Medical Image and Video Quality Assessment in e-Health Applications and Services

Manzoor Razaak and Maria G. Martini
WMN Research Group, Faculty of SEC, Kingston University London, UK
{M.Razaak, M.Martini}@kingston.ac.uk

***Abstract*— Healthcare services are increasingly using medical images and video in their applications. Quality assessment can be performed via image quality evaluation metrics. However, due to the specific nature of the associated data, quality evaluation of medical images poses several issues. In this paper, we provide an insight into medical image quality evaluation, by discussing the issues faced and the current trends in the literature. Based on the analysis of the medical and technical literature, three broad categories of quality evaluation metrics are presented. The different methodologies are compared in different environments and services and recommendations for future research are provided.**

Keywords— medical image quality evaluation; objective metrics; subjective metrics; service science; diagnostic accuracy; quality of experience.

I. INTRODUCTION

Healthcare services are one of the most complex systems of the service sector. Today, the services provided by healthcare encompass various disciplines like medicine, engineering, information technology, finance, management, etc. The emergence of advanced communication technologies is currently addressing various issues related to healthcare, for instance, the increasing demand for quick and reliable delivery of healthcare services over modern communication channels, long distance medical diagnosis/consultations, medical video conferencing, remote medical education, medical data storage, etc. The wide range of healthcare services provided using the modern communication technologies are now popularly termed as telemedicine as a branch of e-Health. This amalgamation of various disciplines of healthcare and the emergence of e-Health throw several challenges while delivering the services to the users. The application of service science in e-Health can help in addressing the challenges faced by the healthcare service sector.

In any given service sector field, it is essential that the service delivered is of high quality. Usually, the quality of service delivered is measured with various measurement techniques relevant to that particular field. In e-Health, the measurement of the quality of service is a way of ensuring that the correct information is received by the end user after it undergoes different processing systems, like data compression or transmission over the communication channel. The quality of the service delivered can be measured according to the various aspects under consideration. For instance, for transmission of medical information over wireless channels the quality of service from the network point of view is measured using various transmission performance metrics like delay, throughput, spectral efficiency, etc. Similarly, in video streaming services, the quality is measured using various metrics like initial delay, re-buffering probability, duration of re-buffering events, playout time, etc. The evaluation of the service quality can cater to the ultimate goal of service science, i.e., to facilitate better design processes and provide scalability to improve service systems for better business and societal purposes [1].

One of the major applications in the field of e-Health is the transmission of medical images and videos over various communication channels. Medical images and videos are used for various purposes like patient diagnosis, broadcast of surgical operations, exchange of medical information, educational purposes, etc. Due to various communication channel constraints and video processing procedures prior to the transmission, the received image/video often suffers loss of information and hence is subjected to degradation in the received video quality. This can be further inferred as degradation in the quality of the service delivered. Usually, various methods are employed to improve the received service quality. However, it is important to measure or quantify the quality of the service received. In the context of image processing, several image quality assessment techniques are generally employed. These image quality metrics give an assessment of the quality of the image after any kind of processing. In e-Health applications, medical image quality assessment is used for quality assessment of the image/video received which in turn is a reflection of the quality of the service provided.

In this paper, we look into the various aspects of medical image quality evaluation and provide a study about the current trends and challenges faced in this particular context. The study involved reviewing recent papers (not more than 5 years old) on medical image quality evaluation. Based on the papers reviewed, a broad classification of medical image quality metrics is presented in this paper. Section II provides an introduction to the medical image quality evaluation topic. Section III discusses about objective quality metrics used in the medical image quality evaluation, followed by Section IV which discusses subjective quality metrics. Section V gives an insight into the hybrid approach used in the context of medical image quality evaluation, and Section VI concludes the paper.

II. QUALITY EVALUATION FOR MEDICAL IMAGES AND VIDEO

The advancements in image and video acquisition technology have led to an increased usage of medical images and videos for the purpose of diagnosis. The field of telemedicine has benefited from the use of modern communication technologies which have enabled reliable transmission of medical videos / data over different types of communication channels. However, the large memory requirements of medical images and videos demand a large storage space and high bandwidth resources for transmission over various communication channels. Image/video compression is one of the solutions used to address the challenges of high storage and transmission bandwidth requirements of medical videos.

