IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received April 10, 2017, accepted May 9, 2017, date of publication May 15, 2017, date of current version June 28, 2017.

Digital Object Identifier 10.1109/ACCESS.2017.2704285

Study of Video Quality Assessment for Telesurgery

LUCIE LÉVÊQUE[1], WEI ZHANG[1], (Student Member, IEEE), CHRISTINE CAVARO-MÉNARD[2], PATRICK LE CALLET[3], (Senior Member, IEEE), AND HANTAO LIU[1], (Member, IEEE)

[1]School of Computer Science and Informatics, Cardiff University, Cardiff CF24 3AA U.K.
[2]LARIS, University of Angers, 49000 Angers, France
[3]IRCCyN, University of Nantes, 44300 Nantes, France

Corresponding authors: Lucie Lévêque (levequel@cardiff.ac.uk) and Hantao Liu (liuH35@cardiff.ac.uk)

ABSTRACT Telemedicine provides a transformative practice for access to and delivery of timely and high-quality healthcare in resource-poor settings. In a typical scenario of telesurgery, surgical tasks are performed with one surgeon situated at the patient's side and one expert surgeon from a remote site. In order to make telesurgery practice realistic and secure, reliable transmission of medical videos over large distances is essential. However, telesurgery videos that are communicated remotely in real time are vulnerable to distortions in signals due to data compression and transmission. Depending on the system and its applications, visual content received by the surgeons differs in perceived quality, which may incur implications for the performance of telesurgery tasks. To rigorously study the assessment of the quality of telesurgery videos, we performed both qualitative and quantitative research, consisting of semi-structured interviews and video quality scoring with human subjects. Statistical analyses are conducted and results show that compression artifacts and transmission errors significantly affect the perceived quality; and the effects tend to depend on the specific surgical procedure, visual content, frame rate, and the degree of distortion. The findings of the study are readily applicable to improving telesurgery systems.

INDEX TERMS Telemedicine, telesurgery, subjective experiment, video quality assessment, statistical analysis.

I. INTRODUCTION

Telemedicine refers to "the use of information and communications technologies (ICT) to provide and support clinical healthcare at a distance" [1]. The ultimate goal of telemedicine is to improve the accessibility, efficiency and effectiveness of clinical processes utilised by healthcare organisations, practitioners and patients. Telemedicine is being conducted to optimise medical resources and compensate for the lack of healthcare provision in places where the number of medical professionals relative to the size of the population is very small and access to advanced healthcare is limited. As a result, the use of telemedicine offers a transformative opportunity for access to and delivery of high quality healthcare in resource-poor settings, particularly but not exclusively in rural areas.

There are four recognised acts of telemedicine, namely tele-consultation, tele-expertise, tele-monitoring and tele-assistance [1]. Tele-consultation involves a remote consultation between a doctor and a patient, which is often conducted via telephoning or videoconferencing. Tele-expertise is concerned about communicating a remote request of clinical advice between medical professionals, such as seeking a second opinion on image-based diagnosis off-line. Telemonitoring enables medical professionals to remotely interpret pre-recorded medical data of patients. Tele-assistance is used to allow a clinical expert to support other clinical professionals during a medical intervention, which is the topic to be investigated in the current paper.

In principal, tele-assistance is conducted in more demanding situations, where the success of the practice partially depends on the effectiveness of the transmission of medical videos in real time. Recently, telesurgery (also known as remote surgery), has emerged as a prevalent form of medical tele-assistance, where the clinical practice involves an expert surgeon assisting a remote (less-experienced) surgeon to complete a surgical procedure [2]. It is useful in the case of a shortage of experienced professionals, and benefits health care services in terms of timely surgery for patients, reduction of costs, and training of young surgeons, etc.

The implementation of telesurgery would not have been possible without the support of modern digital image and video communication systems. Unfortunately, these systems generate as a side effect various types of distortion in visual signals, such as visual impairments caused by lossy data compression and transmission errors due to bandwidth limitations [3]. Therefore, the ultimate visual content received by the end user largely differs in perceived quality depending on the system and its applications. The visual distortions can affect viewers' visual experiences and, consequently, may impact the practice of telesurgery, and thus risk the patient's health. Being able to maintain, control and improve the quality of medical videos has become a fundamental challenge in the design of telesurgery systems [4].

