

Redundancy Removal in Video frames using Luminance Masking

Snehal A. Patil

PG student

Smt. Kashibai Navale College of Engineering
Vadgaon(Bk), Pune, India

Sonal K. Jagtap

Head, Dept. of Elect. & Telecommunication
Engineering Smt. Kashibai Navale College of
Engineering
Vadgaon(Bk), Pune, India

ABSTRACT

Digital video compression techniques have a crucial contribution in the world of telecommunication and multimedia sector where bandwidth is a valued constraint. A large amount of multimedia data has to be stored in a limited storage space. Hence, video compression techniques mainly focus on reducing the volume of information required for picture sequences/streaming pictures without losing much of its quality. Thus, in order to provide an competent compression method for multimedia data, Luminance Masking technique was proposed. In this context, an Intensity Dependant Spatial Quantization (IDSQ) perceptual means is proposed which attempts the intensity masking of the human visual system and perceptually adjusts quantization. IDSQ allows for adaptation to the video characteristics and its design meets low complication implementation requirements The proposed method has been incorporated into the HEVC reference model for the HEVC Range Extensions and its performance was judged by measuring the bit rate reduction against the HEVC Range Extensions.

Keywords

HDR, HEVC, Intensity Dependant Spatial Quantization (IDSQ), Luminance Masking.

1. INTRODUCTION

Only a small part of visual information visible to the human eye is accumulated using common image and video compression formats, such as JPEG or MPEG. Since these formats have been heretofore designed for CRT displays, the quality they provide is often not sufficient to fully utilize capabilities of new display devices. While the recent displays offer contrast of 10,000:1 or even 50,000:1 and brightness exceeding 500cd/m² or 3000cd/m² for prototypes, the digital image and video content is still in the period of CRT displays with the contrast of 1:200 and the maximum brightness of 100cd/m². The 8-bit coding of luminance values are unable to encode the full range of luminance that can be found in the real world and observed by the human eye and thus resulted into image and video with clipped specular highlight, attenuated shadows and missing bright light sources. The recent wide-color-gamut displays also cannot take full advantage of their capabilities as the color spaces used for image and video coding store only a fraction of the visible color gamut. To entirely take advantage of new display technologies and to enhance compatibility of digital content between different devices, image and video formats need to be extended to encode wider color gamut and higher dynamic range.

High dynamic range (HDR) video and image formats encode the complete visible range of luminance and color gamut, thus giving final fidelity, restricted only by the capabilities of the human eye and not by any existing technology. In this paper

focus is on the major advantages of high dynamic range representation and show how it differs from traditional encodings. This paper explained in detail that how existing image and video compression standards can be enlarged to encode HDR content efficiently. This is gained by a custom color space for encoding HDR pixel values that is obtained from the visual evaluation data.

2. RELATED WORK

HDR compression procedures can be divided into two major sections [2]. Backward-capable HDR compression methods which are based on tone-mapping (Fig. 1) are developed so that standard 8-bit codecs, that can only cope with SDR content and low dynamic range (LDR) displays, are still capable to display and decode a tone-mapped version of the HDR content. HDR capable decoders are capable to inverse tone-map the decoded LDR content and, by decoding the entire bit stream, use the HDR-LDR residual to recreate the HDR content.

As given in Fig 1, backward-compatible HDR compression procedures can be divided into four main processing modules. The first module consists of Tone-Mapping Operator (TMO) used for mapping High Dynamic Range (HDR) content down to Low Dynamic Range (LDR) bit depths. The Tone-Mapping Operator (TMO) can be adjusted in order to increase the compression effectiveness by taking into account the coding mode (i.e. intra or inter) selected by the encoder [3]. The HDR-LDR residual frame is calculated by second module. This forms the side information data essential for reconstructing HDR content at the decoder. This residual frame can be evaluated as a ratio between the HDR and tone-mapped LDR image [4] or through a perceptually improved nonlinear mapping function]. The third module is the actual LDR coding/decoding entity that executes a method compatible with existing SDR coding architectures. The final module executes inverse tone-mapping and HDR content reconstruction by decoding the HDR-LDR residual included in the coded bit stream.

In this type of coding scheme, the inverse tone-mapping and tone-mapping algorithms form a important part of the HDR codec as they can have a large influence on its rate-quality evaluation. Subjective judgement of various tone-mapping algorithms are applied to typical HDR images and delivered in numerous controlled and uncontrolled environments and devices was performed in [5].