In simple terms, image/video compression is achieved by the removal of redundant data, which often implies loss of information. Ideally, loss of information is not a desirable feature in any type of signal processing, i.e., data, speech, image and video processing. However, appropriate removal of redundant data has shown to have no or negligible effect on the perceptual quality of the output signal. Perceptual redundancy is exploited in signal processing to achieve high amounts of data compression. Image quality evaluation is a method used to measure the quality of the image obtained.

For medical image quality evaluation, standard objective & subjective image quality metrics are commonly used. However, these quality metrics are considered inappropriate for medical image quality evaluation, mainly because these metrics have not been successful in measuring the diagnostic quality of the processed image. During the medical image compression process there is a possibility of losing information of diagnostic importance which may result in inaccurate diagnosis. Therefore, in medical image processing it is highly essential that the diagnostic importance of the image is maintained with high quality. Unfortunately, there is no objective image quality evaluation metric which measures the diagnostic quality of the image. This lack of an exclusive quality evaluation metric for medical images calls for higher research attention on medical image quality evaluation [2].

Several works in the literature have focussed on medical image quality evaluation. In this paper, works related to medical image quality evaluation from the literature are broadly classified into three categories as shown in Figure 1: objective, subjective and hybrid methods.

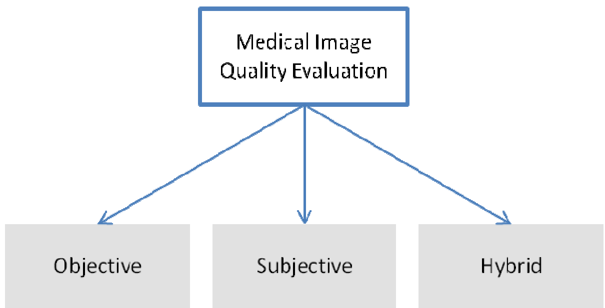

Figure 1. Classification of medical image quality evaluation in this paper

In the following sections, a detailed discussion about the above mentioned categories is given.

III. OBJECTIVE QUALITY MEASUREMENT

In order to evaluate the quality difference between the original and the processed image, several objective image quality metrics are available. These metrics usually are a numerical representation of the quality of the processed images [3] and can be applied on any image, irrespective of its type. Many of these image quality evaluation metrics have been used in the quality evaluation of medical images. Some of the popular state of the art objective image quality metrics for medical quality evaluation used in the literature are listed in Table I.

Objective metrics, in particular MSE, PSNR, and SSIM (defined in Table I) are the most popular metrics used in medical image quality evaluation, for instance to measure the performance of image compression standards or to measure the impact of different image processing techniques, as in [4]-[10].

Apart from these frequently used metrics, several works have explored other image quality evaluation metrics for medical images. The authors in [11] have used 21 different types of objective (statistical) metrics for evaluating the image quality of medical ultrasound compressed images to further propose their tool for evaluation of ultrasound image compression. They categorise the metrics based on the features it measures, which are: distance measures, correlation measures, spectral, contrast and Human Visual System (HVS) measures.

A major criticism of objective quality metrics for medical images is that these metrics do not correlate well with the visual quality of the image. In particular, the objective metric parameters are not a reliable measure of the diagnostic quality of the image [15]. Also, the compression ratios correlate poorly with the quality of image degradation [16]. Therefore, it cannot always be said that a high value objective quality metric ensures a superior diagnostic quality medical image. However, in spite of this shortcoming, objective metrics have been widely used for quality evaluation of medical images [17].

IV. SUBJECTIVE QUALITY MEASUREMENT

In order to overcome the diagnostic non-reliability shortcoming of objective medical image quality evaluation, subjective quality assessment is widely adopted. In subjective assessment, medical field experts give their opinions on the images based on the perceptual quality and the diagnostic information preserved in the processed image. Several methods have been used to perform subjective quality assessment for medical images. Some of them are described below.

A. Mean Opinion Score (MOS)

This is one of the simplest and widely adopted methods for subjective quality assessment. In this method, the medical expert measures the perceptual quality of the image and visually analyses the diagnostic information preserved in the image [18]. Typically, medical experts are presented with a randomized set of images and are asked to rate the quality of the images in a scale of 1 to 5, where 1 is poor and 5 is the highest quality. These results are used to further obtain the average score and other statistical analysis.

