SSIM modification for video error concealment methods quality assessments*

Dmitiry Kulikov
Department of Computational Mathematics and Cybernetics
Moscow State University, Moscow, Russia
dkulikov@graphics.cs.msu.ru

Abstract

The typical problem in video streaming is a packet-loss during signal transmission. Due to this fact different error concealment methods are developed. The best possible way for comparing processing quality of these methods is subjective quality assessments. But due to fact this type of estimation is very difficult and expensive to perform objective quality metrics are widely used. The most well-known are PSNR and SSIM, last one is closer to human perception. The main drawback for this metrics in video quality assessments is per-frame metrics value that does not take into account any temporal artifacts which could be more sensible to the human than static artifacts. This article is devoted to new metric development based on SSIM, that could estimate temporal artifacts as well as spatial. This new metric was tested and compared to original PSNR and SSIM for different error concealment methods with help of SAMVIQ methodology for subjective quality assessments.

Keywords: *video, error concealment, quality assessments, SSIM modification.*

1. INTRODUCTION

The task of automatic quality estimation for video error concealment methods is very important, because subjective quality assessments performing is very expensive and time-consuming task. But this task could not be solved solely with mathematical tools, because many methods during processing do not recover lost information, but change it to most similar in the frame or sequence. And the task of quality assessments has an addition not only to estimate the closeness of the original and processed video sequences, but also take into account visually noticeable artifacts. Quality measurement algorithms are based on human visual system imitation[1]. But human visual system is not fully investigated yet for truly adequate automatic estimation algorithms construction [1, 2]. This situation forces to develop new quality estimation algorithms and to find approaches to increase adequateness of known metrics. The typical work-flow for video processing algorithm quality assessment is to compute some "spatial" metric value for each frame and than to average per-frame values for the whole video sequence. The main quality metrics are PSNR and SSIM[3].

PSNR (Peak to Signal Noise Ratio) estimates inverse mean square error in logarithmic scale actually. The main idea of SSIM (Structural Similarity Index Measure) computation is to consider three main types of distortion: luminance distortions, contrast distortions and structure distortions. Every component is very essential for human perception and its combination could approximate human video or image estimation. Many experiments show that SSIM is more adequate to the human perception than PSNR and it is preferable to use while performing quality assessments tasks. But for video error concealment methods quality estimation even SSIM is not always adequate. To estimate metric adequateness the experiment described in Sec. 3. was performed. To etsimate metric correlation to subjective marks Kendall's τ was used. It shows that when average $\tau = 0.67$ for some test vectors $\tau = 0.42$. The main feature of these vectors is slow motion, processed vectors have artifacts in temporal area, e.g. not monotonous motion of some processed parts of the frames. This low correlation is explained by the fact that per-frame SSIM metric could not estimate such artifacts, but human eye is very sensible to it and sometimes temporal artifacts are more important than spatial. And because of it new metric is proposed in this article to take into account temporal artifacts.

2. PROPOSED METHOD DESCRIPTION

2.1 Algorithm schema

The main idea of proposed algorithm is to perform concealment area characteristics consistency analysis in addition to standard SSIM computation. This analysis is done by the motion vector in concealed blocks deviation power estimation comparing to the reference frame of the source and processed video sequences.

The algorithm schema is shown at Fig. 1. First of all SSIM calculation for two frames (source F_t and processed G_t) is performed. The first intensional step of algorithm is motion vector field homogeneity analysis performed for the source(F_t and F_{t-1}) and processed (G_t and G_{t-1}) video sequences. The second step is distortion coefficient computation. And the final step is SSIM temporal modification (SSIMt) metric value calculation for the whole frame.

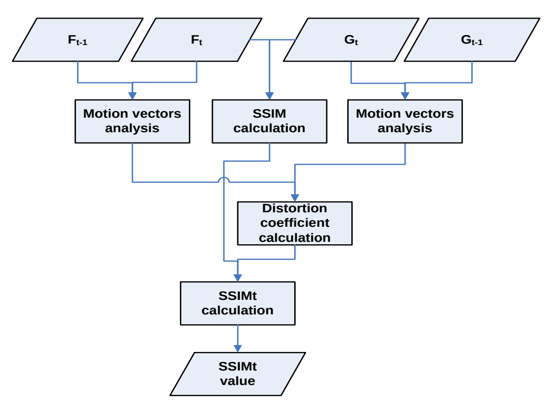

Figure 1: SSIMt calculation schema

*This work was supported by RFBR grant 07-01-00759-a.

2.2 Motion vector field homogeneity analysis

Motion homogeneity for video sequence could be performed by means of motion vector mechanism with the following approach. The homogeneity coefficient is computed as a ratio of average motion vector consistency (comparing to the neighbors) to similar value for motion vector of corresponding block in reference frame. It could be done using next algorithm:

1. Adaptive threshold calculation for horizontal and vertical components of the motion vector. This threshold is proportional to neighbors vectors value dispersion:
$\sigma_c^{TH} = \sqrt{\sum_{mv^j \in \theta(mv^i)} \left(mv_c^{AVG} - mv_c^j\right)^2 / \sum_{mv^j \in \theta(mv^i)} 1}$
where c=x or y, mv_x and mv_y – horizontal and vertical components of motion vectors, mv^{AVG} – average value of the corresponding components of vectors, θ – neighborhood of motion vector.