The second approach for High Dynamic Range (HDR) coding is not backward-compatible; these systems firstly assign a perceptual quantization transformation (e.g. LogLuv) to map HDR data to the highest bit depth supported by the encoder and then compress the input as per the selected video coding standard. There are numerous ways to achieve the preceding

perceptual transformation. The work proposes a Perceptual Quantizer (PQ) to enhance the perceptual quality when generating HDR content with bit depth of 10 or 12 bits.

HDR-specific advancements to a codec must clearly concentrate on HDR related aspects of the HVS. The HVS exhibits a non-linear sensitivity to stimuli due to various masking sensations. These, together with the varying Human Visual System (HVS) distortion sensitivity associated with them, can be expressed in terms of Just Noticeable Distortions (JND) [6]. JND refers to the level of distortion that can be tolerated in an image or video because it is imperceptible by a human observer. Naccari et al. [7] proposed the use of a spatio-

temporal JND model to perceptually modulate the quantization step

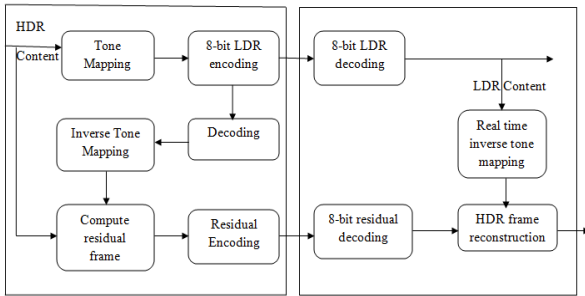


Fig 1: Generalized schematic of backward-compatible HDR codecs.

step allocated to each DCT coefficient in H.264/AVC. To neglect the encoding cost associated with communicating the variable quantization steps to the decoder, the JND model used is estimated at the decoder in [7]. The results show that, at the same objective quality, the proposed method provides an average bit rate reduction of about 30% compared to H.264/AVC-High profile. JND modeling has been also used in [8] to quantify the number of discernible colors provided when a particular color space is applied to HDR content. Zhang et al. proposed the implementation of perceptual filtering prior to HDR encoding, as shown in Fig 2, to exploit the spatial frequency masking behavior of the HVS [8].

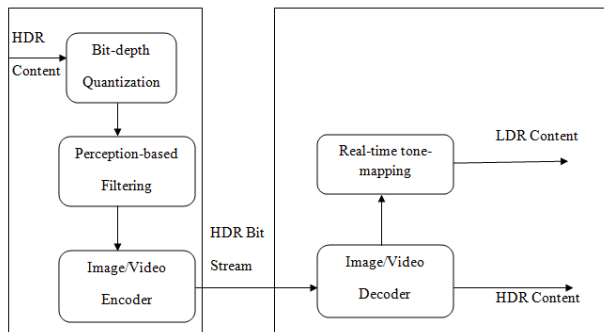


Fig 2: Perception-based HDR codec's structure [8].

3. PROPOSED WORK

The main question for HDR content is to create means for effectively compressing the data without compromising perceptual quality. In the specified approach at first input video frames undergo transform coding and their smoothing followed by objective function minimization. Later on output files are compared with input files and at last output files are converted to output video. In this way, our method compresses the input video and increases its visual quality.

For block-based hybrid video coding technique, transforms are applied to the residual signal resulting from inter- or intra-frame prediction. Main resource of blocking artifacts is the block-transform coding of the prediction error followed by coarse quantization. Two approaches to reduce blocking artifacts are in-loop filtering and post-filtering. Post-filtering is not specific for any of the video coding standard and can be performed, e.g. in the display buffer. The implementer has a choice to propose an algorithm driven by relevance-specific requirements. Loop filters need to be normative to avoid flow between the encoder and decoder as they operate within the encoding and decoding loops. The HEVC draft standard gives two in-loop filters that can be applied consecutively to the reconstructed picture. Block schematics of proposed method is shown

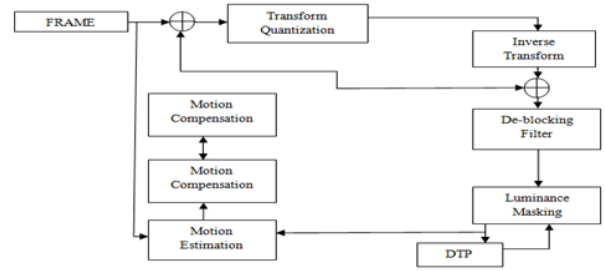


Fig 3: Block schematics of the proposed system.

in Fig. 3. HVS shows a time and space varying sensitivity to the coding distortion introduced in video and images. Intensity Dependent Quantization profile is given as:

$$IDQ(\mu) = \begin{cases} k_1 \cdot (1 - \frac{2\mu}{256})^{\lambda_1} + 1 & \mu \leq 128 \\ k_2 \cdot (\frac{2\mu}{256} - 1)^{\lambda_2} + 1 & \text{otherwise} \end{cases}$$

(1)

Where, (B)=average pixel intensity.