B. Multiple Reader Multiple Case

In this method, the medical experts of various skill levels are made to read multiple cases independently and their opinion is collected. The problem with this method is that it may lead to biases during the assessment procedure by the experts due to their inconsistency in multiple assessments. Hence, this method must involve proper design of the assessment tests [19].

TABLE I. OBJECTIVE IMAGE QUALITY METRICS USED IN THE LITERATURE

Objective Image Quality Metric	Defining Equation and Features considered	
Mean Square Error (MSE)	$MSE = \frac{1}{MN}\sum_{y=1}^{M}\sum_{x=1}^{N}[I(x,y) - I'(x,y)]^2$	Pixel difference between the original and the processed image
Peak Signal to Noise Ratio (PSNR)	$PSNR = 20 * \log\left(\frac{255}{\sqrt{MSE}}\right)$	Gives ratio of signal over the noise, where signal refers to the original image and noise refers to the standard error
Structural Similarity Index Metric(SSIM)	$SSIM = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x{}^2 + \mu_y{}^2 + C_1)(\sigma_x{}^2 + \sigma_y{}^2 + C_2)}$	Measures the structural similarity between two images
Moran Peak Ratio (MPR)	$MPR = \frac{N\sum_{y=1}^{mn}\sum_{x=1}^{mn}\delta_{ij}(x_i - \overline{x})(x_j - \overline{x})}{\sum_{y=1}^{mn}(x_i - \overline{x})^2}$	A window based metric which gives a good measure of the spatial correlation
Universal Quality Index [12]	$Q = \frac{4\sigma_{xy}\bar{x}\,\bar{y}}{((\sigma_x{}^2 + \sigma_y{}^2)[(\overline{x})^2 + (\overline{y})^2]}$	Measures the structural distortion and gives good approximation of perceived distortion
Spatial Frequency Measurement	SFM = $\sqrt{R^2 + C^2}$; R = Row pixel difference, C = Column pixel difference	Its calculation considers the characteristics of the image
Visual Information Fidelity (VIF) [14]	The images are decomposed into wavelets and computation is done using several models which gives a measure of the visual quality	
Visual Signal to Noise Ratio (VSNR) [13]	Contrast thresholds are used to identify distortions. The distortions above the threshold are modeled to measure the image quality	

C. *Double Stimulus Continuous Quality Scale (DSCQS)*

The DSCQS uses a just noticeable difference (JND) approach in which the medical expert is presented with two images side by side, typically an original and processed image. The expert is then asked to rank both the images. The ratings obtained are then used to get the mean scores and other desired statistics. This technique is particularly used to measure the quality of the compressed images. If the compressed image is visually close to the diagnostic quality of the original image then it would imply that "perceptually lossless" compression has been achieved [20]. Usually, the pictures are presented once or twice to the expert who is asked to rate them on a scale of 1 to 5.

An advantage of using subjective quality assessment for medical images is that a reliable assessment of the processed image quality can be obtained by relevant experts. However, a disadvantage of this approach is the inconsistency of the results obtained by these techniques. Different subjects may give largely different ratings for the same image and hence impact the overall average rating. Even in cases where the subject assesses the image multiple times there might be chances of potential bias i.e. the subject may rate the same image differently the second time [21].

V. HYBRID APPROACH

The merit of using subjective assessment is the reliability on the quality measurement of the expert. However, this is often an inconsistent and lengthy procedure. On the other hand, objective assessment is a quicker and an easier methodology, however it is not entirely reliable. Hence, there have been efforts to develop methodologies which can combine both subjective and objective evaluation techniques, to further improve the medical image evaluation efficiency [22]. This approach is referred as "hybrid approach" in this paper.

One of the first efforts was made as early as in 1994 by Cosman et al. [18], who proposed the concept of using a "gold standard" that would be the reference for the diagnostic truth of the original image based on which the diagnostic accuracy of other processed image can be measured. Their work also introduced the usage of Receiver Operating Characteristics (ROC) analysis to measure the diagnostic accuracies of the digitally processed images. The ROC curve analysis is based on signal detection methodologies in which Gaussian noise is added to signals which are then filtered. The samples obtained from the filtered signal are compared to a threshold based on which the presence of signal is detected. To implement ROC analysis in medical image quality evaluations, the threshold is calculated by obtaining the subjective ratings of the experts on a scale (say, 1-5). This threshold is then used for evaluating the feature detection accuracy in the image. The ROC method is used especially in images on which processes like lesion detections are performed. If the quality of the image is higher, then it implies that the probability of correctly detecting the lesion is higher [23].

