Aligned-PSNR (APSNR) for Objective Video Quality Measurement (VQM) in Video Stream over Wireless and Mobile Network

Yoanda Alim Syahbana.[1], Herman.[2], Azizah Abdul Rahman.[3], Kamalrulnizam Abu Bakar.[4]
Universiti Teknologi Malaysia
Faculty of Computer Science and Information System
Johor, Malaysia
[1]asyoanda2@live.utm.my, [2]herman8@live.utm.my, [3]azizahar@utm.my, [4]knizam@utm.my

***Abstract*—Objective Video Quality Measurement (VQM) overcomes inefficiency of subjective Video Quality Assessment (VQA) in time and resources of survey. It uses video attributes such as video signal, noise signal, and encoder parameters to assess user-perceived quality. Peak Signal-to-Noise Ratio (PSNR) as one of objective VQM has been widely used to measure video quality. However, this conventional PSNR demonstrates inaccuracy of measurement while applied to measure video stream over wireless and mobile network. It is due to packet loss issue in the wireless and mobile network. This paper discusses evaluation of MPSNR as preexisting solution for this issue. In this paper, a concept of dynamic *window* size is used to improve accuracy of frame lost detection. The concept is named Aligned-PSNR (APSNR). An experiment is conducted to evaluate accuracy of the APSNR. The result is benchmarked with conventional PSNR based on Pearson product-Moment Correlation Coefficient (PMCC) value.**

Keywords-Objective Video Quality Measurement (VQM); Peak Signal-to-Noise Ratio (PSNR); video streaming; wireless and mobile network; Pearson product-Moment Correlation Coefficient (PMCC)

I. INTRODUCTION

Emerging advanced technologies on wireless and mobile multimedia have encouraged new generation of mobile services which includes video stream service over wireless and mobile network. The service is comprised of three service components i.e. wireless and mobile network infrastructure, video, and mobile device. Wi-Fi, WiMAX, UMTS and HSDPA are some of wireless and mobile device infrastructure that is utilized in the service. The streamed video itself may take a different form based on its variety of service e.g. video call, video teleconference, IP television (IPTV), and Video on Demand (VoD). The streamed video could be video that captured by mobile device in video call or video from webcam in case of video teleconference. It could be also a video file in mobile-VoD service. Mobile-VoD service streams video file that resides in video server (original video) to user by request. User can receive the streamed video using various types of mobile device i.e. notebook, Personal Digital Assistant (PDA), and smart phone.

Quality of the second service component, streamed video quality, is the most influencing factor that determines the service quality [1]. As regards this statement, the service quality can be abstracted from Quality of Service (QoS) and Quality of user Experience (QoE). QoS determines service quality based on how well the streamed video is received by mobile device. It is highly depends on quality of other two service components. On the other side, QoE deals with how well the service quality is perceived by user. In this level, QoE is influenced by user subjective factors.

According to abstraction of service quality, there are two methods that can be used to evaluate streamed video quality [2]. They are subjective Video Quality Assessment (VQA) and objective Video Quality Measurement (VQM). The subjective approach assesses video quality based on user's opinion that is gathered through survey. It provides high accuracy in determining video quality since it is straightly derived from user opinion [2]. However, this kind of assessment is inefficient. It takes time and need a lot of resources for conducting the survey [2] [3] [4] [5]. As an option, the objective VQM provides simpler implementation. The objective approach predicts subjective VQA result by calculating streamed video quality distortion. It measures noise signal, video frame loss, and also some network parameters such as packet loss and delay.

This paper discusses Peak Signal-to-Noise Ratio (PSNR) as one of objective VQM method. Limitation of PSNR method in measuring quality of video stream over wireless and mobile network is evaluated. To address the limitation, an improved algorithm is proposed. The paper is divided into five sections including this first introduction section. Problem background and related work is provided in the next section. The proposed algorithm is described in the third section. At the fourth section, simulation experiment is explained and evaluation of the algorithm is discussed. Finally, conclusions and future work is summarized in the fifth section.

