Screen Content Coding with VP9

by

Petru Lupascu

Bachelor Thesis in Electrical and Computer Engineering

Prof. Dr.-Ing Werner Henkel
Steffen Schulze(LMI)
Name and title of the supervisors

Date of Submission: May 28th, 2019

Jacobs University — Focus Area Mobility

Contents

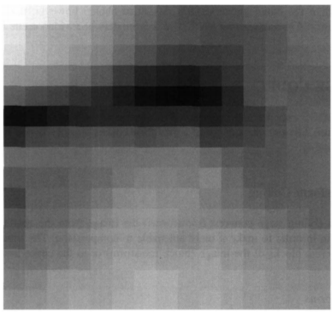
List of Figures

1 Introduction

From the yearly days of digital television, video compression techniques gained increased attention, mainly due to bandwidth always being an expensive asset, fact which is relevant. Throughout the years, video coding techniques played a vital role in reducing the size of the video sequences without significant alteration of its quality. In parallel with advancement in computer performance, video coding allowed for services such as video telephone and digital television to be more accessible, which in turn increased the demand. As a consequence, the development of video coding techniques was incentivised. Straight Forward Pulse Code Modulation(PCM) was one of the first attempt on coding video signals at around 140 Mbits/s. Since then video coding have gone a massive progress, modern codecs being able to code Video Signals as low as 9 Mbits/s for HDV format. A newer generation codec targets to achieve the same performance as the previous generation one at the half bit rate. This is done at the expensive of increasing complexity. Most of coding techniques require hardware implementations for optimized performance which makes standardization essential to ensure compatibility with as large amount of devices as possible.[1]

An Image is a projection of a 3-D scene, characterized by depth, texture and illumination, onto a 2-D plane characterized by texture and illumination without depth information [2, p. 5]. It may be defined as a 2 dimensional signal $f(x, y)$, where x and y are spatial coordinates and f is the intensity at that point, when x, y and f are finite we call this image a Digital Image [3, p. 1]. Therefore, a Video represents a sequence of images over a period of time and can be defined as $f(x, y, t)$, where x, y, f are spatial and intensity values and t is the time. For the sake of simplicity we will call the two dimensional point a pixel and its intensity pixel value and each image in a video sequence frame. Furthermore, an image can be characterized in terms of its resolution and colour format, for the video, additionally there is length. The resolution commonly describes the amount of pixels present in the image, for example: 740x480. The colour format represents a typical arrangement of colours in an image such as Grayscale where the pixels value represents the light intensity(luminance) information, commonly 0 to 255 for an 8 bit image. Other important colour formats are the YUV and RGB, where the image is divided into three sub planes containing luminance, red chrominance and blue chrominance values for YUV and Red, Green and Blue colour intensity values for RGB. Usually 8 bits values per sub plane pixel are used. Generally, all the parameters mentioned depend on the particular application. However, in most of cases, the amount of data required to store or transmit a video or an image tend to be very large. A two-hour Standard Definition(SD) 720x480x24 bits per frame movie, displayed at 30 frames per second must be accessed at $31, 104, 000$ bytes/sec and would require roughly $224GB$ of data and it gets much larger in case of High Definition (HD) videos where the resolution is 1920x1080x24 [3, p. 525-526]

It is clear that storing video data in it's raw form is extremely inefficient, deeming a compression scheme necessary. Such a compression scheme is commonly referred as a codec. Due to the fact that a video tends to have both high spacial redundancy across a frame and temporal redundancy across multiple frames. A group of neighbouring pixels tends to have the same or similar pixel intensity values and can be present in multiple frames across a video sequence. This allows the compression scheme to be optimized beyond typical source coding schemes such as Arithmetic Coding. Therefore, a video codec will efficiently decorelate a video in attempt to remove spacial and temporal redundancies and then perform entropy coding.