Previous studies have shown that the influence of professional expertise on the assessment of medical image quality is significant, and that experts and naïve observers (i.e., without medical imaging or clinical experience) differently assess the quality of medical visual content degraded with distortions [5]. Psychophysical studies have been undertaken to understand the way medical experts perceive image quality aspects in the area of telesurgery [6]–[12]. Chaabouni *et al.* [6], [7], Kumcu *et al.* [8], [9], Martini *et al.* [10], Nouri *et al.* [11] and Shima *et al.* [12] carried out various subjective quality assessment experiments for laparoscopic surgery, using methodologies prescribed by the ITU (International Telecommunication Union) [13], [14]. Chaabouni *et al.* [6], [7], used the DSCQS (Double Stimulus Continuous Quality Scale) method to assess the perceived quality of four otolaryngology excerpts of a real surgical procedure with ENT (Ear, Nose and Throat) surgeons. They found that compression artifacts could be noticeable from compression ratio (CR) = 100:1 up to CR = 270:1 for H.264/AVC compressed videos. Kumcu *et al.* [8], [9], chose the ACR (Absolute Category Rating) method to assess four abdominal sequences from different surgical procedures. They showed that bit rate = 5.5 Mbps (or CR = 111:1 for H.264/AVC) was suitable for a surgical procedure, whereas bit rate = 3.2 Mbps (or CR = 214:1) was too poor to conduct a surgery. Martini *et al.* made use of the DSIS (Double Stimulus Impairment Scale) method to evaluate the effects of video transmission errors on two sequences from a biopsy suture. They found that the medical experts scored low quality for all stimuli, which may be due to the fact that the simulated errors (i.e., various packet loss rates) are annoying and not acceptable for the surgeons. Nouri et al. also made use of the DSCQS method on four videos representing different stages of a laparoscopic surgery. They found a quality threshold at bit rate = 3.2 Mbps (or CR = 90:1 for MPEG2 compression), below which surgeons thought the quality was too low to perform their tasks. Finally, Shima *et al.* used the DSCQS method as well, to assess four source videos representing different types of cancer. They showed that, when videos were transmitted over an LTE (Long Term Evolution) network, the quality decreased for pancreatic cancer videos but not for oesophageal, colon and gastric cancers. These standardised methods as mentioned above are widely used to assess the perceived quality of natural visual content under natural viewing conditions. However, the use of these methods for the assessment of medical images/videos remains an open-end question. It should be noted that clinical practice is rather complex, and issues such as how telesurgery practitioners perceive video quality aspects remain largely unexplored. Due to a lack of integrated methodology for the subjective assessment of quality in the context of telesurgery, attempts have been made in [9] and [12] to adjust the existing experimental methodology, where application-specific aspects, e.g., ''suitability for surgery'' have been integrated into the conventional quality scoring tasks. In view of this, it is worth further investigating dedicated experimental methodologies for video quality assessment in the context of telesurgery.

In this paper, the following contributions have been made towards the study of video quality assessment for telesurgery. First, semi-structured interviews were carried out with abdominal surgeons from Angers University Hospital in France. Qualitative data regarding critical aspects for video quality assessment in the context of telesurgery were obtained. Second, based on a better understanding of the context, a controlled experiment was designed and conducted, where surgeons rated the quality of videos distorted with various levels of degradation. A statistical analysis was performed to provide an insight into the quality perception of telesurgery videos.

II. SEMI-STRUCTURED INTERVIEWS: RELATING QUALITY IN THE CONTEXT OF TELESURGERY

A. PROTOCOL

In order to better understand quality perception in the context of telesurgery, we carried out qualitative research through interviews – an important gathering technique involving verbal communication between the researcher and research subject [15]. Interviews can be structured or semi-structured. For a structured interview, a rigorous set of questions is prepared prior to the interview and usually allows a limited number of answers. A semi-structured interview is more open (i.e., allowing informants the freedom to bring up ideas or express their views during the interview), but with a list of specific topics of interest being thought about well in advance which has to be covered. These interviews are used to collect data on a particular topic [16]. Semi-structured interviews are widely used in qualitative research to provide reliable and comparable qualitative data [17]. We used semi-structured interviews to reveal all relevant aspects of the quality assessment problem, including purpose, context and meaning, in order to get preliminary information linked to the quality attributes present in the videos.

