International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:9, No:8, 2015

Subjective Quality Assessment for Impaired Videos with Varying Spatial and Temporal Information

Muhammad Rehan Usman, Muhammad Arslan Usman, Soo Young Shin

***Abstract*—**The new era of digital communication has brought up many challenges that network operators need to overcome. The high demand of mobile data rates require improved networks, which is a challenge for the operators in terms of maintaining the quality of experience (QoE) for their consumers. In live video transmission, there is a sheer need for live surveillance of the videos in order to maintain the quality of the network. For this purpose objective algorithms are employed to monitor the quality of the videos that are transmitted over a network. In order to test these objective algorithms, subjective quality assessment of the streamed videos is required, as the human eye is the best source of perceptual assessment. In this paper we have conducted subjective evaluation of videos with varying spatial and temporal impairments. These videos were impaired with frame freezing distortions so that the impact of frame freezing on the quality of experience could be studied. We present subjective Mean Opinion Score (MOS) for these videos that can be used for fine tuning the objective algorithms for video quality assessment.

***Keywords*—**Frame freezing, mean opinion score, objective assessment, subjective evaluation.

I. Introduction

IN modern era of communication, wireless portable consumer devices, e.g. smart phones, tablets etc., are becoming more robust day by day. In order to fully utilize the capabilities of these devices, operators have to provide high end-user quality of services (QoS). With these recent advances video streaming has become one of the most popular challenging services for operators, as the modern devices now support fast data rates and high end-user resolutions. Recent digital imaging technologies and efficient wireless transmission systems have increased the demand of high quality video transmission. Videos transmitted through wireless medium, to and from the mobile devices, have accounted for 66% of the global mobile data traffic by 2014 [1]. A key factor in maintaining QoS in video streaming is quality of experience (QoE) which is based on the end-user experience. This challenge for maintaining QoE is more visible in wireless communication as the quality degradation chances increase over a wireless medium because of common issues like packet loss and bandwidth constraints etc. [2]. Further visual medial reaches the end user following through different stages, i.e., capturing, storing, transmission, reproduction and display; the quality may also be degraded due to distortion that may occur at any of these stages.

To overcome the degradation issue or more likely to enhance the QoE, objective and subjective testing methods are used. Currently several automated objective testing algorithms for video quality assessment (VQA) are in use, but to be sure the true judge of video quality is the end-user, i.e., humans. This is because the objective testing algorithms are developed using the input from the subjective testing, as only human eye can identify the problems faced in video streaming. That is why the video quality expert group (VQEG), which is the leading body in VQA formed in 1997, concluded in their report that no objective measure for VQA can replace subjective testing [3]. VQEG has been the major body for conducting evaluations on objective testing using the subjective assessment [4].

Subjective assessment is done in order to record the mean opinion score (MOS) of test subjects. In simple terms, it is a means of knowing the perceptual impact on end users when they view a video. In live video streaming there is a high chance of transmission errors as well as hardware errors. Such kind of errors can result in bad quality of experience for the end users hence reducing the ratings of the network operators. Continuous monitoring of the videos can result in a better understanding of the transmission and hardware related errors so that the network operators can employ mechanisms to overcome such errors. There are objective algorithms that perform the task of assessing the quality of the video but in order to validate their performance subjective tests have to be performed. Most of the objective algorithms assess visual quality by quantifying spatial (intra frame) impairments such as blocking, blurring, ringing and temporal (inter frame) degradations such as jitter or jerkiness [5]. Subjective tests are a meaningful and strong source in order to better understand the behaviour of an error in a digitally transmitted video. The packet loss in a video transmission can result into errors like frame freezing which is a very well-known temporal impairment. A frame freeze occurs in a received video when a frame is copied on the successive frames until a correct frame is received. This kind of behaviour can be due to some error concealment algorithm or it can be a transmission or hardware error. Multiple frames freezing can result in a received video being frozen on multiple instants. A freeze event is defined as a total duration of freezing of a particular frame in a video. In this paper we have discussed the phenomenon of frame freezing and we have performed video quality assessment for videos that contain frame freezes.

