Practical SVC video streaming over partially realiable transport

by

Jaideep More

Master Thesis

Software Engineering Chair
Faculty of Natural Sciences and Technology I
Department of Computer Science
Saarland University

Supervisor
Prof. Anja Feldmann, Ph.D.

Reviewers
Prof. Anja Feldmann, Ph.D.
Dr. Tobias Fiebig

June 26, 2025

UNIVERSITÄT DES SAARLANDES

Declaration of Authorship

Eidesstattliche Erklärung

Ich erkläre hiermit an Eides Statt, dass ich die vorliegende Arbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet habe.

Statement in Lieu of an Oath

I hereby confirm that I have written this thesis on my own and that I have not used any other media or materials than the ones referred to in this thesis.

Einverständniserklärung

Ich bin damit einverstanden, dass meine (bestandene) Arbeit in beiden Versionen in die Bibliothek der Informatik aufgenommen und damit veröffentlicht wird.

Declaration of Consent

I agree to make both versions of my thesis (with a passing grade) accessible to the public by having them added to the library of the Computer Science Department.

Datum/Date: ______________________________

Unterschrift/Signature: ______________________________

SAARLAND UNIVERSITY
Software Engineering Chair
Department of Computer Science

Abstract

Practical SVC video streaming over partially realiable transport
by Jaideep More

Scalable Video Coding (SVC) employs layered coding techniques to encode video, consisting of a base layer for base video quality and enhancement layers for improved quality.

Traditional Dynamic Adaptive Streaming over HTTP (DASH) with Advanced Video Coding (AVC) faces suboptimal quality adaptation. SVC addresses this challenge by enabling in-segment quality adjustments due to its layered structure, allowing for smoother quality degradation.

A detailed study of the SVC layered coding structure shows that base-layer data is essential for continuous playback, whereas enhancement-layer data, while beneficial, is not strictly required.

Our findings indicate that occasional drops in enhancement-layer data has minimal impact on user Quality of Experience (QoE).

Building on these insights, we propose an approach that integrates SVC with a partially reliable transport protocol to optimize video streaming.

By prioritizing base-layer data over reliable transport channels while opportunistically transmitting enhancement layers, our method ensures graceful quality adaptation in response to network fluctuations.

Acknowledgements

TODO: Complete acknowledgements

Content

TODO: We want to put an abbreviations section

Dedicated to WHATEVER.

Chapter 1

Introduction

1.1 Motivation

Discuss advantages of SVC over AVC. Discuss the potentional of SVC and PRT. We will need references here.

