A full-reference laparoscopic video quality assessment algorithm

Hrishikesh Hemant Borate[a], Peter A. Kara[b], Balasubramanyam Appina[a], and Aniko Simon[c]

[a]Indian Institute of Information Technology, Design and Manufacturing Kancheepuram, Chennai, India
[b]Budapest University of Technology and Economics, Budapest, Hungary
[c]Sigma Technology, Budapest, Hungary

ABSTRACT

In this paper, we introduce a full-reference quality assessment model for laparoscopic videos. The perceived quality of medical imaging in general is of upmost importance, as visual degradations may lead to severe negative impacts on diagnostic accuracy. Laparoscopy, also known as keyhole surgery, is a camera-assisted operation performed in the abdomen or the pelvis of the patient. Unlike the conventional utilization of telescopic rod lens systems, digital laparoscopy uses a miniature digital camera at the end of the laparoscope, and therefore, the surgeon fully relies on the quality of the medical video. In our scientific contribution, we utilize different image quality measures for each frame of the laparoscopic videos. We implement a regression neural network architecture on the frame-level features with the associated mean opinion score as labels. Finally, we calculate the average of the predicted frame-level scores to compute the overall quality score. The performance of the proposed model is evaluated on the well-known LVQ laparoscopic video dataset. The evaluation results confirm that our model is competitive with the state-of-the-art 2D full-reference and no-reference supervised algorithms. Furthermore, the model demonstrates robust performance across all distortion types of the dataset.

Keywords: Video quality assessment, laparoscopic video, full-reference algorithm

1. INTRODUCTION

Laparoscopy is a minimally-invasive surgical procedure, which belongs to the medical category of endoscopy. Surgeons insert laparoscopic tools through small incisions of the patient's body and perceive the real-time functioning of human abdominal organs on a video monitor. The extracted information guides the surgeon to perform the operation on malfunctioning organs. This low-risk method is less painful for the patient and recovery is rather fast compared to traditional surgery.

The laparoscopic equipment has a camera, which enters the patient's body through one of the incisions created by the doctor. This camera displays all the surgical activities performed inside the patient's body on a monitor and this information is also recorded for further analysis. The captured video undergoes several processing and acquisition stages. Each processing stage may produce spatial and temporal artefacts, such as smoke, fog, blur, motion blur, ghosting and frame freeze distortions. Additionally, these degradations may also be the result of specific technical issues. Such artefacts effectively reduce the perceptual quality of the video, which creates a visual discomfort to the surgeon and it reflects in the degradation of surgical efficiency. At the time of this paper, surgeons utilize the services of technical staff to overcome this problem. However, this method is neither robust nor reliable due to manual errors, and manual intervention requires additional time. These unfortunate technological circumstances resulted the necessity to analyze and assess the quality of laparoscopic videos in order to enhance the efficiency of surgical performance.

Further author information: (Send correspondence to Balasubramanyam Appina)
Hrishikesh Hemant Borate: E-mail: evd16i006@iiitdm.ac.in
Peter A. Kara: E-mail: kara@hit.bme.hu
Balasubramanyam Appina: E-mail: appina@iiitdm.ac.in
Aniko Simon: E-mail: aniko.simon@sigmatechnology.se

Quality assessment (QA) is a standard approach to assess the perceptual loss at each processing stage. QA models are typically classified into three categories, based on the involvement of reference-quality content. Full-reference (FR) algorithms require the entire reference-quality content to perform the QA task. Reduced-reference (RR) algorithms utilize a small number of the features of the reference-quality content to compute the quality. No-reference (NR) models do not require any kind of information about the degradation-free content to perform the quality evaluation. In this paper, we propose an FR quality assessment model for laparoscopic videos, based on performing the regression neural network on frame-level quality measures of videos.

The remainder of the paper is structured as follows. Section 2 introduces the algorithm in detail. Section 3 analyzes and discusses the performance of the proposed model. The paper is concluded in Section 4.

2. PROPOSED ALGORITHM

The proposed algorithm consists of two stages. In the first stage, we compute the frame-level quality measures of laparoscopic videos. The second stage performs the regression network training on the frame-level quality measures and provides the prediction of the overall quality.

2.1 Frame-level quality measures

Avcıbas and Sankur[1] performed statistical analysis on image quality assessment (IQA) models to classify them into multiple categories, such as pixel-difference-based measurements, correlation-based measurements, edge-based measurements, spectral distance measurements, context measurements and measurements based on the human visual system (HVS). This classification covers a wider spectrum of the off-the-shelf IQA models. These algorithms collect diverse features of images to estimate the quality. We are motivated by the robust performance of these IQA models and we utilize these computations to measure the frame-level quality of laparoscopic videos.

