

MULTILAYER BIT ALLOCATION FOR VIDEO ENCODING

Elavarasan.R¹ and Dr.Sunitha Abburu²

¹Adhiyamaan College of Engineering, Department of Computer Application, Hosur
elava.mca@gmail.com

²Professor and Director, Adhiyamaan College of Engineering, Department of Computer
Application, Hosur
ausunithaa@yahoo.co.in

ABSTRACT

Video compression approach removes spatial and temporal redundancy based on the signal statistical correlation. Bit allocation technique adopts a visual distortion model for a better rate visual distortion video coding. Visual distortion model uses both motion and the texture structures in the video sequences. The existing video coding mechanisms reduces the bit rate for video coding. However to get better video compression ratio there is a need for multilayer compression technique. In this paper we proposed a multilayer bit allocation video coding mechanism. The proposed model reduces the bit allocation for video coding by retaining the same video quality. The experimental results using the proposed model reduced the bit rate by 3% to 4%. The result are promising. Finally we conclude with conclusion and future work.

KEYWORDS

Foveation Model (FM), Spatial Model(SM), Temporal Mode(TM), Advanced Video Coding (AVC), Human Visual System (HVS).

1. INTRODUCTION

The compression from JPEG to advance video coding (AVC) uses the spatial and temporal features for video coding compression [1], [2]. Video compression refers to reducing the quantity of data used to represent digital video images, and is a combination of spatial image compression and temporal motion compensation.

Types of video compression:

1. Lossless compression
2. Lossy compression

Traditional video coding methods are facing a lot of challenges, as it provide the flexibility to adapt various transmission conditions and diversity of customer satisfaction. The commonly used techniques includes transform, motion estimation/compensation, intra/inter prediction, motion compression entropy coding, decoding, etc. To improve the performance of video compression, the human visual system (HVS) needs to be understood and carefully utilized. Video coding methods [3], [4] aim to account for visual properties of the human visual system in the quantization and bits allocation. We have proposed a visual distortion model, exploiting the non-uniform spatio temporal sensitivity characteristics of the HVS, for video coding. Macroblocks (MBs) visual distortion is analyzed based on both motions and textural structures [5]. Few bits are

allocated to MBs in which higher distortion can be tolerated. The bit-rate of the whole video can be reduced according to video coding. This method has been further extending by considering extra spatial and temporal cues based on a better distortion model. These cues include motion attention, spatial velocity visual distortion sensitivity, and visual masking effect. It is well known that human cannot perceive the fine-scale variation of visual signal due to the psycho visual properties of HVS [6] and the visual distortion sensitivity of HVS can be measured by a spatiotemporal contrast sensitivity function (CSF). The spatiotemporal CSF shows the contrast required to detect the masking of different spatial and temporal frequency.

The rest of this paper is organized as follows. Section 2 gives the literature survey. Section 3 multilayer bit allocation for video encoding. Section 4 experimental results. Finally we conclude with conclusion and future work in Section 5.

2. LITERATURE SURVEY

[7] address the key problem in real-time video coding, the Rate-Distortion (R-D) tradeoff investigate that the R-D characteristics of color video signal in traditional transform-based video coding systems should be modeled for the luminance and chrominance components separately. Chen proposes separable R-D models for color video coding. The feedback from the encoder buffer is analyzed by a control-theoretic adaptation approach to avoid buffer overflow and underflow. To achieve smooth video quality and satisfy the delay constraints in real-time applications, a novel R-D tradeoff controller is designed. Both the quality variation and buffer safety are considered.

Chun-Hsien et al, in [8] discussed important service in multimedia communications. Because of the limited bandwidth and high error rates of the wireless channel, the video coding should be designed to have high code efficiency in maintaining acceptable visual quality at low bit rates and robustness to suppress the distortion due to transmission debug. In this method, the coding efficiency of a 3-D sub band video coding is optimized by removing not only the redundancy due to spatial and temporal correlation but also perceptually insignificant components from video coding signals. Unequal error protection is applied to the source coding of different perceptual importance. The method based on noisy channel is assumed to be corrupted by the regular errors depends on the strength of the received wave and the burst errors due to Rayleigh fading.

The early attempts to remove perceptual redundancy sequence in signal compression by exploiting human perception mechanisms discussed in [9] Jayan, develops the notion of perceptual coding based on the concept of distortion masking by the signal being compressed, and describes how the field has progressed as a result of advances in classical video coding concept, modeling of human perception, and digital signal processing (DSP). This approach is based on signals, and while travelling there may be loss of signal due to some disturbance.

