# Exploring Visual Temporal Masking for Video Compression

Velibor ADZIC, Hari KALVA, and Borko FURHT

***Abstract*—In this paper we present work on exploiting visual temporal masking phenomenon applied to video compression. Our results show that it is possible to reduce bitrate of the compressed video sequence without affecting subjective quality and quality of experience (perceptually lossless). The principles we present here are applicable to all modern hybrid coding systems and can be implemented seamlessly with video delivery platforms. Results show up to 6% of additional savings when implemented on top of the state of the art encoder.**

## I. INTRODUCTION

Modern video compression algorithms rely in some part on characteristics of human visual system (HVS). However, there are many findings in psycho-visual studies that haven't been explored in the context of video compression applications. One such finding is the phenomenon of temporal visual masking. Visual masking in temporal and spatial domain has been discovered by psychologists more than a century ago [1], [2]. It occurs when the visibility of target stimulus is reduced by the presence of mask stimulus. This paper focuses on temporal masking – particularly backward masking. It is manifested at abrupt scene changes, when new scene masks certain amount of frames from the previous scene. A number of frames that precede a scene change are essentially erased form higher levels of processing in HVS. Subject is unable to consciously perceive these frames. Although scientific community doesn't have clear explanation for this phenomenon, one of the promising explanations for backward masking could be the variation in the latency of the neural signals in the visual system as a function of their intensity [3]. Detailed overview of models and findings in visual backward masking can be found in [4]. Our goal is to explore visual masking phenomena and its potential benefits for video compression.

## II. RELATED WORK

Although significant amount of research related to visual masking and signal processing has been done over past years, it is mostly focused on spatial masking for image compression [5], [6]. As far as temporal masking is concerned seminal paper by Girod [7] explores forward masking - showing that there is some form of masking effect immediately after scene change. Tam et al. [8] investigated the visibility of MPEG-2 coding artifacts after a scene cut and found significant visual masking effects only in the first subsequent frame. Carney et al. [9] investigated levels of sensitivity of HVS to blur in the first 100-200 milliseconds (ms) after scene cut. However, in our initial tests we couldn't confirm any usefulness of forward masking, since subjects began to notice distortions even in the first frame after scene cut. Hence, we focused on backward masking trying to achieve better results.

For the reasons that are not apparent, much less work has been done towards application of backward masking to video coding. One reason could be that backward masking requires buffering to detect scene changes and hence not suitable for realtime broadcast applications. Majority of video services over the web today are on-demand services delivering pre-coded video and are best suited to exploit backward masking. Pastrana-Vidal et al. [10] studied the presence of backward and forward temporal masking based on visibility threshold experiments using video material in common intermediate format (CIF) resolution (352x288 pixels). They simulated a single burst of dropped frames near a scene change, for different impairment durations from 0 to 200 ms. The transitory reduction of the HVS sensibility was reported to be significant in the first 160ms for forward masking and up to 200ms for backward masking. Study by Huynh-Thu and Ghanbari [11] also showed that backward masking is more significant than forward masking. They used burst of frozen frames as stimulus and scene cut as mask. These papers, however, do not explore the impact of masking on video bitrate and bandwidth reduction. Our work is the first in which possibilities of bitrate savings based on backward masking are explored.

## III. EXPERIMENTS AND RESULTS

Our experiments were aimed at discovering how bitrate can be saved by introducing distortions or impairments in the frames just before scene change. We tested both frame freezing and our proposal (more aggressive quantization). In order to confirm our hypothesis we conducted experiments with sequences obtained using process flow shown in Figure 1. The process is similar to traditional two-pass coding. However, algorithm can be applied immediately without the need for first pass and parsing if we know positions of the scene changes. Such list of I-frames can be supplied as metadata with original sequence. We used "x264" open source encoder to produce H.264 bitstream sequence. Our source dataset contained 20 video sequences with standard definition

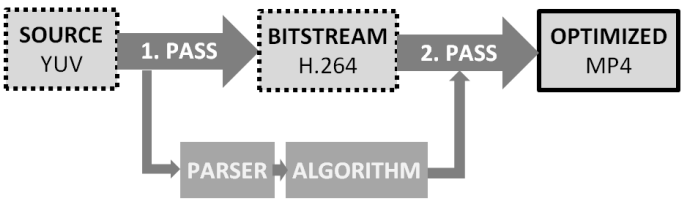


Fig. 1. Flow of the experimental setup.

978-1-4673-1363-6/13/$31.00 ©2013 IEEE

TABLE I
BITRATE SAVINGS FOR DIFFERENT METHODS COMPARED TO BASELINE

| CBR (bitrate) | SAVINGS Freezing, N=3 | Freezing, N=5 | TMFQ, M=10 |
|---|---|---|---|
| 900kbps | 1.54% | 3.48% | 5.66% |
| 1300kbps | 1.53% | 3.46% | 5.82% |
| 1900kbps | 1.54% | 3.45% | 5.58% |

resolution (SD, 720x480p) obtained from DVD sources. Videos are 30 second long clips from popular feature and animated movies and music videos – in general, content that is very popular and generates most of the traffic on the Internet. All videos were presented at 25 frames per second (fps) on the 20 inch monitors, in a setting that complies with ITU-R recommendation BT.500-11. Subjects were 5 students with normal or corrected to normal vision.

