

(c) 2014 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other users, including reprinting/ republishing this material for advertising or promotional purposes, creating new collective works for resale or redistribution to servers or lists, or reuse of any copyrighted components of this work in other works.

A Study on Quality Assessment for Medical Ultrasound Video Compressed via HEVC

Manzoor Razaak, Maria G. Martini, *Senior Member, IEEE*, and Ketty Savino

Abstract—The Quality of Experience (QoE) and Quality of Service (QoS) provided in the healthcare sector are critical in evaluating the reliable delivery of the healthcare services provided. Medical images and videos play a major role in modern e-health services and have become an integral part of medical data communication systems. The quality evaluation of medical images and videos is an essential process, and one of the ways of addressing it is via the use of quality metrics. In this paper, we evaluate the performance of seven state of the art video quality metrics with respect to compressed medical ultrasound video sequences. We study the performance of each video quality metric in representing the diagnostic quality of the video, by evaluating the correlation of each metric with the subjective opinions of medical experts. The results indicate that the Visual Information Fidelity (VIF), Structural Similarity Index Metric (SSIM), and Universal Quality Index (UQI) metrics show good correlation with the subjective scores provided by medical experts. The tests also investigate the performance of the emerging video compression standard, High Efficiency Video Coding (HEVC), for medical ultrasound video compression. The results show that, using HEVC, a diagnostically reliable compressed ultrasound video can be obtained for compression with values of the quantization parameter, QP, upto 35.

Index Terms—Medical video quality evaluation, HEVC, service science, objective & subjective video quality assessment, medical ultrasound videos, video compression.

I. INTRODUCTION

RECENT advancements in technology have enabled the delivery of new healthcare services. In particular, the emergence of advanced communication systems has empowered the field of telemedicine. Today, communication systems have led to the growth of innovative services like remote patient monitoring and diagnosis, medical video conferencing, long distance consultations, live surgery broadcast for educational purposes, ambient assisted living for old and cognitive impaired people, etc. Further, the amalgamation of healthcare services with various other disciplines like engineering, information technology, and business management has converted healthcare into a major service sector in the modern world. Hence, the application of service science in e-health can facilitate a more efficient and reliable delivery of healthcare services.

Modern communication systems are now an integral part of the e-health sector and reliable transmission of medical data is now possible due to advancements in communication

technologies. Further, the emergence of efficient image and video compression standards has enabled efficient storage of medical images and videos. However, the growth rate in data storage and communication technologies is still lower than the rate of increase in demand for storage and transmission of medical data, in particular medical images and videos. Also, bandwidth is a precious resource in communication systems and compression is often needed before transmission, in particular over wireless systems [1], [2]. Hence, compressing images and videos is still a requirement for most applications and services, and the effect of compression and transmission is often a reduced quality. Therefore, it is essential that the quality of the medical images and videos received is monitored via Image Quality Assessment (IQA) and Video Quality Assessment (VQA) techniques, so that, along with quality evaluation, it can also facilitate the design of future medical multimedia services and applications [3] [4].

The quality evaluation of medical videos is particularly challenging since most objective video quality metrics are essentially based on the principle of measuring the visual quality of the video. Since medical videos carry sensitive information, an impairment on the video may result in loss of information of diagnostic importance. Therefore, the objective metrics which primarily measure the visual quality of a video may not give a reliable measure of the diagnostic accuracy of the impaired video. Hence, it is important that the video quality metrics used are tested for their reliability in approximating the diagnostic quality.

In this paper, we delve into the topic of VQA in the context of medical ultrasound videos, extending the work presented in [5], where an insight into the VQA topic in the context of medical videos is given. Video quality evaluation techniques are mainly classified into *objective* and *subjective* methods. In subjective VQA methods, a subject rates the quality of the video based on their perception, whereas in objective VQA, mathematical algorithms are employed to numerically represent the quality of the video. In many cases, the objective VQA method is preferred for quality evaluation since it is impractical to always assess the video quality via subjects. However, subjective VQA better represents the actual Quality of Experience (QoE) and is also useful to assess the performance of the objective VQA metrics [6]. An efficient objective quality metric is expected to correlate well with the subjective scores, since a good correlation is an indication of the metric's reliability in representing the quality of the video. In this paper, we test the performance of some of the widely used state-of-the-art video quality metrics, in the context of medical ultrasound videos, by testing their correlation with

M. Razaak and M.G. Martini are with Wireless and Multimedia Networking research group, Kingston University, London. E-mail: {m.razaak, m.martini}@kingston.ac.uk.

K. Savino, MD, is senior cardiologist at the University Hospital of Perugia, Italy.

the subjective scores of medical experts and non-experts.

Subjective quality methodologies are widely used in the literature for validating objective quality metrics and for designing and validating assessment models. For instance in [7] an index to measure the decrease in diagnostic content of compressed echocardiograms was developed based on the inputs from cardiologists. [8] recommended acceptable compression ratios using MPEG-2 for surgical videos. Very few works in the literature have focused on evaluating the performance of state of the art objective video quality metrics in the context of medical videos, in particular medical ultrasound videos. The works in [9], [10], [11] have considered video quality assessment of compressed medical videos. However, their tests have employed a limited number of videos and video quality metrics and do not consider many popular state-of-the-art VQA metrics. In this work, we try to fill the gap in the literature on the evaluation of VQA metrics for medical ultrasound videos. Our work considers seven state of the art objective video quality metrics and tests their performance in giving a reliable approximation of diagnostic accuracy on ultrasound video sequences. Further, we also evaluate the compression performance of the HEVC standard for medical ultrasound videos.

The paper is organized as follows. Section II provides an introduction to the medical video quality evaluation topic. Section III describes the experimental setup, followed by discussion of results in Section IV. Finally, Section V concludes the paper.

II. QUALITY EVALUATION FOR MEDICAL IMAGES AND VIDEOS

In medical VQA both *Objective* and *Subjective* VQA methods are widely used. A brief overview of these methods is given below.

A. Objective VQA

Over the years, several objective video quality metrics have been developed. Different metrics have employed different approaches for the computation of video quality. The evolution of objective video quality metrics can be broadly classified into two categories: *Statistical measures*, and *Human Visual System (HVS) measures*.

- *Statistical Measures*: Most of the early objective metrics are based on statistical measures of one or more specific features of the video pixel information. For instance, one of the most popular metrics, Peak Signal to Noise Ratio (PSNR), measures the signal to noise ratio, wherein signal is the original video and noise is the Mean Square Error (MSE) between the original and distorted video. Similarly, the Moran Peak Ratio (MPR) metric measures the spatial correlation in the video, and the Spatial Frequency Measurement (SFM) metric computes the video quality as a measure of row and column pixel difference [12]. In spite of low complexity computation, these metrics at some occasions show low correlation with the perceptual quality of the video. An attempt to overcome this shortcoming of objective metrics was made

by developing metrics which consider the impact of video features on the HVS.

- *HVS measures*: In the last decade, several metrics based on the response of the HVS to particular image and video features were developed. One of the landmark IQA metrics developed is the SSIM. The SSIM metric measures the structural similarity between the reference and impaired videos by means of luminance, contrast and structural comparison. The philosophy behind SSIM is to represent video quality degradation as a measure of the changes in the structural information [13]. Other metrics like Visual Signal to Noise Ratio (VSNR), VIF were also developed on models based on the response of the HVS [14]. The metrics developed using the HVS approach showed better correlations with the perceptual quality of the video. This advantage was further exploited by researchers which eventually led to the increased adoption of HVS measures in developing metrics for VQA.

Objective quality metrics may not always ensure the diagnostic quality of the medical video. Hence, an effective way to evaluate the diagnostic quality of medical video is via the opinion of a medical expert. This is referred as subjective quality evaluation.

B. Subjective VQA

In subjective quality assessment, medical experts give their opinions on the videos based on the perceptual quality and the diagnostic information preserved in the processed video. Subjective VQA methods involve obtaining the Mean Opinion Score (MOS). To obtain MOS, medical subjects are presented with a randomized set of videos and are asked to rate the quality of the videos on a given scale. The ratings obtained from all the subjects are collected and their mean result in MOS. Several techniques for subjective quality measurements are recommended by International Telecommunications Union (ITU), for instance, Single Stimulus Continuous Quality Scale (SSCQS), Absolute Category Rating (ACR), Double Stimulus Impairment Scale (DSIS), Double Stimulus Continuous Quality Scale (DSCQS), etc. [15].

In practical terms, the method of subjective VQA is not always feasible. Moreover, subjective VQA is often an inconsistent and lengthy procedure. On the other hand objective VQA is a quicker and an easier methodology, however its reliability must be tested. Therefore, it is important to evaluate the performance of objective VQA in terms of its ability to give a better approximation of the diagnostic accuracy of the processed video. One effective method to evaluate the performance of objective VQA is to assess its correlation with the subjective scores, a method which is explored in this paper.

III. MATERIALS AND METHODS

A. Video Sequences

The performance of the objective VQA metrics is evaluated on nine original medical ultrasound videos, each compressed at eight different quality levels, with a frame resolution of 640×416 . Each video sequence has 100 frames, encoded at 25

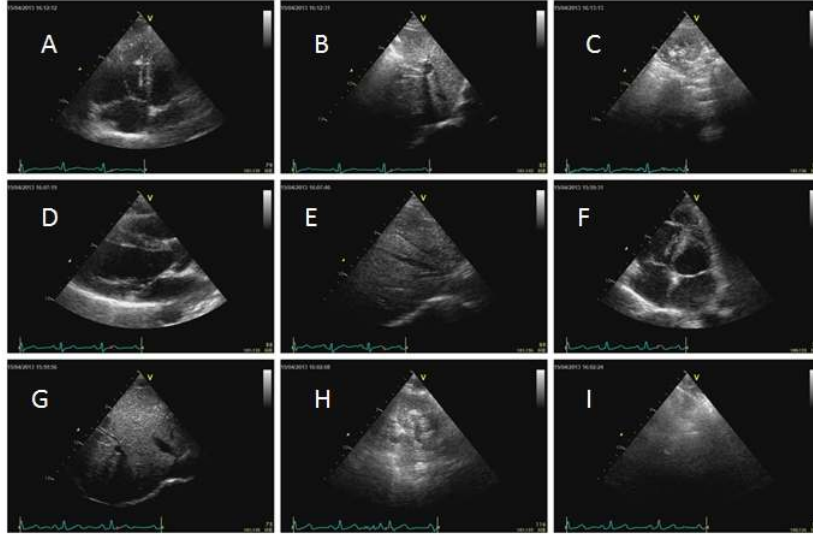


Fig. 1. An example frame of each of the sequences used in the tests. A: Echocardiography in 4 chambers view. The right ventricle is dilated; B: Echocardiography: subcostal view displays the liver and the inferior vena cava; C: Renal ultrasound: cortical and medullary view; D: Echocardiography: parasternal long axis view, displaying left atrium and ventricle, aorta and mitral valves; E: Echocardiography: subcostal view displays the liver, the inferior vena cava and hepatic veins; F: Echocardiography in 4 chambers view: both atria and ventricles are visualized; G: Echocardiography in subcostal view: the liver and hepatic veins are visualized. H: Renal ultrasound: cortical and medullary view; I: Ultrasonography of the lungs. Normal lung, no edema.

TABLE I
OBJECTIVE IMAGE QUALITY METRICS USED IN THE TESTS

Objective Image Quality Metrics	Defining Equation	Features considered
PSNR	$\text{PSNR} = 20 * \log_{10}\left(\frac{255}{\sqrt{\text{MSE}}}\right)$ (1)	Gives ratio of signal over the noise, where signal refers to the original image and noise refers to the standard error.
SSIM [13]	$\text{SSIM} = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}$ (2)	Measures the structural similarity between two images. The constant values are chosen to be $C_1 = 0.01$, $C_2 = 0.03$.
UQI [16]	$Q = \frac{4\sigma_{xy}\mu_x\mu_y}{(\sigma_x^2 + \sigma_y^2)(\mu_x^2 + \mu_y^2)}$ (3)	Measures the structural distortion and gives good approximation of perceived distortion.
Video Quality Metric (VQM) [17]	A standardised metric developed by the National telecommunications & Informations Administration (NTIA), USA. VQM is also recommended by ITU and is widely adopted. VQM measures the quality based on seven parameters to assess the quality. The software is freely available at [17].	
Noise Quality Metric (NQM) [18]	NQM is a weighted signal to noise ratio measure between the original and the processed image. NQM considers variation in contrast sensitivity and local luminance mean, and contrast measures.	
VIF [19]	The images are decomposed into wavelets and computation is done using several models which gives a measure of the visual quality.	
VSNR [14]	Contrast thresholds are used to identify distortions. The distortions above the threshold are modeled to measure the image quality.	

frames per second (fps). Of the nine ultrasound videos, three videos are related to the heart and liver each, two for kidney, and one video is related to the lung. An example frame and a detailed description of each medical video sequence used in the tests is shown in Figure 1.

To obtain the distorted video sequences, the medical videos are compressed using the emerging video compression standard, HEVC. HEVC is the successor to the widely popular H.264/AVC video compression standard. HEVC gives a 50% higher compression than the H.264 standard for the same visual quality [20]. Since HEVC is considered as the future video compression standard, this work also tests its perfor-

mance for medical ultrasound videos.

The compression of the sequences is done at eight different Quantization Parameter (QP) levels using the HM reference software provided by the Joint Collaborative Team on Video Coding (JCT-VC) team [21]. The QP values chosen are 27, 29, 31, 33, 35, 37, 39, and 41. As the QP value increases, the compression ratio increases which in turn leads to lower quality videos. Four out of the eight QP values [i.e. QP = 31, 35, 37, and 41] are chosen as per the testing recommendations of the Video Quality Experts Group (VQEG) [22]. In the tests, we used then 72 impaired medical video sequences i.e., 9 video sequences, compressed at 8 different QPs.

B. Subjective Test

The compressed video sequences were subjectively evaluated for the visual and diagnostic quality by both medical experts and non-medical experts who provided their opinion scores on a scale of a specified range. The subjective evaluation was done using the Double Stimulus Continuous Quality Scale (DSCQS)- type II, which is one of the methodologies recommended by the International Telecommunication Union (ITU) in the document ITU-R BT.500-11 [15]. The DSCQS method is adopted in our tests because in this method the subjective scores are less sensitive to the context, i.e., the ordering and the level of impaired sequences has less influence on the subjective ratings [23]. The DSCQS method is widely used in medical video subjective quality evaluation, for instance, in [24], [8], [25].

The DSCQS methodology uses a Just Noticeable Difference (JND) approach in which the medical expert is presented with two videos side by side, typically the original and a processed video. The subject is asked to rank both the videos. One of the sequences is the reference video, i.e. unimpaired video, whereas the other sequence is impaired. The subject is asked to rate both the sequences on two separate scales of 1 to 5, where 1 corresponds to the lowest and 5 to the highest quality. The subject is unaware of which one is the reference video (the reference video is displayed randomly either at the left or at the right end side). In our tests, the video sequences were displayed on a Liquid Crystal Display (LCD) monitor. The original resolution of the videos was maintained for display in order to avoid any distortions due to scaling. For tests with medical specialists, the evaluation was performed in a room which the specialists use to visualize video sequences and perform diagnosis accordingly. For tests with non expert viewers, the light conditions in the room were those suggested in ITU recommendation [26].

The video sequences were separated in two batches. The subjects were allowed to take a break between the two batches so that fatigue could not affect their judgement. Prior to the actual tests, the subjects underwent a short training session to familiarize with the testing process. During the actual tests, if the subjects were not able to judge the quality of the video in the first viewing, they were provided with an option to play the video sequence again until they were confident of providing a score. The Moscow State University (MSU) perceptual quality tool [27] was used to document the score obtained in the subjective study. The ratings obtained were then used to get the mean scores and other desired statistics.

C. Subjective Scores

The subjective evaluation was done by four medical experts and sixteen non-medical experts. The experts rated the video sequences for their diagnostic quality, whereas the non-experts are more likely to rate based on the visual quality. In the DSCQS method, for each video sequence, two ratings were obtained. One of the scores corresponds to the reference video and the other to the impaired video. If $Ref_{i,j}$ is the rating given to the reference sequence of the j^{th} video by subject i , and $IQ_{i,j}$ is the rating given to the impaired sequence of

the j^{th} video by subject i , then the Differential Opinion Score (DOS) for the j^{th} video by the subject i is given by:

$$DOS_{i,j} = Ref_{i,j} - IQ_{i,j}. \quad (4)$$

The $DOS_{i,j}$ for each video j is obtained for $i = 1, 2, \dots, N$ subjects. The scores of all the subjects were tested for reliability and interobserver variability via the subject rejection procedure mentioned in [15]. The screening procedure methodology is based on determining the normal distribution of the scores by computing the Kurtosis coefficient of the scores. The scores are accepted if the Kurtosis value of the scores is between 2 and 4. If the standard deviation of the subject's scores fall outside the 95% confidence interval range from the mean score then it accounts for large inter-observer variability and makes the scores unreliable, subsequently resulting in the rejection of the subject's scores. In our tests, out of the 16 non-expert subjects, the scores of one subject were rejected since they were out of the expected confidence interval range. None of the expert scores got rejected in the screening procedure. The accepted $DOS_{i,j}$ scores were further used to obtain the mean score i.e. Differential Mean Opinion Score (DMOS) for video sequence j , given by:

$$DMOS_j = \sum_{i=1}^N DOS_{i,j} \quad (5)$$

D. Performance Evaluation Tests

To test the performance of a given objective VQA metric, the correlation of the objective metric with the DMOS is evaluated. A higher correlation would imply that the given objective metric correlates well with the subjective quality and hence is more reliable. Therefore, a correlation analysis between the objective VQA metrics and the DMOS is carried out according to the methodology recommended in [28]. The objective VQA metrics used in our tests are tabulated in Table I.

The correlation between the objective and subjective scores is evaluated using the Pearson Linear Correlation Coefficient (PLCC) and the Spearman Rank Order Correlation Coefficient (SROCC). We obtain the correlation with the expert DMOS and then with the non-expert DMOS, and finally with the combined scores of both experts and non-experts. Further, a non linear regression analysis using a 4-parameter logistic function is performed on the objective metrics in order to improve prediction accuracy and correlation with the DMOS [6]. The 4-parameter logistic function is described in (6):

$$IQ'_j = \beta_2 + \frac{\beta_1 - \beta_2}{1 + \exp(-(\frac{IQ_j - \beta_3}{|\beta_4|}))}. \quad (6)$$

The β values are obtained by implementing (6) using the *nlinfit* tool in MATLAB. The fitted objective values IQ'_j are tested for their correlation with DMOS using the PLCC method.

TABLE II
PLCC AND SROCC OF THE OBJECTIVE SCORES WITH THE DMOS SCORES. THE FIRST TWO ROWS REPORT PLCC RESULTS BEFORE AND AFTER NON-LINEAR REGRESSION. THE THIRD ROW REPORTS THE SROCC RESULTS

	CC		State of the art objective metrics					
		SSIM	VSNR	VIF	UQI	PSNR	VQM	NQM
Experts	PLCC	0.9264	0.8925	0.9258	0.9292	0.9109	0.8080	0.8961
	$PLCC_{Nlin}$	0.9366	0.9150	0.9417	0.9309	0.9261	0.8309	0.9127
	SROCC	0.9375	0.9139	0.9382	0.9251	0.9331	0.8368	0.9090
Non-Experts	PLCC	0.9208	0.8888	0.9431	0.9521	0.8896	0.8146	0.9233
	$PLCC_{Nlin}$	0.9279	0.9200	0.9688	0.9521	0.9116	0.8440	0.9435
	SROCC	0.9383	0.9277	0.9663	0.9495	0.9280	0.8606	0.9464
Experts & Non-Experts combined	PLCC	0.9350	0.9022	0.9532	0.9612	0.9064	0.8249	0.9312
	$PLCC_{Nlin}$	0.9427	0.9321	0.9770	0.9612	0.9272	0.8532	0.9510
	SROCC	0.9570	0.9456	0.9801	0.9633	0.9475	0.8789	0.9587

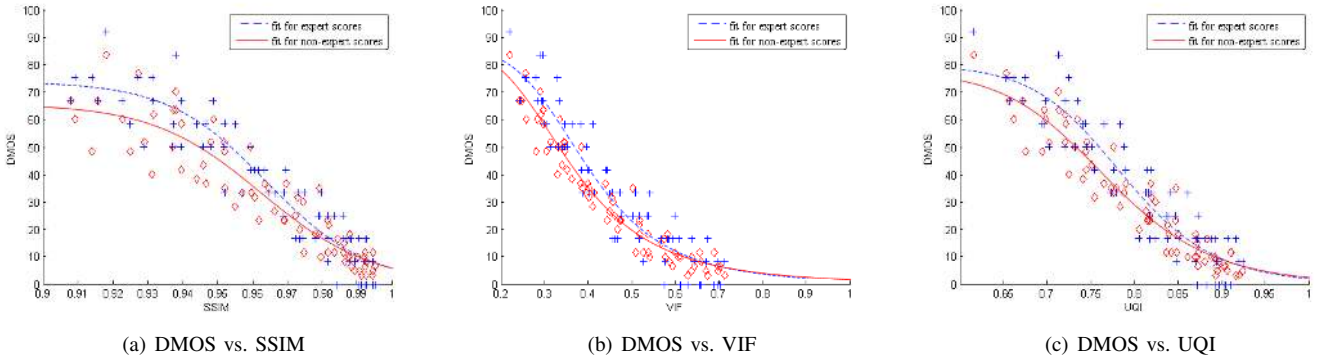


Fig. 2. Scatter plots of DMOS vs. VQA metrics along with logistic fit for both expert and non-expert scores. Blue "+" markers indicate expert scores, and red diamonds indicate non-expert scores.

IV. RESULTS AND DISCUSSION

A. Results

Tables II report the correlation of objective VQA metrics with the subjective scores of experts, non-experts, and all subjects, respectively, with the top performing metric in each correlation test highlighted. In each sub-table, the first two rows present the PLCC measures of the objective metrics with the DMOS for all video sequences before and after non-linear regression analysis. It can be noted that the correlation improves slightly after non linear regression analysis. The third row reports the SROCC measure after non linear regression analysis, a measure of the monotonicity of the prediction.

Figure 3 shows the scatter plots of the expert DMOS against each objective VQA metric considered in the work. The plots show the performance of each metric with respect to the subjective scores of the expert. Further, the plots are logistic fitted with the data to give a better approximation of the metric performance. Figure 2 shows the scatter plot of the metrics SSIM, VIF, and UQI vs. DMOS for both the expert and non-expert scores. Figure 4 shows the variation of DMOS across the QP values for all the nine sequences considered in the tests.

B. Discussion of Results

From Table II, it can be seen that the metrics, VIF, UQI, and SSIM show the highest correlations with the DMOS scores

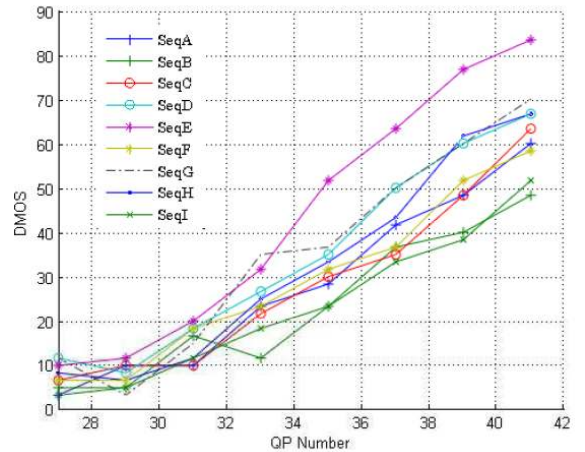


Fig. 4. Variation of DMOS across different QP values. The nine curves represent the nine sequences used in the tests.

of the experts. The performance of the metrics is further illustrated in Figure 3. However, the correlation results of non-expert scores slightly vary. The metric NQM shows a better correlation than the SSIM metric for non-expert scores. The variation in the logistic fit between the non-expert and expert scores can be seen in Figure 2. For the overall scores, i.e., when both expert and non-expert scores are considered, UQI,

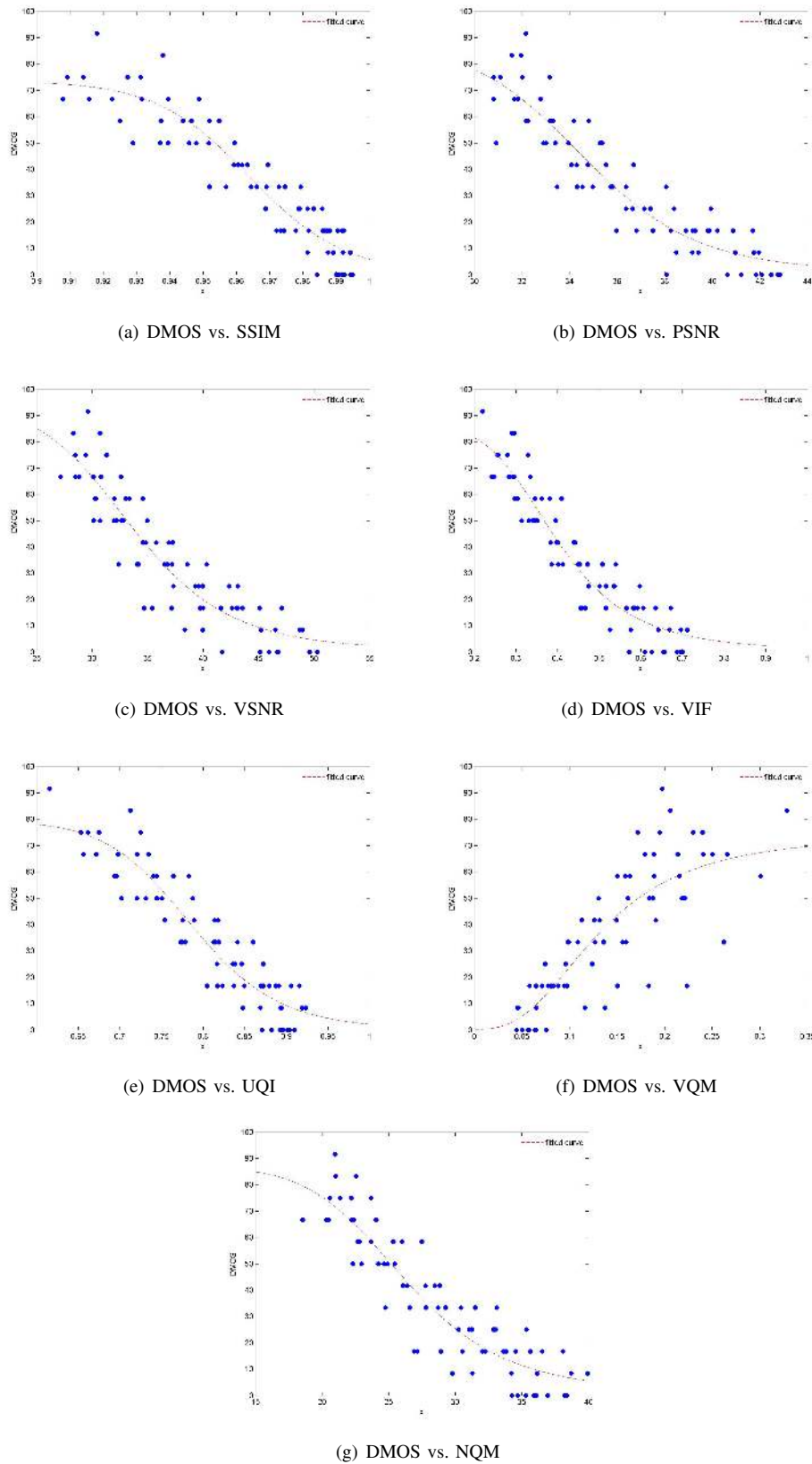


Fig. 3. Scatter plots of DMOS of experts vs. VQA metrics along with logistic fit.

VIF, and SSIM show the highest correlations.

The correlation results from our tests indicate that the met-

rics VIF, UQI, and SSIM consistently show good correlations with both expert and non-expert scores. To further interpret

the results, it can be inferred that these metrics provide a reliable assessment of both, diagnostic and visual quality. The VQM metric shows the lowest correlation amongst the metrics considered.

Figure 4 shows the variation in the DMOS values across the QP range of 27 to 41. As the value of the QP increases, there is also an increase in the DMOS values, which indicates the degradation in the video quality as perceived by the subject. From the plot, it can be seen that considerable degradation in the video quality for most sequences approximately starts around $QP = 35$. In our tests, it was observed that HEVC is able to compress diagnostically reliable ultrasound videos with low bitrate requirements. This implies that a high compression ratio without a compromise on the video quality can be obtained for medical videos using the HEVC standard.

V. CONCLUSION

In this paper, we have evaluated the performance of state of the art VQA metrics for medical ultrasound videos. The correlation of the objective VQA metrics with the subjective scores of both experts and non-experts is used as a measure to indicate the efficiency of the VQA metrics in predicting the diagnostic as well as the visual quality of the processed video. The results showed that VIF, UQI, and SSIM are the best metrics for medical ultrasound video quality evaluation among the considered metrics in this paper. Further, our tests showed that HEVC can compress the considered medical ultrasound videos at low bitrates without compromising on the diagnostic accuracy.

ACKNOWLEDGMENT

The research leading to these results have received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement N°288502 (CONCERTO). We would also like to thank the cardiologists at the University Hospital of Perugia, Italy, in particular Dr. Elisabetta Bordoni and Dr. Clara Riccini.