A video bit allocation technique adopting a visual distortion sensitivity model for better rate-visual distortion coding control is proposed by Tang [10]. Instead of applying complicated semantic understanding, the automatic distortion sensitivity analysis process analyzes both the motion and the texture structures in the video sequences in order to achieve better bit allocation for rate-constrained video coding. The technique evaluates the perceptual distortion sensitivity on a macro block basis, and allocates fewer bits to regions permitting large perceptual distortions for rate reduction.

X.Yang et al, in [11] discussed Human eyes cannot sense any changes below the JND around a pixel due to their underlying spatial-temporal masking effect properties. JND model can significantly help to improve the performance of video coding. From the viewpoint of signals compression, smaller variance of signal results in low objective distortion sensitivity of the reconstructed signal for a given bit-rate. A new JND estimator for video coding is devised in

image-domain with the non-linear additivity model for masking (NAMM) and is incorporated into a motion-compensating the residue signal preprocessor for variance reduction towards coding quality enhancement.

C.-H. Chou et al, in [12] represented an image of high perceptual quality with the lowest possible bit rate. An effective image compression algorithm should not only remove the redundancy due to statistical correlation but also the perceptually insignificant components from image signal. In this approach, perceptually tuned sub-band image coding scheme is presented, where a just-noticeable distortion (JND) or minimally noticeable distortion (MND) profile is employed to quantify the perceptual coding redundancy. The JND profile provides each signal being coded with a visibility of distortion, below which reconstruction errors are rendered imperceptible. Based on a perceptual model that incorporates the threshold sensitivities due to the background and texture masking effect, the JND profile is estimated from analyzing local properties of image signals. According to the sensitivity of human visual perception to spatial, the full-band JND/MND profile is decomposed into component JND/MND profiles of different frequency sub-bands. This approach is based on image compression, it takes more time to compress the image.

3. MULTILAYER BIT ALLOCATION FOR VIDEO ENCODING

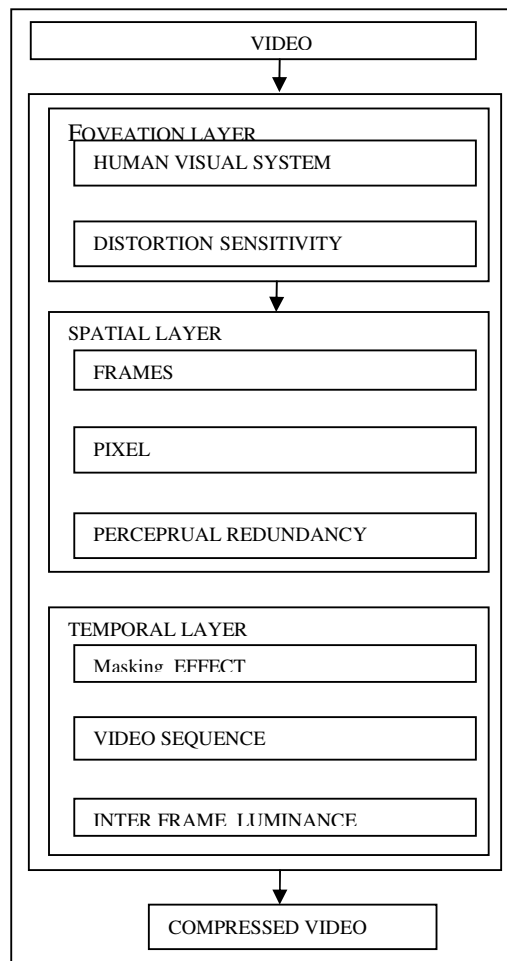


Figure 1. System Architecture of Multilayer Bit Allocation for Video Encoding.

The existing video coding approaches are using large number of bit rate for a video. Bit-rate plays a key role in a high quality of video encoder. The target is to achieve the better perceptual picture quality at a given bit rate through proper bit allocation process. In order to achieve good visual sensitivity quality across different area with optimal bit allocation, psychophysical model has been taken into account in the bit allocation process. A video coding bit-rate allocation technique based on visual sensitivity analysis, directs the video coder to assign few bits to regions that tolerates larger distortions model, and accordingly, the bit-rate saving is achieved. This approach makes use of the concept of visual masking effects in the human visual system. The proposed algorithm improves the performance by reducing the bit rate of video coding.

3.1. Encrypted Video Coding

In this paper we propose a multilayer bit allocation video coding mechanism. The System architecture of multilayer bit allocation for video encoding is shown figure 1. The proposed approach consists of three layers Foveation layer, Spatial layer and Temporal layer.

Foveation layer representing the human visual system (HVS) is highly space-variant in sampling, processing and understanding of visual information. Distortion sensitivity distinguishes the background luminance. Perceptual redundancy in the spatial domain is mainly based on the visual sensitivity of the HVS due to luminance contrast, which is been delt in spatial layers. Masking effect is done by temporal layer.