1.2 Research Questions

1.2.1 RQ1: Feasibility of Virtual Quality Levels in SVC

1.2.2 RQ2: Frame Importance using SVC Dependencies

1.3 Contributions

1.4 Outline

Chapter 2

Background

2.1 H.264/SVC Basics

Basic introduction of SVC features

2.2 Temporal Scalability

Hierarchical prediction structure for temporal scalability

2.3 Quality Scalability

Inter layer prediction

2.4 Spatial Scalability

Inter layer prediction

2.5 SVC Bitstream Structure

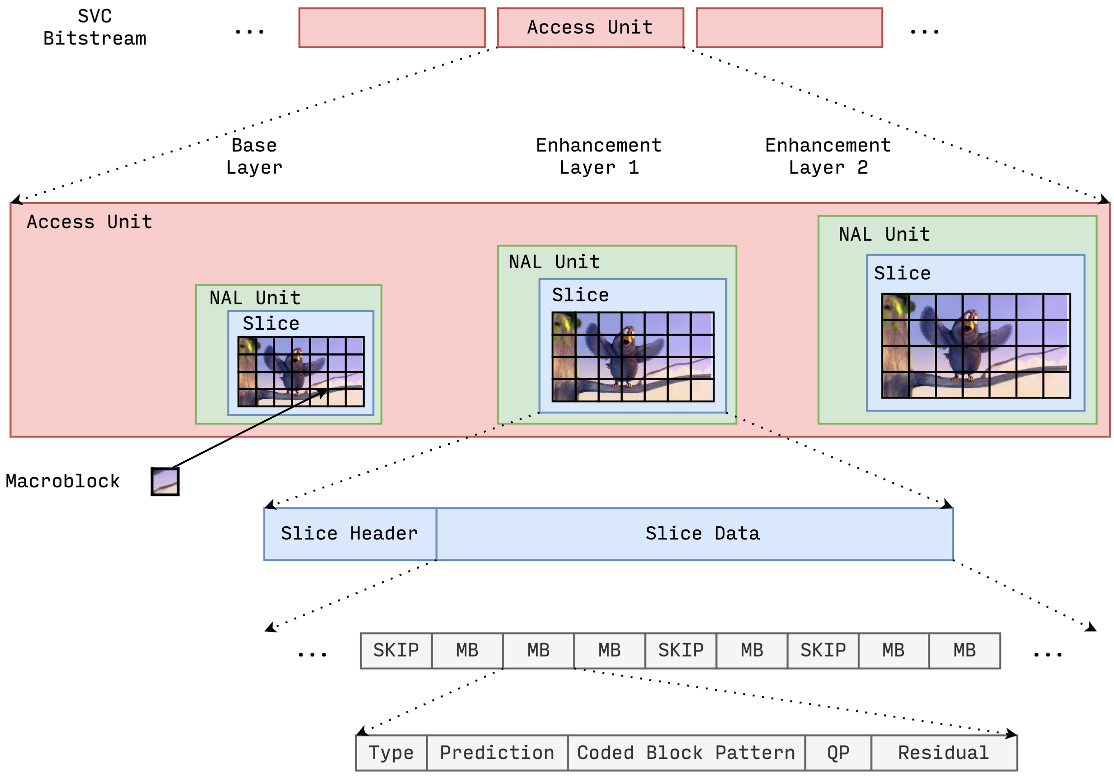


FIGURE 2.1: Bitstream Structure

2.6 Quality Metrics (SSIM, PSNR)

Chapter 3

Methodology

3.1 Constructing Virtual Quality Levels by Skipping Frame Data

In this section, we investigate whether intermediate quality levels—beyond the explicitly defined scalability layers—can be created by partially omitting enhancement-layer data.

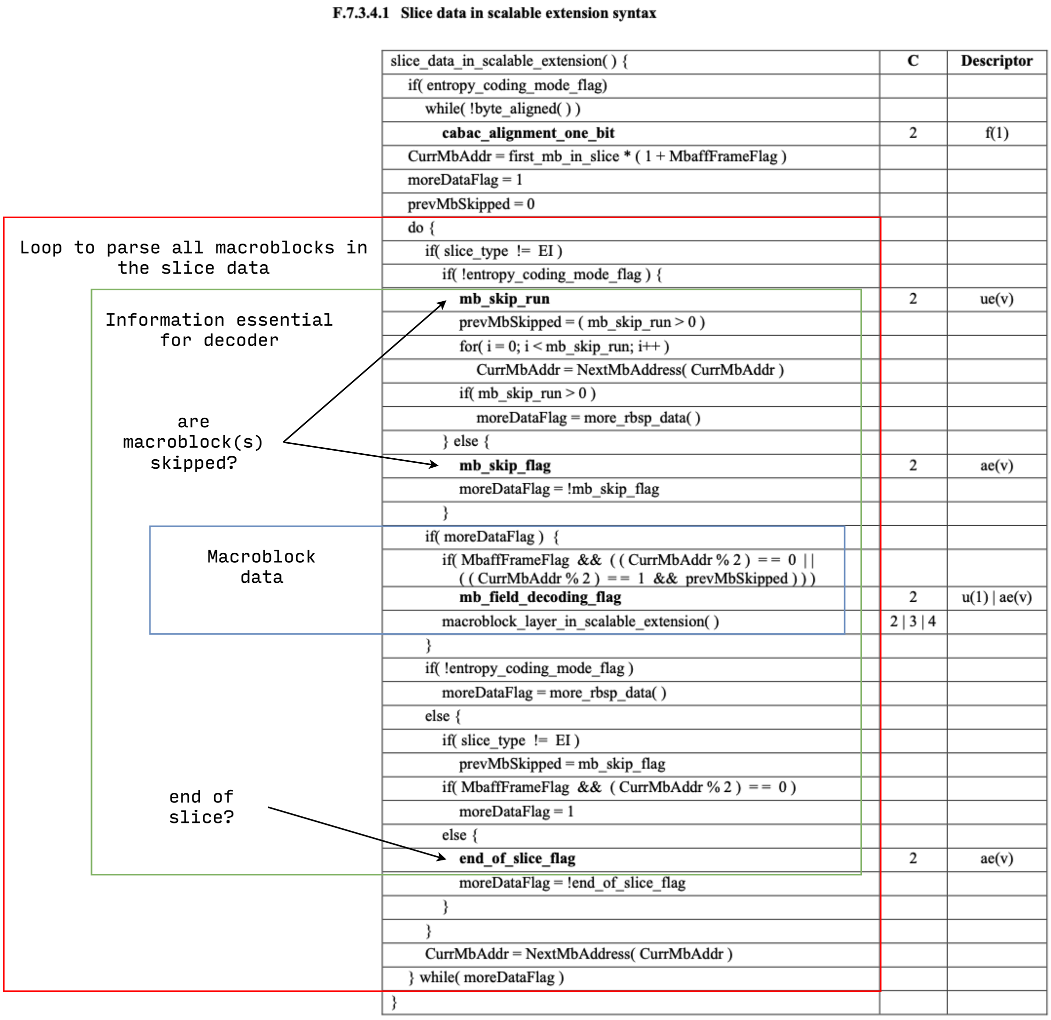
As described in Section 2.5, the SVC bitstream is composed of access units, each containing slices corresponding to different enhancement layers. Each slice includes a slice header and a slice data section, which holds a list of macroblocks. These macroblocks are the smallest coding units in the video and contain motion, prediction, and residual data used by the decoder to reconstruct the frame.

