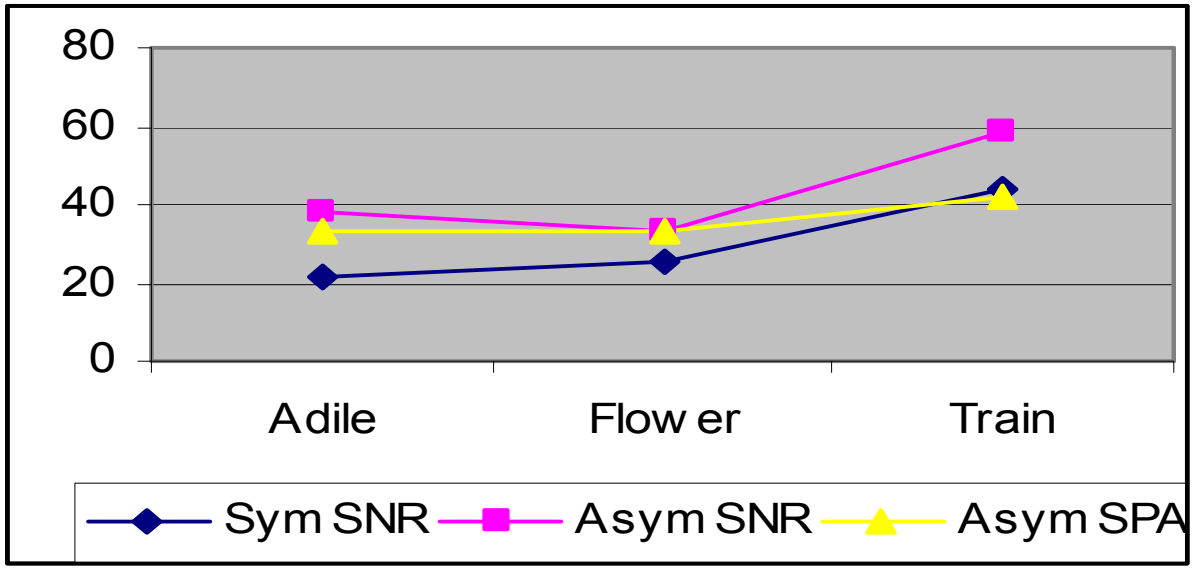
Table 4.4: PSNR and bitrate values for symmetric & asymmetric coding below threshold PSNR

	PARALLAX			PROJECTOR		
	Adile	Flower	Train	Adile	Flower	Train
Asym SNR	34.4 ± 3.6	26.4 ± 3.5	41.4 ± 8.0	38.4 ± 3.3	33.1 ± 2.7	58.1 ± 4.3
Asym SPA	20.6 ± 7.3	15.4 ± 5.6	24.6 ± 6.1	33.4 ± 5.1	33.1 ± 7.1	41.9 ± 6.8
Sym SNR	8 ± 2.9	18.0 ± 6.8	31.7 ± 6	21.6 ± 6	25.3 ± 3.6	43.6 ± 7.2

Table 4.5: Normalized subjective test scores for symmetric & asymmetric coding below threshold PSNR



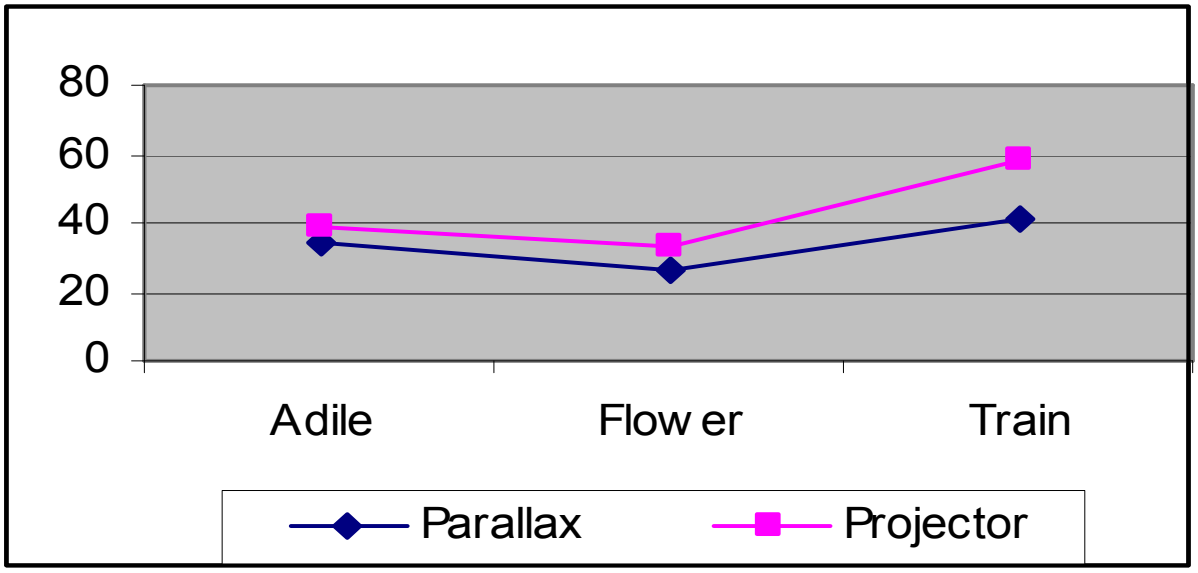
(a)



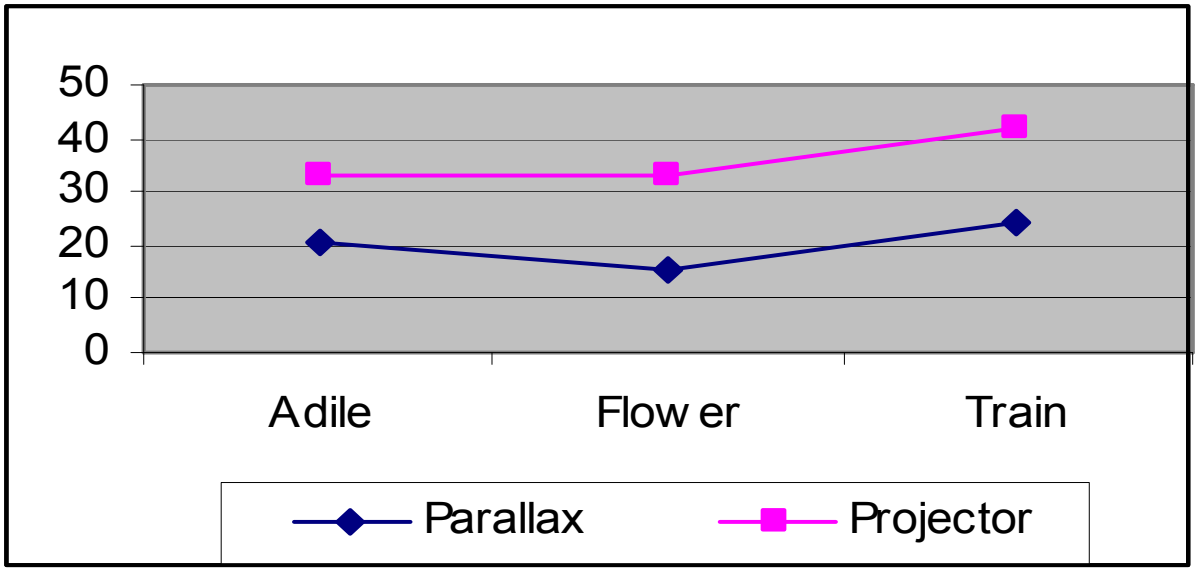
(b)

Fig. 4.3: Normalized subjective test scores for symmetric & asymmetric coding below threshold PSNR (a) Parallax, (b) Projector