TABLE II. WORKS IN THE LITERATURE IMPLEMENTING THE CORRELATION APPROACH

Authors	Evaluation criteria	Outcomes
B. Kumar, et al. [24]	Used PSNR & SSIM for SPIHT algorithm & correlated with MOS	PSNR & SSIM showed 0.86-0.98 correlation
A. Panayides, et al. [25]	SROCC & PLCC correlation approaches were used to correlate various metrics with MOS	Best correlation were obtained for (ROI) PSNR and WSNR
A. Przelaskowski [26]	A hybrid vector measure uses Hosaka plots, MSE and PSNR kind of metrics, Picture Quality Scale (PQS) distortion measure and correlation to subjective scores	Their proposed measure shows a correlation of 0.9
J. F. Pambrun, et al. [28]	JPEG 2000 compressed images are assessed using VSNR & subjective metrics and correlated	VSNR shows high correlation with subjective assessment
H. R. Sheikh, et al. [29]	JPEG & JPEG 2000 compressed images were evaluated using various objective metrics	PSNR shows least correlation to MOS scores, whereas VIF metric gives high correlation
K. J. Kim, et al. [30]	Correlation was done between various objective metrics with subjective metrics	PSNR shows worst performance while HDR-VDP metric gives best correlation
J. Oh, et al. [31]	Decompose image to different spatial frequency bands an use Perceptual Quality Rating (PQR) metric for correlation	PQR metric correlates well with the radiologists assessment

The Correlation Approach

A correlation based medical image quality evaluation approach is increasingly used in the literature. This approach involves obtaining statistical correlations between the objective and subjective quality evaluations. The objective metric which shows high correlation with the subjective metric is thus considered to give a higher diagnostic quality oriented evaluation. The outcome of this approach is that it helps in choosing and designing an appropriate objective quality metric for medical images. In the next paragraph one of the works based on the correlation approach is discussed and a few related works are presented in Table II.

The authors in [27] use their proposed wavelet weighted bold vessel distortion measure (WBVDM) metric to analyse the medical image quality. The proposed metric was compared with other objective metrics like PSNR, MSE, SSIM, VIF, etc. Their methodology distorts the image using different wavelet distortion methods like JPEG, SPIHT, Gaussian blur, etc. The images obtained from these methodologies were evaluated using both subjective and objective quality evaluations. In order to obtain a correlation between the subjective and objective evaluations, the authors perform a non linear regression function on the objective measures to transform them into a set of predicted MOS values. This predicted MOS values are then compared with the original MOS values by finding a correlation using various correlation methods. The results showed that their proposed WBVDM method gives a measure closer to the subjective rating. Additionally, their results also indicated that metrics like VIF, VSNR shows improved performance with respect to the traditional MSE, PSNR metrics. Several other works have used similar methodologies for medical image quality evaluation and the results have indicated that non traditional objective metrics show a better correlation with subjective measures than the traditional ones like PSNR, MSE, etc.

A benefit of this “hybrid” approach is that it also considers the characteristics of HVS. This is beneficial because the resulting analysis will be based on the diagnostic importance of the image which is essential in medical image applications. HVS models consider various factors like contrast sensitivity, luminance, masking effect, and frequency decomposition. The metrics usually correlate well with human visual performance and hence enhance the quality evaluation efficiency [32]. Also, an interesting observation from the reviewed works is that the traditional metrics like PSNR, SSIM do not perform well in the context of medical images [33]. Other less known metrics like VIF, VSNR give a more appropriate evaluation.

VI. CONCLUSION

In this paper, an overview of quality evaluation of medical images is given. Several objective and subjective image quality assessment metrics exist, however there are no specific quality metrics valid in general for medical images/videos. The existing generic image quality metrics have been used in the literature for the quality evaluation of medical images. Due to the sensitive nature of the information in medical images it is essential that the image quality metrics primarily consider the diagnostic quality of the image instead of other statistical criterion for evaluation.