The structure of slice data for SVC bitstream is shown in Figure 3.1. Within the `slice_data_in_scalable_extension()` syntax, the `mb_skip_flag` is used to indicate that a macroblock is skipped.

When a macroblock is skippped, motion vectors or residuals are not explicitly coded. Thus, `macroblock_layer_in_scalable_extension()` which contains the macroblock information is empty.

This skipping mechanism applies only to P and B macroblocks, which use inter prediction. I macroblocks, which rely solely on intra prediction, cannot be skipped and must always be fully encoded.

The macroblock skip mode in H.264 is primarily intended to improve compression efficiency by exploiting spatial and temporal similarities between frames. When the content of a macroblock closely matches a corresponding region in a reference frame, the encoder may mark it as skipped. Instead of encoding new data, the decoder reconstructs the macroblock using prediction. This often requires only a single bit to signal skipping via the `mb_skip_flag`, reducing the bitrate with minimal impact on visual quality.



F.7.3.4.1 Slice data in scalable extension syntax

slice_data_in_scalable_extension( ) {	C	Descriptor
if(entropy_coding_mode_flag)		
while(!byte_aligned( ))		
cabac_alignment_one_bit	2	f(1)
CurrMbAddr = first_mb_in_slice * (1 + MbaffFrameFlag)		
moreDataFlag = 1		
prevMbSkipped = 0		
do {		
if(slice_type != EI)		
if(!entropy_coding_mode_flag) {		
mb_skip_run	2	ue(v)
prevMbSkipped = (mb_skip_run > 0)		
for(i = 0; i < mb_skip_run; i++)		
CurrMbAddr = NextMbAddress(CurrMbAddr)		
if(mb_skip_run > 0)		
moreDataFlag = more_rbsp_data( )		
} else {		
mb_skip_flag	2	ae(v)
moreDataFlag = !mb_skip_flag		
}		
if(moreDataFlag) {		
if(MbaffFrameFlag && ((CurrMbAddr % 2) = = 0 || ((CurrMbAddr % 2) = = 1 && prevMbSkipped)))		
mb_field_decoding_flag	2	u(1) | ae(v)
macroblock_layer_in_scalable_extension( )	2 | 3 | 4	
}		
if(!entropy_coding_mode_flag)		
moreDataFlag = more_rbsp_data( )		
else {		
if(slice_type != EI)		
prevMbSkipped = mb_skip_flag		
if(MbaffFrameFlag && (CurrMbAddr % 2) = = 0)		
moreDataFlag = 1		
else {		
end_of_slice_flag	2	ae(v)
moreDataFlag = !end_of_slice_flag		
}		
}		
CurrMbAddr = NextMbAddress(CurrMbAddr)		
} while(moreDataFlag)		
}		

Figure 3.1: Slice Data Syntax in H.264/SVC

In our approach, we take advantage of this mechanism to selectively skip macroblocks in the enhancement layers. For each macroblock marked for omission, we set the `mb_skip_flag` and remove the associated `macroblock_layer_in_scalable_extension()` data from the bitstream. This enables fine-grained control over which parts of the enhancement data are retained, allowing us to simulate continuous quality variations rather than relying solely on predefined SVC layer switching.

These selectively modified bitstreams result in what we refer to as *virtual quality levels*—intermediate representations that provide smoother adaptability and more flexible quality trade-offs. In the following sections, we evaluate the visual impact of these virtual quality levels using objective metrics such as SSIM, PSNR, and VMAF, and assess their potential to improve streaming performance under network constraints.