Freezing was implemented by repeating last selected frame until the scene change; i.e., a set of N frames immediately before a scene change are replaced by an identical frame. Temporally masked frame quantization (TMFQ) was implemented by raising quantizing parameter (QP) for target window of M frames immediately before a scene change. Last two frames were quantized with QP = 51 (maximum allowed in H.264). For the rest of preceding frames (M-2) we implemented sigmoid-like ramp that gracefully lowered the QP increase.

The first set of experiments showed that freezing can be applied with limited success for frames in the range of 100 – 200 ms before scene change. At least one subject noticed freezing impairments for $N \geq 3$. For more than 5 frames impairments were noticeable in 100% of cases. Freezing was reported as being most annoying for the frames at the end of a scene that contain high motion activity.

For second set of experiments we targeted perceptually lossless optimization using TMFQ. We hypothesized that high quantization is not going to impair the whole motion flow, and hence will have higher threshold of noticeability. Subject reports were analyzed in order to find the limit at which there are 0% of reported distortions (perceptually lossless). We were able to achieve this for M = 10 frames before scene cut, using the ramp described earlier. TMFQ allowed for additional distortions in more frames than freezing. Our hypothesis for better results with TMFQ was confirmed. Achieved savings are presented in Table 1 (Freezing with N = 3 and 5 and TMFQ with M = 10). Savings are calculated compared to constant bitrate H.264 coding (CBR). We benchmarked CBR as baseline because it is used in platforms such as adaptive streaming which are reported to contribute the most to video traffic on the Internet.

TMFQ can be implemented together with other techniques, such as optimized adaptive streaming [12] to introduce improved bitrate savings. These savings (~6%) are not negligible given that recent study estimated that the sum of all forms of video will exceed 86% of global consumer traffic by the year 2016 [13].

## IV. CONCLUSIONS

Using cues from HVS studies can be very beneficial for modern video compression optimization. Once we reach limits of traditional hybrid coding the most promising path of improvement is further exploration of psycho-visual and perceptual studies. Our experiments have demonstrated that visual temporal masking can be used to achieve savings in bitrate.

Our algorithm can be used in conjunction with all modern video coding standards and on top of popular platforms for the delivery of on demand content (such as adaptive streaming over HTTP). Furthermore, all principles explored in this paper are expected to work on top of future standards (i.e. High Efficiency Video Coding - HEVC).

Algorithm can be implemented for live video streaming in the scenarios which allow short delay. The only information that is needed in advance is the position of scene change.

Implementation of psycho-visual algorithms such as the one that we introduced here can have significant impact on bandwidth usage optimization, given the trend of fast growing video content related traffic on the Internet.

## REFERENCES


[1] C.S. Sherrington, "On the reciprocal action in the retina as studied by means of some rotating discs," *J. Physiology* 21, 1897, p. 33–54.

[2] W. McDougall, "The sensations excited by a single momentary stimulation of the eye," *Brit J. Psychol* 1, 1904, p. 78–113.

[3] A.J. Ahumada Jr., B.L. Beard and R. Eriksson, "Spatio-temporal discrimination model predicts temporal masking function," *Proc. SPIE Human Vision and Electronic Imaging*, vol. 3299, 1998, pp. 120–127.

[4] B.G. Breitmeyer and H. Ogmen, "Recent models and findings in visual backward masking: A comparison, review, and update," *Percept Psychophys* 62, 2000, pp. 1572–1595.

[5] A.N. Netravali and B. Prasada, "Adaptive quantization of picture signals using spatial masking," *Proceedings of the IEEE* , vol.65, no.4, pp. 536- 548, April 1977.

[6] M. Naccari and F. Pereira, "Comparing spatial masking modelling in just noticeable distortion controlled H.264/AVC video coding," *11th International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS)*, 2010 , vol., no., pp.1-4.

[7] B. Girod, "The information theoretical significance of spatial and temporal masking in video signals," *Proc. SPIE Human Vision, Visual Processing and Digital Display*, vol. 1077, 1989, pp. 178–187.

[8] W.J. Tam, L.B. Stelmach, L. Wang, D. Lauzon and P. Gray, "Visual masking at video scene cuts," *Proc. SPIE Human Vision, Visual Processing and Digital Display*, vol. 2411, 1995, pp. 111–119.

[9] Q. Hu, S.A. Klein and T. Carney, "Masking of high-spatial-frequency information after a scene cut," *Society for Informational Display 93 Digest*. n. 24, 1993, p. 521-523.

[10] R.R. Pastrana-Vidal, J.-C. Gicquel, C. Colomes and H. Cherifi, "Temporal Masking Effect on Dropped Frames at Video Scene Cuts," *Proc. SPIE Human Vision and Electronic Imaging IX*, vol. 5292, 2004, pp. 194-201.

[11] Q. Huynh-Thu and M. Ghanbari, "Asymmetrical temporal masking near video scene change," *ICIP 2008. 15th IEEE International Conference on Image Processing*, vol., no., pp.2568-2571.

[12] Adzic, V.; Kalva, H.; Furht, B.; , "Optimizing video encoding for adaptive streaming over HTTP," *Transactions on Consumer Electronics, IEEE* , vol.58, no.2, pp.397-403, May 2012.

[13] Cisco, "Visual Networking Index Services Adoption (VNI SA) Forecast, 2011-2016," *Whitepaper*, 30 May 2012.