The “hybrid” approach which links the objective metrics with subjective assessment has proven to give a better diagnostic-oriented quality evaluation. However, the research in this approach is limited, and hence needs more attention. The increased adoption of medical videos in healthcare is facilitated by the rapid advancements in the field of e-Health and the evolution of healthcare as a service industry. This enhances the importance of the need for medical video quality evaluation metrics, not just for image quality evaluation, but also for the quality evaluation of the service provided as a whole. The current research could focus more on evaluating the existing and upcoming image / video compression standards like JPEG 2000, H.264, HEVC, etc., using various quality metrics with a goal of achieving diagnostic-quality oriented assessment. An appropriate quality metric would enable not only performance evaluation, but also the design of future medical multimedia services and applications [34] in line the service goal (diagnosis, treatment, etc.) and the relevant quality requirements.

REFERENCES

[1] P. Maglio, and J. Spohrer, "Fundamentals of service science," *Journal Of The Academy Of Marketing Science,* vol. 36, no. 1, pp. 18-20, 2008.

[2] D. Ivetic, and D. Dragan, "Quality Evaluation of Medical Image Compression: What to Measure?" in *IEEE 8th International Symposium on Intelligent Systems and Informatics*, Serbia, pp. 37-42, 2010.

[3] A. N. Netravali, and B. G. Haskell, "Digital pictures: representation, compression and standards", *New York: Plenum*, 2000.

[4] S. Udomhunsakul, and K. Hamamoto, "Wavelet filters comparison for ultrasonic image compression", in *IEEE Region 10 Conference TENCON*, 2004.

[5] S. Arpah, B. Ahmad, M. N. Taib, N. E. A. Khalid, and H. Taib, "Analysis of image quality based on dentists' perception cognitive analysis and statistical measurements of intra-oral dental radiograph", *in International Conference on Biomedical Engineering*, 2012, pp. 379-384.

[6] R. S. H. Istepanian, N. Philip, M. G. Martini, N. Amso, and P. Shorvon, "Subjective and objective quality assessment in wireless teleultrasonography imaging", *in 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2008.

[7] K. Vidhya and S. Shenbagadevi, "Performance Analysis of Medical Image Compression", *in International Conference on Signal Processing Systems*, 2009.

[8] S. E. Ghrare, M. A. M. Ali, M. Ismail, and K. Jumari, "Diagnostic Quality of Compressed Medical Images: Objective and Subjective Evaluation", *in Second Asia International Conference on Modeling & Simulation*, 2008.

[9] S. Finn, M. Glavin, and E. Jones, "Echocardiographic speckle reduction comparison", *IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control*, vol. 58, no. 1, pp. 82-101, 2011.

[10] I. Nakajima, K. Nakada, Y. Tomioka, H. Juzoji, and T. Kitano, "Methods for performing video quality evaluations in the field of medicine", *in IEEE Region 10 Conference TENCON*, 2010.

[11] C. Delgorge, C. Rosenberger, G. Poisson, and P. Vieyres, "Towards a New Tool for the Evaluation of the Quality of Ultrasound Compressed Images", *in IEEE Transactions on Medical Imaging*, 2006.

[12] P. Loizou, C. Pattichis, M. Pantziaris, T. Tyllis, and A.Nicolaides, "Quality evaluation of ultrasound imaging in the carotid artery based on normalization and speckle reduction filtering", *Medical and Biological Engineering and Computing,* vol. 44, no. 5, pp. 414-426, 2006.

[13] D. M. Chandler, and S. S. Hemami. "VSNR: A wavelet-based visual signal-to-noise ratio for natural images." *IEEE Transactions on Image Processing,* vol.16, no. 9 2007.

[14] Y. Han, Y. Cai, Y. Cao, and X. Xu, "A new image fusion performance metric based on visual information fidelity", *Information Fusion*, Vol. 14, no. 2, pp. 127-135, April 2013.

[15] B. J. Erickson, "Irreversible compression of medical images", *Journal of Digital Imaging,* vol. 15, no. 1, pp. 5-14, 2002.

[16] A. Fidler, and B. Likar, "What Is Wrong with Compression Ratio in Lossy Image Compression?", *Radiology*, vol. 245, no. 1, pp. 299, 2007.

[17] S. Sudha, R. Sukanesh, and G. R. Suresh, "Performance evaluation of shape adaptive discrete wavelet transform based magnetic resonance images coding", *in Proceedings of International Conference on Future Computer and Communication*, 2009.

