Reconstruction of incomplete decoded videos for use in objective quality metrics

Sedano, Inigo; Kihl, Maria; Brunnström, Kjell; Aurelius, Andreas

Published in:
[Host publication title missing]

2012

RECONSTRUCTION OF INCOMPLETE DECODED VIDEOS FOR USE IN OBJECTIVE QUALITY METRICS

Iñigo Sedano[1], Maria Kihl[2], Kjell Brunnström[3], Andreas Aurelius[3]

[1]TECNALIA, Telecom Unit, Zamudio, Spain, [2]Dept. of Electrical and Information Technology, Lund University, Sweden, [3]Acreo AB, Networks and Systems Lab (NETLAB), Kista, Sweden

ABSTRACT

Many full-reference objective metrics require that the original and the degraded video contain the same number of frames. Most codecs are not able to decode the video properly when the video is subject to packet losses and produce incomplete video files with lower number of frames than the original video. In this paper we present a simple method to reconstruct the degraded videos so that it has the same length as the original. Information regarding the frame numbers is inserted into the original uncompressed video and used later on to identify the missing frames and reconstruct the degraded video.

Index Terms**—** Objective evaluation techniques, decoder, transmission distortions, reconstruction, video quality

1. INTRODUCTION

The quality of a video can be evaluated using objective metrics or subjective evaluations by a panel of observers. The objective metrics are very useful because they are quicker and cheaper compared to the subjective assessments.

The objective metrics can be classified in three different types: no-reference, reduced-reference and full-reference. In the no-reference type only the degraded video is available. In the reduced-reference the degraded video and some reduced information related to the original video are available. In the full-reference both the original video and degraded videos are available.

Most of the full-reference objective metrics compare the videos frame by frame and later on an overall quality index is computed. Thus most of the metrics require that the original video and the degraded video contain the same number of frames. PSNR, SSIM [1] [2] and MS-SSIM [3] are some examples.

The decoders are able to decode the video properly when it has been compressed but no data has been lost. When packet losses are introduced most codecs are not able to decode corrupted video files properly. In some cases the decoder produces fewer frames (skip frames in some parts) while in others the decoder crashes.

The scenario in which transmission distortions are introduced is interesting because videos are increasingly being delivered over IP networks and the packet losses are inherent to those networks.

In this paper we describe a method to reconstruct the degraded video so that it has the same length as the original. The method has been implemented in language C.

Finally we compare our algorithm with the algorithm developed by S. Wolf [4] that is capable of estimating variable video delays. His algorithm determines the best matching original (or input) video frame for each processed (or output frame). Thus it could be used to remove the variable frame delays from the processed sequence as a calibration step before computing a full-reference quality measurement.

2. RECONSTRUCTION

Ideally the video quality metrics should be able to realign the frames and evaluate the videos considering the position and duration of the missing frames. Even if the video quality metrics could be applied when the number of frames of the original video and degraded (not reconstructed) are different, the shift due to the missing frames would introduce significant error in the video quality evaluation, as shown in Fig. 1.

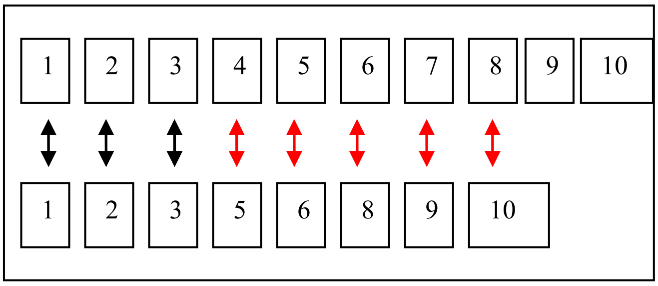

Fig. 1. Error in quality metrics due to shift caused by missing frames (original video sequence and degraded video).

The error in the video quality evaluation is minimized when the reconstructed video is used instead, as shown in Fig. 2.

This work was supported by Fundación Centros Tecnológicos – Iñaki Goenaga. The work was also partly financed by the CELTIC project IPNQSIS, with the Swedish Governmental Agency for Innovation Systems (VINNOVA) supporting the Swedish contribution.

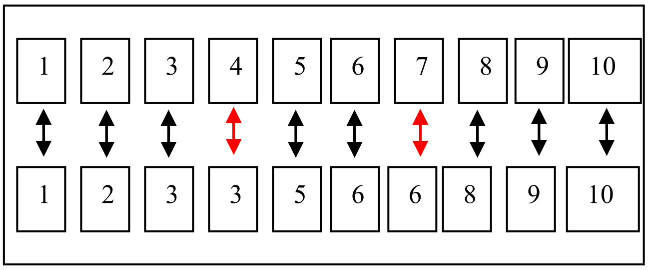

Fig. 2. Error in quality metrics is minimized using the reconstructed video (original video sequence and reconstructed video).