Knowledge management is a methodology commonly used to conduct semi-structured interviews. The principle of knowledge management is to make the know-how of experts explicit (i.e., the transition from a tacit know-how to an explicit knowledge) [18]. CTA (Cognitive Task Analysis) [19] is a knowledge capitalisation method whose purpose is to elicit knowledge and skills used by an expert during a task involving a strong cognitive activity (such as making a decision, solving a problem, and being aware). This method is made of five steps: collection of preliminary knowledge, identification of knowledge representations, application of knowledge elicitation methods, analysis of acquired data and formatting results. This method is the extension of traditional task analysis techniques to underline goal generation, decision-making and judgments [20], which are essential to study quality assessment in the context of telesurgery.

As recommended by the CTA methodology, a literature review was made on the topic of interest (telemedicine and surgery) to set up the interviews. We then made a list of questions and topics to be explored with the surgeons. The topics can be divided into two categories: telemedicine and field of expertise on the one hand, video quality on the other hand. Examples of questions are: 'What do you generally expect in a setting of telesurgery?'; 'What are the most important aspects you would look in this kind of exam?'; 'What is the most challenging part of this procedure?'; and 'What are the necessary knowledge and material for the conduct of such a procedure?'. Three surgeons were interviewed, and each interview lasted about one hour. During each interview, several videos of abdominal surgery (i.e., open and laparoscopic surgeries) were shown. These videos were pre-captured in Angers University Hospital, with a Sony Handycam HDR-CX280 model for open surgeries and a Stryker's endoscope for laparoscopic surgeries. All interviews were tape-recorded.

B. RESULTS

The recordings of the interviews were analysed using Strauss and Corbin's coding methodology [21], which is widely used in the literature to analyse qualitative data. It consists of the following stages: gathering qualitative data (i.e., semi-structured interviews), organising the data (i.e., transcribing), fragmenting the data (i.e., open coding), categorising the data (i.e., axial coding), and linking the data (i.e., selective coding).

After transcribing the interviews, every sentence was analysed using open coding. The similar open codes were then grouped into categories to define the axial codes. To achieve this step, key words and key ideas were first grouped using synonyms or lexical fields, and then into concepts. Finally, two core concepts were defined during the selective coding: information about medical tele-assistance in general and information more specifically linked to the video characteristics. Tables 1 and 2 represent the key concepts defined from the interviews. All the elements are detailed further.

In terms of the clinical environment of the telesurgery practice, surgeons do not need assistance during the whole procedure but only for critical phases. To help a remote surgeon, the expert has to know the degree of emergency of the procedure on the one hand, and the surgeon's level and potential difficulties on the other hand. They also need information about the clinical case of the patient, such as context, medical history and drug therapy. Finally, the expert has to be in a quiet room with low light to watch the videos, like in a routine clinical practice.

TABLE 1. Generalities on telesurgery defined from the interviews.

Concepts	Axial codes	Open codes
Use	Remote assistance Teaching	Critical phases Second opinion Future surgeons
Requirements	Patient clinical case	Context Drug history Potential problems
	Visualisation	Quiet room Low light conditions Large screen

TABLE 2. Video characteristics information from the interviews.

Concepts	Axial codes	Open codes
Image	Features	Centre Colours Contrasts Edges Textures
	Shooting	Abdominal cavity for open surgery Whole scene for laparoscopic surgery
Video	Transmission	Real time
	Audio	Real time Both ways

In terms of the perception of video material, experts only watch the central area of the picture during a surgical procedure if acquisition was correctly made; therefore it can be inferred that distortions at the periphery of the picture are less annoying than that in the centre. The expert can approximate the size of the organs by comparing them to the instruments. Some elements appeared to be very important to locate the organs, such as colour, edge, texture, and contrast. These attributes have to be of a sufficient quality in the context of remote assistance in surgery. A point was raised during each interview that the sound and interaction were essential. However, this point is considered outside of the scope of our study.