The IDQ profile provides a sign of the utmost amount of distortion that can be tolerated for a given intensity level. In order to exploit an IDQ profile for enhancing coding efficiency there should be a connection between quantization step increments and that particular IDQ profile. This link can be found by studying the quantization error (eq) for a given frequency coefficient C in a coding block B being quantized by a uniform quantizer with step Δ. In particular, eq is given by:

$$e_q = C - C'$$

(2)

Where, C' denotes after inverse quantization (i.e. reconstructed) coefficient. For a uniform quantizer, eq is always confined in the interval [-Δ/2, Δ/2]. For the maximum value for eq (i.e. Δ/2),

$$\frac{\Delta}{2} \leq IDQ(\mu(B))$$

(3)

Therefore, the quantization step related with a given IDQ profile value is:

$$\Delta = 2 \cdot IDQ(\mu(B))$$

(4)

Using the quantization step suggested by (3.4) will present the highest amount of quantization noise to a coded video that will stay unnoticeable to a human observer. The final goal of tone mapping is to maintain the display of the HDR content. Therefore a good tone-mapping operator should retain features such as visual attention fixation points and intensity contrast. The tone-mapping curve offers finer granularity where the luma histogram is denser while coarse mapping is applied to regions where the luma values are more sparse. Based on this approach following relationship between tone-mapping curve and IDQ profile may be inferred:

- Fine tone-mapping~ high sensitivity to distortion.
- Coarse tone-mapping~ low sensitivity to distortion.

Let f denotes the tone-mapping function that maps each HDR intensity level I_{HDR} into the corresponding LDR intensity level

I_{LDR} :

$$I_{LDR} = f(I_{HDR})$$

(5)

Computing the differential for this correlation gives:

$$d I_{LDR} = f' I_{HDR} \cdot d I_{HDR}$$

Where, f' shows the derivate of f with respect to I_{HDR} .

Therefore, by getting the differential intensities corresponding to the JND (i.e. the IDQ value), (6) becomes:

$$IDQ_{LDR} = f'(I_{HDR}, IDQ_{HDR})$$

(7)

Finally, the IDQ profile for HDR content is given as:

$$IDQ_{HDR} = \frac{IDQ_{LDR}}{f'(I_{HDR})}$$

(8)

Where, IDQ_{LDR} can be any profile.

Algorithm 1: for derivation of the HDR IDQ profile

1. Function: derivation of HDR IDQ profile
2. Selection of any IDQLDR profile
3. Selection of a tone-mapping operator
4. For each video frame X of the HDR sequence do
5. Compute the tone-mapping curve for frame X , f ;
6. Compute the derivate of $f(f')$ either in closed form or numerically;
7. for each intensity I_{HDR} in the HDR domain do
8. Compute the LDR intensity I_{LDR} using (3.5);
9. Compute corresponding IDQ value as
10. $IDQ_{LDR} = IDQ(I_{LDR})$;
11. Compute the IDQ_{HDR} value using (3.8);
12. End
13. End

Stepwise execution of the proposed method:

- Take input video file
- Convert video to frame
- Perform pre-processing on image frames
- Luminance Masking algorithm
- Convert frame to video

- Comparison of various parameters

The performance evaluation will be decided based on following parameters such as:

1. Compression ratio:

Compression is defined as the methodology to reduce the redundant features and the unwanted pixels that are present in the image to reduce the need of bandwidth. The better image compression leads to the improved compression ration which is calculated as follows:

$$\text{Compression Ratio} = \frac{\text{Compressed Size}}{\text{Uncompressed Size}}$$

(9)

2. Peak signal to noise ratio:

Peak signal-to-noise ratio (PSNR) is used to represent the index of media quality analysis and substitutes the average MSE of frames in the PSNR computing equation to obtain the PSNR value of the media segment, expressed as,

$$PSNR = 20 \log_{10} \left(\frac{255}{\sqrt{MSE}} \right)$$

(10)

Where,

$$MSE = \frac{1}{n} \sum_{i=1}^n (\bar{Y}_i - y_i)^2$$

(11)

YI = vector of n predictions

y = Vector of observed values

n = sum over variable

Using the intensity masking phenomenon of the HVS leads to a modified perceptual quantization step ΔIDQ given by the following equation, where Δ is the original quantization step:

$$\Delta^{IDQ} = \Delta \cdot IDQ(\mu)$$

(12)

To avoid the multiplication in (12), the relationship between the Quantization Parameter (QP) and Δ in the HEVC quantizer is considered:

$$\Delta = 2^{(QP-4)/6}$$

(13)

Using (12) and inverting (13) the QP associated with can be computed.