[18] P. C. Cosman, R. M. Gray, and R. A. Olshen, "Evaluating Quality of Compressed Medical Images: SNR, Subjective Rating and Diagnostic Accuracy", *Proceedings of the IEEE* , vol. 82, no. 6, pp. 919-932, 1994.

[19] L. L. Pesce, C. E. Metz, and K. Doi, "Experimental design and data analysis in receiver operating characteristic studies: Lessons learned from reports in Radiology from 1997 to 2006", *Radiology*, vol. 253, no. 3, pp. 822-830, 2009.

[20] B. Kim, K. H. Lee, K. J. Kim, R. Mantiuk, H. R. Kim, and Y. H. Kim, "Artifacts in slab average-intensity-projection images reformatted from JPEG 2000 compressed thin-section abdominal CT Data sets", *American Journal of Roentgenology*, vol. 190, no. 6, pp. 342-350, 2008.

[21] P. Duraisamy, X. Yuan, A. E. Saba, and S. Palanisamy, "Contrast enhancement and assessment of OCT images", *in International Conference on Informatics, Electronics & Vision*, pp. 91-95, 2012.

[22] C. Rosenberger, G. Poisson, P. Vieyres, and C. Delgorge, "Evaluation of the quality of ultrasound image compression for a robotic teleechographic system", *in Proceedings of 32rd Annual Inernational Conference on Computing in Cardiology*, 2005.

[23] L. Zhang, C. Cavaro-Ménard, and P. L. Callet, "Key issues and specificities for the objective medical image quality assessment*", in Proceeding of Sixth International Workshop on Video Processing and Quality Metrics for Consumer Electronics*, 2012.

[24] B. Kumar, S. P. Singh, A. Mohan, and H. V. Singh, "MOS Prediction of SPIHT Medical Images Using Objective Quality Parameters", *in International Conference on Signal Processing Systems*, 2009.

[25] A. Panayides, M. S. Pattichis, C. S. Pattichis, C. C. Loizou, M. Pantziaris, and A. Pitsillides, "Atherosclerotic Plaque Ultrasound Video Encoding, Wireless Transmission, and Quality Assessment Using H.264", *in IEEE Transactions on Information Technology in Biomedicine*, vol. 15, no. 3, pp. 387-397, 2009.

[26] A. Przelaskowski, "Vector quality measure of lossy compressed medical images" *Computers in biology and medicine*, vol. 34, no. 3, pp. 193-207, 2004.

[27] S. R. Nirmala, S. Dandapat, and P. K. Bora, "Performance Evaluation of Distortion Measures for Retinal Images", *International Journal of Computer Applications*, vol. 17, no. 6, pp. 17-23, 2011.

[28] J. F. Pambrun and R. Noumeir, "Perceptual quantitative quality assessment of JPEG2000 compressed ct images with various slice thicknesses", *in IEEE International Conference on Multimedia and Expo (ICME)*, 2011.

[29] H. R. Sheikh, M. F. Sabir, and A. C. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms", *in IEEE Transactions on Image Processing*, 2006.

[30] K. J. Kim, K. H. Lee, H. S. Kang, S. Y. Kim, Y. H. Kim, B. Kim, J.Seo, and R. Mantuik, "Objective index of image fidelity for JPEG2000 compressed body CT Images", *Medical Physics*, vol. 36, no. 7, 2009.

[31] J. Oh, S. I. Woolley, T. N. Arvanitis, and J. N. Townend, "A multistage perceptual quality assessment for compressed digital angiogram images*", IEEE Transactions on Medical Imaging*, vol. 20, no. 12, pp. 1352-1361, 2001.

[32] C. Cavaro-Menard, L. Zhang, and P. L. Callet, "Diagnostic quality assessment of medical images: Challenges and trends*", in 2nd European Workshop on Visual Information Processing (EUVIP)*, 2010.

[33] K. J. Kim, R. Mantiuk, K. H. Lee, and W. Heidrich, "Calibration of the visual difference predictor for estimating visibility of JPEG2000 compression artifacts in CT images", *Journal Of Optics Nouvelle Revue D'Optique*, 2010.

[34] M. G. Martini and M. Mazzotti, "Quality driven wireless video transmission for medical applications", *28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, (EMBS'06),* New York, August 2006.