2.1.1 Pixel-difference-based IQA measurements

Mean Square Error (MSE), Mean Absolute Error (MAE), Peak Signal-to-Noise Ratio (PSNR) and Modified Infinity Norm (MIN) are well-known metrics that measure image quality at pixel level. We perform MSE, MAE, PSNR and MIN computations between the reference-quality and the distorted frames of laparoscopic videos to measure the pixel-difference-based frame-level quality.

$$Q_1 = \frac{1}{K}\sum_{k=1}^{K}\frac{1}{NM}\sum_{i=0}^{N-1}\sum_{j=0}^{M-1}(A_k(i,j) - \hat{A}_k(i,j))^2, \tag{1}$$

$$Q_2 = \frac{1}{K}\sum_{k=1}^{K}\frac{1}{NM}\sum_{i=0}^{N-1}\sum_{j=0}^{M-1}|A_k(i,j) - \hat{A}_k(i,j)|, \tag{2}$$

$$Q_3 = \frac{1}{K}\sum_{k=1}^{K}(20\log_{10}(max(A_k)) - 10\log_{10}(Q_1)), \tag{3}$$

$$Q_4 = \sqrt{\frac{1}{K}\sum_{k=1}^{K}\frac{1}{R}\sum_{r=1}^{R}|A_k(i,j) - \hat{A}_k(i,j)|}, \tag{4}$$

where N and M represent frame dimensions; i and j are spatial indexes; K represents the chrominance channel ($K = 3$); max represents the maximum pixel value; R indicates the pixel count threshold; r varies between 0 and R; A and $\hat{A}$ indicate reference-quality and distorted laparoscopic video frames, respectively; and Q_1, Q_2, Q_3 and Q_4 are frame-level MSE, MAE, PSNR and MIN measurements, respectively.

2.1.2 Correlation-distance-based IQA measurements

Correlation-distance-based algorithms quantify the similarity between the reference and the distorted image. We compute the structural content score, the mean of the angle difference, the mean of the combined angle magnitude difference, the normalized cross correlation measurement and the Czenakowski distance between reference and distorted laparoscopic video frames to represent the frame-level similarity measurements.

$$Q_5 = \frac{1}{K}\sum_{k=1}^{K}\frac{\sum_{i=0}^{N-1}\sum_{j=0}^{M-1}A_k(i,j)^2}{\sum_{i=0}^{N-1}\sum_{j=0}^{M-1}\hat{A}_k(i,j)^2}, \tag{5}$$

$$Q_6 = 1 - \frac{1}{NM}\sum_{i=0}^{N-1}\sum_{j=0}^{M-1}\left(\frac{2}{\pi}\cos^{-1}\frac{\langle \mathbf{A}(i,j), \hat{\mathbf{A}}(i,j)\rangle}{\|\mathbf{A}(i,j)\|\|\hat{\mathbf{A}}(i,j)\|}\right), \tag{6}$$

$$Q_7 = \frac{1}{NM}\sum_{i=0}^{N-1}\sum_{j=0}^{M-1} 1 - \left[1 - \frac{2}{\pi}\cos^{-1}\frac{\langle \mathbf{A}(i,j), \hat{\mathbf{A}}(i,j)\rangle}{\|\mathbf{A}(i,j)\|\|\hat{\mathbf{A}}(i,j)\|}\right] \times \left[1 - \frac{\|\mathbf{A}(i,j) - \hat{\mathbf{A}}(i,j)\|}{\sqrt{3 \times 255^2}}\right], \tag{7}$$

$$Q_8 = \frac{1}{K}\sum_{k=1}^{K}\frac{\sum_{i=0}^{N-1}\sum_{j=0}^{M-1}A_k(i,j) \times \hat{A}_k(i,j)}{\sum_{i=0}^{N-1}\sum_{j=0}^{M-1}A_k(i,j)^2}, \tag{8}$$

$$Q_9 = \frac{1}{NM}\sum_{i=0}^{N-1}\sum_{j=0}^{M-1}\left(1 - \frac{2\sum_{k=1}^{K}\min\left[A_k(i,j), \hat{A}_k(i,j)\right]}{\sum_{k=1}^{K}\left[A_k(i,j) + \hat{A}_k(i,j)\right]}\right), \tag{9}$$

where Q_5, Q_6, Q_7, Q_8 and Q_9 represent the structural content score, the mean of the angle difference, the mean of the combined angle magnitude difference, the normalized cross correlation measurement and the Czenakowski distance of the distorted laparoscopic video frame, respectively.