3.2 Foveation Model

In Foveation Model human observes the video picture to identify the regions where the video compression mechanism can to be applied. It consists of two phases they are 1) Human visual system, 2) Distortion sensitivity (Distortion sensitivity is a quality or condition of being sensitive). The foveation model is used to identifying the detectable region of video as a function of the coordinates of the fixation point (the point on the image which is under direct observations of human visual system) and the viewing distance of the observer from the video.

The larger video contrast can be eliminated by using video coding. We consider both the distortion sensitivity and properties of the human visual system to find the frame to which video compression technique should be applied. The foveation describes the relative sensitivity of the HVS and visual distortion sensitivity at different frequencies.

The perceptible luminance difference of a stimulus depends on the surrounding luminance level, which is the HVS is sensitive to luminance contrast rather than the absolute luminance level. For example, the visual sensitivity distortion is most visible against a gray background color. The frame with very bright or very bark background can be distorted. Foveation model measures the visibility of the HVS according the characteristics of the visual sensitivity signal.

$$FM(r, s, t, v, e) = f(SM(r, s), TM(r, s, t), F(r, s, v, e)) \quad (1)$$

Where $FM(r, s, t, v, e)$, $SM(r, s)$, $TM(r, s, t)$, and $F(r, s, v, e)$ denote FM, SM and TM, respectively. t is the frame index, v is the viewing distance, and e is the eccentricity for the point (r, s) relative to the fixation point (rf, sf) . v is the viewing distance, and e is the eccentricity for the point (r, s) relative to the fixation point (rf, sf) .

3.3 Spatial Model

The spatially-encoded video representation differs from conventional video streams and compression algorithms in different ways. The frames are organized in multiple dimensions that include position and orientation. The frames are divided into number of frames to distortion. Further, our encoding algorithm utilizes spatial coherence between the impostors and model-based distortion sensitivity. The perceptual redundancy in the spatial method is mainly based on

the visual sensitivity of the HVS due to luminance contrast and spatial masking effect. This model has been developed to reduce the bit rate. In this layer it identifies where more contrast is available. If the contrast is high in a frame, then this layer reduces the pixel rate. Pixel value is reduced in the spatial model to get the low bit rate. JND models were built in spatial (pixel) domain. The discrete cosine transform (DCT), and wavelet domains, we have to use the spatial function to frames and pixel.

$$SM(r, s) = \max\{f1(bg(r, s), mg(r, s)), f2(bg(r, s))\} \quad (2)$$

Where $f1(bg(r, s), mg(r, s))$ and $f2(bg(r, s))$ are functions to estimate the spatial masking and luminance contrast, respectively. The quantity $f1(bg(r, s), mg(r, s))$ is defined as

$$f1(bg(r, s), mg(r, s)) = mg(r, s) \times \alpha(bg(r, s)) + \beta(bg(r, s)) \quad (3)$$

Where $mg(r, s)$ is the maximum weighted average of luminance differences derived by calculating the weighted average of luminance changes around the pixel (r, s) in four directions as

$$mg(r, s) = \max\{|grad(r, s)|\} \quad (4)$$

$$grad(r, s) = \frac{1}{16} \sum_{i=1}^{55} \sum_{j=1}^{55} p(r-3+i, y-3+i) \times Gk(i, j) \quad (5)$$

$$bg(r, s) = \frac{1}{32} \sum_{i=1}^{55} \sum_{j=1}^{55} p(r-3+i, y-3+i) \times B(i, j) \quad (6)$$

The function $f2(bg(r, s))$ computes the visibility threshold from the luminance contrast.

3.4 Temporal Model

Temporal compression layer is used to achieve the highest compression rate in video coding. Temporal compression is a technique of reducing compressed video size by not encoding each frame as a complete image. The frames that are encoded completely (like a static image) are called *key frames*. All other frames in the video are represented by data specifying the change since the last frame. The masking effect and the temporal masking effect should also be considered to remove replications. Usually, large inter-frame luminance difference results in larger temporal masking effect. A video sequence has been constructed in which a square of luminance level moves horizontally over a background of luminance other level. Noise has been randomly added or subtracted to each pixel in small regions. The distortion sensitivity thresholds have been determined as a function of the inter-frame color difference and background color.

$$TM(r, s, t) = \max(t, H/2 \exp(r+s+t) + r) \leq 0 \quad (7)$$

Where $H = 8$ and $t = 0.8$ starting that the scale factor should be reduced to 0.8 as $(x, y, t) < 5$, in order to minimize the allowable distortion in stationary regions. The temporal model are used to

4. EXPERIMENTAL RESULTS

H.264 is a next-generation video coding compression format. H.264 is also well known as MPEG AVC. H.264 maintains high quality at lower file sizes than MPEG-2, MPEG-4. The proposed technique is implemented using Java Swing. In the proposed approach the video compression get the same original video quality. Using proposed video compression approach the bit rate is reduced by 3% to 4% by maintaining the original video quality. The experiments are conducted

by taking tennis video. The video streams are passed through the foveation model, spatial model, and temporal model for computing better reduced bit rate. The video motion vectors used in the attention model are generated using the full search motion estimation algorithm.