2. Motion vector matching value computation comparing to its neighbors using calculated thresholds:

$$\beta(mv_i) = \sqrt{\beta_x^2 + \beta_y^2}, where \quad (1)$$
$$\beta_x = \sigma_x^{TH} - \alpha|mv_x^{AVG} - mv_x^i| \quad (2)$$
$$\beta_y = \sigma_y^{TH} - \alpha|mv_y^{AVG} - mv_y^i|, \quad (3)$$

where α – weighting coefficient.

3. Motion vector matching coefficients for the current and reference frame ratio calculation:

$$\beta = \left(\beta(mv_i(t)) + \varepsilon\right) / \left(\beta(mv_i(t-1)) + \varepsilon\right), \quad (4)$$

where ε – small coefficient to prevent devision by zero.

2.3 Distortion coefficient and metric value calculation

During concealed area (block) temporal visibility estimation the analysis have to be performed in comparison to original undistorted video sequence because of the fact that even strong differences in temporal area could be internal video sequence characteristics and do not connect with video processing algorithm.

And because of it distortion coefficient is calculated as a ratio between accordance coefficients for source video sequence B_f and reference – B_g:

$$\lambda(B) = (\min(\beta(B_f), \beta(B_g)))/(\max(\beta(B_f), \beta(B_g))) \quad (5)$$

SSIMt metric value calculation is performed using SSIM value and calculated distortion coefficient:

$$SSIMt(B) = \alpha SSIM(B) + (1-\alpha)\lambda(B), \quad (6)$$

where α – empiric weighting coefficient.

For metric value for the whole frame average values for the all blocks are used and for the video sequence metric value average per-frame values are used.

3. EXPERIMENTAL RESULTS

To perform metrics adequateness estimation comparing to subjective marks the next experiment was carried out. The test set of video sequences was chosen, it contains 4 different video sequences with different motion patterns, each sequence was corrupted and restored using 7 different video error concealment methods. Thus there are 32 test vectors including original ones. 10 experts participate the experiment.

The typical situation during the experiment is the next: two different methods A and B results are very close visually and by objective metrics at one frame, but result of method B is temporal inconsistent, for example it produces block flicking. These changes are not quite big and not estimated by frame by frame comparison, but produce strong temporal visual artifacts.

For SSIMt metric adequateness to human perception and comparing it to original SSIM and PSNR the next approach is used – for every test vector subjective mark is received using SAMVIQ methodology for subjective quality assessments from European Broadcasting Union[4]. To measure the degree of correspondence between objective and subjective marks the Kendall τ rank correlation coefficient is used. The average τ coefficients for three different are $\tau_{PSNR} = 0.6786$, $\tau_{SSIM} = 0.6964$ and $\tau_{SSIMt} = 0.7143$, comparing to average subjective marks received during test. The proposed approach leads to rank coefficient increase that means more adequate metric in the terms of correlation between objective and subjective tests.

4. CONCLUSION AND FUTURE WORK

In this article the new approach for quality metric is proposed. It is based not only on spatial metrics calculation, but also on temporal artifacts estimation using motion vector field analysis. This approach leads to more adequate metrics value in the terms of human perception (using subjective tests) comparing to original SSIM and PSNR and could be used in different video processing methods quality estimation, video error concealment methods. On of future work directions is to perform complex comparison to other motion-based SSIM modifications like [5]. The goal of new metric construction is to produce metric for video error concealment methods evaluation that is close to subjective marks. And proposed approach could be a good basis for new metrics.

5. REFERENCES

[1] Marcus J. Nadenau, David Alleysson, and Murat Kunt, "Human vision models for perceptually optimized image processing – a review," in *submitted to Proc. of the IEEE*, September 2000.

[2] Bernd Girod, "What's wrong with mean-squared error?," *Digital images and human vision*, pp. 207–220, 1993.

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

[4] Steinmann V. Sunna P. Wyckens E Kozamernik, F., "Samviq – a new ebu methodology for video quality evaluations in multimedia," *SMPTE motion imaging journal*, vol. 114, no. 4, pp. 152–160, 2005.

[5] Qiang Li and Zhou Wang, "Video quality assessment by incorporating a motion perception model," in *Image Processing, 2007. ICIP 2007. IEEE International Conference on*, 2007, vol. 2, pp. 173–176.

ABOUT THE AUTHOR

Dmitriy Kulikov is a Ph.D. student at Moscow State University, Department of Computational Mathematics and Cybernetics. His contact email is `dkulikov@graphics.cs.msu.ru`.