II. BACKGROUND AND RELATED WORKS

Peak Signal-to-Noise Ratio (PSNR) is one of widely used method to measure streamed video quality. It is categorized into Full Reference (FR) VQM that requires existence of original video in its measurement process. It compares difference between every frame in streamed video against every frame in original video. It is defined as [6]

$$PSNR = 10log_{10}\frac{(2^M-1)^2}{MSE}. \qquad (1)$$

2^M-1 is maximum pixel value for M-bit frame while MSE is abbreviation of Mean Square Error. For a video with frame size of (x, y) pixels, MSE is calculated as

$$MSE = \frac{1}{XYT}\sum_{t=1}^{T}\sum_{y=1}^{Y}\sum_{x=1}^{X}[p(x,y,t) - p'(x,y,t)]^2. \quad (2)$$

The p(x, y, t) represents the original video frame (O-frame), while the p'(x, y, t) represents the streamed video frame (S-frame). t denotes frame index in these two videos.

Substituting (2) into (1) results on PSNR value for the streamed video. Large value of PSNR in decibel indicates high streamed video quality while the small value represents low streamed video quality. The interpretation of PSNR value in terms of quality is presented in Table 1 [7].

This conventional PSNR has limitation while applied to measure streamed video quality that transmitted over wireless and mobile network. Fig.1 illustrates this conventional PSNR limitation. It is due to packet loss characteristic in the network that impacts on video frame to become distorted and lost. In case of frame loss, the conventional PSNR limitation is arisen. Frame loss makes total frame in streamed video less than total frame in original video. In this circumstance, the conventional PSNR will blindly compare frame in streamed video against frame in original video without considering correspondences between the video frames.

This noncorresponding frame issue has become a concern in some previous works. On different case from previous limitation, [8] concerns on streamed video that experiences constant delay during the streaming. [8] works to locate first frame in streamed video to its corresponding frame in original video. Since the streamed video is influenced by constant delay, all next frames to the first-located frame are automatically aligned to their corresponding frames.

Work in [8] has inspired [9] to propose modification of PSNR (MPSNR) regarding noncorresponding frame issue. The concept is to match each frame in streamed video against frame in original video until it is found maximum PSNR value among the pairs. It is based on the assumption that the corresponding frame pair has higher value of PSNR than noncorresponding frame pairs.

TABLE I. PSNR INTERPRETATION IN TERMS OF STREAMED VIDEO QUALITY

PSNR Value	Quality
PSNR > 33 dB	Excellent Quality
33 dB > PSNR > 30 dB	Fair Quality
PSNR < 30 dB	Poor Quality

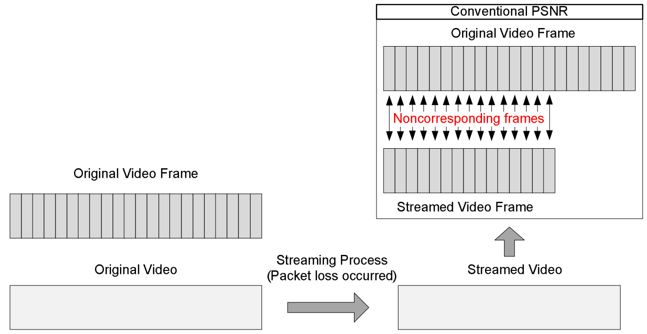

Figure 1. Illustration of conventional PSNR.

In finding the corresponding frames, MPSNR defines a boundary called *window*. *Window* determines limit of frame loss searching so that the algorithm only need to find corresponding frames in total of *window* size. For their experiments, [9] defines static value of five for the *window* size. However, there is shortcoming in this static *window* concept. If there are five or more consecutive frame losses, MPSNR will not find correct frame pair. Fig. 2 illustrates this circumstance.