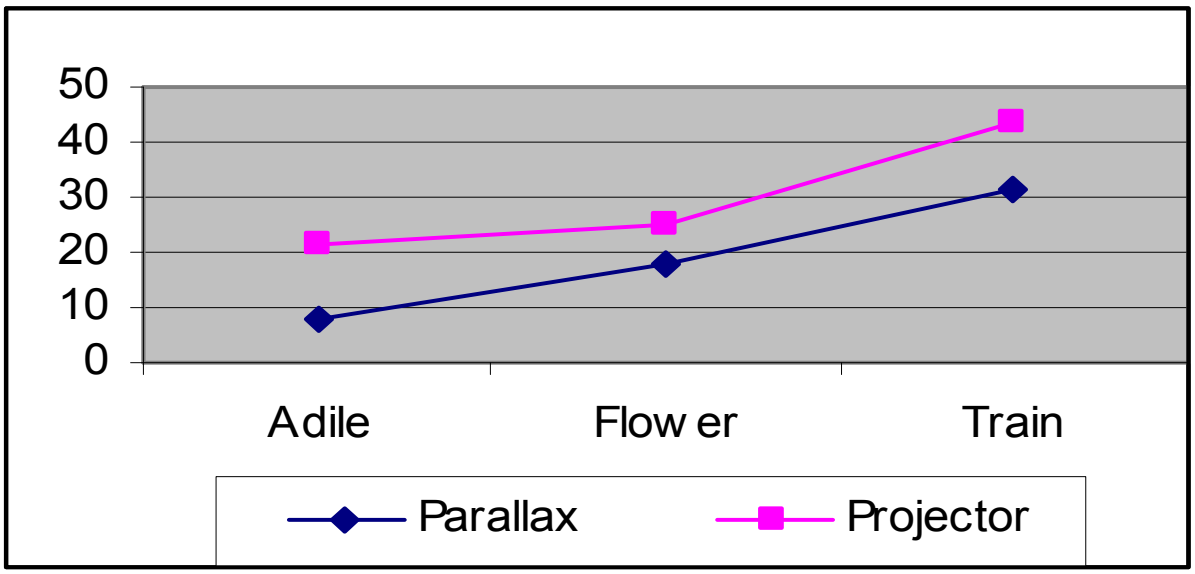
The scores also show the error concealment effect of parallax barrier. Since lower grades means better perceived quality, the grades of test sequences displayed in parallax barrier is lower than projector displays'. Figure 4.4 presents the subjective grades of test sequences for both display technologies for better visualization.



(a)



(b)



(c)

Fig. 4.4: Normalized subjective test scores for parallax barrier & projector display technologies below threshold PSNR (a) Asym SNR, (b) Asym SPA and (c) Sym SNR

4.3 Generation of Quantitative Metric

Peak Signal to Noise Ratio (PSNR) is frequently used and widely accepted as a quantitative metric for assessing the quality of impaired monoscopic video. However, in the stereoscopic case, averaging over the individual PSNR values of the views can yield misleading results. It is well known that HVS is tolerant to views at unequal quality and can neglect quality degradations in single view. For this reason, it is more difficult to correlate the PSNR values of a stereo content with the perceived quality. Additionally, experimental results show that the 3D display technology has an impact on the perceived visual quality especially if views are encoded at different spatial resolutions, which is disregarded if one uses simple averaging of PSNR values to evaluate visual quality.

In our previous works, we propose a simple yet effective formula to better assess the perceived quality of stereoscopic 3D content. Visual tests suggest that if PSNR of both views are close to each other then the perceived 3D video quality (Q) can be better reflected with the higher PSNR value. This observation is captured by a ratio test in Eqn. 1. The view with lower quality starts to degrade perceived 3D quality of the content as the difference between PSNRs grows. In this case, a weighting parameter (β) is used similar to [5]

$$Q(\alpha,\beta)=\begin{cases} PSNR_{HI} & \dfrac{PSNR_{LOW}}{PSNR_{HI}}<\alpha \\ (\beta)PSNR_{HI}+(1-\beta)PSNR_{LOW} & otherwise \end{cases} \quad (1)$$

We proposed that β depends on two criteria, the display technology and spatial resolution of the auxiliary view. In an autostereoscopic system, the emitted light is either blocked by parallax barrier or diverted by a lens array. That operation decreases the effective spatial resolution of the 3D display. This factor can conceal the blurring effect caused by upsampling a spatial scaled auxiliary view. However, projector based systems display the content at full resolution. Therefore, the blurring artifact is exposed clearly. The threshold value (α) is observed to be 0.85. Table 4.6 provides the suggested β parameter based on our test results.

