Video Compression Using Block By Block Basis Salience Detection

[1] P.Muthuselvi M.Sc.,M.E., , [2] T.Vengatesh MCA.,M.Phil.,(Ph.D).,

[1] *Assistant Professor,VPMM Arts & Science college for women, Krishnankoil 626149,Virudhunagar(Dt),Tamilnadu,India.*

[2] *Assistant Professor,VPMM Arts & Science college for women,Krishnankoil-626149,Virudhunagar(Dt),Tamilnadu,India.*

---***---

Abstract - *This dissertation explores the concept of visual saliency a measure of propensity for drawing visual attention and presents a variety of novel methods for utilization of visual saliency in video compression and transmission. While the region-of-interest (ROI)-based video coding, ROI parts of the frame are encoded with higher quality than non-ROI parts. It was reducing salient coding artifacts in non-ROI parts of the frame in order to keep user's attention on ROI. Unfortunately, the salient coding of non-ROI compression parts is uneasy to find the block-by-block basis. The proposed method aims at reducing salient coding artifacts in non-ROI parts by optimizing the saliency-related Lagrange parameter, possibly on a block-by-block basis. Experimental results indicate that the proposed method is able to improve visual quality of encoded video relative to conventional rate distortion optimized block-by-block basis video coding, as well as two state-of-the art perceptual video coding methods.*

***Key Words*: ROI-based video coding, attention-grabbing coding artifacts, non-ROI parts, salient coding artifact**

1.INTRODUCTION

Lossy image and video encoders are known to produce undesirable compression artifacts at low bit rates [1], [2]. Blocking artifacts are the most common form of compression artifacts in block-based video compression. When coarse quantization is combined with motion-compensated prediction, blocking artifacts propagate into subsequent frames and accumulate, causing structured high-frequency noise or motion-compensated edge artifacts that may not be located at block boundaries, and so cannot be attenuated by de-blocking filters that mostly operate on block boundaries [2]. Such visual artifacts may become very severe and attention-grabbing (salient), especially in low-textured regions.

Recently, region-of-interest (ROI) coding of video using computational models of visual attention [3] has been recognized as a promising approach to achieve high-performance video compression [4]–[7]. The idea behind most of these methods is to encode an area around the predicted attention grabbing (salient) regions with higher quality compared to other less visually important regions. Such a spatial prioritization is supported by the fact that only a small region of 2–5◦ of visual angle around the center of gaze is perceived with high spatial resolution due to the highly non-uniform distribution of photoreceptors on the human retina [4].

ROI-based processing can also be employed in the context of video transmission to combat the effects of transmission channel errors. For instance, ROI parts of the frame can be protected heavily (e.g., by using stronger channel codes) than non-ROI parts of the frame, so that in the case of channel errors or losses, important parts of the frame can still be decoded correctly. In this case, also, ROI could be detected either based on direct eye-tracking measurement or based on visual saliency models.

2. RELATED WORK

A. The IKN Saliency Model

Among the existing bottom-up computational models of visual attention, the Itti-Koch-Niebur (IKN) model [3] is one of the most well-known and widely used. In this model, the visual saliency is predicted by analyzing the input image through a number of pre-attentive independent feature channels, each locally sensitive to a specific low-level visual attribute, such as local opponent color contrast, intensity contrast, and orientation contrast. More specifically, nine spatial scales are created using dyadic Gaussian pyramids, which progressively low-pass filter and down-sample the input image, yielding an image-sizereduction factor ranging from 1:1 (scale zero) to 1:256 (scale eight) in eight octaves [3].

B. Rate-Distortion Optimization (RDO)

The H.264/AVC video coding standard supports various block coding modes such as INTER16 × 16, INTER16 × 8, INTER8 × 16, INTER8 × 8, INTRA16 × 16, INTRA4 × 4, etc. [11]. The rate-distortion optimization (RDO) process proposed in H.264/AVC minimizes the following Lagrangian cost function for coding mode selection of each macro block (MB) [8]:

$$J(\psi|Q,\lambda_R) = D_{MSE}(\psi|Q) + \lambda_R R(\psi|Q) \quad (1)$$

where Q is the quantization step size, $D_{MSE}(\psi|Q)$ and $R(\psi|Q)$ are, respectively, the Mean Squared Error (MSE) and bit rate for coding the current MB in the coding mode ψ with quantization step size Q, and λ_R is the Lagrange multiplier, which quantifies the trade-off between the rate and distortion [13].

3. PROPOSED METHODOLOGY

The Saliency-Aware Video Compression system is a new tool to give the solution for security of templates as well as authentication of the individual users. In this paper, video compression coding is acting as keys method attempts to reduce attention-grabbing coding artifacts and keep viewers' attention in areas where the quality is highest. The method allows visual saliency to increase in high quality parts of the frame, and allows saliency to reduce in non-ROI parts. The overview of the proposed work is shown in Fig.1. This work consists of two phases; the first one is a video feature extraction and second one is a saliency compression phase.

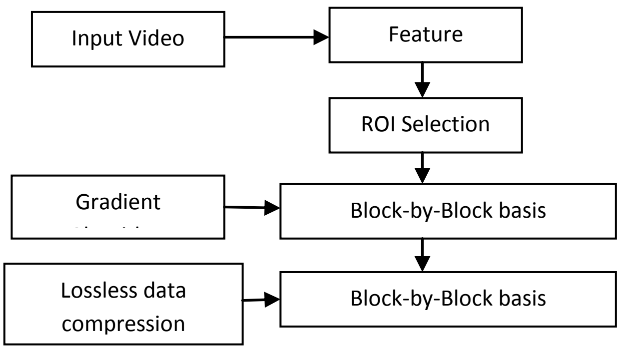

Figure 1: Proposed Architecture Diagram

A. Video Feature Extraction

Video summarization is a compact representation of a video sequence. It is useful for various video applications such as video browsing and retrieval systems. A video summarization can be a preview sequence which can be a collection of key frames which is a set of chosen frames of a video. Key-frame-based video summarization may lose the spatio-temporal properties and audio content in the original video sequence; it is the simplest and the most common method. When temporal order is maintained in selecting the key frames, users can locate specific video segments of interest by choosing a particular key frame using a browsing tool. Key frames are also effective in representing visual content of a video sequence for retrieval purposes. Video indexes may be constructed based on visual features of key frames, and queries may be directed at key frames using image retrieval techniques. Video frames reduce the amount of data required in video indexing and provides framework for dealing with the video content.

B. ROI-Based Video Coding

After frame feature extraction conversational video, Region of Interest parts are detected using the direct frame difference and skin-tone information. After detecting ROI parts, several coding parameters including quantization parameter (QP), coding modes, the number of reference frames, the accuracy of motion vectors, and the search range for motion estimation, were adaptively adjusted at the macro block (MB) level according to the relative importance of each MB and a given target bit rate.

In ROI-based video coding, more bits are typically assigned to ROI parts of the frame compared to non-ROI parts. There are two main aspects that differentiate existing ROI coding methods: (1) the way in which ROI is decided, and (2) the way in which bits are allocated to ROI vs. non-ROI. The proposed method differs from the existing work in the area in both aspects. First, to detect ROI, a novel saliency estimation method is employed. It consists of a convex approximation to the spatial Itti-Koch-Niebur (IKN) saliency combined with a global motion compensated temporal saliency estimate. The new saliency estimate is computationally more efficient, yet competitive in terms of accuracy with the IKN model for video, which is known to be highly accurate in terms of gaze prediction on video. Second, the bit allocation procedure in the proposed method takes into account the way in which saliency is affected by compression. We propose a model for this process and employ it in the RDO procedure to control the way in which saliency is allowed to change after compression.

C. Block-by-block Basis Separation using conjugate gradient algorithm

The Block-by-block basis Conjugate Gradient (BCG) Method is an iterative method for finding the minimum of a quadratic cost function. The objective of the method here, is the minimization of a cost function. The BCG method originates from the so-called conjugate direction (CD) method. The main idea in the CD method is to obtain a set of linearly independent direction vectors which are conjugate with respect to R so that the vector that minimizes the objective function can be expressed as a linear combination of these vectors. The weights of the linear combination are calculated by imposing R-conjugation between the direction vectors. The set of the R-conjugate direction vectors are generated by a set of *M* linearly independent vectors. In the Conjugate Gradient (CG) method these M vectors are the successive gradients of the cost function, obtained as the method progresses. The main task is to derive the block-by-block updating the weight coefficient vector, based on the basic CG algorithm. Initially, it is essential to focus on obtaining the block update recursion of the autocorrelation and the cross correlation vectors by using the autocorrelation method of data windowing method.

D. Video Compression using Lossless data compression algorithms

The video compression process on the audiovisual content should produce a kind of lossless video compression that is imperceptible to the end user of the audiovisual content while simultaneously maintaining the benefits of keeping all of the information of the source and also the benefits of compression during the production process. Lossless compression offers the best of both worlds, with mathematically identical content to an uncompressed file while simultaneously enjoying the benefits of a smaller file size in increased file storage versus uncompressed files, and the benefits of faster transportation of the compressed file."Lossless" means that the output from the de-compressor is bit-for-bit identical with the original input to the compressor. The decompressed video stream should be completely identical to original. In a non-video context, there exist a number of examples of lossless formats.The importance of lossless compression has an incontrovertible value to the archival community. Wavelet compression is one of the most effective methods of image compression. The algorithm is based on multi-resolution analysis. Like the traditional JPEG compression, wavelet compression algorithm presents an image as sets of real coefficients. Most of the wavelet coefficients of a typical image are nearly zero, and the image thus is well-approximated with a small number of large wavelet coefficients. The advantage of wavelet compression is that, in contrast to JPEG, wavelet algorithm does not divide image into blocks, but analyze the whole image. The characteristic of wavelet compression allows the process to get best compression ratio, while maintaining the quality.

The two options have been proposed:

- Moving and storing uncompressed files or streams which are absolutely identical to the original, but may require large amounts of bandwidth and storage space;
- Compressing moving image files or bit-streams in order to reduce storage requirements and transmission times

4. EXPERIMENTAL RESULTS

The experiment was run in a quiet room with 15 participants (14 male, 1 female, aged between 18 and 30). All participants had normal or corrected to normal vision. A 22-inch Dell monitor with brightness 300 cd/m^2 and resolution 1680 × 1050 pixels was used in our experiments. The brightness and contrast of the monitor were set to 75%. The actual height of the displayed videos on the screen was 185 millimeters.The results for the comparison are shown in Table I. In Table I we show the number of responses that showed preference for the foveated just noticeable-difference (*FJND*) model [9] method vs. the proposed method.Using the EWPSNR metric given by [10], we can now define an equivalent eye-tracking-weighted PSNR as follows

$$EWPSNR = 10 log_{10}\left(\frac{255^2}{EMMSE}\right) \quad (2)$$

Table 1: Comparing Various Eye-Tracking-Weighted

Method	EWPSNR	PNSR	SSIM
FJND	+0.40	0.15	0.003
Proposed method	+1.55	0.21	0.002

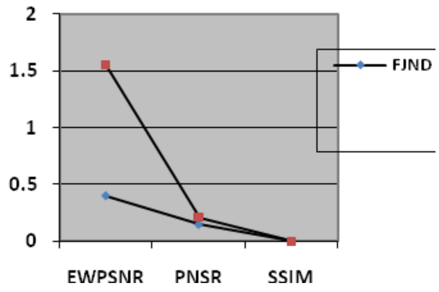

Figure 2: Comparison Chart

5. CONCLUSION

Error concealment in loss-corrupted streaming video is a challenging under-determined problem. The proposed saliency estimation method uses the spatial component from the convex approximation mentioned above, but improves temporal saliency estimation via global motion compensation. Overall, this method is not convex, but is more accurate than the IKN saliency model on certain sequences with camera motion. This method uses motion vectors to accomplish global motion compensation and subsequently temporal saliency estimation, so it can be very attractive for video compression applications.

In our proposed saliency-aware video compression method, we combined a saliency distortion term with the conventional MSE distortion metric. As a possible future work, the conventional MSE distortion metric can be replaced by a more perceptually-relevant distortion metric such as SSIM to achieve even better result

REFERENCES

[1] M. A. Robertson and R. L. Stevenson, "DCT quantization noise in compressed images," IEEE Trans. Circuits Syst. Video Technol., vol. 15, no. 1, pp. 27–38, Jan. 2005.

[2] A. Leontaris, P. C. Cosman, and A. R. Reibman, "Quality evaluation of motion-compensated edge artifacts in compressed video," IEEE Trans. Image Process., vol. 16, no. 11, pp. 943–956, Apr. 2007.

[3] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," IEEE Trans. Pattern Anal. Mach. Intell., vol. 20, no. 11, pp. 1254–1259, Nov. 1998.

[4] L. Itti, "Automatic foveation for video compression using a neurobiological model of visual attention," IEEE Trans. Image Process., vol. 13, no. 10, pp. 1304–1318, Oct. 2004.

[5] Z. Chen, N. K. Ngan, and W. Lin, "Perceptual video coding: Challenges and approaches," in Proc. IEEE ICME, Jul. 2010, pp. 784–789.

[6] Z. Li, S. Qin, and L. Itti, "Visual attention guided bit allocation in video compression," Image and Vision Computing, vol. 29, no. 1, pp. 1–14, 2011.

[7] Y. Liu, Z. Li, and Y. C. Soh, "Region-of-interest based resource allocation for conversational video communications of H.264/AVC," IEEE Trans. Circuits Syst. Video Technol., vol. 18, no. 1, pp. 134–139, Jan. 2008.

[8] I. E. Richardson, The H.264 Advanced Video Compression Standard. New York, NY, USA: Wiley, 2010.

[9] Z. Chen and C. Guillemot, "Perceptually-friendly H.264/AVC video coding based on foveated just-noticeable-distortion model," IEEE Trans. Circuits Syst. Video Technol., vol. 20, no. 6, p. 806–819, Jun. 2010.

BIOGRAPHIES

First Author Mrs.P.Muthuselvi was born Srivilliputtur,Virudhunagar Dt ,Tamilnadu,India. She had completed her B.Sc (CS) from Madurai Kamaraj University,Madurai in 2011. She had completed her M.Sc(CS&IT) from Madurai Kamaraj University,Madurai in 2013.She achieved her Master of Engineering(M.E (cse)) from Anna University,Chennai 2015 . She is the Assistant professor of Department of Computer Science in VPMM Arts & Science College for Women, Krishnankoil, Srivilliputtur(Tk), Virudhunagar(Dt),Tamilnadu, India since 2016.

Second Author Mr.T.Vengatesh was born in Virudhunagar, Tamilnadu,India. He had completed his BCA from Madurai Kamaraj University.Madurai in 2009. He achieved his Master of Computer Application Degree(MCA) from Anna University,Chennai in 2012. He