The breakdown of this paper is as follows: Section II contains the survey of related works. The subjective tests and their results have been discussed in Section III followed by the conclusion in Section IV.

Muhammad Rehan Usman, Muhammad Arslan Usman, Soo Young Shin (+82-10-9192-4998), are with the WENS Lab. Department of IT Convergence, Kumoh National Institute of Technology, Gumi, South Korea (e-mail: rehanusman@gmail.com, arslanusman@msn.com, wdragon@kumoh.ac.kr).

TABLE I
ABSOLUTE CATEGORY RATING (ACR)

Category Rating (Quality)	Mean Opinion Score (MOS)
Excellent	5
Good	4
Fair	3
Poor	2
Bad	1

II. RELATED WORK

Some efforts of subjective video quality evaluation have also been conducted such as [6]-[8] but unfortunately the test conditions and subjective results are not shared to public hence limiting their use to a particular group. Also in [9], subjective quality assessment was performed for the H. 264/ AVC encoded videos. The video sequences used in this work were low resolution videos. In [2], the authors have performed a vast subjective study of the successor of H. 264 known has H.265/ HEVC and they have shown the results for a number of video sequences. Also, they have made a comparison with a few objective video quality algorithms with their results.

In our paper we have followed the similar trend of providing subjective evaluation of video sequences with different temporal and spatial aspects.

TABLE II
DESCRIPTION OF USED VIDEOS FOR SINGLE AND MULTIPLE FREEZE EXPERIMENT

Spatial Resolutions / Temporal Resolutions	Video Sequences	Total Duration (s)	Number of Freeze Events & respective durations in seconds (Duration of Event (s) × No. of Events)				
			Single Freeze	2	3	5	8
Native: VGA 640 × 480 @ 30 Frames per second	City	12	(0.12, 0.2, 0.52, 1, 2, 3) × 1	(0.067, 0.133, 0.2, 0.4) ×2	(0.067, 0.133, 0.267, 0.533) ×3	(0.067, 0.133, 0.33, 0.66) × 5	(0.067, 0.133, 0.2, 0.4) × 8
	Crew	12	(0.12, 0.2, 0.52, 1, 2, 3) × 1	(0.067, 0.133, 0.2, 0.4) × 2	(0.067, 0.133, 0.267, 0.533) ×3	(0.067, 0.133, 0.33, 0.66) × 5	(0.067, 0.133, 0.2, 0.4) × 8
	Harbour	12	(0.12, 0.2, 0.52, 1, 2, 3) × 1	(0.067, 0.133, 0.2, 0.4) × 2	(0.067, 0.133, 0.267, 0.533) ×3	(0.067, 0.133, 0.33, 0.66) × 5	(0.067, 0.133, 0.2, 0.4) × 8
	Ice	10	(0.12, 0.2, 0.52, 1, 2, 3) × 1	(0.067, 0.133, 0.2, 0.4) × 2	(0.067, 0.133, 0.267, 0.533) ×3	(0.067, 0.133, 0.33, 0.66) × 5	(0.067, 0.133, 0.2, 0.4) × 8
Native: VGA 640 × 480 @ 25 Frames per second	Mother & Child	12	(0.12, 0.2, 0.52, 1, 2, 3) × 1	(0.08, 0.16, 0.24, 0.48) × 2	(0.08, 0.16, 0.24, 0.64) × 3	(0.08, 0.16, 0.4, 0.8) × 5	(0.08, 0.16, 0.24, 0.48) × 8
	Children	12	(0.12, 0.2, 0.52, 1, 2, 3) × 1	(0.08, 0.16, 0.24, 0.48) × 2	(0.08, 0.16, 0.24, 0.64) × 3	(0.08, 0.16, 0.4, 0.8) × 5	(0.08, 0.16, 0.24, 0.48) × 8