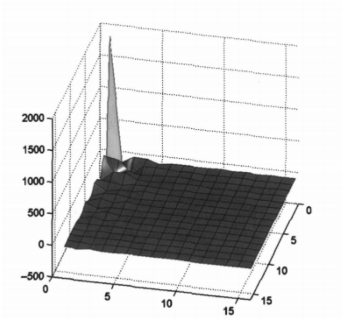


Figure 1: Typical video codec block diagram[2, p. 44]

Since a video is just a sequence of frames one might want to compress each frame indvidually by decorelating the image by aplying a 2-D Transform such as Karhunen-Loeve transform(KLT) to blocks of the image. KLT has the nice property that it's coeffictionts are decorelated and the energy of the block is packed into the minimal amount of coeffictionts. However it is computationally inefficient since the functions required to out the transform must be calculated in advance and transmitted to the decode, rendering its use impractical. Another transform that performs nearly as well as the KLT, but is much more conputationally efficient, is the on Discrete Cosine Transform (DCT). The DCT represents a discrete signal in terms of a sum of cosines of different frequencies and the energy is concentrated at the few top left coeffictionts.

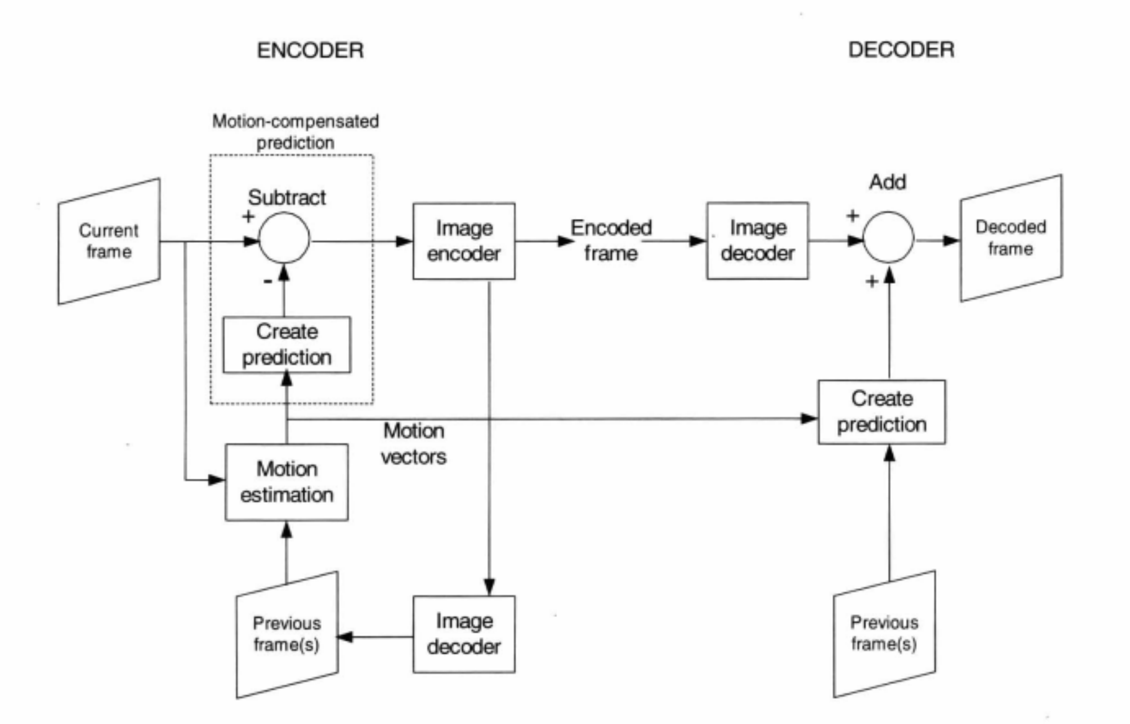


Figure 2: Typical video codec block diagram[2, p. 44]

2 Background and literature review

Bibliography

References

[1] M. Ghanbari, *Standard codecs : image compression to advanced video coding*. Institution of Engineering and Technology, 2011, ISBN: 9780863419645.

[2] I. Richardson, *Video codec design : developing image and video compression systems*. Wiley, 2002, ISBN: 9780471485537.

[3] R. Gonzalez, *Digital image processing*. Prentice Hall, 2008, ISBN: 9780131687288.

Acknowledgements

I would like to express my sincere gratitude to my thesis supervisor Prof. Werner Henkel for all the guidance and support throughout the research process and for providing me with all the necessary facilities and equipment. I am also very thankful to Mr. Uwe Pagel for all his technical support and expertise and for his help with extra required materials. Lastly, I am grateful to my parents, Sokol Shyti and Ermira Shyti, and brother, Arbi Shyti, for all their encouragement along the journey and for proofreading my thesis.