		Auxiliary View Resolution	
		Full	Scaled
Display Resolution	Full	2/3	1/2
	Scaled	2/3	2/3

Table 4.6: Suggested β values

However our final result show us that this formula strongly relies on the quality of the reference view and the content. Although our first threshold level is well presented by this formula, it is unable to explain why symmetric coding performs better than asymmetric coding between the two thresholds and also why the asymmetric coding performs better than symmetric coding after the second threshold level. Therefore we better explain the performances of coding methods by thresholds rather than a mathematical formulation.

Chapter 5

CONCLUSION

5.1 Summary

According to HVS theory, human eye can compensate the absence of high frequency information of one eye, as long as that information is present at the other view. Therefore we expect that asymmetric coding needs to achieve better perceived quality degrees under the same bandwidth limitations than symmetric coding. However the degree of asymmetry is a crucial factor since HVS has limitations and we can not assume that a sufficient 3D video quality can still be perceived by the viewers when one of the views is extremely or entirely distorted.

In this work, we concentrated on finding the threshold levels of distortions until where asymmetric coding succeeds better perceived quality degrees than symmetric coding. Additionally we tried to answer the question of which scalability option should be preferred with asymmetric coding above and below these thresholds. Our subjective test results show that as long as the reference view is encoded at sufficiently high quality (~38dB) and the auxiliary view is encoded at 32 dB, asymmetric coding with SNR scaling performs better than asymmetric coding with spatial scaling and symmetric coding. Therefore we set our first threshold level to 32dB (33 dB for projector display, 31 dB for parallax barrier). We have also observed that viewers tend to give higher subjective test grades to test sequences displayed on parallax barrier than projector display. This behavior is induced by the technology of the displays with parallax barrier. The perceived spatial resolution is halved in parallax barrier technology, therefore the distortions on the auxiliary view is reduced. As we continue to decrease the rate of the generated test streams, the SNR scaled auxiliary view of the asymmetric coded test sequence starts to suffer blockings whereas spatial scaled one starts to suffer from blur. At the levels well below the threshold, we observe that spatial scaled asymmetric coding starts to achieve better perceived quality levels than both symmetric coding and SNR scaled asymmetric coding. The PSNR level where this event occurs points a second threshold level where asymmetric coding should be preferred over symmetric coding. Between these two levels, we suggest to use symmetric coding for rate adaptation purposes.

There are many multi description generation models. We choose 4 different models and observe their R-D performances under varying channel packet loss probability rates. 3 of these models uses the same amount of bandwidth, in other words the amount of redundancy inside each of these models are equal. However extra redundancy is added to develop 4th model in order to increase the error resilience of the resulting descriptions. This redundancy is around %40 of the total rate of descriptions. The channel simulation results show that there isn't a significant difference between the first 3 models in terms of R-D performance under %3, %5 and %10 channel packet loss probabilities. Nevertheless they are always have better performance than model 4 even when the packet loss probabilities of the two channels are at %20. In other words, model 4 is infeasible to use under networks with packet loss probability rates up to %20. Still model 4 achieves better performance in terms of standard deviation rates which means model 4 shows more steady behavior under varying network congestion situations than other models.

5.2. Future Work

As the transportation of 3D video becomes popular in the near future, rate allocation techniques will be more crucial. Also people will need to transmit multi-view video (more than two views) which means the congestions will frequently occur. Our work targets stereoscopic video and although asymmetric coding strategy is sufficient for rate adaptation for today's network infrastructure, it won't be enough for multi-view video transportation. Therefore we need to expand our results by utilizing asymmetric MVC to our subjective tests. Moreover we should compare subjective test scores of asymmetric coding and asymmetric MVC with scores of approaches using depth map.