For open surgeries, the expert does not need a video of the operating room but only of the abdominal cavity, whereas a video of the whole scene is necessary for laparoscopic surgeries because the trocars have to be correctly positioned. For laparoscopy, both videos (i.e., the abdominal cavity and the scene) are not needed at the same time. We also learned some additional information during the interviews. For instance, during a planned surgery, a surgeon can take a break of up to 4 to 5 minutes if there is a technical problem during video transmission (e.g., Internet disruption). Laparoscopic surgery is never used for emergency cases whereas open surgery can be.

The results of the semi-structured interviews are used to develop the methodology for the subjective assessment of video quality in the context of telesurgery, which is detailed below.

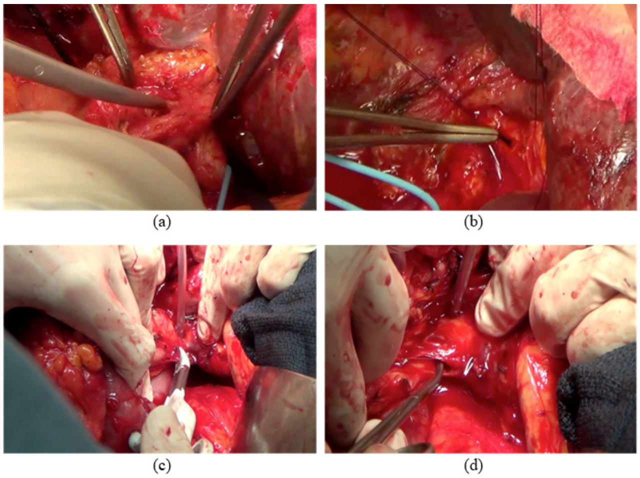


FIGURE 1. Illustration of one frame from each of the four open surgery videos used in our experiment: (a) content 1 (hepatic artery dissection), (b) content 2 (small wires used to tie a knot), (c) content 3 (use of a clamp to stop a blood flow), and (d) content 4 (important bleeding during the procedure).

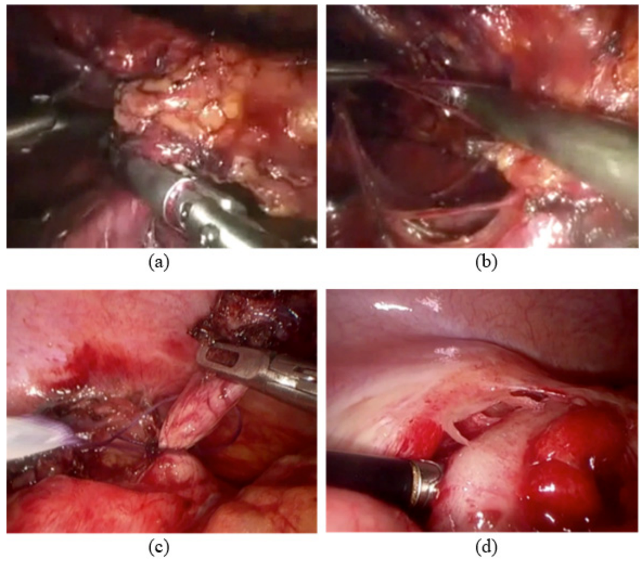


FIGURE 2. Illustration of one frame from each of the four laparoscopic surgery videos used in our experiment: (a) content 1 (dissection of the mesenteric artery), (b) content 2 (before the artery dissection), (c) content 3 (use of an Endoloop ligature to suture), and (d) content 4 (different textures during the procedure).

