

# Abstract

Determining perceptually irrelevant and redundant information from human point of view is one of the fundamental problems today that is limiting the performance of current video compression algorithms. The performance of the existing video compression standards is based on minimizing the cumulative sum of objective distortion, namely mean squared error (MSE), measured for each pixel. Recently there have been quite a few advancements made to understand human visual models and apply them for a compact representation at very low bitrates. However, most of these approaches offer advantages over a very limited range of input sequences using predefined models for analysis of static scene, human head, and human body.

The existing video compression standards typically aim to increase the spectral flatness measure of the residue signal, by increasing the number of both spatial and temporal predictors. With the increase in the choices of predictors, the corresponding bits, required to convey the choice of the predictor to the decoder, also increases. This mandates the need for jointly optimizing the distortion and the required side information for a given quantization factor using special rate distortion measures. This thesis is aimed at suggesting alternative solution of removing perceptual redundancy without increasing the number of predictors using two approaches. The first one is to increase the spectral flatness measure by removing perceptually irrelevant residual information. The second one is to model the perceptually relevant residual information loss due to quantization and parameterize the same for synthesizing it at the decoder end. This basically evolves around two analytical and estimation problems. The first problem is to identify the perceptually irrelevant quantization noise and remove it from the resulting source. The second problem is to model the perceptually relevant quantization noise.

The first contribution of this dissertation is to classify regions into homogenous / non-homogenous and rigid / non-rigid, based on different perceptual cues like variance, edge, color, and motion. Quantization noise for each region is shaped differently to ensure minimal perceptual quality degradation. At very low bitrates, the rigid regions with small residue errors results in AC coefficients which are small in magnitude. These coefficients, which typically get quantized to zero value, are regenerated / synthesized at the decoder end using statistical characteristics of the temporal predictors. The regions are coarsely segmented based on edge, color, and motion descriptors. Regions with rigid texture are more optimized for rate compared to distortion using higher values of quantization parameter.

The second contribution of this dissertation is identification and representation of non-rigid textured regions like grass, flowing water etc. with a dense motion vector field (DMVF) instead of conventional motion compensated signal. The analysis part contains identification of such regions and classification of macroblocks into rigid and non-rigid homogenous textures. The DMVF is computed only for the macroblocks classified under non-rigid textured regions. A replacement technique is used to substitute a block of texture pixels with a block of motion vectors which are then differentially coded using causal neighbors and context adaptive binary arithmetic coding (CABAC). As a part of texture synthesis, the decoder then simply decodes these

motion vectors, regenerates the DMVF and compensates each pixel individually using the regenerated DMVF.

The remaining macroblocks which are not classified as homogenous texture (rigid or non-rigid) are coded using conventional H.264 encoder. Although the underlying techniques are generic enough to be augmented with any video standard, we specifically picked H.264 video compression standard considering it is the current state-of-the-art. We compare coding approaches using NTIA model for objective measure of subjective quality. Comparing our techniques with H.264 standard compliant JM encoder developed by JVT (Joint Video Technology) committee members, we got a bit-rate savings of around 15%.

The chapters of this dissertation are organized as follows. An introduction to the H.264 standard features and improvements made over several years over existing video standards like MPEG-2 and H.263 are presented in Chapter 1. It also consists of highlighting some of the techniques published to reduce the computation complexity for enabling real-time implementation of encoders. A literature survey of existing techniques which use perceptual criterion for video coding is presented in Chapter 2. Chapter 3 highlights some of the limitations of schemes mentioned in the literature and is followed by the contributions made in the present work to overcome these limitations. Experimental results are presented in Chapter 4 and the thesis is concluded in Chapter 5 highlighting some of the future work which could be carried out in this direction.