This perceptual QP can be stated with respect to the base QP acquiring the intensity differential quantization parameter (idQP):

$$QP^{IDQ}(\mu) = 6 \cdot \log_2(\Delta \cdot IDQ(\mu)) + 4$$

$$= 6 \cdot \log_2(\Delta) + 4 + 6 \cdot \log_2(IDQ(\mu))$$

(14)

$$QP^{IDQ}(\mu) - QP = 6 \cdot \log_2(IDQ(\mu))$$

(15)

Where, $idQP(\mu) = QP^{IDQ}(\mu) - QP$

To limit QP and idQP rounding is applied such as:

$$idQP(\mu) = [6 \cdot \log_2(IDQ(\mu)) + 0.5]$$

(16)

The average intensity estimate $\hat{\mu}$ is computed for each TU by averaging all the samples pertaining to the spatial or temporal predictor used for the unit:

$$\hat{\mu} = \frac{1}{M \cdot N} \sum_{x=y=0}^{M \cdot N} P(x, y)$$

(17)

Where, M,N=no. of rows and columns forming Transform Unit(TU)

P=Predictor used to compute residual

The estimate calculated using (17) avoids the transmission of the original μ values to the decoder and allows intensity dependent quantization to be performed for each transform unit.

The profile used in intensity-dependent quantization needs to be transmitted to the decoder to confirm correct decoding of the compressed video. The whole μ range [0, 2N-1] is separated into bins whereby each bin (b) is characterized by its width (w) and the quantity δ is given as:

$$\delta = idQP(b_m) - idQP(b_{m-1})$$

(18)

The average pixel intensity μ used for calculating the idQP value of a given coding block is calculated by computing the average intensity over all the pixels belonging to the predictor (temporal or spatial) of that coding block. At the encoder side, forward quantization uses a QP value equal to QP + idQP.

At the decoder side, inverse quantization needs calculation of the value of idQP, which in turn requires the decoded pixel values of the block predictor. This analogy on the availability of the block predictor may introduce prohibitive delay in some practical decoder implementations that make use of parallel processing. For example, considering the estimation of the motion compensated predictor in the decoder's pipeline, the predictor may become available only before actual reconstruction, i.e. after inverse transform and inverse quantization.

4. RESULTS AND ANALYSIS

Visual masking model is useful in detecting the masking effect of human visual system. Visual masking model consists of 3 main

elements such as:

1. Just Noticeable Difference Model (JND)

It gives the information about maximum distortion the human visual system cannot perceive.

2. Visual Attention Model

It calculates the information about the attention point of the eye in video/image based on features such as texture, luminance, motion



Fig 4: Original Image

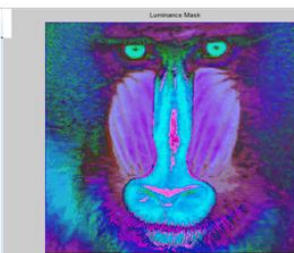


Fig 5: Luminance Mask

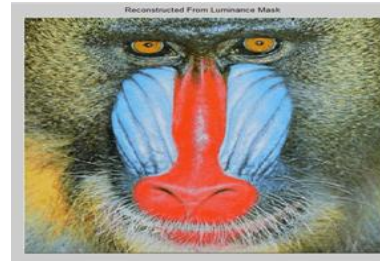


Fig 6: Reconstructed From Luminance Mask

and object based characteristics.

3 Weighing Model

For HVS highest sensitivity and spatial resolution is present at the point of fixation, calculation of fixation point is included into the overall model using modulating weighing map by foveation method.

JND model considers a pixel domain edge detection using canny edge detection and results of edge detection are used for block type

separation. For coming step the model works in frequency domain by applying a discrete cosine transform on the input image and includes a spatial-temporal contrast sensitivity function, contrast masking, luminance adaptation and the effect of eye movement.