III. CONTROLLED EXPERIMENT: RATING QUALITY IN THE CONTEXT OF TELESURGERY

A. METHODOLOGY

For each type of surgical procedures, namely open and laparoscopic surgeries, four relevant videos of twenty second each were extracted from real surgical acts by a senior surgeon from Angers University Hospital who did not participate in the later stages of the subjective experiment. The choice of these excerpts was justified by the model created as a result of the semi-structured interviews; and they represent the space of possibilities. Amongst these videos, some videos represent content of tiny details whereas other videos contain colourful regions. These aspects were considered as representative attributes by the surgeons during the interviews.

Fig. 1 shows one representative frame from each of the four videos of open surgery. Fig. 2 illustrates one representative frame from each of the four videos of laparoscopic surgery. Table 3 describes the technical specifications of these videos.

To simulate a realistic telesurgery scenario, e.g., in satellite communication, various levels of compression artifacts and transmission errors were generated on the source videos. We used H.264 [22] to compress videos, resulting in typical artifacts such as blocking, blur, ringing and motion compensation mismatches. The transmission errors were simulated using packet loss generated by an internal tool (i.e., based on a Gilbert-Elliott model [23]). We also varied the frame rate of the videos. This yielded 32 distorted videos for each surgical procedure. Table 4 details the configuration of the dataset. Fig. 3 shows an example of how the video content is distorted in open surgery.

We conducted a visual perception experiment using a single-stimulus method [13], where human subjects were requested to score video quality aspects for each stimulus in the absence of a reference. The questionnaire is illustrated in Fig. 4, which reflects the elements extracted from the semi-structured interviews. For each video, the participants were asked to express their opinions on the general quality of the video with a task of helping a remote surgeon in mind

TABLE 3. Technical characteristics of the videos used in our experiment.

Procedure	Content	Resolution	Frame rate
Open surgery	1	872×480	30 fps
	2	872×480	30 fps
	3	960×720	30 fps
	4	960×720	30 fps
Laparoscopy	1	352×288	25 fps
	2	352×288	25 fps
	3	720×576	25 fps
	4	720x576	25 fps

TABLE 4. Distortion parameters used in our experiment.

Procedure	Condition	Bit rate	Frame rate	Packet loss rate
Open surgery	1	128 kbps	30 fps	0%
	2	256 kbps	30 fps	0%
	3	350 kbps	30 fps	0%
	4	512 kbps	30 fps	0%
	5	1 Mbps	30 fps	0%
	6	1 Mbps	15 fps	0%
	7	1 Mbps	30 fps	1%
	8	1 Mbps	30 fps	3%
Laparoscopy	1	128 kbps	25 fps	0%
	2	256 kbps	25 fps	0%
	3	350 kbps	25 fps	0%
	4	512 kbps	25 fps	0%
	5	1 Mbps	25 fps	0%
	6	1 Mbps	12.5 fps	0%
	7	1 Mbps	25 fps	1%
	8	1 Mbps	25 fps	3%

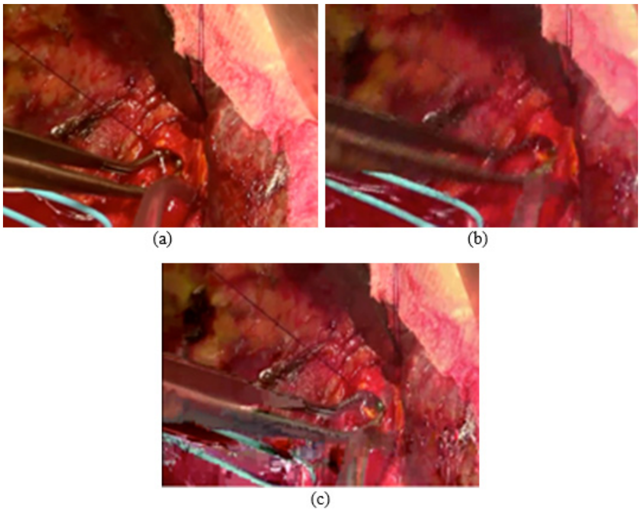


FIGURE 3. **Illustration of different types of distortions for the same content (i.e., open surgery, content 2): (a) reference (un-distorted video, 30fps) (b) bit rate distortion (128kbps, 30fps), and (c) packet loss distortion (1Mbps, 30fps).**