It would be possible to reconstruct the video using different methods instead of repeating the previous frame to the missing one. For example it would be possible to repeat the next frame to the missing one or to insert a white or black frame. The algorithm proposed in this paper has been applied in the development of no-reference metrics, confirming the validity of our approach [5].

In order to reconstruct the distorted video, it is necessary to first detect which frames are missing. The detection of the missing frames is performed by inserting the frame numbers into the original uncompressed sequence. The frame numbers are inserted in the images in different positions. The video is then compressed, packet losses are introduced and the video is uncompressed, resulting in an incomplete video file due to the decoder problems. The frame numbers are then extracted from the frames and used to reconstruct the video. The Fig. 3 shows the whole process.

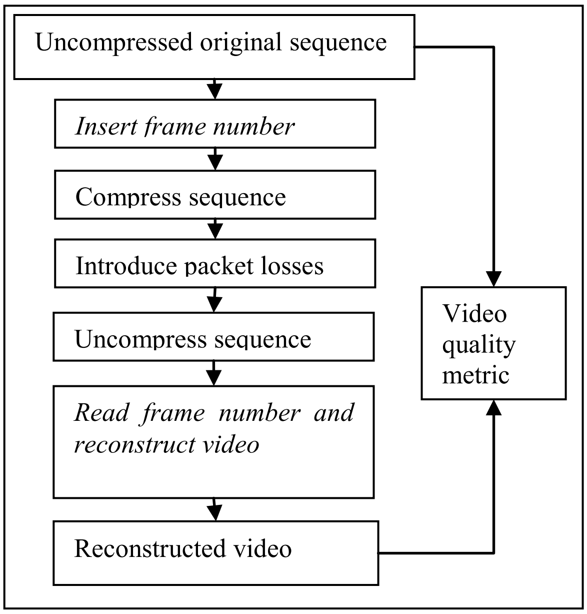

Fig. 3. The frame number is inserted in the frames and used in the reconstruction process.

3. FRAME NUMBER INSERTION

The original uncompressed video is in planar format YUV 4:2:0. YUV formats fall into two distinct groups, the packed formats where Y, U (Cb) and V (Cr) samples are packed together into macropixels which are stored in a single array, and the planar formats where each component is stored as a separate array, the final image being a fusing of the three separate planes. The planar YUV 4:2:0 is an 8 bit Y plane followed by 8 bit 2x2 subsampled U and V planes. An NxN Y plane is followed by (N/2)x(N/2) U and V planes.

TABLE I
YUV 4:2:0 SUBSAMPLING

	Horizontal	Vertical
Y sample period	1	1
U sample period	2	2
V sample period	2	2

The frame number is inserted into the Y plane. The frames are numbered starting from 1. The frame number is represented by 16 bits, being the most significant to the left. Each of the bits is mapped to one pixel. If the bit is 0 the luminance byte takes the value 0, and if the bit is 1 the luminance byte takes the value 255. The XOR of the frame number is also calculated and inserted into the luminance information of the video in the same way.

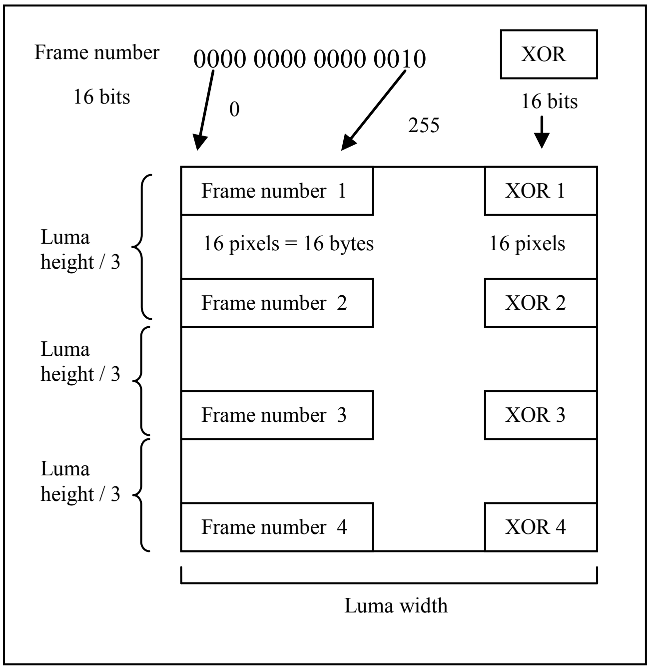

Fig. 4. Frame number and XOR positions.

