Journal of Medical Imaging

MedicalImaging.SPIEDigitalLibrary.org

CUQI: cardiac ultrasound video quality index

Manzoor Razaak
Maria G. Martini

SPIE.

Manzoor Razaak, Maria G. Martini, "CUQI: cardiac ultrasound video quality index," *J. Med. Imag.* **3**(1), 011011 (2016), doi: 10.1117/1.JMI.3.1.011011.

Journal of Medical Imaging 3(1), 011011 (Jan–Mar 2016)

CUQI: cardiac ultrasound video quality index

Manzoor Razaak* **and Maria G. Martini***
Kingston University, Wireless and Multimedia Networking Research Group, Penhryn Road, Kingston upon Thames, KT1 2EE, London, United Kingdom

Abstract. Medical images and videos are now increasingly part of modern telecommunication applications, including telemedicinal applications, favored by advancements in video compression and communication technologies. Medical video quality evaluation is essential for modern applications since compression and transmission processes often compromise the video quality. Several state-of-the-art video quality metrics used for quality evaluation assess the perceptual quality of the video. For a medical video, assessing quality in terms of "diagnostic" value rather than "perceptual" quality is more important. We present a diagnostic-quality–oriented video quality metric for quality evaluation of cardiac ultrasound videos. Cardiac ultrasound videos are characterized by rapid repetitive cardiac motions and distinct structural information characteristics that are explored by the proposed metric. Cardiac ultrasound video quality index, the proposed metric, is a full reference metric and uses the motion and edge information of the cardiac ultrasound video to evaluate the video quality. The metric was evaluated for its performance in approximating the quality of cardiac ultrasound videos by testing its correlation with the subjective scores of medical experts. The results of our tests showed that the metric has high correlation with medical expert opinions and in several cases outperforms the state-of-the-art video quality metrics considered in our tests. [DOI: 10.1117/1.JMI.3.1.011011]

Keywords: cardiac ultrasound; ultrasound videos; quality assessment; objective metric; motion information; edge information.

Paper 15149SSRR received Jul. 15, 2015; accepted for publication Feb. 16, 2016; published online Mar. 14, 2016.

1 Introduction

Medical imaging systems are increasingly used for diagnosis in healthcare services. Advancements in communication systems have facilitated medical images and videos in becoming a significant part of modern telemedicine systems. Medical videos are transmitted over communication channels for various applications such as remote patient monitoring and diagnosis, medical consultations, ambient assisted living, educational purposes, and so on. It is expected that the significant ongoing developments in medical image acquisition and processing techniques would see an increasing adoption of telemedicine technologies in healthcare.

Medical images, especially for the recent high-definition images, demand huge amounts of storage space and network capacity for transmission. To overcome this demand, video compression is often used. The compression process reduces the file size of a given image or video. The compression ratio is the ratio of a given video file size before compression over the video file size after compression. The compression ratio depends on the compression codec used. For instance, the latest video compression standard high-efficiency video coding (HEVC) is known to provide a 50% higher compression ratio over its predecessor, compression standard H.264, and in some cases can possibly reduce the file size by a factor of over 100 and still keep an acceptable perceptual quality. Generally, it is ensured that the extent of the compression applied does not introduce any noticeable impairments on perception of the image. In some cases, despite low compression ratios, it is possible that some impairments are introduced that lead to a compromise in the quality of medical videos for diagnosis. Further, during transmission over communication channels, videos may suffer transmission impairments due to factors such as bandwidth limitations and network congestion, which may also compromise the video quality. Due to the sensitive nature of the information they carry, quality is a crucial factor for medical images and videos. Hence, it is important to ensure, via video quality evaluation, that the video quality is not affected after compression and transmission.[1]

The process of video quality evaluation can be of two types: "subjective" and "objective" quality evaluation. Subjective video quality evaluation is an approach in which a pool of observers rate the quality of a video on a given rating scale. The average of the observers' ratings is computed and used as a measure of the video quality. Subjective quality evaluation is often a reliable approach since the quality is rated by a pool of observers. However, the disadvantages of this approach include possible inconsistencies in the ratings obtained, costs to run the tests, and challenges to implement in real time. Moreover, in the context of medical videos, subjective quality evaluation is more difficult, since observers have to be medical experts, which restricts the choice of observers. On the other hand, objective video quality evaluation is an approach in which mathematical models are used to measure the quality of a possibly impaired video with respect to the original one. Most video quality metrics (VQMs) are designed and tested on natural videos with a goal of representing the perceptual quality of the video. The state-of-the-art VQMs, although not specifically designed for medical applications, are still popularly used in medical video quality evaluation. In our previous work,[2] we tested popular state-of-the-art VQMs for medical video quality assessment and evaluated their performance for quality evaluation of the medical videos under test. In the study, it was observed that some of the VQMs do not perform as well as in natural videos for medical video quality evaluation. Several

*Address all correspondence to: Manzoor Razaak, E-mail: m.razaak@kingston.ac.uk; Maria G. Martini, E-mail: m.martini@kingston.ac.uk

state-of-the-art quality metrics considered in the test had low correlation with the subjective scores of medical experts, indicating weaker performance in the context of medical ultrasound videos. It was also observed that the quality metrics were mainly designed by considering the image properties of natural videos and are not necessarily well suited for medical video quality evaluation. For medical videos, the "quality" factor is dependent on the suitability of the video for reliable diagnosis rather than its perceptual quality. Therefore, it is important that the quality metrics for medical videos be designed with a focus on analyzing the characteristics that influence the diagnostic suitability of the videos.

Toward this approach, we propose a VQM designed for quality evaluation of cardiac ultrasound videos termed as cardiac ultrasound video quality index (CUQI). The proposed metric is based on the features that cardiologists consider as important features for performing a diagnosis via cardiac ultrasound: the identification of the boundaries between organs and of the walls of the heart, as well as the motion/contraction pattern of the heart. In particular, we consider the phenomenon of motion perception and edge detection of human visual system (HVS). The HVS is known to respond to static and motion stimuli separately. The motion of objects in a video helps the HVS in object identification. Further, edge detection by the HVS enables it to process semantic information from an image.

In this regard, CUQI considers the motion characteristics and the edge information of the cardiac video to evaluate its video quality. In this work, we test and compare the performance of CUQI with other state-of-the-art VQMs. The results show that the proposed metric exhibits consistently high performance and in some cases outperforms several objective quality metrics considered in our tests.

2 Background

In this section, a brief discussion of motion and edge perception by HVS is provided, followed by a brief insight into optic flow methods. An overview of different compression artifacts is provided, followed by a discussion of the related work.

2.1 *Motion and Edge Perception*

Motion in a video acts as a stimulus to the HVS. Classical literature has studied the response of the HVS to motion stimuli and several theories have been suggested in the scientific literature to explain its behavior.[3–7]

Motion perception is important for several reasons. Motion enables shape perception, object speed and direction judgments, navigation, and activity recognition.[7–9] Motion in videos can make objects more detectable. In the context of medical videos, for structure detection, ultrasound videos are preferred over static images, as they make the detection process easier. This is because motion in ultrasound videos provide more information over static images and reduces independent noise significantly. Also, visual motion enables better perception of structure boundaries and reduces patient motion artifacts.[10] Video processing may distort the apparent motion present in the video and may influence the diagnosis process of observers, especially during the defect detection processes.[11]

Similarly, several studies have shown that the perception of an image by the HVS highly relies on the structural information of the image and that the edge information in an image can have significant impact on visual perception of the image.[12–14] In the context of medical videos, edge information plays an important role and is predominantly used in applications like image segmentation, region of interest detection, classification, image interpolation, and noise removal.[15–17] The edges of an image represent a crucial part of the image that can be effectively used in VQA, and the medical doctors we interacted with confirmed its importance for performing a diagnosis.

2.2 *Motion Estimation Using Optic Flow Methods*

Motion estimation involves computing motion vectors, which represent the predicted motion in the next frame with respect to the current frame. Some of the methods for motion estimation are the block-matching, phase correlation, and optical flow methods. Optical flow is one of the most widely used methods for motion estimation. Optical flow measures the spatiotemporal changes and provides information about the speed, direction, and path of an object in an image/video.[18] Optical flow methods explore the time-varying intensity of the pixels across successive video frames for motion estimation. If $I(x, y, t)$ represents the intensity of a pixel at position (x, y) at time t on a two-dimensional (2-D) plane, then at time $t + \delta t$, the pixel with the same intensity will be in location $x + \delta x$, $y + \delta y$.

$$I(x, y, t) = I(x + \delta x, y + \delta y, t + \delta t). \tag{1}$$

Assuming constant intensity and using Taylor's expansion, the optical flow equation is given as

$$I_x u + I_y v + I_t = 0, \tag{2}$$

where $u = dx/dt$ and $v = dy/dt$ represent the motion vectors in the horizontal and vertical directions. A solution to Eq. (2) is obtained by introducing additional constraints.[19] The type of constraint introduced defines the type of optical flow computation.

2.3 *Compression Artifacts*

In simple terms, image/video compression is achieved by removal of redundant data, which often implies loss of information, typically with negligible impact on the perception quality. Various steps in compression codecs, such as motion compensation, quantization, and transform coding, may distort the edge and motion information present in the video. On occasions, when compression distortions cross the perceptual threshold, visible artifacts could be seen on the images, which in turn impacts the image visual quality. Some of the most common compression artifacts that introduce spatial and temporal distortions include "blockiness"[20] due to the block-based approach used by codecs, "blurriness and ringing artifacts"[21] due to quantization process, "mosquito noise"[22] due to the interframe coding used by compression codecs, "motion-compensated edge artifacts" due to motion prediction errors by the codec, "ghostiness,"[23] and "jerkiness."[24]

2.4 *Related Work*

Image/video quality assessment is an active area of research. In the literature, several objective VQMs have been presented. Mean square error (MSE) and peak signal-to-noise ratio (PSNR) are two of the most widely used metrics for image and video quality assessment. Both MSE and PSNR are often used to measure the loss in signal quality, which can be a one-dimensional audio signal or a 2-D image signal. Therefore, these

statistical quality models do not necessarily give a good approximation of the impact of image quality degradation on the perception of the image and are often shown to correlate poorly with human perception.

In recent years, several image and VQMs based on HVS response models to particular image and video features have been proposed. One such popular metric is the structural similarity index metric (SSIM), which measures the structural similarity between reference and impaired image/videos by means of luminance, contrast, and structural comparison.[25] Similarly, other metrics such as visual signal-to-noise ratio (VSNR),[26] visual information fidelity (VIF),[27] and several others were developed based on the response of the HVS to different feature stimuli. The metrics developed by considering HVS behaviors have shown better correlations with the perceptual quality of the image/video.[28]

Several VQMs were originally designed for image quality assessment, then later extended to measure video quality. In recent years, a few research works have developed metrics for video quality evaluation by considering the spatial and temporal features of the video. The VQM from the National Telecommunications and Information Administration uses several spatial and temporal distortions from a video to measure the quality. The main components of the VQM metric are spatial distortions such as edge shifts, spatial information loss, color losses, and so on. The only temporal distortion measured by this metric is the measure of frame differences.[29] A few VQMs based on motion information have been developed by researchers. For instance, a VQM was presented based on spatial-temporal distortion assessment.[30] The motion quality was evaluated along the motion trajectories of the video by constructing the Gabor filter responses from the reference and distorted videos. A human visual speed perception model was developed[31] to estimate the motion information, the perceptual uncertainty, and an associated spatiotemporal weight for quality assessment of videos. Also, a few other works[32–36] have used quality assessment strategies involving motion information.

There is a dearth of VQMs specifically designed for medical videos. The state-of-the-art objective VQMs may not be suitable for medical videos, since they are mainly designed to evaluate the perceptual quality, which may not ensure the diagnostic reliability of the video. To the best of our knowledge, there is currently no VQM that considers both motion and edge information as the main criteria for medical video quality assessment. The proposed metric is an approach for quality assessment of cardiac ultrasound video sequences, as it considers two important video characteristics: motion information and edge information.

3 Quality Metric for Cardiac Ultrasound Videos

To illustrate how compression may impact the motion information in a cardiac ultrasound video, the motion vector trajectories of a single pixel of an original video and its distorted version across the frames of the video are shown in Fig. 1. The distorted video is obtained by compressing the original video with a quantization parameter (QP) of 41. It can be observed that there are significant distortions in the pixel trajectory of the distorted video and in most frames the pixel belonging to the distorted video suffers significant reduction in the vector magnitude values. Such distortion to the motion information may have significant impact on the video perception.

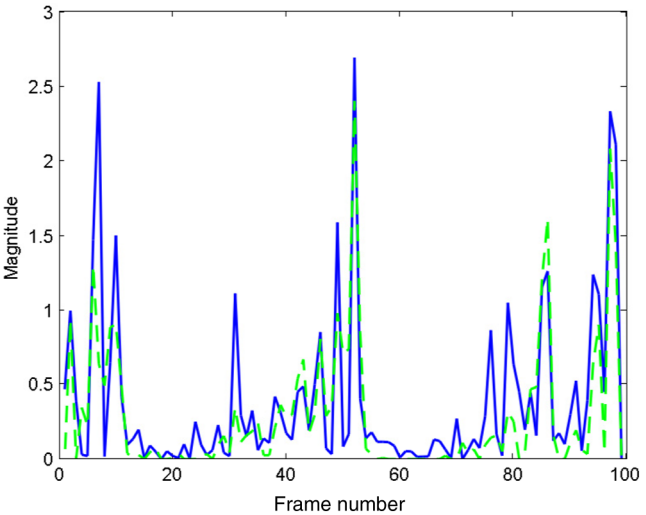

Fig. 1 The motion vector trajectory of a single pixel across all frames of both original and the impaired videos. Blue curve presents the pixel trajectory of original cardiac video. Dashed green curve presents the pixel trajectory of the distorted cardiac video.

The proposed CUQI metric measures the distortion to the motion information along with the edge distortions between the reference and the distorted video and uses them to quantify the quality of the cardiac video. In the first stage, the classic Horn and Schunck optical flow method[37] is employed to estimate motion in cardiac videos. Next, a window-based Gaussian filter weighting function is used to assign weights to the motion vectors of each pixel of the original and distorted videos. The weighted motion vectors of the original and distorted videos are used to compute the motion distortion in the cardiac video. In the second stage, edge information quality preserved in the compressed video is computed using Laplacian of Gaussian (LoG) edge filter and correlation methods. Finally, a quality score is obtained for the cardiac videos by combining both the motion and edge quality scores.

3.1 *Motion Vector Estimates*

The Horn and Schunck optical flow method is used to estimate the motion between two successive frames in the video. For each pixel i, two motion vectors in the horizontal and vertical directions are derived. Let u_i be the horizontal motion vector and v_i be the vertical motion vector of a pixel i in a given frame. The root sum of squares of vectors u_i and v_i gives the resultant magnitude vector

$$m_i = \sqrt{u_i^2 + v_i^2}. \tag{3}$$

The horizontal and vertical motion vectors for each pixel of the original and the distorted image frame are computed and the resultant magnitude vector defined by Eq. (3) for each pixel is obtained, resulting in a matrix M for each frame. Figure 2 shows estimated motion vectors of an example frame from an ultrasound cardiac sequence.

3.2 *Weighting Function*

A Gaussian weighting function is applied to the magnitude vectors matrix M of both the original and distorted frame to

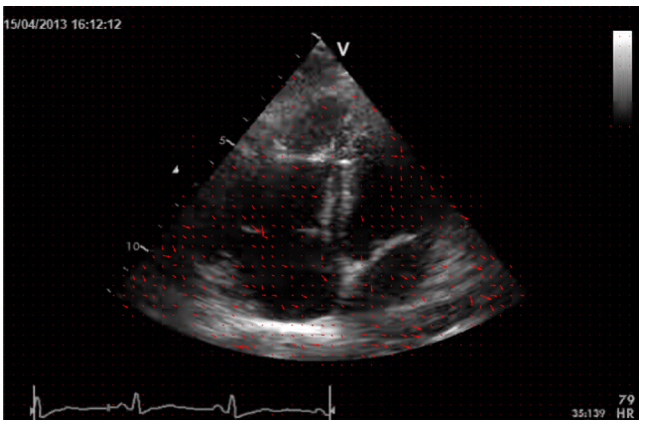

Fig. 2 The motion vectors on frame 35 of a cardiac ultrasound sequence. The motion vectors are illustrated for every 10 pixels.

compute the Gaussian weights. A window-based approach is used. Let P be an $n \times n$ window of the motion vector matrix M. If μ is the mean of the magnitude vectors in the window P and σ is the standard deviation, then the Gaussian weight for each motion vector $i = 1,2,\ldots,n^2$ in window P can be computed as follows:

$$w_{M_i} = \exp^{\frac{-(M_i-\mu)^2}{2\sigma^2}}. \tag{4}$$

We define W as an $n \times n$ matrix of the weights, and the weight for each pixel indicates the distance of the vector magnitude of the pixel from the mean vector magnitude of the window. Pixels whose vector magnitude values are closer to the μ value get higher weights. A sliding window is used on matrix M to compute a weight for each pixel of the frame. Then, a Gaussian weighted response of matrix M can be derived by taking the product of the element of matrices obtained from Eqs. (3) and (4):

$$M_g = W * M. \tag{5}$$

The matrix M_g is composed of elements $M_{g_i} = w_i M_i$ where each pixel i is tuned to its corresponding Gaussian weight. The matrix M_g is nothing but a Gaussian filtered response that reduces the noise in the vector magnitude matrix M and provides a smooth version of M.

3.3 *Edge Detection*

To compute edge distortions, we first extract edge information from the videos. For edge detection, a LoG detector proposed by Marr and Hildreth is used. This detector uses a Gaussian filter along with a Laplacian operator, which is nothing but the second derivative of the filter and detects edges at the point of zero-crossings. The LoG edge detector can be represented as shown in Eq. (6), where x and y are the Cartesian coordinates of the pixels in the image.

$$\mathrm{LoG}(x,y) = -\frac{1}{\pi\sigma^4}\left(1 - \frac{x^2+y^2}{2\sigma^2}\right)\exp^{\frac{-(x^2+y^2)}{2\sigma^2}}. \tag{6}$$

3.4 *Quality Metric*

The CUQI metric is a full reference metric where the "distorted" version of the video is compared with its original video, termed the "reference" video. CUQI has two components, "motion quality" and "edge quality," computed using the motion and edge measures obtained in Eqs. (5) and (6), respectively.

Let R be a frame of the reference video and D be the corresponding frame from the compressed video. Using Eq. (5), the Gaussian weighted response for each frame of the reference video and the corresponding frame of the distorted video can be represented by $R_g(i,f)$ and $D_g(i,f)$, respectively, where $i = 1,2,\ldots N$ is the pixel index in the frame f, and N is the total number of pixels in the frame. Further, to restrict the quality metric in the range between 0 and 1, the condition of "boundedness" is applied

$$f(\alpha) = \frac{1}{\alpha^2+1}; 0 \le f(\alpha) \le 1. \tag{7}$$

The error between $R_g(i,f)$ and $D_g(i,f)$ is computed using the MSE

$$e_m(f) = \frac{1}{N}\sum_{i=1}^{N}[R_g(i,f) - D_g(i,f)]^2. \tag{8}$$

Implementing the bounding function of Eq. (7)

$$E_M(f) = \frac{1}{N}\sum_{i=1}^{N}\left[\frac{1}{R_g(i,f)^2+1} - \frac{1}{D_g(i,f)^2+1}\right]^2. \tag{9}$$

The error index given in Eq. (9) is between a single frame of the reference and the compressed video. Thus, it is a measure of change in the motion quality of a frame based on the change in motion estimates between the reference and the impaired videos.

The motion information error across the video is calculated by taking the mean across all k frames of the video. The measure obtained by (9) is a measure of the motion error of a frame f. Subtracting the error from 1 would give a measure of the motion quality. Therefore, the motion quality of a given impaired cardiac video is given as

$$Q_M = \sum_{f=1}^{k}[1 - E_M(f)]. \tag{10}$$

To compute the second component, "edge quality," the edge detector defined by Eq. (6) is implemented on both the reference video frame and the corresponding compressed video frame. If $R_{\mathrm{edg}}(i,f)$ and $D_{\mathrm{edg}}(i,f)$ are the edge maps of the corresponding reference and compressed video frame f with $i = 1,2,\ldots,N$ pixels, then the edge quality measure can be obtained by computing the correlation between $R_{\mathrm{edg}}(i,f)$ and $D_{\mathrm{edg}}(i,f)$. The correlation score would give a measure of closeness between the edge map of reference and the compressed video frame. Therefore, the edge quality measure can be defined as

$$\mathrm{EdgeErr}(f) = \frac{\sum_{i=1}^{N} R_{\mathrm{edg}}(i,f)D_{\mathrm{edg}}(i,f) - \bar{R}_{\mathrm{edg}}(i,f)\bar{D}_{\mathrm{edg}}(i,f)}{(N-1)S_{R_{\mathrm{edg}}}S_{D_{\mathrm{edg}}}}. \tag{11}$$

In Eq. (11), $\bar{R}_{\mathrm{edg}}(i,f)$ and $\bar{D}_{\mathrm{edg}}(i,f)$ are the mean of edge information across the reference and impaired video frame, respectively. Similarly, $S_{R_{\mathrm{edg}}}$ and $S_{D_{\mathrm{edg}}}$ are its corresponding standard

deviation for N number of pixels. Further, by taking the mean across all k frames of the video, the Edge Quality measure can be obtained as

$$Q_E = \sum_{f=1}^{k} \text{EdgeErr}(f). \tag{12}$$

Finally, the CUQI metric measure is obtained by multiplying the "motion quality" index by the "edge quality" index. This approach gives equal weight to both motion and edge quality measures and hence captures any degradation in either of the quality indices. Therefore, CUQI is computed as

$$\text{CUQI} = Q_M \times Q_E. \tag{13}$$

4 Metric Implementation and Evaluation

4.1 *Metric Implementation*

The CUQI metric was implemented on MATLAB®. The Horn and Schunck optical flow algorithm was implemented on two successive frames (i.e., the current and the next frames) to obtain the motion vectors of the current frame. Thus for F frames of a video, the motion vectors were computed for $F-1$ frames.

An $n \times n$ sliding window-based approach was used for implementing the Gaussian weights. The Gaussian kernel size was chosen to be 32×32, as it gave a good approximation of the quality for less complexity than smaller windows. Our tests showed that increasing the block size to 64×64 would increase the computation speed; however, the approximation of quality was weaker. Using a smaller block size of 15×15 or 7×7 decreases the computation speed without significant changes to the quality approximation when compared to 32×32 window.

The LoG edge detector was implemented in MATLAB® to extract the edge map. For the LoG edge detector, it is important to define the threshold and sigma values. Choosing lower threshold and sigma values may result in wrong detection of speckle noise as edge information, whereas higher threshold values might sometimes miss detecting certain edges from the video. Therefore, in our experiments, we tested several threshold and sigma values and found that a threshold value of 0.0035 and a sigma value of 2.25 for the filter gave the best edge map extraction. The correlation measure between the reference and the impaired cardiac video frames was done using the Pearson correlation as defined in Eq. (11). Also, since the motion quality is measured for $F-1$ frames only, the edge quality was also measured for the same number of frames.

4.2 *Evaluation Setup*

4.2.1 *Video sequences*

The CUQI metric was tested on cardiac ultrasound sequences. In our tests, we used three cardiac ultrasound sequences consisting of 100 frames with a frame resolution of 640×416 and frame rate of 25 frames per second. The video sequences were compressed at eight different quality levels using the latest video compression standard, HEVC.[38] The quality levels were determined by the value of the QP used. Therefore, eight different QP values were used to impair the video at eight different quality levels. The QP values chosen were 27, 29, 31, 33, 35, 37, 39, and 41. The compression ratios achieved were in the range of 100:1 to 550:1 depending on the QP value used. For instance, the original video sequence, SeqA (refer to Fig. 3), was of file size 84,000 KB. At a QP level of QP = 27, the file size was reduced to 834 KB. In total in the tests, we used 27 video sequences, i.e., three video sequences compressed at eight different quality levels. Figure 3 shows an example frame of the cardiac sequences considered in our tests.

4.2.2 *Subjective test*

The compressed video sequences were subjectively evaluated for the diagnostic quality by medical experts. The subjective evaluation setup in our tests followed the double stimulus continuous quality scale (DSCQS) approach—a type II methodology that is one of the methodologies recommended by the International Telecommunication Union (ITU) in the document ITU-R BT.500-11.[39] The DSCQS methodology is widely used in medical video subjective quality evaluation, for instance, in Refs. 40–42. The subjective scores obtained by this methodology are less sensitive to the context; i.e., the ordering and the level of impaired sequences have less influence on the subjective ratings.[43] The context effect occurs when the ratings given by the subject are influenced by the severity and order in which the impaired video sequences are placed. This effect might lead to memory-based biases from previously viewed sequences that may impact the ratings provided by the subjects to the video under test. In DSCQS methodology, the context effects are minimized, as pairs of videos are shown in a randomized order so that the subjects are less likely to be biased in their ratings from the previously viewed sequences.[43]

In the DSCQS methodology, the medical expert is presented two videos side by side, typically the reference and the impaired

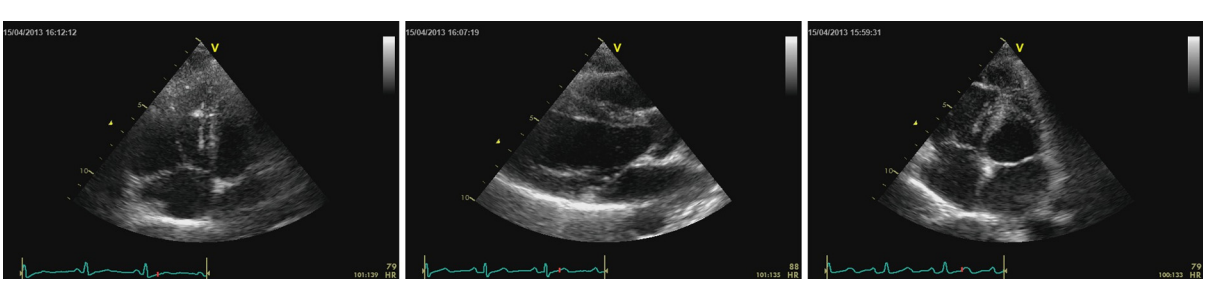

Fig. 3 An example frame of each of the sequences used in the tests. Right to left. Seq A: Echocardiography in four chambers view. The right ventricle is dilated. Seq B: Echocardiography: parasternal long axis view, displaying left atrium and ventricle, aorta and mitral valves. Seq C: Echocardiography in four chambers view: both atria and ventricles are visualized.

videos. The subject is asked to rate both video sequences on two separate scales of 1 to 5, where 1 corresponds to the lowest and 5 to the highest quality. The quality, for the medical experts, is defined as the suitability of the medical video for a reliable diagnosis. In other words, the experts are asked to rate the video based on how reliable they find the video to be used for diagnosis. The DSCQS is a blind test, i.e., the subject is unaware of which one is the reference video. The video sequences were displayed in their original resolution, i.e., 640×416, on a liquid crystal display monitor, and the tests were performed in a room that specialists use to visualize video sequences and perform diagnoses. The monitor display settings and the ambient lighting conditions were set at natural conditions under which the medical experts perform their daily diagnosis activities. In other words, the subjective test was conducted in an ambience similar to the ambience the medical experts use to view cardiac videos on a regular basis.

Prior to the actual tests, the subjects underwent a short training session to familiarize them with the testing process. During the actual tests, the subjects had an option to replay the video sequence in case they were not able to judge the quality of the video in the first viewing; however, they were not able to manipulate the videos or do operations such as zoom/pan and others. The Moscow State University perceptual quality tool[44] was used for conducting the subjective tests and documenting the score obtained in the subjective study. The ratings obtained were then used to get the mean scores and other desired statistics.

The subjective tests involving medical experts were conducted at the University Hospital of Perugia, Italy. The ethics approval to conduct the subjective tests was obtained from the Faculty Research Ethics Committee (in SEC Faculty) in Kingston University and Ethics Committee of Hospital of Perugia, Italy.

4.2.3 *Subjective scores*

The subjective evaluation was done by four medical experts (three cardiologists and a radiologist) who subjectively rated the video sequences based on their diagnostic and perceptual quality. In the DSCQS method, for each video sequence, two ratings were obtained, where one score corresponds to the score given to the reference video and the other to the impaired video. If $\mathrm{Ref}_{i,j}$ is the rating given to the reference sequence of the j'th video by subject i, and $\mathrm{Imp}_{i,j}$ 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 as

$$\mathrm{DOS}_{i,j} = \mathrm{Ref}_{i,j} - \mathrm{Imp}_{i,j}. \tag{14}$$

The $\mathrm{DOS}_{i,j}$ for each video j is obtained for each subject i. The scores of all the subjects were tested for reliability and interobserver variability via the subject rejection procedure adopted from Ref. 39. The subject screening methodology is based on determining the normal distribution of the scores by computing the Kurtosis coefficient of the scores. The scores are considered to be normally distributed and accepted if the Kurtosis value of the scores is between 2 and 4. In cases where the scores are not normally distributed and if the standard deviation of the subject's scores falls outside the 95% confidence interval range from the mean score, then it accounts for large interobserver variability and makes the scores unreliable, subsequently resulting in the rejection of the subject's scores. In our tests, none of the expert ratings were rejected. The accepted $\mathrm{DOS}_{i,j}$ scores were further computed to obtain the mean score, i.e., differential mean opinion score (DMOS) for video sequence j, given as

$$\mathrm{DMOS}_j = \sum_{i=1}^{N} \mathrm{DOS}_{i,j}. \tag{15}$$

4.2.4 *Performance evaluation*

To evaluate the performance of the metric, the correlation analysis approach is used. The correlation between the CUQI metric scores and the DMOS scores is computed. A high correlation implies that the metric shows good performance. Further, we compare the performance of the CUQI metric with seven other state-of-the-art metrics also considered in our previous study.[2]

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). Further, a nonlinear regression analysis using a four-parameter logistic function is performed on the objective metrics in order to improve prediction accuracy and correlation with the DMOS.[45] The four-parameter logistic function is described as

$$\mathrm{VQ}'_j = \beta_2 + \frac{\beta_1 - \beta_2}{1 + \exp\left[-\left(\frac{IQ_j - \beta_3}{|\beta_4|}\right)\right]}. \tag{16}$$

The β values are obtained by implementing Eq. (16) using the *nlinfit* tool in MATLAB®. The fitted objective values VQ'_j are tested for their correlation with DMOS using the PLCC method.

Table 1 The PLCC and SROCC of the objective scores with the DMOS scores. The first two rows report PLCC results before and after nonlinear regression. The third row reports the SROCC results.

	State of the art objective metrics							
CC	SSIM	VSNR	VIF	UQI	PSNR	VQM	NQM	CUQI
PLCC	**0.9304**	0.9040	0.9214	0.9233	0.8917	0.9236	0.9179	0.9005
$\mathrm{PLCC}_{\mathrm{Nlin}}$	0.9335	0.9261	0.9317	0.9245	0.9098	0.9258	0.9278	**0.9415**
SROCC	0.9222	0.9217	0.9279	0.9257	0.9125	0.9301	0.9160	**0.9311**

Note: The best performing metric in each row is highlighted in bold.

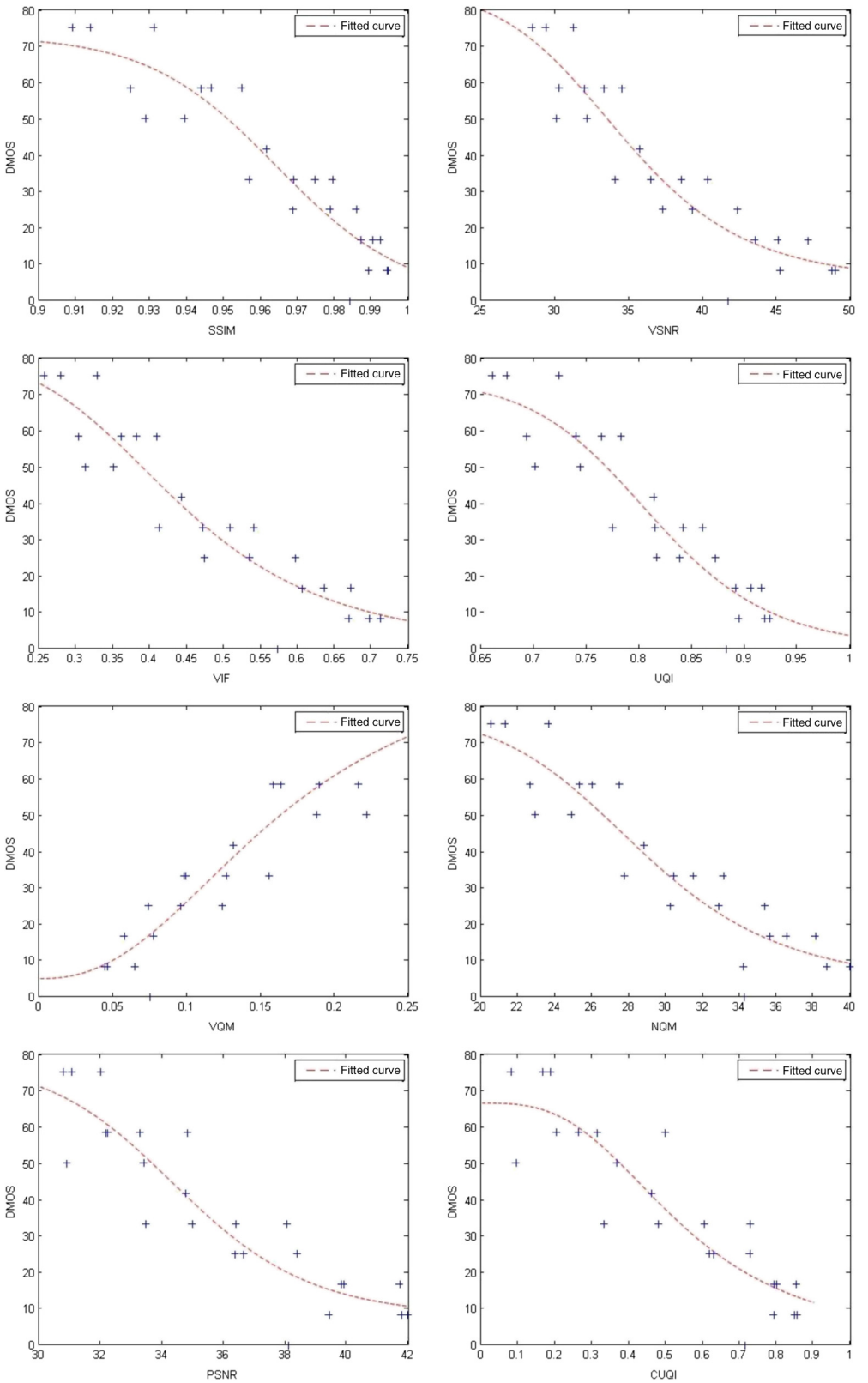

Fig. 4 Scatter plots of DMOSs of experts versus objective quality metrics along with logistic fit.

5 Results

Table 1 presents the correlation scores of the metrics considered with the subjective scores of the medical experts. The first row shows the PLCC scores, the second row presents the PLCC scores after nonlinear regression analysis, and the third row presents the SROCC scores. The best-performing metric in each row is highlighted. It can be seen that the CUQI metric shows consistently high correlation with the subjective scores.

Table 2 The F-test results evaluating the performance of the CUQI with other objective metrics. The $F_{critical}$ value is 2.72. The first row shows the F_{Ratio} value obtained when CUQI performance is evaluated with other metrics. In the second row, "I" indicates the CUQI is statistically inferior to the metric of that column, "S" indicates the CUQI is statistically superior to the metric of that column and the symbol "–" indicates the performance of the CUQI and the corresponding metric is equal.

	State-of-the-art objective metrics						
CUQI ($F_{critical} = 2.72$)	SSIM	VSNR	VIF	UQI	PSNR	VQM	NQM
F_{Ratio}	3.51	2.70	1.73	2.41	1.46	4.59	2.70
Result	I	—	S	S	S	I	—

In some instances, the CUQI metric performs better than some of the state-of-the-art metrics considered in our tests. In particular, under PLCC analysis after nonlinear fitting and under SROCC analysis, the CUQI shows high correlation. It can be observed that the SSIM, VQM, and VIF metrics show very high performance, too. The performance of the metric is further evaluated statistically using the F-test in Sec. 5.1.

The performance of the metric is further shown in Fig. 4. The objective scores of all the quality metrics considered in our tests are plotted against the DMOS values. A best fitting logistic curve is also shown in the plot. The logistic fit curve corresponds to the four-parameter logistic function used for nonlinear regression analysis of objective scores to improve the prediction accuracy as described by Eq. (16) in Sec. 4.2.4. It can be observed that the majority of quality predictions from the CUQI metric are close to the fitted curve. This indicates that CUQI gives quality predictions very close to the subjective scores of the medical experts.

5.1 *Statistical Significance Test*

The performance of the metric is further tested using the F-test statistical significance test. The F-test statistical evaluation is an ITU recommendation to evaluate the performance of one objective quality metric over the other.[46]

The "null hypothesis" considered for the F-test is that the variance of the residual error of two objective quality metrics are equal. Based on this null hypothesis, the performance of the CUQI metric is evaluated with other metrics at 1% significance level. The "degrees of freedom" for the test is one unit lower than the number of test videos, i.e., $24 - 1 = 23$. The results of the F-test are shown in Table 2.

In Table 2, the F_{Ratio} values of the CUQI with each metric are given in the first row. When the F_{Ratio} value exceeds the $F_{critical}$ value, the performance of the CUQI is statistically insignificant. It can be observed that the CUQI metric performance is statistically superior to VIF, UQI, and PSNR. Further, the performance of the CUQI is statistically equal to VSNR and NQM. The SSIM and NQM metrics' performance is statistically superior to the CUQI.

The results of our tests show that the proposed approach of considering the motion and edge information for quality evaluation of cardiac ultrasound sequences is an efficient method of quality evaluation. The CUQI metric also shows slightly better performance than popular VQMs such as PSNR, SSIM, VQM, VIF, and others.

Cardiac ultrasound sequences are significantly characterized by rapid, repetitive motions and distinct structural information. Most state-of-the-art VQMs give little to no consideration to the apparent motion of the video sequences for quality evaluation. Therefore, in the CUQI metric, we compute the change in motion and edge integrity of the cardiac ultrasound sequences to perform quality evaluation. This approach proves to better represent the diagnostic quality of the cardiac sequences than other state-of-the-art metrics.

6 Discussion and Conclusion

A VQM for cardiac ultrasound videos is proposed. In cardiac videos, motion and edge information are two significant features. Distortions to the motion and edge information may significantly affect the perception of the video, which in turn affects the diagnostic quality of the video. Therefore, a VQM that can primarily measure the motion and edge distortions in cardiac videos would provide a suitable approach for quality assessment of cardiac videos. In the literature, to the best of our knowledge, there has been very little work that has focused on an approach that considers combined motion and edge distortions measurement for video quality assessment. Further, we were not able to find any VQM that is specifically used for quality assessment of cardiac videos. To address the lack of specific VQMs for cardiac ultrasound videos, the CUQI metric is proposed. The proposed CUQI metric can be considered a diagnostic-quality oriented metric since it evaluates the quality based on the image features, which has significant potential to affect video perception for diagnosis. To validate our claim, the results of our tests show that the motion and edge information in cardiac ultrasound videos can be effectively utilized for a reliable objective evaluation of diagnostic quality. The correlation with expert opinion scores and the statistical significance test validate the performance of the CUQI metric.

Further, the proposed metric can have significant applications clinically. The relevant medical staff such as cardiologists and radiologists could use the metric to ensure that the cardiac video does not have significant impairments that might have affected the diagnostic information in the video. Also, in applications such as remote consultations or other telemedicinal applications where the videos are compressed and transmitted, the metric can be applied at the receiver end to verify the video integrity. In several countries, hospitals are legally required to archive medical videos in their servers for a specific period of time for which compression is often used to reduce the storage costs. The faculty responsible for archiving the videos could use the CUQI metric to verify that the compression process does not introduce unacceptable impairments. The metric can also be used to identify acceptable compression ratios for cardiac videos based on its objective quality assessment. To be able to identify an acceptable compression ratio is important because acceptable compression ratios may differ for each video. A particular

compression ratio may have a different impact on video quality for two different videos. Therefore, the CUQI could be used to adapt the compression ratio specifically for each cardiac video depending on its impact on the quality. Finally, the CUQI metric can help automate the process of quality validation of cardiac videos and reduce the need to subjectively validate the video quality, which is often a time-consuming and operationally expensive process.

The presented study could be further strengthened by expanding the subjective study to include more relevant expert subjects. A limited number (4 to 5) of expert observers for medical subjective evaluation is an accepted norm since recruiting expert observers is a difficult and expensive process. Additionally, in our studies, three cardiac videos were considered for quality evaluation. The study can be further expanded by considering more cardiac ultrasound videos at different impairment levels, thereby using a larger dataset for performance evaluation of the CUQI metric. The main aim of the metric evaluation was to assess the quality of compressed cardiac videos. However, the study can be further expanded to evaluate the metric performance in quality assessment of different distortions such as packet losses during the transmission process. The CUQI metric is a "full-reference" metric, implying that the original uncompressed video is required for quality assessment. In scenarios where the original video is unavailable, the CUQI metric cannot be used in the current form. This limitation could be overcome by developing "reduced-reference" or "no-reference" versions of the CUQI metric that would not require the original video for quality assessment. The CUQI metric can also be modified if necessary and tested on other types of medical videos that have significant motion and edge characteristics.

To summarize, our study showed that the approach of using specific characteristics of the medical video can enable design and development of more diagnostic-quality oriented VQMs. The topic of medical video quality evaluation remains a challenge. There is a need for development of content-aware VQMs for various types of medical videos. This can be achieved by a better understanding of medical video features affecting the diagnostic quality and by the participation of medical specialists in the development of quality metrics.

Acknowledgments

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).

References

1. M. Razaak and M. G. Martini, "Medical image and video quality assessment in e-health applications and services," in *2013 IEEE 15th Int. Conf. on e-Health Networking, Applications & Services (Healthcom)*, pp. 6–10 (2013).
2. M. Razaak, M. G. Martini, and K. Savino, "A study on quality assessment for medical ultrasound video compressed via HEVC," *IEEE J. Biomed. Health Inf.* **18**(5), 1552–1559 (2014).
3. Z. L. Lu and G. Sperling, "The functional architecture of human visual motion perception," *Vision Res.* **35**(19), 2697–2722 (1995).
4. E. H. Adelson and J. R. Bergen, "Spatiotemporal energy models for the perception of motion," *J. Opt. Soc. Am. A* **2**(2), 284–299 (1985).
5. Z. L. Lu and G. Sperling, "Three-systems theory of human visual motion perception: review and update," *J. Opt. Soc. Am. A* **18**(9), 2331–2370 (2001).
6. M. A. Georgeson and T. M. Shackleton, "Monocular motion sensing, binocular motion perception," *Vision Res.* **29**(11), 1511–1523 (1989).
7. A. A. Stocker and E. P. Simoncelli, "Noise characteristics and prior expectations in human visual speed perception," *Nat. Neurosci.* **9**(4), 578–585 (2006).
8. S. P. McKee, "A local mechanism for differential velocity detection," *Vision Res.* **21**(4), 491–500 (1981).
9. A. B. Watson and A. J. Ahumada, "Model of human visual-motion sensing," *J. Opt. Soc. Am. A* **2**(2), 322–342 (1985).
10. S. M. Pizer and B. M. ter Haar Romeny, "Fundamental properties of medical image perception," *J Digit Imaging* **4**(1), 1–20 (1991).
11. F. Massanes and J. G. Brankov, "Motion perception in medical imaging," *Proc. SPIE* **7966**, 796610 (2011).
12. D. Marr and E. Hildreth, "Theory of edge detection," *Proc. R. Soc. B: Biol. Sci.* **207**(1167), 187–217 (1980).
13. J. J. Koenderink, "Theory of edge-detection," in *Analysis for Science, Engineering and Beyond*, pp. 35–49, Springer, New York (2012).
14. K. Ghosh, S. Sarkar, and K. Bhaumik, *The Theory of Edge Detection and Low-level Vision in Retrospect*, INTECH Open Access Publisher, Rijeka, Croatia (2007).
15. D. Gupta, R. S. Anand, and B. Tyagi, "Edge preserved enhancement of medical images using adaptive fusion based denoising by shearlet transform and total variation algorithm," *J. Electron. Imaging* **22**(4), 043016 (2013).
16. D. Y. Tsai, E. Matsuyama, and Y. Lee, "Quantitative images quality evaluation of digital medical imaging systems using mutual information," in *4th Int. Conf. on Biomedical Engineering and Informatics (BMEI '11)*, Vol. **3**, pp. 1515–1519, IEEE (2011).
17. P. Wayalun et al., "Quality evaluation for edge detection of chromosome G-band images for segmentation," *Appl. Med. Inf.* **32**(1), 25–32 (2013).
18. R. Sekuler, S. N. J. Watamaniuk, and R. Blake, "Perception of visual motion," in *Stevens' Handbook of Experimental Psychology*, H. Pashler et al., eds., John Wiley Publishers, New York (2004).
19. D. Fleet and Y. Weiss, "Optical flow estimation," in *Handbook of Mathematical Models in Computer Vision*, pp. 237–257, Springer, New York (2006).
20. S. Lee and S. J. Park, "A new image quality assessment method to detect and measure strength of blocking artifacts," *Signal Process. Image Commun.* **27**(1), 31–38 (2012).
21. T. Liu et al., "A novel video quality metric for low bit-rate video considering both coding and packet-loss artifacts," *IEEE J. Sel. Top. Signal Process.* **3**(2), 280–293 (2009).
22. K. Zeng et al., "Characterizing perceptual artifacts in compressed video streams," *Proc. SPIE* **9014**, 90140Q (2014).
23. A. Leontaris, P. C. Cosman, and A. R. Reibman, "Quality evaluation of motion-compensated edge artifacts in compressed video," *IEEE Trans. Image Process.* **16**(4), 943–956 (2007).
24. Y. Yang and N. P. Galatsanos, "Removal of compression artifacts using projections onto convex sets and line process modeling," *IEEE Trans. Image Process.* **6**(10), 1345–1357 (1997).
25. Z. Wang et al., "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.* **13**(4), 600–612 (2004).
26. D. M. Chandler and S. S. Hemami, "VSNR: a wavelet-based visual signal-to-noise ratio for natural images," *IEEE Trans. Image Process.* **16**(9), 2284–2298 (2007).
27. Y. Han et al., "A new image fusion performance metric based on visual information fidelity," *Inf. Fusion* **14**(2), 127–135 (2011).
28. K. Seshadrinathan et al., "Study of subjective and objective quality assessment of video," *IEEE Trans. Image Process.* **19**(6), 1427–1441 (2010).
29. M. H. Pinson and S. Wolf, "A new standardized method for objectively measuring video quality," *IEEE Trans. Broadcast.* **50**(3), 312–322 (2004).
30. K. Seshadrinathan and A. Bovik, "Motion tuned spatio-temporal quality assessment of natural videos," *IEEE Trans. Image Process.* **19**, 335–350 (2010).
31. Z. Wang and Q. Li, "Video quality assessment using a statistical model of human visual speed perception," *J. Opt. Soc. Am. A* **24**(12), B61–B69 (2007).
32. A. B. Watson, J. Hu, and J. F. McGowan, "Digital video quality metric based on human vision," *J. Electron. Imaging* **10**(1), 20–29 (2001).
33. B. Ortiz Jaramillo et al., "A full reference video quality measure based on motion differences and saliency maps evaluation," in *Int. Conf. on*

Computer Vision Theory and Applications (VISAPP '14), Vol. **2**, pp. 714–722, SCITEPRESS Digital Library (2014).
34. M. Caramma, R. Lancini, and M. Marconi, "Subjective quality evaluation of video sequences by using motion information," in *Int. Conf. on Image Processing, (ICIP '99)*, Vol. **2**, pp. 313–316, IEEE (1999).
35. M. Ries, O. Nemethova, and M. Rupp, "Motion based reference-free quality estimation for H.264/AVC video streaming," in *2nd Int. Symp. on Wireless Pervasive Computing (ISWPC '07)*, (2007).
36. O. A. Lotfallah, M. Reisslein, and S. Panchanathan, "A framework for advanced video traces: evaluating visual quality for video transmission over lossy networks," *EURASIP J. Adv. Signal Process.* **2006**, 042083 (2006).
37. B. K. Horn and B. G. Schunck, "Determining optical flow," *Proc. SPIE* **0281**, 319 (1981).
38. G. J. Sullivan et al., "Overview of the high efficiency video coding (HEVC) standard," *IEEE Trans. Circuits Syst. Video Technol.* **22**(12), 1649–1668 (2012).
39. ITU-R BT, Recommendation, "500-11, Methodology for the subjective assessment of the quality of television pictures," Technical Report, International Telecommunication Union (2002).
40. H. Yu, Z. Lin, and F. Pan, "Applications and improvement of H.264 in medical video compression," *IEEE Trans. Circuits Syst. Regul. Pap.* **52**(12), 2707–2716 (2005).
41. N. Nouri et al., "Subjective MPEG2 compressed video quality assessment: application to tele-surgery," in *Proc. IEEE Int. Symp. Biomed. Imag.: From Nano to Macro*, pp. 764–767, Rotterdam, Netherlands (2010).
42. Y. Shima et al., "Qualitative and quantitative assessment of video transmitted by DVTS (digital video transport system) in surgical telemedicine," *J. Telemed. Telecare* **13**(3), 148–153 (2007).
43. M. H. Pinson and S. Wolf, "Comparing subjective video quality testing methodologies," *Proc. SPIE* **5150**, 573 (2003).
44. MSU, "MSU perceptual video quality tool," http://www.compression.ru/video/quality_measure/perceptual_video_quality_tool_en.html (2013).
45. K. Brunnstrom et al., "VQEG validation and ITU standardization of objective perceptual video quality metrics [standards in a nutshell]," *IEEE Signal Process Mag.* **26**, 96–101 (2009).
46. I. T. U. T, Tutorial, *Objective Perceptual Assessment of Video Quality: Full Reference Television*, ITU-T Telecommunication Standardization Bureau, Geneva, Switzerland (2004).

Manzoor Razaak is currently a PhD student at the Wireless Multimedia Networking Research group, Kingston University, London. He received his BE degree in electronics and communication engineering from Sambhram Institute of Technology, Bangalore, India, in 2010, and his MSc degree in embedded systems from Kingston University, London. His research interests include medical image and video processing and compression, medical video quality assessment, and video transmission over wireless networks.

Maria G. Martini is a professor in the Faculty of Science, Engineering and Computing in Kingston University, London, where she also leads the Wireless Multimedia and Networking (WMN) Research Group. Her research interests include wireless multimedia networks, cross-layer design, joint source and channel coding, 2-D/3-D error resilient video, 2-D/3-D video quality assessment, and medical applications. She is the author of over 100 international scientific articles and book chapters, and the inventor of several patents on wireless video.