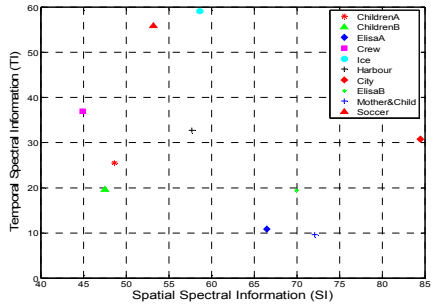

Fig. 1 Raw Results of SITI Plotting

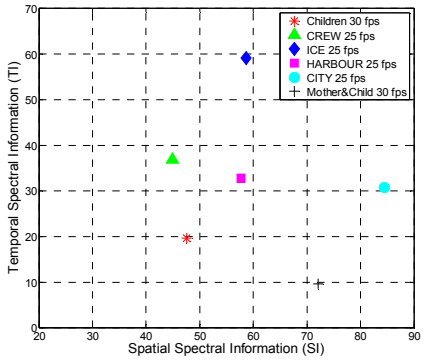

Fig. 2 Refined Results of SITI Plotting

III. SUBJECTIVE VIDEO QUALITY ASSESSMENT TESTS

To create a meaningful subjective VQA database, the spectral contents of the videos, under consideration for quality assessment (QA) tests, must be characterized using varying amounts of spatial and temporal spectral information.

This section contains the description of the spectral characterization indices, i.e., spatial spectral information (SI) and temporal spectral information (TI), and provides the description and configuration of the subjective VQA tests.

A. Spatial Spectral Information (SI)

It is generally a measure of the amount of spatial detail of a picture. For more spatial complex scenes it is usually considered to be higher. It is neither the measure of entropy nor it is connected to the information defined in the communication theory [4], [10].

In more technical terms, SI is based on the sobel filter (also called as sobel operator). It is a filter that uses edge detection algorithms and creates images which emphasizes regions of high spatial frequency that corresponds to edges [4], [10]. The process of SI measurement is as follows [4], [10].

At first, each frame of the video is filtered with the sobel filter according to a specific time. Then, each sobel-filtered frame's standard deviation is computed over the pixels. For each frame in the video sequence the described operation is repeated and as a result a time series of spatial information of

the scene is created. A maximum value is chosen in the time series to represent the SI content of the scene. This whole process is summarized in (1) taken from [4], [10].

$$SI = max_{time}\{std_{space}[Sobel(fn)]\} \quad (1)$$

where, max_{time} is the maximum value in time series, std_{space} is the standard deviation over the pixels, *Sobel* represents the sobel filter and *fn* denotes the frames over time *n*.

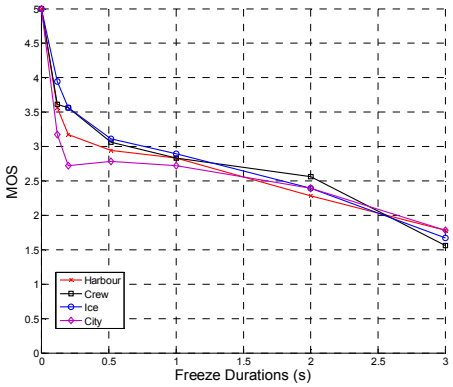

Fig. 3 Subjective Quality Assessment Results of Single Freeze Tests for videos with 25 fps

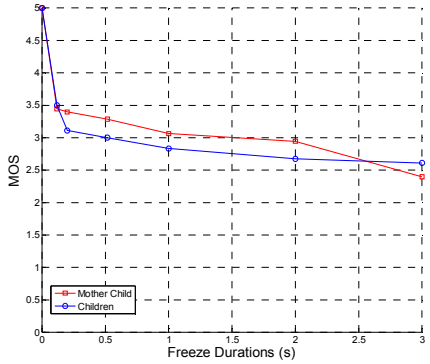

Fig. 4 Subjective Quality Assessment Results of Single Freeze Tests for videos with 30 fps

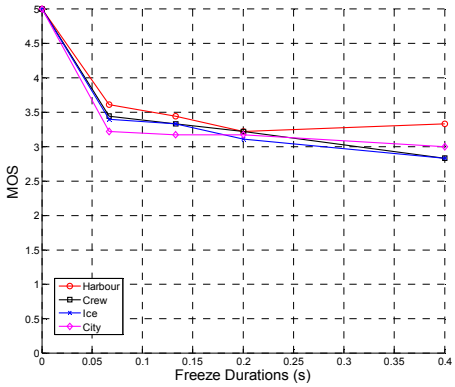

Fig. 5 Subjective Quality Assessment Results of Multiple Freeze Tests for videos with 3 freeze events @ 25 fps