3.2 Motion and Residual and Upsampling

Chapter 4

Evaluation

4.1 Experimental Setup

4.1.1 Test Videos and Encodings

We selected four publicly available test videos from the Xiph.org video dataset 1, varying in content and motion complexity. All videos are available in the raw YUV format (YUV420p, 1080p, 24 fps), therefore PSNR and SSIM values of the encoded videos can be calculated accordingly.

Title	Length	Type
blue_sky	217 frames	Movie
pedestrian_area	375 frames	Animation
rush_hour	500 frames	Animation
riverbed	250 frames	Animation

Table 4.1: Video Dataset Used for Evaluation

Each video was encoded into three SVC variants, with each variant using a different type of scalability, to independently study the impact of enhancement layer loss for each type: **Quality**, **Temporal**, and **Spatial**.

- `ONLY_QUALITY` variant uses one base layer and two quality enhancement layers. All layers are encoded at a fixed resolution of 360p. The base layer employs coarse quantization to provide a low-quality baseline, while the enhancement layers use SNR scalability to progressively improve visual quality.

- `ONLY_SPATIAL` variant uses spatial scalability, where each layer increases the resolution of the video. The base layer is encoded at 360p, followed by enhancement layers at 720p and 1080p.
- `ONLY_TEMPORAL` variant uses temporal scalability, where all layers are encoded at a fixed resolution of 360p, and each layer increases the frame rate of the video. The base layer is encoded at 6 fps, followed by enhancement layers at 12 fps and 25 fps.

Configuration	Layers	Resolution	Frame Rate(fps)	Bitrate
	BL	360p	25	–
`ONLY_QUALITY`	EL1	360p	25	Quality
	EL2	360p	25	Quality
	BL	360p	25	–
`ONLY_SPATIAL`	EL1	720p	25	Spatial
	EL2	1080p	25	Spatial
	BL	360p	6	–
`ONLY_TEMPORAL`	EL1	360p	12	Temporal
	EL2	360p	25	Temporal

Table 4.2: Layer Structure of Test Variants

4.1.2 Evaluation Methodology

Our evaluation consists of four main stages: generating layered bitstreams, applying controlled degradation, decoding and assessing video quality, and aggregating the results.

TODO: we need a diagram to show this methodology

Figure 4.1: Evaluation pipeline overview

We begin by creating separate scalable video bitstreams, each representing a different enhancement configuration. The enhancement configuration is shown in Table 4.3. These configurations allow us to evaluate how the presence of additional enhancement layers affects video quality under loss.

Stream Id	Stream Configuration
`BL`	BL
`BL_EL1`	BL + EL1
`BL_EL1_EL2`	BL + EL1 + EL2

Table 4.3: Bitstreams extracted for evaluation

To simulate unreliable network transmission, we randomly skip macroblocks from the top-most enhancement layer in each configuration. We vary the amount of skipped macroblocks from 0-100%, increasing in steps of 10%.

Each degraded version of the bitstream is decoded using the JSVM Decoder v9.19. We then evaluate the quality of the reconstructed video using two commonly used objective metrics: Structural Similarity Index Measure (SSIM) and Peak Signal-to-Noise Ratio (PSNR). The original, unaltered video serves as the reference for comparison.

To account for the randomness of macroblock drops, each experiment is repeated multiple times per drop percentile. We then average the SSIM and PSNR scores across these runs to get stable quality estimates. Finally, we plot the average scores against the drop percentiles for each bitstream configuration

4.2 RQ1: Feasibility of Virtual Quality Levels in SVC

Discuss visual quality with dropped data

4.3 RQ2: Frame Importance using SVC Dependencies

Discuss visual quality with dropped data based on frame importance

4.4 Discussion of Findings

Chapter 5

Related Work

TODO: complete this section

Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet.

Citations look as follows: [1].

Figure A.1 shows an example of a Figure.

Table A.1 shows an example of a Table.

Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea

This is a figure

Figure 5.1: Example of a figure

a	b
c	d

Table 5.1: Example of a table

rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet.

Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet.

Chapter 6

Conclusion

TODO: complete this section

List of Figures

List of Tables

Bibliography

[1] N. G. De Bruijn. Lambda calculus notation with nameless dummies, a tool for automatic formula manipulation, with application to the church-rosser theorem. *INDAG. MATH*, 34:381–392, 1972.

Appendix A

This is my appendix

Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet.

Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet.

Citations look as follows: [1].

Figure A.1 shows an example of a Figure. Table A.1 shows an example of a Table.

Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur

This is a figure

FIGURE A.1: Example of a figure

a	b
c	d

Table A.1: Example of a table

sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet.

Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet.