TELKOMNIKA, Vol.16, No.2, April 2018, pp. 909~914
ISSN: 1693-6930, accredited **A** by DIKTI, Decree No: 58/DIKTI/Kep/2013
DOI: 10.12928/TELKOMNIKA.v15i4.9044

Reduced-reference Video Quality Metric Using Spatio-temporal Activity Information

Farah Diyana Abdul Rahman*[1], Ahmad Imran Ibrahim[2], Dimitris Agrafiotis[3]
[1,2]Department of Electrical and Computer Engineering, International Islamic University Malaysia
[3]Department of Electrical and Electronic Engineering, University of Bristol, BS8 1UB, UK
*Corresponding author, e-mail: farahdy@iium.edu.my

Abstract

Monitoring and maintaining acceptable Quality of Experience is of great importance to video service providers. Perceived visual quality of transmitted video via wireless networks can be degraded by transmission errors. This paper presents a reduced-reference video quality metric of very low complexity and overhead that makes use of frame based spatial (SI) and temporal (TI) activity levels to monitor the effect of channel errors on video transmitted over error prone networks. The performance of the metric is evaluated relative to that of a number of full and reduced reference metrics. The proposed metric outperforms some of the most popular full reference metrics whilst requiring very little overhead.

Keywords*: wireless LAN, quality of service, video signal processing, reduced-reference metric, objective video quality metric*

1. Introduction

Strong demand for digital video and expectations among consumers for good quality has made the assessment of the end-user video quality an important issue that needs to be addressed. This is even more so in the case of error prone wireless video transmission and, in particular, multicast wireless video transmission. In multicast transmission, automatic repeat requests are not allowed, making the transmitted video stream more prone to channel errors. In addition, multiple users connected to the same session can experience different levels of video quality depending on the channel conditions that each one of them encounters.

Accurate video quality assessment can be conducted through time-consuming subjective video quality tests which are impossible to conduct in real-time. Objective quality metrics are thus usually employed for estimating the perceived quality of the received video in real-time. In this time-sensitive scenario, reduced-reference (RR) and no-reference (NR) metrics are more suitable for assessing the received video quality than full-reference (FR) because they require limited or no information from the original encoded or transmitted video [1].

Peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) are two of the most popular full reference objective quality metrics [2]. More advanced FR metrics that take into account more perceptual aspects include the visual signal-to-noise ratio (VSNR) [3], video quality model (VQM) [4], motion tuned spatial-temporal quality assessment method (MOVIE) [5], spatio-temporal most apparent-distortion (STMAD) [6] and perception-based video quality metric (PVM) [7].

RR video quality metrics offer a promising solution for monitoring the effects of the transmission channel on the video quality as they tend to offer better performance than NR metrics and can have fairly low computational and overhead requirements. A number of RR metrics [8-10] extract spatial and temporal features as the means to form the RR information. The STRRED method [11] employs spatio-temporal entropic differences for performing the quality assessment. STIS-SSIM [12] combines spatio-temporal selection with a modified SSIM-based framework.

In this paper we examine the possibility of extracting video quality information at the receiver by analysing the spatial and temporal information of a transmitted video stream at both the transmitter (encoder) and receiver (decoder). In particular we make use of frame based spatial activity (SI) and temporal activity (TI) values to form the Spatio-Temporal Information Reduced Reference Metric (STIRR).

Received September 10, 2017; Revised December 18, 2017; Accepted March 29, 2018

Spatial activity has been used before in [13-14] to form an RR method that tries to estimate the PSNR of a received sequence. This RR metric was extended to NR video quality estimation in [15]. The method of [17] employs perceptual weighting parameters in order to estimate the quality of the received video through activity difference values between the transmitted and received videos. This method is fairly complex and produces sizeable side information (one value per block of pixels). The rest of the paper is organised as follows. Section 2 describes the proposed RR metric (STIRR). Section 3 describes the evaluation procedure followed and presents the results collected, including performance comparisons. Finally, conclusions and suggestions for future work are given in Section 4.

2. Research Method

One of the factors affecting the perceptual video quality is the amount of spatial and temporal details in a video [16]. ITU Recommendation P.910 specifies how to calculate SI and TI for the purpose of characterizing video complexity. In order to calculate the value of SI for one video frame, a Sobel filter is first applied on the luminance values. The SI value of frame F_n at time n is then equal to the standard deviation of the image resulting from convolving frame F_n with the Sobel kernel:

$$\mathrm{SI} = \left\{\mathrm{std}_{\mathrm{space}}[\mathrm{Sobel}(\mathrm{F_n})]\right\} \quad (1)$$

TI is calculated by subtracting two successive frames and taking the standard deviation of the resulting residual frame:

$$\mathrm{TI} = \left\{\mathrm{std}_{\mathrm{space}}[\mathrm{M_n(i,j)}]\right\} \quad (2)$$

where

$$M_n(i,j) = F_n(i,j) - F_{n-1}(i,j) \quad (3)$$

$F_n(i,j)$ is the pixel value at row i and column j of the n_{th} frame. At the receiver end the SI and TI values of the received video are calculated. The STIRR value for each nth frame is equal to the Euclidean distance between the two feature vector (SI,TI) of the transmitted video frame and that of the received video frame:

$$STIRR(n) = \sqrt{\sum (q - p)^2} \quad (4)$$

where $q=(q_1,q_2)$ are the coordinates of the received frame's (SI,TI) vector and $p=(p_1,p_2)$ are those of the transmitted. Frame by frame STIRR values are averaged over the length of a Group of Pictures (GOP) with the IDR frames being excluded.

Table 1. Tested IEEE 802.11n PHY Modes.

Transmission Mode	Modulation Scheme	Code rate	Data rate (Mbps)	Video bit rate (Mbps)
1	BPSK	1/2	27	8
3	QPSK	3/4	81	24
5	16-QAM	3/4	162	48

3. Results and Analysis

We are interested in the use of the STIRR metric as an indicator of the effects that packet errors (missing packets) have on the quality of compressed video. We assume that the video is transmitted over error prone wireless channels and that the video decoder at the receiver end has error concealment capabilities. The performance of the decoder's error

concealment module depends on the actual concealment method, the affected video content, the error resilience of the compressed video, and the severity of the errors (packet error rate and nature of errors). In effect we wish to be able to estimate if the quality of the video after concealment is acceptable or not so that the network link adapts to a more robust mode when the latter is the case.

3.1. Simulation Setup

We simulated wireless video transmission over IEEE 802.11n wireless networks by dropping packets according to error patterns produced by a compliant IEEE 802.11n, PHY-layer simulator [17]. The received video streams were decoded and concealed using previous frame copy (PFC) as well as motion copy (MC) concealment. The resolution of the test sequences used (*CrowdRun*, *PrincessRun* and *DanceKiss*) was 1920x1080 at 50 frames per second. Figure 1 shows a plot of the spatio-temporal activity indicators of the three test sequences (HD). *DanceKiss* at the bottom-left has the lowest spatio-temporal activity while *PrincessRun* at the top-right has the highest. All sequences contained 500 frames and were encoded using the JM H.264/AVC reference software (JM18.0-high profile) with an IPPP GOP of size 10. IDR frames were assumed to be error free, and one slice was set to be equal to one row of blocks.

The transmission modes and the video bit rates tested are summarized in Table 1. The setting used for the IEEE 802.11n simulation were as follows: MMSE detection, 800ns guard interval, channel model B (Non Line-of-Sight residential environment). For each of the three transmission modes, we tested three packet error rates-1%, 2%, and 4%, - corresponding to three different channel signals to noise ratios. For each packet error rate ten simulation runs (ten error patterns) were performed with a different starting point for the errors.

3.2. Results

The hypothesis behind this experiment is that increases in the distortion of the received video due to channel errors would result in increased differences between the STIRR values of the transmitted and received (concealed) video. To test this hypothesis we compared the STIRR difference values with three established objective quality metrics: PSNR, SSIM and VIFP. More specifically we measured the correlation (Pearson correlation coefficient) between the STIRR difference values and the quality of the received video as measured by the three selected metrics.

Table 2 and Table 3 show correlation results for the case of previous frame copy concealment and motion copy concealment respectively. The average correlation for all sequences and all metrics was 0.8 for PFC and 0.78 for MC, with values ranging from highs of 0.952 (*CrowdRun*, SSIM, MC) to lows of 0.515 (*PrincessRun*, SSIM, MC).

We additionally evaluated the performance of STIRR with LIVE Video Quality Database [18-19] (wireless transmission errors, motion copy concealment). Six different reference videos were used (*Station*, *Tractor*, *River Bed*, *Shield*, *Mobile & Calendar* and *Blue Sky)* with four error patterns per reference video. These videos are distorted according to manually adjusted strengths of wireless distortion, in order to ensure that the distorted videos are separated by different levels of perceptual distortion. The SI and TI values of these sequences are also shown in Figure 1.

Table 2. Pearson correlation between STIRR difference values and objective quality metrics for the case of previous frame copy (PFC) concealment

STIRR		PER 1%	PER 2%	PER 4%	Average
CrowdRun (PFC)	*PSNR*	0.892	0.953	0.958	0.934
	SSIM	0.891	0.960	0.958	0.936
	VIFP	0.906	0.954	0.953	0.938
PrincessRun (PFC)	*PSNR*	0.758	0.534	0.673	0.655
	SSIM	0.707	0.551	0.581	0.613
	VIFP	0.790	0.552	0.669	0.670
DanceKiss (PFC)	*PSNR*	0.807	0.793	0.790	0.797
	SSIM	0.843	0.883	0.876	0.867
	VIFP	0.820	0.837	0.777	0.811
Average		0.824	0.780	0.804	0.802

Table 3. Pearson correlation between STIRR differences (transmitted and received) and objective quality metrics for the case of motion copy (MC) concealment

STIRR		PER 1%	PER 2%	PER 4%	Average
CrowdRun (MC)	*PSNR*	0.947	0.953	0.951	0.950
	SSIM	0.968	0.958	0.931	0.952
	VIFP	0.953	0.954	0.944	0.950
PrincessRun (MC)	*PSNR*	0.758	0.534	0.596	0.629
	SSIM	0.633	0.516	0.395	0.515
	VIFP	0.790	0.552	0.467	0.603
DanceKiss (MC)	*PSNR*	0.816	0.805	0.813	0.811
	SSIM	0.782	0.836	0.863	0.827
	VIFP	0.817	0.833	0.775	0.808
Average		0.829	0.771	0.748	0.783

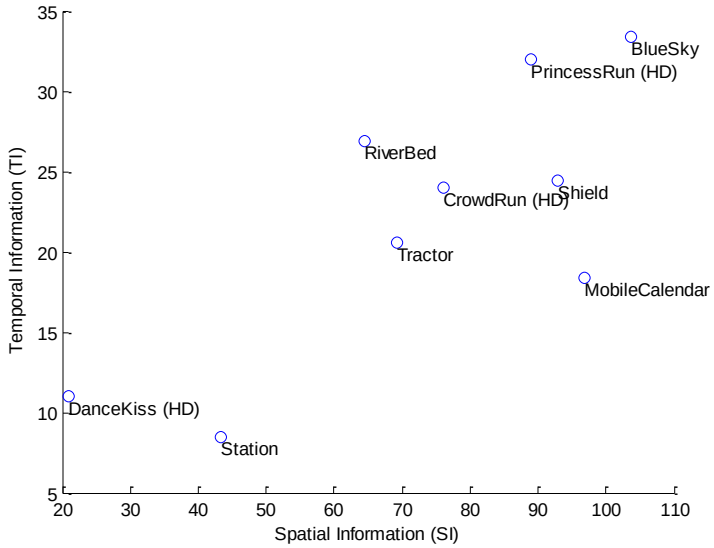

Figure 1. Spatio-temporal activity (SI-TI) indicators of the six test sequences from the LIVE database as well as the three HD test sequences used in the experiments.

Table 4 presents a summary of the performance results obtained with the tested FR and RR quality metrics using the LIVE database. The results show that despite its very low complexity STIRR is able to outperform some of the FR metrics tested (PSNR, SSIM, VFIP). Reduced reference metrics STRRED and STIS-SSIM perform better than STIRR but generate significantly more side information and thus incur much more overhead. Overhead is normalised with regards to the number of pixels in one frame (P). In addition our method exhibits very little complexity relative to all other methods (except PSNR) as shown in Table IV. Complexity was measured as the average execution time on an Intel i7-2600 CPU @ 3.40GHz PC and was normalised relative to the execution time of PSNR. All test metrics were realised in Matlab except MOVIE, which is realised in C.

Table 4. Comparison of the performance of vqa algorithms for wireless distortion (LIVE DATABASE)

Prediction Model	VQA	LCC	SROCC	Complexity	No. of scalars per frame
PSNR	FR	0.468	0.433	1	P
SSIM	FR	0.540	0.523	13	P
VIFP	FR	0.549	0.551	49	P
VQM	FR	0.733	0.721	681	P/25
MOVIE	FR	0.839	0.811	2206	P
STRRED	RR	0.804	0.786	97	P/576
STIS-SSIM	RR	0.806	0.829	9	P/256
STIRR	RR	0.623	0.624	3	P/331776

4. Conclusion

In this paper we described STIRR a very low redundancy reduced reference metric that makes use of the spatiotemporal activity values of a transmitted sequence in order to estimate the quality of the received video in the presence of errors. STIRR was found to correlate adequately with quality values estimated by a number of full reference objective quality metrics. STIRR was also shown to outperform some full reference metrics when tested on the wireless distortion part of the LIVE video database. Future work will concentrate on improving the performance of the metric through the use of further information regarding the channel and the SI/TI levels of the transmitted sequence.

Acknowledgements

The authors acknowledge the financial assistance of this research which is supported by the Research Initiative Grant Scheme (RIGS) with the grant number RIGS16-087-0251 and International Islamic University Malaysia (IIUM).

References

[1] S Chikkerur, V Sundaram, M Reisslein, LJ Karam. Objective Video Quality Assessment Methods: A Classification, Review, and Performance Comparison. *Broadcasting, IEEE Transactions.* 2011; 57(2): 165-182.
[2] W Zhou, AC Bovik, HR Sheikh, EP Simoncelli. Image quality assessment: from error visibility to structural similarity. *Image Processing, IEEE Transactions.* 2004; 13(4): 600-612.
[3] DM Chandler, SS Hemami. VSNR: A wavelet-based visual signal-to-noise ratio for natural images. *Image Processing, IEEE Transactions*. 2007; 16(9): 2284-2298.
[4] MH Pinson, S Wolf. A new standardized method for objectively measuring video quality. *Broadcasting, IEEE Transactions.* 2004; 50(3): 312-322.
[5] K Seshadrinathan, AC Bovik. Motion Tuned Spatio-Temporal Quality Assessment of Natural Videos. *Image Processing, IEEE Transactions*. 2010; 19(2): 335-350.
[6] PVVu, CTVu, DM Chandler. *A spatiotemporal most-apparent-distortion model for video quality assessment.* Image Processing (ICIP), 2011 18th IEEE International Conference. 2011: 2505-2508.
[7] F Zhang, DR Bull. *Quality assessment methods for perceptual video compression.* Image Processing (ICIP), 2013 20th IEEE International Conference. 2013: 39-43.
[8] K Zeng, Z Wang. *Temporal motion smoothness measurement for reduced-reference video quality assessment.* Proc. IEEE Int. Conf. Acoustic Speech Signal Process. 2010: 1010-1013.
[9] M Rohani, A Nasiri Avanaki, S Nader-Esfahani, M Bashirpour. *A Reduced Reference Video Quality Assessment method based on the human motion perception.* Telecommunications (IST), 2010 5th International Symposium. 4-6 Dec. 2010: 831-835.
[10] IP Gunawan, M Ghanbari. Reduced-Reference Video Quality Assessment Using Discriminative Local Harmonic Strength With Motion Consideration. *Circuits and Systems for Video Technology, IEEE* Transactions. 2008: 18(1): 71-83.
[11] R Soundararajan, AC Bovik. Video Quality Assessment by Reduced Reference Spatio-Temporal Entropic Differencing. *Circuits and Systems for Video Technology, IEEE* Transactions. 2013: 23(4): 684-694.
[12] M Wang, F Zhang, D Agrafiotis. *A very low complexity reduced reference video quality metric based on spatio-temporal information selection.* Image Processing (ICIP), 2015 IEEE International Conference. 2015: 571-575.
[13] T Yamada, Y Miyamoto, M Serizawa. *End-user video-quality estimation based on a Reduced-Reference model employing activity-difference for IPTV services*. Consumer Electronics, 2009. ICCE '09. Digest of Technical Papers International Conference. 10-14 Jan. 2009: 1-2.
[14] T Yamada, Y Miyamoto, Y Senda, M Serizawa. Video-Quality Estimation Based on Reduced-Reference Model Employing Activity-Difference. *IEICE transactions on fundamentals of electronics, communications and computer sciences.* 2009; 92(12): 3284-3290.
[15] T Yamada, T Nishitani. *No-reference quality estimation for compressed videos based on inter-frame activity difference*. Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference. 25-30 March 2012: 2325-2328.
[16] GE Legge, JM Foley. Contrast masking in human vision. *JOSA.* 1980; 70(12): 1458-1471.
[17] A Doufexi, S Armour, M Butler, A Nix, D Bull. *A Study of the Performance of HIPERLAN/2 and IEEE 802.11 a Physical Layers.* in Vehicular Technology Conference, 2001. VTC 2001 Spring. IEEE VTS 53rd. 2001: 668-672.
[18] K Seshadrinathan, R Soundararajan, AC Bovik, LK Cormack. Study of Subjective and Objective Quality Assessment of Video. *Image Processing, IEEE* Transactions. 2010; 19(6): 1427-1441.

[19] K Seshadrinathan, R Soundararajan, AC Bovik, LK Cormack. A subjective study to evaluate video quality assessment algorithms. *IS&T/SPIE Electronic Imaging*. 2010: 75270H-75270H-10.
[20] Ch Subrahmanyam, Venkata Rao. Low bit Rate Video Quality Analysis Using NRDPF-VQA Algorithm. *International Journal of Electrical and Computer Engineering (IJECE)*. 2015; 5(1): 71-77.
[21] Feng Xiao, Guo Li, Guo Lina. A New Sub-pixel Edge Detection Method of Color Images. *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS).* 2014; 12(5): 3609-3615.
[22] Chen Ning, Xiao-ping Song, Yi Liu. Edge detection based on biomimetic pattern recognition. *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS).* 2014; 12(9): 6965-6968.