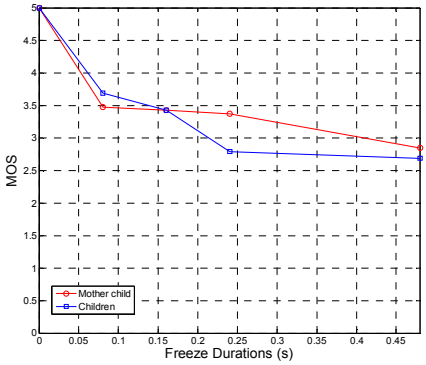

Fig. 6 Subjective Quality Assessment Results of Multiple Freeze Tests for videos with 3 freeze events @ 30 fps

B. Temporal Spectral Information (TI)

It is generally a measure of the amount of temporal changes of a video sequence. For high motion sequences it is usually considered to be higher. It is neither the measure of entropy nor it is connected to the information defined in the communication theory [10], [11].

In more technical terms, TI is based on the motion difference feature. This feature is the difference between the pixel values at the same location in space but at successive frames of a video [10], [11]. The motion difference feature is the function of time *n* given by [10], [11]:

$$M_n(i,j) = F_n(i,j) - F_{n-1}(i,j) \quad (2)$$

where, *Fn (i,j)* is the pixel at i^{th} row and j^{th} column of the n^{th} frame in time. The TI is measured as the maximum over time of the standard deviation over space of the motion difference between pixel values for sequential frames over all *i* and *j* the TI equation is given as [9], [10]:

$$TI = max_{time}\{std_{space}[M_n(i,j)]\} \quad (3)$$

where, max_{time} is the maximum value in time series, std_{space} is the standard deviation over the space and $M_n(i,j)$ motion difference between pixel values for sequential frames and (i,j) is the pixel location. Equation (3) shows that higher the notion in the adjacent frames higher will be the TI value [10].

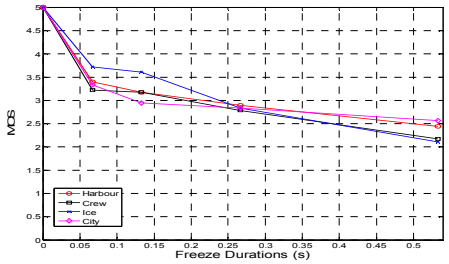

Fig. 7 Subjective Quality Assessment Results of Multiple Freeze Tests for videos with 2 freeze events @ 25 fps

TABLE III
DESCRIPTION OF USED VIDEOS

Video Resolution	Frame Rate of Video (fps)	Name of Video Sequence	Content	Duration of Videos (s)	Spatial Information (SI)	Temporal Information (TI)
Native: VGA 640 × 480	30 Frames per second	City	A Camera panning over a City from Above	10	High	Low
		Crew	A space shuttle crew walking through the aisle	10	Low	High
		Harbour	View of Boats moving in a Harbour	10	High	High
		Ice	People skating in an Ice skating arena	9	Low	Low
Native: VGA 640 × 480	25 Frames per second	Mother & Child	A talking mother and a child sitting in her lap	12	High	Low
		Children	Two children talking and playing with toys	12	Low	High

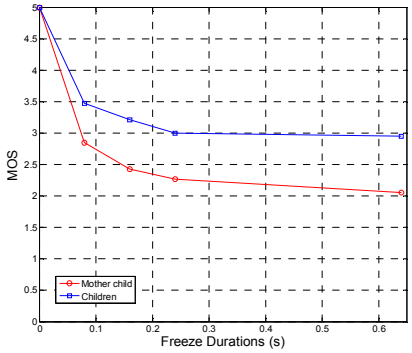

Fig. 8 Subjective Quality Assessment Results of Multiple Freeze Tests for videos with 2 freeze events @ 30 fps

For the subjective tests, we chose a set of ten videos with different Spatial and Temporal content in them. The description about the videos is given in Tables II and III. The videos chosen for this study have both temporal and spatial content in them as shown in the Fig. 1. Four of the videos had to be skipped as with this much number of videos, it would be hard to conduct the subjective tests within the recommended time by ITU [10], [11]. A few of the videos were showing almost the same SI and TI, so they were skipped. So, the subjective tests were performed on the remaining six videos. The SI and TI plot for these videos can be seen in the refined results of Fig. 2. These videos were further altered using the MATLAB and their respective descriptions are given in Table II. For performing the subjective tests we have followed the recommendation by ITU-R BT 500 [11]. We used the absolute category rating (ACR) scale for assessing the videos. The ACR scale has been explained in the Table I. Where, a rating of 5 means excellent quality video and a rating of 1 means a very poor quality video.

C. Test Environment and Results

For the setup of the tests, we booked a lab which was a cubical room with white background and no extra furniture other than the one required. For temporal aspects we have used videos with different frame rates, 25 frames per second (fps) and 30 fps, and for the spatial aspect we have used the common VGA resolution i.e. 640 × 480. The bitrates of all the videos were kept unaltered. Also, the videos are not compressed by any compression tool, as we kept them in their native resolutions as explained in Table II.

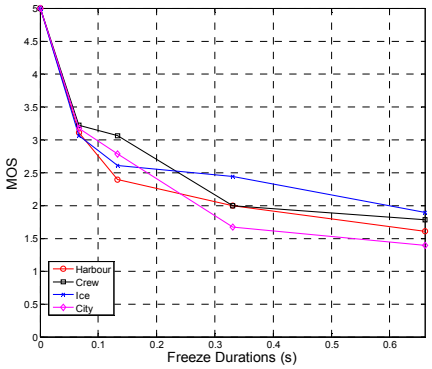

Fig. 9 Subjective Quality Assessment Results of Multiple Freeze Tests for videos with 5 freeze events @ 25 fps

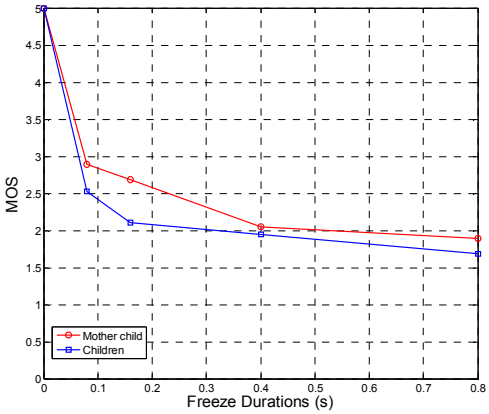

Fig. 10 Subjective Quality Assessment Results of Multiple Freeze Tests for videos with 5 freeze events @ 30 fps

A total of 22 test subjects took part in the tests. Out of them 12 were from the expert category, 6 were from the non-expert category and the remaining 6 were from the youth (less than 20 years of age) category. The recommendations of ITU [10], [11] have been strictly followed while conducting the tests. The participants were from mixed gender i.e. both male and female.

The electronic equipment used for the tests included a 19 inch desktop TFT monitor with RGB colour scale. The desktop system used was from *Asus* and had a processing speed of 3.2 GHz along with 6 GB of RAM. The distance between the test subjects and the monitor was kept higher than 3 times of the height of the display size of the video. The seating arrangement for the test subjects was kept as quiet and comfortable as possible to avoid any fatigue, keeping in view the duration of the tests. The video sequences were displayed on the monitor screen for a particular duration as explained in Table II for every test subject and then a standard interval of 10 seconds was provided in which the participant was asked to perform the grading of the video according to the ACR scale. Only one test subject was invited to perform the test at a time.

The tests were broken down into 2 categories, single and multiple freeze tests, for the ease of the participants. Each test was designed to be completed within 30 minutes of time as recommended by [10], [11]. The results for the single freeze and multiple freeze experiment have been discussed as follows.

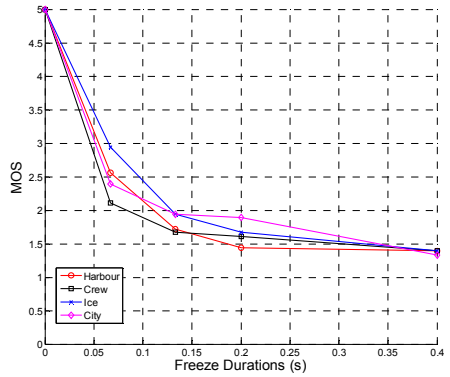

Fig. 11 Subjective Quality Assessment Results of Multiple Freeze Tests for videos with 8 freeze events @ 25 fps

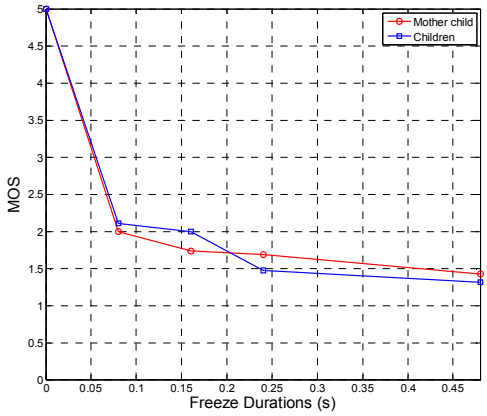

Fig. 12 Subjective Quality Assessment Results of Multiple Freeze Tests for videos with 8 freeze events @ 30 fps

1) Single Freeze Test

The single freeze test was conducted within 2 days of time. The first day was dedicated for the experts category i.e. the participants were aware of the ongoing research on subjective video quality assessment. Then the proceeding day the remaining two categories performed the tests. The results in Fig. 3 shows that the videos named Harbour and City show a better MOS as compared to Crew and Ice as the former have low motion content in them. For a test subject it is hard to detect a frame freeze if the motion content of the video is lower.

Similar trend can be seen in the results shown in Fig. 4 for the videos with a frame rate of 30 fps. The description of the content of the videos can be seen from Table III.

2) Multiple Freeze Test

These tests were conducted in a total of 4 days as the number of videos generated for these tests was much larger than the single freeze test. First 2 days were dedicated for the experts group and then the proceeding days were dedicated for the other 2 categories.

We can see a similar trend for the videos containing 2 and 3 freeze events from Figs. 5-8. But as the number of freeze events are increased to 5 and 8, we can see that the test subjects have assessed the videos with low motion content, Harbor and City, the same manner as they have assessed the other videos as shown in Figs. 9-12. The high motion content is more likely to have a lower MOS if the videos are impaired as there is a bigger chance for data loss. Low motion content videos have higher MOS, even though the duration of frame freezing is same, because there is less difference between the motion intensity of consecutive frames. For the 30 fps videos, the results are comparable to 25 fps videos. The videos with higher frame rate are considered to be much better in terms of quality. So it can be noticed here that the test subjects have given a higher MOS for the 30 fps videos compared to the 25 fps videos.

Finally, we can move to the conclusion of this paper with all the results in hand.

IV. Conclusion

In this paper subjective tests have been provided for videos with different spatial and temporal content. These subjective test results can be used in order to validate and check the objective video quality algorithms. Also, these tests can be extended further with higher spatial resolutions and different temporal resolutions i.e. for higher frame rates.

The results of these subjective tests cover all types of video contents, from low motion content videos to high motion content videos. The contribution of this research is to provide a MOS database to the fellow researchers from the same field.

Acknowledgment

This research was supported by the MSIP (Ministry of Science, ICT and Future Planning), Korea, under the CITRC (Convergence Information Technology Research Center) support program (IITP-2015-H8601-15-1011) supervised by the IITP (Institute for Information & communications Technology Promotion).

References

[1] Index, Cisco Visual Networking. "Global mobile data traffic forecast update, 2010-2015," *White Paper,* Feb. 2011.

[2] J. Nightingale, Qi Wang, C. Grecos, and S. Goma, "The impact of network impairment on quality of experience (QoE) in H. 265/HEVC video streaming," *IEEE Transactions on Consumer Electronics,* vol.60, issue.2, pp. 242-250, May 2014.
[3] Video Quality Experts Group VQEG, "Final rep. from the video quality experts group on the validation of objective models of video quality assessment VQEG," 2000. (Online). Available: www.vqeg.org
[4] Tutorial, I. T. U. T. "Objective perceptual assessment of video quality: full reference television." *ITU-T Telecommunication Standardization Bureau,* 2004. (Online). Available www.itu.int/ITU-T.
[5] Yuen, Michael, and H. R. Wu. "A survey of hybrid MC/DPCM/DCT video coding distortions." *Signal processing* 70.3 (1998): 247-278.
[6] Winkler, Stefan, and Ruth Campos. "Video quality evaluation for Internet streaming applications." *Electronic Imaging 2003*. International Society for Optics and Photonics, 2003.
[7] Muntean, Gabriel-Miro, Philip Perry, and Liam Murphy. "Subjective assessment of the quality-oriented adaptive scheme." *Broadcasting, IEEE Transactions on*51.3 (2005): 276-286.
[8] Zhai, Guangtao, et al. "Cross-dimensional perceptual quality assessment for low bit-rate videos." *Multimedia, IEEE Transactions on* 10.7 (2008): 1316-1324.
[9] M. Shahid, A. K. Singam, A. Rossholm, and B. Lovstrom, "Subjective quality assessment of H. 264/AVC encoded low resolution videos," *IEEE 5th International Congress on Image and Signal Processing (CISP)*, pp. 63-67., Oct. 2012.
[10] ITU-T RECOMMENDATION, P, "Subjective video quality assessment methods for multimedia applications," pp. 34-35, 1999.
[11] "ITU-R Radio communication Sector of ITU, Recommendation ITU-R BT.500-12," 2009.

M. Rehan Usman was born in Lahore City, Pakistan in 1986. He received the B.S. degree in Electrical Engineering from COMSATS Institute of Information Technology, Lahore, Pakistan, in 2008 and holds two M.S. Degrees, i.e., 1) M.S. Electrical Engineering with Specialization in Telecommunications from Blekinge Tekniska Hogkola (BTH), Karlskrona, Sweden, in 2010 and 2) M.S. Project Management and Operational Development from Kungliga Tekniska Hogskola (KTH), Stockholm, Sweden, in 2013. His research interests include coexistence among networks, cognitive radio, MIMO, OFDM, NOMA, small cells and next generation mobile wireless networks (4G/5G).

He was a Lecturer in Electrical Engineering department at University of South Asia Pakistan from July 2012 to January 2013. Then he joined as a lecturer in Electrical Engineering department of Superior University Lahore, Pakistan, from January 2013 to March 2014. After serving in Superior University he is now a PhD research scholar at WENS Lab in School of Electronics in Kumoh National Institute of Technology since March 2014.

M. Arslan Usman was born in Okara City, Pakistan in 1988. He received his B.S. degree in Electrical Engineering from COMSATS Institute of Information Technology, Lahore, Pakistan, in 2010 and M.S. Degrees in Electrical Engineering with Specialization in Signal Processing from Blekinge Tekniska Hogkola (BTH), Karlskrona, Sweden, in 2013. His research interests include coexistence among networks, cognitive radio, MIMO, OFDM, NOMA, small cells, audio / visual signal processing and next generation mobile wireless networks (4G/5G).

He has worked in Alcatel-Lucent Nigeria as a wireless pre-sales engineer (4G-LTE campaign for West-Central Africa), from 2013–2014, in the department of technical sales. After serving in Alcatel-Lucent for a year he is now a PhD research scholar at WENS Lab in School of Electronics in Kumoh National Institute of Technology since September 2014.

Soo Young Shin was born in 1975. He received his B.S., M.S., and Ph. D degrees in Electrical Engineering and Computer Science from Seoul National University, Korea in 1999, 2001, and 2006, respectively. His research interests include wireless LAN, WPAN, WBAN, wireless mesh network, sensor networks, coexistence among wireless networks, industrial and military network, cognitive radio networks, MIMO, OFDM, mmWave, NOMA and next generation mobile wireless broadband networks (4G/5G).

He was a visiting scholar in FUNLab at University of Washington, US, from July 2006 to June 2007. After 3 years working in WiMAX design lab. of Samsung Electronics, he is now an assistant professor in School of Electronics in Kumoh National Institute of Technology since September 2010.