5. CONCLUSION AND FUTURE SCOPE

High dynamic range video compression is the essential functionality resides in the image processing techniques which in turn used to compress the videos with the consideration of the different system parameters. Video compression has widely been adopted by the document storage, multimedia film making, industry and video conferencing or organization because of its wide applicability in real life and this has been growing more and more. There are many existing methods for video compression all of them having some merits and demerits. In future, focus will be on implementing masking algorithm for video frames and convert compress frames into video.

6. ACKNOWLEDGMENTS

This is the acknowledgement of the comprehensive drive and technical competence of many individuals who have contributed to the success of the project. We are grateful to Electronics and Telecommunication department of Smt. Kashibai Navale College of Engineering Vadgaon(Bk) Pune for endowing us with moral support and infrastructure facilities.

7. REFERENCES

- [1] Yang Zhang, Matteo Naccari, Dimitris Agrafiotis, Marta Mrak, and David R.Bull, "High Dynamic Range Video Compression Exploiting Luminance Masking," *IEEE Trans. Circuits and Syst. For Video Technol.*, vol. 13, no. 99, pp. 560-574, 27.Apr. 2015.
- [2] Y. Zhang, D. Agrafiotis, and D. Bull, "High dynamic range image & video compression a review," in *Proceedings of IEEE International Conference on Digital Signal Processing (DSP)*, pp. 1-7, 2013.
- [3] A. Le Dauphin, R. Boitard, D. Thoreau, Y. Olivier, E. Francois, and F. LeL'annec, "Prediction-guided quantization for video tone mapping," in *SPIE 8499, Applications of Digital Image Processing XXXVII*.

- International Society for Optics and Photonics, p. 92170B, 2014.
- [4] G. Ward and M. Simmons, "JPEG-HDR: A backwards-compatible, high dynamic range extension to JPEG," in Proceedings of Color Imaging Conference, pp. 283–290, 2005.
- [5] P. Korshunov and T. Ebrahimi, "A JPEG backward-compatible HDR image compression," in Proceedings of SPIE Conference on Applications of Digital Image Processing XXXV, p. 84990J, 2012.
- [6] H. R. Wu and K. R. Rao, Digital video image quality and perceptual coding. CRC Press, 2005.
- [7] M. Naccari and F. Pereira, "Advanced H.264/AVC-based perceptual video coding: architecture, tools, and assessment," IEEE Transactions on Circuits and Systems for Video Technology, vol. 21, no. 6, pp. 766–782, 2011.
- [8] Y. Zhang, E. Reinhard, D. Agrafiotis, and D. Bull, "Image and video compression for HDR content," in Proceedings of SPIE Conference on Applications of Digital Image Processing XXXV. SPIE, pp.84 990H1–84 990H13, 2012.
- [9] Yu-Cheng Fan, Shu-Fen Wu, Bing-Lian Lin. Three-Dimensional depth map motion estimation and compensation. IEEE Transactions on Magnetics, vol.47, no.3, pp. 691-695, 2011.
- [10] Anmin Liu, Weisi Lin, Manoranjan Paul, Fan Zhang, Chenwei Deng. Optimal compression plane for efficient video coding. IEEE Transactions On Image Processing. vol. 20, no.10, pp.2788-2799, 2011.
- [11] Zicong Mai, Hassan Mansour, Rafal Mantiuk, Panos Nasiopoulos, Rabab Ward, Wolfgang Heidrich. Optimizing a tone curve for Backward-compatible high dynamic range image and video compression. IEEE Transactions on Image Processing. vol.20, no.6, pp.1558-1571, 2011.
- [12] Christoph Posch, Daniel Matolin, and Rainer Wohlgenannt. A QVGA 143 dB Dynamic Range Frame-Free PWM Image Sensor With Lossless Pixel-Level Video Compression and Time-Domain CDS. IEEE Journal Of Solid-State Circuits. vol. 46,no.1, pp.259-275,2011.
- [13] T. E. R. Chaminda Hewage, G. Maria Martini. Edge-Based Reduced-reference quality metric for 3-D video compression and Transmission. IEEE Journal of Selected Topics in Signal Processing. vol. 6,no.5, pp.471-82,2012.
- [14] Zicong Mai, Hassan Mansour, Panos Nasiopoulos, Rabab Kreidieh Ward. Visually favorable Tone-mapping with high compression performance in Bit-depth scalable video coding. IEEE Transactions on Multimedia. vol.15,no.7,pp.1503-1518,2013.