Based on Fig. 2, O-frame number three to seven is sequential frame loss position in the S-frame (five sequential frame losses). In other words, O-frame number three to seven will not have its pairs in the S-frame. The processed S-frame (S-frame number x) should correspond with O-frame number eight. However, due to fixed *window* size of five, the will not measure PSNR between processed S-frame number x and O-frame number eight. As a consequence, the measured PSNR fails to include this corresponding frame pair. This shortcoming also influences the next processed S-frame.

Works by [9] also mentions approach to evaluate the newly proposed objective VQM. [9] uses subjective VQA result, due to its accuracy, as a reference. The evaluation measures how well the newly proposed objective VQM predicts subjective VQA result.

Subjective VQA has five Mean Opinion Scale (MOS) in representation of streamed video quality. MOS scale is shown in Table. 2.

TABLE II. MOS INTERPRETATION IN TERMS OF STREAMED VIDEO QUALITY [10]

MOS	Description
5	Excellent
4	Very Good
3	Good
2	Fair
1	Poor

The evaluation uses Pearson product-Moment Correlation (PMCC) to compare result of objective VQM and result of subjective VQA. PMCC is formulated as

$$r = \frac{\sum_{i=1}^{n} x_i y_i - \frac{1}{n}\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{\sqrt{\sum_{i=1}^{n} x_i^2 - \frac{1}{n}(\sum_{i=1}^{n} x_i)^2}\sqrt{\sum_{i=1}^{n} y_i^2 - \frac{1}{n}(\sum_{i=1}^{n} y_i)^2}}. \quad (3)$$

x and y are result of objective VQM and subjective VQA while n is total of evaluated video. r is the PMCC value that states accuracy of objective VQM in predicting subjective VQA. r value will be founded in range of zero to one which value of one represents perfect prediction of objective VQM towards subjective VQA.

Considering PSNR as objective VQM method that is discussed in this paper and this PMCC evaluation procedure, PSNR to MOS conversion need to be included to the previous Table 1. This conversion is aimed to consider nonlinearity between PSNR interpretation and MOS interpretation towards streamed video quality. Therefore, after receiving PSNR result from (1), it should be normalized into MOS score. Table 3 shows the conversion from PSNR interpretation in terms of quality to equivalent MOS score.

TABLE III. PSNR INTERPRETATION IN TERMS OF STREAMED VIDEO QUALITY AND MOS

PSNR Value	Quality	MOS
PSNR > 33 dB	Excellent Quality	5
33 dB > PSNR > 30 dB	Fair Quality	2
PSNR < 30 dB	Poor Quality	1

III. PROPOSED ALGORITHM

Instead of use fixed constant *window* size as used in MPSNR, aligned-PSNR (APSNR) uses different approach to define the *window* size. APSNR defines dynamic *window* size as

$$w = sumFL + 1. \quad (4)$$

Where
w = *window* size
sumFL = total of frame loss

In APSNR, the *window* size will proceed as shown in the Fig. 3. Fig. 3 depicts same case as in previous Fig. 2 case. However, in Fig. 3, the *window* size is not defined as a fixed value. The *window* size is depending on sum of frame loss in the S-frame. In the previous case, there are total of five frame losses. Based on previous formula, the *window* size is six.

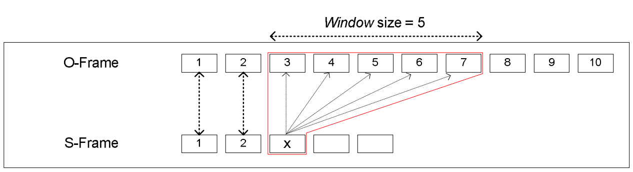

Figure 2. Possibility problem with static *window* size.