2.1.3 Spectral-distance-based IQA measurements

Frequency-based distance measures are useful to understand the deviation of the geometric properties between reference-quality and distorted images. We use spectral phase distortion strength and weighted spectral magnitude and phase distortion strength calculations as frequency-based distortion measures. These measures are computed from the magnitude and phase responses of the discrete Fourier transform (DFT).

$$Q_{10} = \frac{1}{NM}\sum_{u=0}^{N-1}\sum_{v=0}^{M-1}|\varphi(u,v) - \hat{\varphi}(u,v)|^2, \tag{10}$$

$$Q_{11} = \frac{1}{NM}\left(\lambda\sum_{u=0}^{N-1}\sum_{v=0}^{M-1}|\varphi(u,v) - \hat{\varphi}(u,v)|^2 + (1-\lambda) \times \sum_{u=0}^{N-1}\sum_{v=0}^{M-1}|M(u,v) - \hat{M}(u,v)|^2\right), \tag{11}$$

where u and v indicate frequency coordinates; λ is a weighting factor and it is equal to 2.5×10^{-5}; φ and $\hat{\varphi}$ are phase responses estimated from the DFT of the reference and the distorted video frames, respectively; M and $\hat{M}$ are magnitude responses estimated from the DFT of the reference and the distorted frames, respectively; Q_{10} and Q_{11} are the spectral phase distortion strength and the weighted spectral magnitude and phase distortion strength of a video frame, respectively.

2.1.4 Context-based IQA measurements

This measurement aims to estimate the amount of information loss of a distorted image. We compute the entropy score of distorted frames to measure the loss of information of a laparoscopic video.

$$Q_{12} = -\sum_{x} p_x \times \log_2 p_x, \tag{12}$$

where x represents the number of bins and p denotes the normalized histogram counts of a frame.

2.1.5 HVS-based IQA measures

The behavioural response of the HVS can be modeled as a bandpass filter. The transfer function (in polar coordinates) of a bandpass filter is defined as

$$H(\rho) = \begin{cases} 0.05e^{\rho^{0.554}}, & \rho < 7, \\ e^{-9[|\log_{10}\rho - \log_{10} 9|]^{2.3}}, & \rho >= 7, \end{cases} \tag{13}$$

where $\rho = (u^2 + v^2)^{1/2}$.

We perform a preprocessing step by computing the aforementioned bandpass filter response of a frame. Then, we calculate the inverse discrete cosine transform (IDCT) of the DCT with spectral masked bandpass responses of the reference and distorted frames to measure the error in the bandpass response characteristics.

$$Q_{13} = \frac{1}{K}\sum_{k=1}^{K} \frac{\sum_{i=0}^{N-1}\sum_{j=0}^{M-1} \left| U\left\{A_k(i,j)\right\} - U\left\{\hat{A}_k(i,j)\right\}\right|}{\sum_{i=0}^{N-1}\sum_{j=0}^{M-1} \left| U\left\{A_k(i,j)\right\}\right|}, \tag{14}$$

$$Q_{14} = \frac{1}{K}\sum_{k=1}^{K} \left[\frac{1}{NM}\sum_{i=0}^{N-1}\sum_{j=0}^{M-1} \left| U\left\{A_k(i,j)\right\} - U\left\{\hat{A}_k(i,j)\right\}\right|^2\right]^{1/2}, \tag{15}$$

where U represents the IDCT of the DCT with the spectral masked bandpass response of a frame, and Q_{13} and Q_{14} represent error measurements, which are estimated by using HVS properties.

2.2 Supervised learning using regression neural network

We compute the aforementioned quality measurements of each frame to form the feature vector of a video frame.

$$Q_z = [Q_{1_z}, Q_{2_z}, Q_{3_z}, \ldots, Q_{14_z}], \tag{16}$$

where z represents the frame sequence of a video.

An earlier work of Appina *et al.*[2] highlighted that the Mean Opinion Score (MOS) of a video highly correlates with the frame-level perceptual opinion score during short temporal durations. We are motivated by this finding to use video-level MOS score as the ground truth quality representative of a frame. Therefore, we perform the regression of the frame-level quality measurements and the video-level MOS scores (P) as its label. The feature vector of a video V is

$$Q_z^V = [Q_{1_z}^V, Q_{2_z}^V, Q_{3_z}^V, \ldots, Q_{14_z}^V], \tag{17}$$

with the corresponding label P^V. The video-level feature vector and the associated labels are used to train the regression neural network to perform supervised learning.