There are many models for multi description coding and in our work we have used only 4 of them. First of all, we should investigate R-D performances of other models. Additionally we need to prepare suitable subjective tests to visualize the performance of different models under varying channel packet loss probability rates.

Bibliography

[1] G. I Conklin, G.S. Greenbaum, K. O. Lillevold, A.F. Lippman, and Y. A. Reznik, "Video coding for streaming media delivery on the internet," *IEEE Trans. On Circ. and Syst. For Video Techn., March 2001, vol 11, no. 3, pages: 269-281.*

[2] W. Tan and A. Zakhor, "Video multicast using layered FEC and scalable compression," IEEE Trans. *On Circ. and Syst. For Video Techn., March 2001, vol. 11, no. 3, pages: 373-387.*

[3] M.G.Perkins, "Data compression of stereo pairs," *IEEE Trans. on Communications*, vol. 40, no. 4, pp.684-696, April 1992.

[4] K. Mueller, P. Merkle, H. Schwarz, T. Hinz, A. Smolic, T. Oelbaum, "Multi-view video coding on H.264/AVC using hierarchical B-frames," *in Proc. of the Picture Coding Symposium*, Beijing, China, Apr. 2006.

[5] Ravi Krishnamurthy, Bing-Bing Chai, Hai Tao, and Sriram Sethuraman, "Compression and transmission of depth maps for image-based rendering," *International Conference on Image Processing, ICIP 2001, vol. 3, pages: 828-831.*

[6] Krzysztof Klimaszewski, Krzysztof Wegner, Marek Domanski, "Distortions of synthesized views caused by compression of views and depth maps," *3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video, 2009 pages: 1 - 4*

[7] Philipp Merkle, Ajoscha Smolic, Karsten Müller, and Thomas Wiegand, "Multi-view video plus depth representation and Coding," *International Conference on Image Processing, ICIP 2007, vol. 1, pages: 201-204.*

[8] A. Aksay, S. Pehlivan, E. Kurutepe, C. Bilen, T. Ozcelebi, G Bozdagi Akar, M. R Civanlar,and A. M. Tekalp, "End-to-end stereoscopic video streaming with content-adaptive rate and format control," Signal Proc.: Image Communication, vol 22, pp 157-168, Feb 2007.

[9] N. Ozbek and A. M. Tekalp, "Unequal inter-view rate allocation using scalable stereo video coding and an objective stereo video quality measure," in *Proc. IEEE ICME*, Germany, June 2008.

[10] C. Fehn, P. Kauff, S. Cho, H. Kwon, N. Hur, and J. Kim, "Asymmetric coding of stereoscopic video for transmission over T-DMB," in *Proc. 3DTV-CON*, Kos, Greece, May 2007.

[11] L. Stelmach, W. Tam, D. Meegan, and A. Vincent, "Stereo image quality: effects of mixed spatio-temporal resolution," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 10, no. 2, pp. 188–193, 2000.

[12] Nukhet Ozbek, A. Murat Tekalp and Turhan Tunalı, "Rate allocation between views in Scalable Stereo Video Coding using an objective stereo video quality measure," *Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference.*

[13] ITU-R Rec.BT.500-11, "Methodology for the subjective assessment of the quality of television pictures," 2002.

[14] Truang Cong Thang, Jae-Gon Kim, Jung Won Kang, Jeong-Ju Yoo, "SVC adaptation: Standard tools and supporting methods," *Signal Processing: Image Communication vol. 24, issue 3, March 2009, pages 214-228.*

[15] L.B. Stelmach, W.J. Tam, D.V. Meegan, A. Vincent, and P. Corriveau, "Human perception of mismatched stereoscopic 3D inputs" *Proc.* IEEE Int. Conf. Image Proc. (ICIP), pp:5 – 8, 2000.

[16] J. Reichel, H. Schwarz, M. Wien (eds.), "Scalable Video Coding – Working Draft 1, " Joint Video Team (JVT), Doc. JVT-N020, Hong-Kong, China, Jan 2005.