Fig. 4 shows in which positions of the frame the frame number and the XOR are inserted (the same number is inserted four times and the same XOR is inserted four times). The objective is to maximize the probability that the number is correctly read despite the degradation of the image. The XOR is used to verify if the number read is correct.

4. READ NUMBER AND RECONSTRUCT FILE

The reconstruction program receives as input parameters the YUV input file, the YUV output file, the frame resolution and the total number of images.

The frame number and XOR are read from four different positions, as shown in Fig. 4. In the process of reading the numbers and XORs a bit is 0 if the pixel value is lower than 127 and 1 if the pixel value is higher than 127. The number and XOR combinations verified are summarized in the table II.

TABLE II
COMBINATIONS VERIFIED

Number	XOR
1	2, 3 and 4
2	1, 3 and 4
3	1, 2 and 4
4	1, 2 and 3

The number 1 is considered valid if the combination frame number 1 XOR 2 is correct, or the combination frame number 1 XOR 3 is correct or the combination frame number 1 XOR 4 is correct and the number is equal or lower than the total number of images in the video. The same idea applies to the rest of the numbers.

Fig. 5 explains the complete algorithm. Once the whole video has been read if there are more frames to be inserted the last frame inserted is repeated until completing the total number of frames. On the other side, if the first frame is missing the first processed frame is repeated.

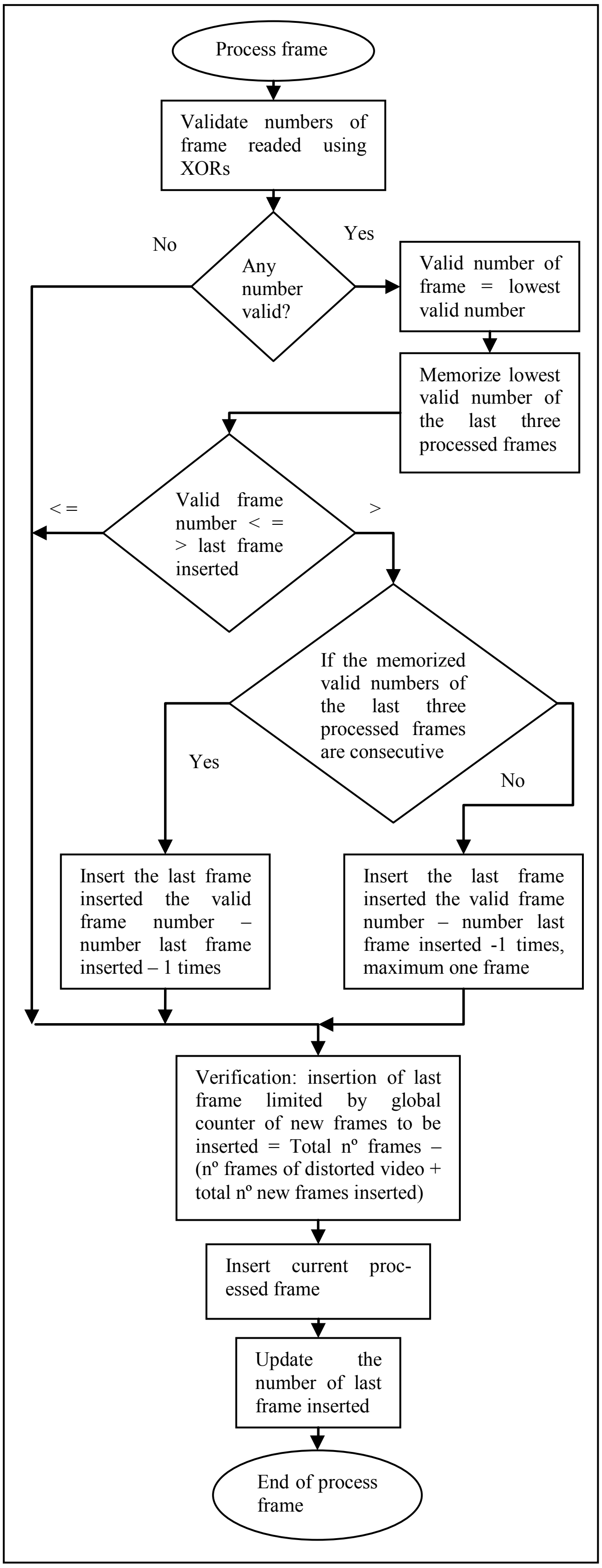

Fig. 5. Reconstruction tool algorithm.