We use a 6-layered hidden neural network architecture to perform the regression on the given input features. The dimension of hidden layers are 1024, 512, 256, 128, 64 and 32, and these are connected in the feed-forward model. We use the ReLU activation function and the initial learning rate is fixed to 0.0001. The network is trained for 450 epochs with the gradient descent along with the Adam optimization algorithm. The MAE computation is used as a cost function to optimize the network weights and biases. Finally, the regression network performs the quality prediction of each frame of a video. We compute the average of frame-level quality prediction scores to estimate the overall quality score of a laparoscopic video.

3. RESULTS AND DISCUSSION

The efficacy of the proposed algorithm is evaluated on the publicly available Laparoscopic video quality (LVQ) dataset.[8] The LVQ dataset consists of 10 reference videos with a resolution of 512 × 288. Each video sequence is 10 seconds long and the frame rate is 25 fps. The videos are available in `.avi` container. There is a total of 200 test stimuli, and the distorted video sequences are affected by a combination of motion blur, defocus blur, smoke,

Table 1: The efficacy of the proposed algorithm in terms of LCC measurements on the LVQ dataset (with expert subjective scores).

	Noise	Defocus Blur	Motion Blur	Uneven Illumination	Smoke	Overall
PSNR	0.9939	0.8146	0.8226	0.9452	0.9777	0.6853
SSIM[3]	0.9706	0.7358	0.8827	0.9847	0.9116	0.5732
VIF[4]	0.9896	0.9806	0.9708	0.9878	0.9808	0.5909
BRISQUE[5]	0.9761	0.9623	0.4208	0.2973	0.4009	0.4434
NIQE[6]	0.9741	0.9883	0.7836	0.6655	0.4301	0.4407
VIIDEO[7]	0.8658	0.3498	0.5136	0.4035	0.4195	0.3744
Proposed algorithm	0.9987	0.9396	0.9673	0.9743	0.9854	0.9454

Table 2: The efficacy of the proposed algorithm in terms of SROCC measurements on the LVQ dataset (with expert subjective scores).

	Noise	Defocus Blur	Motion Blur	Uneven Illumination	Smoke	Overall
PSNR	0.9579	0.7836	0.7977	0.9530	0.9478	0.6914
SSIM[3]	0.9435	0.7320	0.8802	0.9580	0.8817	0.5653
VIF[4]	0.9592	0.9555	0.9376	0.9534	0.9459	0.5642
BRISQUE[5]	0.9527	0.9355	0.3994	0.2634	0.4355	0.3842
NIQE[6]	0.9594	0.9443	0.7028	0.5605	0.3382	0.3674
VIIDEO[7]	0.8822	0.3023	0.3915	0.4281	0.4416	0.3334
Proposed algorithm	0.9701	0.9701	0.9461	0.9286	0.9333	0.9309

noise and uneven illumination. The subjective assessment study involved experts and non-expert subjects, and the authors published the MOS scores of both groups as quality representatives of the dataset; the data of the two test subject groups are available separately.

We partitioned the dataset in a 80 : 20 proportion. This means that 80% of the videos is used to train the neural network and the remaining 20% is used to perform the regression analysis. It is important to emphasize that the training and the testing sets do not overlap. The network is trained with the frame-level features and estimates the regression score at frame level. Finally, we compute the average of the frame-level regression scores to estimate the overall quality score of a laparoscopic video.

We compute the Linear Correlation Coefficient (LCC), the Spearmann Rank Order Correlation Coefficient (SROCC), and the Root Mean Square Error (RMSE) statistics to indicate the performance of the proposed algorithm. The LCC represents the linear relationship between predicted scores and the MOS, the SROCC measures the monotonic relationship between the components and the RMSE measures the error between the estimates and the MOS scores. These statistics are reported after performing a 4-parameter non-linear logistic fit.[9]

Tables 1 and 2 show the efficacy of the proposed algorithm on the LVQ dataset with expert MOS scores. Tables 3 and 4 show the efficacy of the proposed algorithm on the LVQ dataset with non-expert MOS scores. From these tables, it is clear that the proposed algorithm correlates well with both expert and non-expert MOS scores. Also, we compare the proposed algorithm's performance with off-the-shelf FR and NR quality assessment models. PSNR, SSIM[3] and VIF[4] are 2D FR IQA models. BRISQUE[5] and NIQE[6] are 2D NR IQA models. The IQA algorithms are applied on each frame and the average score is computed to estimate the final quality score of the video. VIIDEO[7] is a video quality assessment model. From the results, it is clear that the proposed algorithm shows robust and state-of-the-art performance numbers on the different MOS scores.