Figure 2. Original image Figure 3. Image of reduced bit rate

Figure 2 shows the original image and figure 3 shows the reduced bit rate image, but there are barely any visual code differences between the two images. This is because few bit-rates are allocated to visually less sensitivity place, while more number bit-rate are assigned to the distortion-sensitive place. It turns over that there is no perceptual visual quality degradation even though the whole video sequence bit rate decreases. The inadequacy distortion measure is shown clearly through these experiments results. For example, for tennis video bit rate is reduced from the original image bit rate.

Our major goal is to compare the viewing experience between the H.264 with rate control and ordinary bit rates. Since these two coded sequences having the same bit rates, it is obvious from the experiments results that our proposed method produces good visual quality.

5. CONCLUSION AND FUTURE WORK

The conventional novel video coder bit allocation technique based on visual distortion sensitivity analysis used to reduce the bit rate. The Spatial and Temporal model is improved by taking into account the foveation properties of the HVS. The visibility given by the foveation model video code bit allocation technique based on visual distortion sensitivity. The subjective test shows the performance of the proposed scheme. Bitrates are saved with practically no perceptual quality degradation.

Future Work:

- Using encoded pictures as references in a much more flexible way than in past standards, allowing up to 16 reference frames (or 32 reference fields, in the case of interlaced encoding) to be used in some cases.
- Weighted prediction, allowing an encoder to specify the use of a scaling and offset when performing motion compensation, and providing a significant benefit in performance in special cases—such as fade-to-black, fade-in, and cross-fade transitions.
- Logarithmic step size control for easier bit rate management by encoders and simplified inverse-quantization scaling
- Frequency-customized quantization scaling matrices selected by the encoder for perceptual-based quantization optimization.

REFERENCES

- [1] Information Technology: Digital Compression and Coding of Continuous Tone Still Images, ISO/IEC IS 10918-1 | ITU-T Rec. T.81, 1994.
- [2] Draft ITU-T Recommendation and Final Draft International Standard of Joint Video Specification (ITU-T Rec. H.264|ISO/IEC 14496-10 AVC), document JVT-G050.doc, Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG, 2003.
- [3] R. J. Safranek and J. D. Johnston, "A perceptually tuned sub-band image coder with image dependent quantization and post quantization data compression," in Proc. IEEE Int. Conf. Acoust., Speech, Signal Process., Pacific Grove, CA, May 1989, pp. 1945–1948.
- [4] Information Technology: JPEG 2000 Image Coding System, ISO/IEC 15444-1, 2000.
- [5] Yosi Keller and Amir Averbuch, "Fast Gradient Methods Based on Global Motion Estimation for Video Compression" IEEE Trans. Circuits Syst. Video Technol., vol. 13, no. 4, April 2003.
- [6] V. V. Gohokar and Dr. V. N. Gohokar, "Optimum Bit Allocation Using Human Visual System for Video Compression" International Conference on Computational Intelligence and Multimedia Applications 2007.
- [7] Z. Chen and K. N. Ngan, "Toward rate-distortion tradeoff in real-time color video coding," IEEE Trans. Circuits Syst. Video Technol., vol. 17, no. 2, Feb. 2007, pp. 158–167.
- [8] C.-H. Chou and C.-W. Chen, "A perceptually optimized 3-D sub-band codec for video communication over wireless channels," IEEE Trans. Circuits Syst. Video Technol., vol. 6, no. 2, Apr. 1996, pp. 143–156.
- [9] N. Jayant, J. Johnston, and R. Safranek, "Signal compression based on models of human perception," Proc. IEEE, vol. 81, no. 10, Oct. 1993, pp. 1385–1422.
- [10] C.-W. Tang, C.-H. Chen, Y.-H. Yu, and C.-J. Tsai, "Visual sensitivity guided bit allocation for video coding," IEEE Trans. Multimedia, vol. 8, no. 1, Feb. 2006, pp. 11–18.
- [11] X. Yang, W. Lin, Z. Lu, E. P. Ong, and S. Yao, "Motion-compensated residue preprocessing in video coding based on just-noticeable-distortion profile," IEEE Trans. Circuits Syst. Video Technol., vol. 15, no. 6, Jun. 2005, pp. 742–752.
- [12] C.-H. Chou and Y.-C. Li, "A perceptually tuned sub-band image coder based on the measure of just-noticeable-distortion profile," IEEE Trans. Circuits Syst. Video Technol., vol. 5, no. 6, Dec. 1995, pp. 467–476.