5. RELIABILITY

In order to evaluate the reliability of the tool the following framework was considered (shown in Fig. 6).

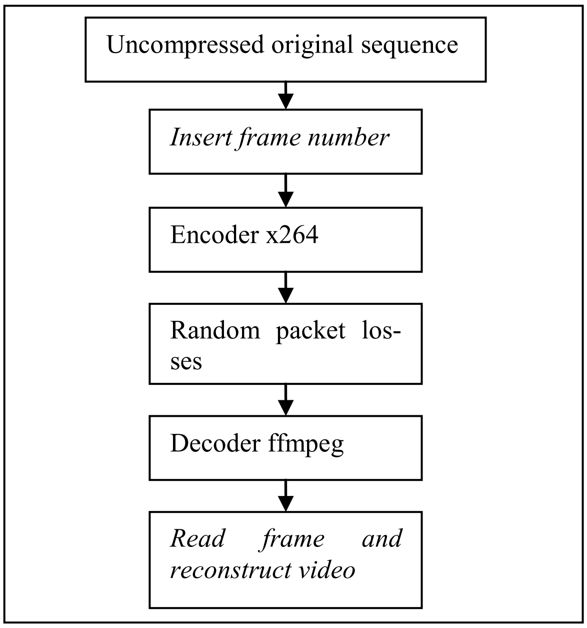

Fig. 6. Framework for the evaluation of the reliability of the method.

The frame number was inserted into the original uncompressed sequence. The sequence was then encoded with the encoder x264 with a Quantization Parameter between 26 and 44. The maximum interval between IDR-frames was set to 12, 36, 60 and 84. Random packet losses were inserted using a packet loss simulator [6]. The video was then decoded using ffmpeg, producing a video with missing frames in some parts. The video was finally reconstructed using the frame numbers.

Each frame number and XOR combination was manually checked and compared with the other combinations. The verification of the frame numbers and XOR combinations in different parts of the image decreases the probability of reading erroneous numbers. The method is reliable for packet loss rates of around 10% (ffmpeg decoder used). The reliability increases with lower packet loss rates.

6. COMPARISON WITH OTHER METHODS

The method by S. Wolf [4] is a full-reference method for estimating variable frame delays. The best matching original (or input) video is determined for each processed (or output) video frame. The algorithm involves the computation of the Mean Square Errors (MSEs) between each processed frame and the set of original frames that are within a user-specified temporal search window. The algorithm only utilizes the luminance images of the video clips.

If we compare both methods, we appreciate that our method is better suited for the case of still or nearly still video or video with repetitive motion. In both methods the position of the frames is identified. However, in our method we include the reconstruction of the distorted video, providing the possibility of directly applying video quality metrics that require videos of the same length, like PSNR or SSIM. Both methods require access to the original uncompressed reference. However, in our method the access to the original uncompressed reference is necessary prior to the processing of the video (compression, introduction of packet losses and decoding) in order to introduce the frame number in the video. The method described by S. Wolf implies more complexity than our method.

7. CONCLUSIONS

The method enables the use of full-reference video quality metrics when the decoder is not able to produce the same number of frames as in the original uncompressed video. The algorithm was implemented, tested and was found to be reliable for packet losses up to 10%. Future work includes the improvement of the accuracy for packet losses greater than 10%.

8. REFERENCES

[1] Z. Wang, A. C. Bovik, H. R. Sheikh and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600-612, Apr. 2004.

[2] The Structural SIMilarity (SSIM) index author's home page [Online]. Available: http://www.ece.uwaterloo.ca/~z70wang/research/ssim/

[3] Z. Wang, E. P. Simoncelli and A. C. Bovik, "Multi-scale structural similarity for image quality assessment," in *Proc. IEEE Asilomar Conference Signals, Systems and Computers*, Nov. 2003.

[4] S. Wolf, "A Full Reference (FR) Method Using Causality Processing For Estimating Variable Video Delays", NTIA Technical Memorandum TM-10-463 [Online]. Available: http://www.its.bldrdoc.gov/pub/ntia-rpt/10-463/10-463_ocr.pdf

[5] I. Sedano, K. Brunnström, M. Kihl and A. Aurelius, "Full-reference video quality metric assisted development of no-reference video quality metrics for real time network monitoring," *IEEE Journal of Selected Topics in Signal Processing*, 2012. Submitted.

[6] JVT-Q069 [Y. Guo, H. Li, Y.-K. Wang] SVC/AVC loss simulator [Online]. Available: http://wftp3.itu.int/av-arch/jvt-site/2005_10_Nice/