Table 3: The efficacy of the proposed algorithm in terms of LCC measurements on the LVQ dataset (with non-expert subjective scores).

	Noise	Defocus Blur	Motion Blur	Uneven Illumination	Smoke	Overall
PSNR	0.9968	0.8166	0.8199	0.9561	0.9811	0.6054
SSIM[3]	0.9690	0.7388	0.8861	0.9926	0.9165	0.6123
VIF[4]	0.9925	0.9764	0.9713	0.9919	0.9853	0.6267
BRISQUE[5]	0.9803	0.9646	0.4090	0.3142	0.3735	0.4593
NIQE[6]	0.9783	0.9880	0.7704	0.6618	0.3238	0.4242
VIIDEO[7]	0.8749	0.3549	0.4998	0.3983	0.4214	0.3842
Proposed algorithm	0.9972	0.9391	0.9639	0.9809	0.9928	0.9401

Table 4: The efficacy of the proposed algorithm in terms of SROCC measurements on the LVQ dataset (with non-expert subjective scores).

	Noise	Defocus Blur	Motion Blur	Uneven Illumination	Smoke	Overall
PSNR	0.9594	0.7773	0.8163	0.9372	0.9439	0.5775
SSIM[3]	0.9509	0.7157	0.8941	0.9502	0.8987	0.5914
VIF[4]	0.9636	0.9417	0.9433	0.9391	0.9316	0.6228
BRISQUE[5]	0.9571	0.9322	0.3564	0.2980	0.4041	0.4304
NIQE[6]	0.9640	0.9514	0.6101	0.5416	0.3589	0.3731
VIIDEO[7]	0.8600	0.3138	0.379	0.3888	0.3866	0.3416
Proposed algorithm	0.9762	0.8982	0.8982	0.9461	0.9000	0.9325

4. CONCLUSION

In this paper, we proposed a full-reference quality assessment algorithm for laparoscopic videos, based on computing the multiple-image quality measures of a frame. The regression neural network architecture was performed on frame-level features to predict the distorted video quality. The proposed algorithm was tested on the LVQ dataset and the results show robust performance with both expert and non-expert MOS scores. It delivered state-of-the-art performance compared to the other FR and NR image and video quality assessment models. In the future, we plan to extend these models to propose supervised and unsupervised quality assessment models.

ACKNOWLEDGMENTS

The work was supported by the National Research Development and Innovation Fund based on the charter of bolster issued by the National Research Development and Innovation Office under the auspices of the Ministry for Innovation and Technology, Hungary.

REFERENCES

[1] Avcibas, I., Sankur, B., and Sayood, K., "Statistical evaluation of image quality measures," *Journal of Electronic Imaging* **11**(2), 206–223 (2002).

[2] Appina, B., Jalli, A., Battula, S. S., and Channappayya, S. S., "No-reference stereoscopic video quality assessment algorithm using joint motion and depth statistics," in [*International Conference on Image Processing (ICIP)*], 2800–2804, IEEE (2018).

[3] Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P., "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing* **13**(4), 600–612 (2004).

[4] Sheikh, H. R. and Bovik, A. C., "Image information and visual quality," *IEEE Transactions on Image Processing* **15**(2), 430–444 (2006).

[5] Mittal, A., Moorthy, A. K., and Bovik, A. C., "No-reference image quality assessment in the spatial domain," *IEEE Transactions on Image Processing* **21**(12), 4695–4708 (2012).

[6] Mittal, Anish, R. S. and Bovik, A. C., "Making a "completely blind" image quality analyzer," *IEEE Signal processing letters* **20**(3), 209–212 (2012).

[7] Mittal, A., Saad, M. A., and Bovik, A. C., "A completely blind video integrity oracle," *IEEE Transactions on Image Processing* **25**(1), 289–300 (2016).

[8] Khan, Z. A., Beghdadi, A., Cheikh, F. A., Kaaniche, M., Pelanis, E., Palomar, R., Fretland, Å. A., Edwin, B., and Elle, O. J., "Towards a video quality assessment based framework for enhancement of laparoscopic videos," in [*Medical Imaging 2020: Image Perception, Observer Performance, and Technology Assessment*], **11316**, International Society for Optics and Photonics (2020).

[9] "VQEG. (2003). Final Report From the Video Quality Experts Group on the Validation of Objective Models of Video Quality Assessment, Phase II. [Online]. Available: http://www.its.bldrdoc.gov/vqeg/projects/frtv-phase-ii/frtv-phase-ii.aspx;accesseddate:19/05/2021."