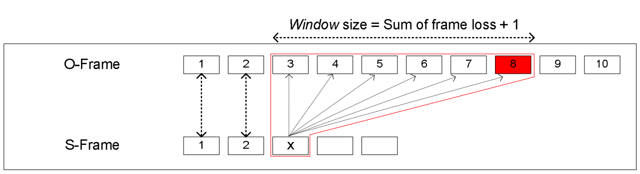

Figure 3. *Window* size in APSNR.

Therefore, the correspondences between processed S-frame and O-frame number eight will be found. In addition, to deal with long computation time, APSNR decreases *window* size value as much as number of discovered frame loss.

APSNR design is divided into three parts. There are searching, shifting, and calculating. Searching part is started by determining size of the *window*. Then, it is continued to pairing frame in streamed video to frame in original video. After that, the algorithm will check maximum PSNR value among the pairs. If the maximum pair is not founded in the first pair then there is indication of frame loss in current searching window. If this condition is occurred, the second part will shift the streamed video frame (shifting part). If it is not, the searching part will continue to the next streamed video frame. This first and second part will be repeated until last frame of streamed video is processed. This first and second part is illustrated in the Fig. 4 at the end of this paper. At final, the third part will sum entire PSNR pairs. According to PSNR formula, if the streamed video frame is distortion-free compared to original video frame, the PSNR value of this perfect match is infinity. For this perfect match case, the calculation assigns value of 100 dB as has been mentioned in [9]. Finally, the entire calculation generates APSNR value for the streamed video.

IV. EXPERIMENTS AND EVALUATIONS

A. Experiment Component

In order to simulate video stream over wireless and mobile network, the experiment is run under NS2 2.28 simulation environment. The experiment uses MYEvalvid_RTP that derived from proposed video streaming framework in [11] to simulate video streaming process. This streaming simulation is enhanced version of EvalVid [12] that improved by audio streaming. The experiment also simulates under HSDPA network infrastructure based on European Commission 5th framework project SEACORN, EURANE [13]. The HSDPA network infrastructure is depicted in Fig. 5. In order to varying the network condition, the experiment uses three value of CQI, 8, 15, and 22, as experimented in [14]. CQI value represents channel condition that perceived by mobile device.

The experiment uses "ANSI T1.801.01 vtc1nw" video as master video [15]. The experiment uses this standard test sequence depicting woman sitting reading a news story for twelve seconds. This master video is configured into twelve videos by varying its frame rate (15 and 30 fps), video bitrate (96 and 512 kbps), and resolution (QCIF, CIF, and 4CIF). At the end of entire simulation process, there are total 36 streamed videos that will be measured by APSNR. VI is denoted Video Index of the streamed video.

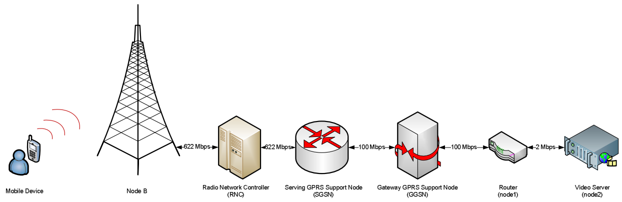

Figure 5. Network infrastructure of EURANE.

B. *Experiment Result*

Table 4, Table 5, Table 6, Table 7, Table 8, and Table 9 show APSNR and conventional PSNR results for the 36 streamed videos.

TABLE IV. RESULT FOR 4CIF RESOLUTION AND 512 KBPS VIDEO RATE

CQI	Resolution (4CIF) and Video Rate (512 kbps)			
	Frame Rate (30 fps)		*Frame Rate (15 fps)*	
	APSNR (dB)(VI= 1)	*PSNR (dB) (VI= 1)*	*APSNR (dB) (VI= 2)*	*PSNR (dB) (VI= 2)*
22	29.24	32.2	27.93	35.3
15	30.02	32.51	26.84	35.6
8	17.19	18.01	14.94	17.28

TABLE V. RESULT FOR 4CIF RESOLUTION AND 96 KBPS VIDEO RATE

CQI	Resolution (4CIF) and Video Rate (96 kbps)			
	Frame Rate (30 fps)		*Frame Rate (15 fps)*	
	APSNR (dB)(VI= 3)	*PSNR (dB)(VI= 3)*	*APSNR (dB)(VI= 4)*	*PSNR (dB) (VI= 4)*
22	30.68	34.36	27.42	34.12
15	29.8	33.89	38.47	38.47
8	23.31	24.28	21.63	25.22

TABLE VI. RESULT FOR CIF RESOLUTION AND 512 KBPS VIDEO RATE

CQI	Resolution (CIF) and Video Rate (512 kbps)			
	Frame Rate (30 fps)		*Frame Rate (15 fps)*	
	APSNR (dB) (VI= 5)	*PSNR (dB)(VI= 5)*	*APSNR (dB)(VI= 6)*	*PSNR (dB) (VI= 6)*
22	44.82	44	47.27	47.27
15	34.65	43.57	45.01	42.96
8	15.9	17.28	20.59	25.16

TABLE VII. RESULT FOR CIF RESOLUTION AND 96 KBPS VIDEO RATE

CQI	Resolution (CIF) and Video Rate (96 kbps)			
	Frame Rate (30 fps)		*Frame Rate (15 fps)*	
	APSNR (dB)(VI= 7)	*PSNR (dB)(VI= 7)*	*APSNR (dB) (VI= 8)*	*PSNR (dB) (VI= 8)*
22	33.62	41.16	35.06	44.13
15	33.62	41.16	34.79	44.15
8	31.83	38.61	33.28	40.83

TABLE VIII. RESULT FOR QCIF RESOLUTION AND 512 KBPS VIDEO RATE

CQI	Resolution (QCIF) and Video Rate (512 kbps)			
	Frame Rate (30 fps)		*Frame Rate (15 fps)*	
	APSNR (dB) (VI= 9)	*PSNR (dB)(VI= 9)*	*APSNR (dB) (VI= 10)*	*PSNR (dB) (VI= 10)*
22	48.82	48.82	48.48	48.48
15	48.88	48.88	48.49	48.49
8	31.98	39.23	48.48	48.48

TABLE IX. RESULT FOR QCIF RESOLUTION AND 96 KBPS VIDEO RATE

CQI	Resolution (QCIF) and Video Rate (96 kbps)			
	Frame Rate (30 fps)		*Frame Rate (15 fps)*	
	APSNR (dB)(VI= 11)	*PSNR (dB) (VI= 11)*	*APSNR (dB) (VI= 12)*	*PSNR (dB) (VI= 12)*
22	33.45	47.32	48.32	48.32
15	33.45	47.32	48.32	48.32
8	31.73	47.32	48.32	48.32

Fig. 6, Fig. 7, and Fig. 8 show graphs of the experiment results that compare APSNR and PSNR result for each CQI value.

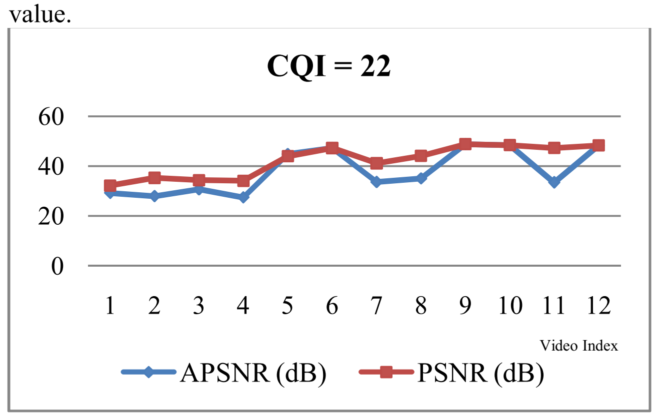

Figure 6. Experiment result for CQI value of 22.