[17] Philipp Merkle, Aljoscha Smolic, Karsten Müller, and Thomas Wiegand, "Efficient prediction structures for multiview video Coding," *IEEE Trans. On Circuits and Systems for Video Tech., Vol. 17, No. 11, November 2007.*

[18] Nukhet Ozbek, A. Murat Tekalp, "Scalable multi-view video coding for interactive 3-D TV," *Multimedia and Expo, 2006 IEEE International Conference on, 9-12 July 2006, pages:213-216*

[19] Ying Chen, Shujie Liu, Ye-Kui Wang, Miska M. Hannuksela, Houqiang Li, Moncef Gabbouj, "Low-complexity asymmetric multiview video coding," *Multimedia and Expo, 2008 IEEE International Conference on, June 23 2008- April 26 2008, Pages: 773-776.*

[20] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the scalable video coding Extension of the H.264/AVC Standard", IEEE *trans. On circuits and systems for video delivery,* vol. 17, no. 9, September 2007.

[21] E. Kohler, M. Handley, and S. Floyd, "Datagram congestion control protocol (DCCP)." Internet Engineering Task Force, RFC 4340, March 2006.

[22] M. Handley, S. Floyd, J. Padhye, and J. Widmer, "TCP friendy rate control (TFRC): Protocol specification," IETF, RFC 3448, January 2003.

[23] Y. Wang, A. R. Riebman, and S. Lin "Multiple description coding for video delivery," *Proc. Of IEEE*, vol. 93, no. 1, January 2005.

[24] V. K. Goyal, "Multiple description coding: Compression meets the network," *IEEE Signal Processing Mag.*, vol. 18, pp. 74-93, Sep. 2001.

[25] E. Akyol, A. M. Tekalp, M. R. Civanlar, "A flexible multiple description coding framework for adaptive peer-to-peer video streaming, " *IEEE Journal of Selected Topics in Signal Process.,* vol. 1,no. 2, August 2007.

[26] D. Comas, R. Singh, A. Ortega , and F. Marques, "Unbalanced multiple description video coding based on a rate-distortion optimization," *EUROSIP J. Appl. Signal Process.*, vol. 2003, no. 1, pp. 81-90, Jan. 2003.

[27] A. Norkin, A. Aksay, C. Bilen, G. B. Akar, A. Gotchev, and J. Astola, "Schemens for multiple description coding of stereoscopic video," *MRCS 2006,LNCS 4105,*Springer, pp. 730-737, Sept. 2006.

[28] A. Norkin, M. O. Bici, A. Aksay, C. Bilen, A. Gotchev, G. Bozdagi Akar, K. Egiazarian, and J. Astola, " Multiple description coding and its relevance to 3DTV". In Haldun M. Ozaktas, and Levent Onural (Eds.), *Three-Dimensional Television: Capture, Transmission , and Display (ch.11*), Springer Verlag, 2007.

[29] G. Saygili, G. Gurler, A. M. Tekalp, "3D Display Dependent Quality Evaluation and Rate Allocation Using Scalable Video Coding", *Proc.* IEEE Int. Conf. Image Proc. (ICIP), pp:717 – 720, 2009.

[30] T. B. Abanoz and A. M. Tekalp, "SVC-based scalable multiple description video coding and optimization of encoding configuration", *Elsevier Science Image Communication*, Vol. 24, Issue 9, pp. 691-701, October 2009.

[31] D. V. Meegan, L. B. Stelmach and W. J. Tam, "Unequal weighting of monocular inputs in binocular combination: implications for the compression of stereoscopic imagery," *Journal of Experimental Psychology:* Applied Vol 7(2), Jun 2001, 143-153.

[32] T. Wiegand, G. J. Sullivan, J. Reichel, H. Schwarz, and M.Wien, "Joint Draft 11 of SVC Amendment, Joint Video Team," *Doc. JVT-X201*, Jul. 2007.

VITA

Görkem Saygılı was born in İzmir on September 12 1985. He received his B.S. degree in Electrical and Electronics Engineering from Bilkent University, Ankara in 2007. From January 2008 to January 2010, he worked as a teaching and research assistant in Koç University, Turkey. He published chapters about his research to ICIP 2009 and also submitted a paper to ICIP 2010.