REFERENCES

- [1] M. G. Martini, R. S. Istepanian, M. Mazzotti, and N. Y. Philip, "Robust multilayer control for enhanced wireless telemedical video streaming," *IEEE Trans. Mobile Comput.*, vol. 9, no. 1, pp. 5–16, 2010.
- [2] R. S. Istepanian, N. Y. Philip, and M. G. Martini, "Medical QoS provision based on reinforcement learning in ultrasound streaming over 3.5 g wireless systems," *IEEE J. Sel. Areas Commun.*, vol. 27, no. 4, pp. 566–574, 2009.
- [3] M. G. Martini and M. Mazzotti, "Quality driven wireless video transmission for medical applications," in *28th Ann. Int. Conf. of the IEEE Engineering in Medicine and Biology Society*, New York, USA, 2006, pp. 3254–3257.
- [4] M. G. Martini, "Wireless broadband multimedia health services: current status and emerging concepts," in *IEEE 19th Int. Sym. on Personal, Indoor and Mobile Radio Commun. - PIMRC*, Nantes, France, Sep. 2008, pp. 1–6.
- [5] M. Razaak and M. G. Martini, "Medical image and video quality assessment in e-health applications and services," in *IEEE HEALTHCOM - The 1st Int. Work. on Service Science for e-Health*, Lisbon, Portugal, Oct. 2013, pp. 6–10.
- [6] K. Brunstrom, D. Hands, F. Speranza, and A. Webster, "VQEG validation and ITU standardization of objective perceptual video quality metrics [standards in a nutshell]," *IEEE Signal Process. Mag.*, vol. 26, no. 3, pp. 96–101, May 2009.
- [7] A. Alesanco, C. Hernandez, A. Portoles, L. Ramos, C. Aured, M. García, P. Serrano, and J. García, "A clinical distortion index for compressed echocardiogram evaluation: recommendations for Xvid codec," *Physiol. Meas.*, vol. 30, no. 5, p. 429, 2009.
- [8] N. Nouri, D. Abraham, J. Moureaux, M. Dufaut, J. Hubert, and M. Perez, "Subjective MPEG2 compressed video quality assessment: Application to tele-surgery," in *Proc. IEEE Int. Symp. Biomed. Imag.: From Nano to Macro*, Rotterdam, Netherlands, April 2010, pp. 764–767.
- [9] A. Kumcu, L. Platiša, B. Goossens, and W. Philips, "Subjective and objective quality evaluation of compressed medical video sequences," in *Third Int. Work. on Quality of Multimedia Experience (QoMEX)*, Mechelen, Belgium, 2011.
- [10] B. Tulu and S. Chatterjee, "Internet-based telemedicine: An empirical investigation of objective and subjective video quality," *Decis. Support Syst.*, vol. 45, no. 4, pp. 681 – 696, 2008.
- [11] E. Cavero, A. Alesanco, L. Castro, J. Montoya, I. Lacambra, and J. Garcia, "SPIHT-based echocardiogram compression: clinical evaluation and recommendations of use," *IEEE J. Biomed. Health. Inform.*, vol. 17, no. 1, pp. 103–112, 2013.
- [12] C. Delgorge, C. Rosenberger, G. Poisson, and P. Vieyres, "Towards a new tool for the evaluation of the quality of ultrasound compressed images," *IEEE Trans. Med. Imag.*, vol. 25, no. 11, pp. 1502–1509, 2006.
- [13] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, 2004.
- [14] D. M. Chandler and S. S. Hemami, "VSNR: A wavelet-based visual signal-to-noise ratio for natural images," *IEEE Trans. Image Process.*, vol. 16, no. 9, pp. 2284–2298, 2007.
- [15] ITU-R BT, Recommendation, "500-11, Methodology for the subjective assessment of the quality of television pictures," International Telecommunication Union, Tech. Rep., 2002.
- [16] Z. Wang and A. C. Bovik, "A universal image quality index," *IEEE Signal Process. Lett.*, vol. 9, no. 3, pp. 81–84, 2002.
- [17] M. H. Pinson and S. Wolf, "A new standardized method for objectively measuring video quality," *IEEE Trans. Broadcast.*, vol. 50, no. 3, pp. 312–322, 2004.
- [18] N. Damera-Venkata, T. D. Kite, W. S. Geisler, B. L. Evans, and A. C. Bovik, "Image quality assessment based on a degradation model," *IEEE Trans. Image Process.*, vol. 9, no. 4, pp. 636–650, 2000.
- [19] Y. Han, Y. Cai, Y. Cao, and X. Xu, "A new image fusion performance metric based on visual information fidelity," *Inf. Fusion*, 2011.
- [20] G. J. Sullivan, J. Ohm, W.-J. Han, and T. Wiegand, "Overview of the High Efficiency Video Coding (HEVC) standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1649–1668, 2012.
- [21] JCT-VC. (2013) High Efficiency Video Coding (HEVC). [Online]. Available: <http://hevc.hhi.fraunhofer.de/>
- [22] J. Ohm, G. J. Sullivan, H. Schwarz, T. K. Tan, and T. Wiegand, "Comparison of the coding efficiency of video coding standards—including high efficiency video coding (HEVC)," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1669–1684, 2012.
- [23] M. H. Pinson and S. Wolf, "Comparing subjective video quality testing methodologies," *SPIE Vis. Commun. and Imag. Process.*, pp. 573–582, 2003.
- [24] H. Yu, Z. Lin, and F. Pan, "Applications and improvement of H.264 in medical video compression," *IEEE Trans. Circuits Syst. I*, vol. 52, no. 12, pp. 2707–2716, 2005.
- [25] Y. Shima, A. Suwa, Y. Gomi, H. Nogawa, H. Nagata, and H. Tanaka, "Qualitative and quantitative assessment of video transmitted by DVTS (digital video transport system) in surgical telemedicine," *J. Telem. Telecare*, vol. 13, no. 3, pp. 148–153, 2007.
- [26] ITU-T Recommendation, P.910, "Subjective video quality assessment methods for multimedia applications," International Telecommunication Union, Tech. Rep., 2008.
- [27] MSU. (2013) MSU perceptual video quality tool. [Online]. Available: http://compression.ru/video/quality_measure/perceptual_video_quality
- [28] K. Seshadrinathan, R. Soundararajan, A. C. Bovik, and L. K. Cormack, "Study of subjective and objective quality assessment of video," *IEEE Trans. Image Process.*, vol. 19, no. 6, pp. 1427–1441, 2010.