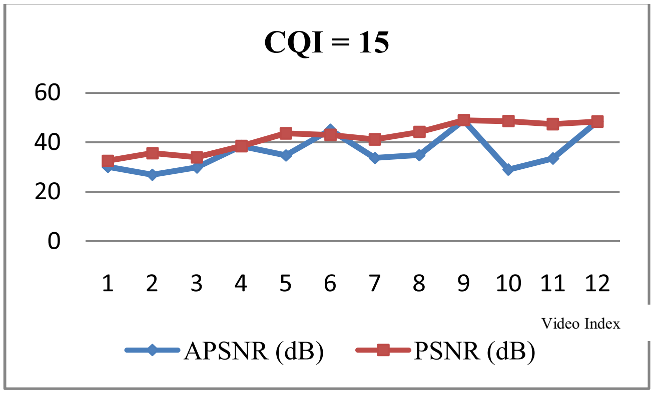

Figure 7. Experiment result for CQI value of 15.

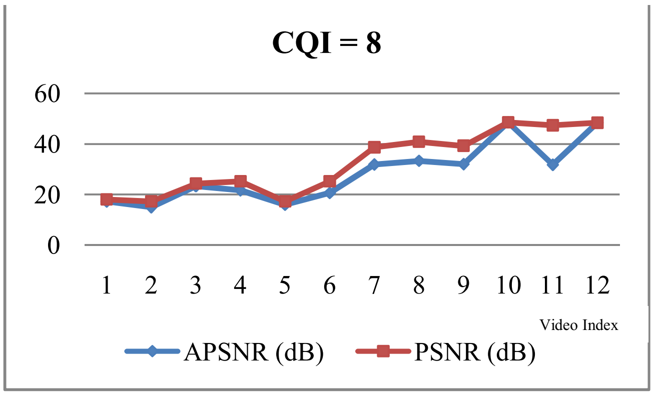

Figure 8. Experiment result for CQI value of 8.

C. Evaluation

The result shows three graphs of APSNR measurement result in comparison with conventional PSNR result. In general, the three graphs show that large CQI value results on high quality of streamed video while small CQI value results on low quality of streamed video. It is due to high packet loss occurrence for the small CQI value. In consequences, streamed videos with high configurations (streamed video number one to four) suffer more quality degradation than streamed video with low configurations (streamed video number nine to twelve).

From the experiment, the packet loss occurrence results on 5-21 consecutive frame loss. As previously mentioned in Section 2, this implies on shortcoming of MPSNR that unable to find correct frame pairs. In addition, [9] also does not mention how to keep number of remaining frame in original video is not less than remaining frame in streamed video as MPSNR is run. So that, the experiment cannot simulate MPSNR since it is constrained by condition of the remaining frame in original video less than remaining frame in streamed video.

From the three graphs, conventional PSNR measures higher value of streamed video quality than APSNR. Then, this result is processed by PMCC formula in order to evaluate accuracy of APSNR. PMCC results for conventional PSNR towards subjective VQA is 0.79 while PMCC results for APSNR towards subjective VQA is 0.81.

V. Conclusions and Future Work

This paper has proposed algorithm, APSNR, to overcome conventional PSNR limitation in measuring quality of video stream over wireless and mobile network. The experiments and evaluations of APSNR also have been exposed. Based on the experiment result, it is shown that APSNR has more accurate measurement than conventional PSNR in determining streamed video quality. It can be seen that APSNR value is close to value of one than conventional PSNR value.

For further research, it can be noticed that the video quality measurement still concern only on the video quality only. This research can be enhanced by considering audio quality in the quality measurement. It is due to customary combination of audio and video that streamed through wireless and mobile network in mobile-VoD. Therefore, there will be more accurate measurement for quality of the video stream service over wireless and mobile network.

References

[1] S. Winkler and P. Mohandas, "The Evolution of Video Quality Measurement: From PSNR to Hybrid Metrics," IEEE Transactions on Broadcasting, vol. 54(3), Jun. 2008, pp. 660-668, doi: 10.1109/TBC.2008.2000733.

[2] Kwangjin Choi, Jun kyun Choi, Jae Hwan Hong, Gyeong Ju Min and Jongkuk Lee, "Comparison of Video Streaming Quality Measurement Methodologies," 10th International Conference on Advanced Communication Technology (ICACT), 17-20 Feb. 2008, pp. 993-996, doi: 10.1109/ICACT.2008.4493933.

[3] K. Piamrat, C. Viho, J. -M. Bonnin and A. Ksentini, "Quality of Experience Measurements for Video Streaming over Wireless Networks," Sixth International Conference on Information Technology: New Generations, IEEE, 27-29 Apr. 2009, pp. 1184-1189, doi: 10.1109/ITNG.2009.121.

[4] U. Reiter and J. Korhonen, "Comparing Apples and Oranges: Subjective Quality Assessment of Streamed Video with Different Types of Distortion," International Workshop on Quality of Multimedia Experience (QoMEx), IEEE, 29-31 Jul. 2009, pp. 127-132, doi: 10.1109/QOMEX.2009.5246963.

[5] Khan, Lingfen Sun and E. Ifeachor, "Content Clustering Based Video Quality Prediction Model for MPEG4 Video Streaming over Wireless Networks," International Conference on Communications (ICC), IEEE, 14-18 Jun. 2009, pp. 1-5, doi: 10.1109/ICC.2009.5198850.

[6] M. Vranjes, S. Rimac-Drlje and K. Grgic, "Locally Averaged Psnr As A Simple Objective Video Quality Metric," 50th International Symposium ELMAR, IEEE, 10-12 September 2008, pp. 17-20.

[7] Y. Ennaji, M. Boulmalf and C. Alaoui, "Experimental Analysis of Video Performance over Wireless Local Area Networks," International Conference on Multimedia Computing and Systems (ICMCS), IEEE, 2-4 Apr. 2009, pp. 488-494 doi: 10.1109/MMCS.2009.5256645.

[8] S. Wolf and M. H. Pinson, "Reference Algorithm for Computing Peaksignal to Noise Ratio (PSNR) of A Video Sequence with A Constant Delay," ITU, ITU-T Contribution COM9-C6-E, Feb. 2009.

[9] An Chan, Kai Zeng, P. Mohapatra, Sung-Ju Lee, S. Banerjee, "Metrics for Evaluating Video Streaming Quality in Lossy IEEE 802.11 Wireless Networks," Proceedings IEEE INFOCOM, IEEE, 14-19 Mar. 2010, pp. 1-9, doi: 10.1109/INFCOM.2010.5461979.

[10] O. Bradeanu, D. Munteanu, I. Rincu and F. Geanta, "Mobile Multimedia End-User Quality of Experience Modeling," ICDT'06 International Conference on Digital Telecommunications, IEEE Press, Aug. 2006, doi: 10.1109/ICDT.2006.49.

[11] Chia-Yu Yu et al., “MyEvalvid_RTP: A New Simulation Tool-set Toward More Realistic Simulation,”. Future Generation Communication and Networking (FGCN), IEEE, 6-8 Dec. 2007, pp. 90-93, doi: 10.1109/FGCN.2007.167.

[12] J. Klaue, B. Rathke and A. Wolisz, "EvalVid – A Framework for Video Transmission and Quality Evaluation," Proc. of the 13th International Conference on Modelling Techniques and Tools for Computer Performance Evaluation, Sep. 2003, pp. 255-272, doi: 10.1.1.10.3918.

[13] SEACORN project. Available at http://eurane.ti-wmc.nl/eurane/. 2004.

[14] F. Brouwer, I. de Bruin, J. C. Silva, N. Souto, F. Cercas, A. Correia, "Usage of Link-Level Performance Indicators for HSDPA Network-Level Simulations in E-UMTS," Eighth International Symposium on Spread Spectrum Techniques and Applications, IEEE 30 Aug.-2 Sept. 2004, pp. 844- 848, doi: 10.1109/ISSSTA.2004.1371820.

[15] Film by NTIA/ITS (National Telecommunications and Information Administration/The Institute for Telecommunication Sciences (ITS). Available at http://www.cdvl.org/find-videos/details.php?id=18.

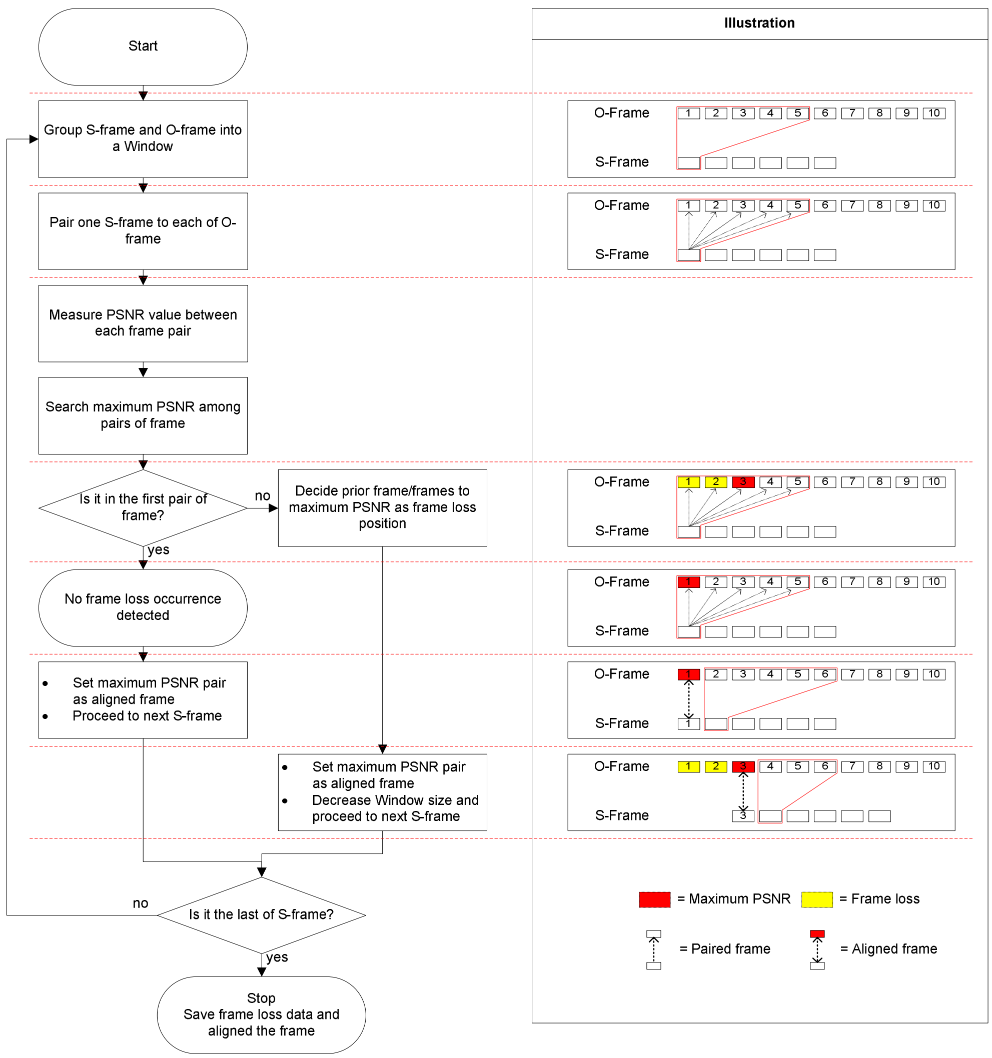

Figure 4. *First and second part of APSNR algorithm.*