Question 1: What do you think of the colours in this video (with a view to help a remote surgeon)? (0: very bad, 10: excellent)

0 1 2 3 4 5 6 7 8 9 10

Question 2: What do you think of the contrasts (details, edges) in this video (with a view to help a remote surgeon)? (0: very bad, 10: excellent)

0 1 2 3 4 5 6 7 8 9 10

Question 3: What do you think of the 3D (relief) in this video (with a view to help a remote surgeon)? (0: very bad, 10: excellent)

0 1 2 3 4 5 6 7 8 9 10

Question 4: What do you think of the textures in this video (with a view to help a remote surgeon)? (0: very bad, 10: excellent)

0 1 2 3 4 5 6 7 8 9 10

Question 5: What do you think of the overall quality of this video (with a view to help a remote surgeon)? (0: very bad, 10: excellent)

FIGURE 4. **Illustration of the questionnaire used in our experiment.**

(i.e., with an embedded view to assess the usability), using a discrete rating scale from 0 for ''bad'' to 10 for ''excellent''. They were also asked to assess other relevant aspects of videos, i.e., colour perception, details/edges of visual content, degree of relief and textures of objects. These application-specific attributes were identified in our qualitative study in Section II.

The experiments were conducted in a standard office environment [24] at Angers University Hospital. The stimuli were displayed on a Dell 27-inch wide-screen liquid-crystal display with a native resolution of 1920 × 1200 pixels. The viewing distance was approximately 60 cm. For each surgical procedure, we recruited four surgeons (note that the sample size used is considered *adequate* in the area of medical image perception mainly due to the high degree of consistency among medical professionals [25], [26]) from Angers University Hospital, having from four to twenty-seven years of expertise. Before the start of the actual experiment, each participant was provided with a briefing on the goal and procedure of the experiment, including the form of the questionnaire, scoring scale and timing. A training session was conducted in order to familiarise the participants with the distortions in videos and how to use the scale for scoring. The video stimuli used in the training were different from those used in the real experiment. After training, all test stimuli were shown one by one in a different random order to each participant in a separate session.

B. RESULTS FOR OPEN SURGERY

To process the raw data, an outlier detection and subject exclusion procedure was applied to the scores [27], [28]. An individual score given for a video quality aspect (i.e., a particular question in Fig. 4) would be considered an outlier if it was outside an interval of two standard deviations around the mean score for that aspect. A subject would be rejected if more than 20% of their scores were outliers. As a result of the above procedure, none of the scores was detected as an outlier and, therefore, no surgeon was excluded from the analysis.

A correlation analysis was performed by calculating the Pearson linear correlation coefficient between the overall quality (i.e., Question 5 in Fig. 4) and one of the identified video quality attributes (i.e., Question 1 to 4 in Fig. 4). The coefficient is 93% between overall quality and ''colour'', and is 96%, 96% and 97% for ''contrast'', ''relief'' and ''texture'', respectively. The results quantitatively confirm the significance of the video quality attributes identified in the qualitative study. We now focus on the statistical analysis using the overall quality scores. *Note*, an extensive discussion on modelling specific quality attributes and their relative impact on the overall quality is outside the scope of this paper, and will be treated in a separate contribution.

Fig. 5 illustrates the mean opinion score (MOS) averaged over all surgeons for each distorted video in our experiment. It shows that compression artifacts, transmission errors and frame rate affect the overall quality. In addition, the effect tends to depend on the video content. The observed tendencies are further statistically analysed with an ANOVA (Analysis of Variance) using the software package SPSS version 23. Based on the configuration of the stimuli as illustrated in Table 4, we perform three individual factorial ANOVA tests: 1) evaluation of the effect of bit rate on videos (based on conditions 1-5 in Table 4), 2) evaluation of the effect of frame rate on videos (based on conditions 5 and 6), and 3) evaluation of the effect of packet loss rate on videos (based on conditions 5, 7 and 8). For each case above, the video content is included in the ANOVA.

To evaluate the effect of bit rate with an ANOVA, the perceived quality is selected as the dependent variable, the video content and bit rate as fixed independent variables, and the participant as random independent variable. The 2-way

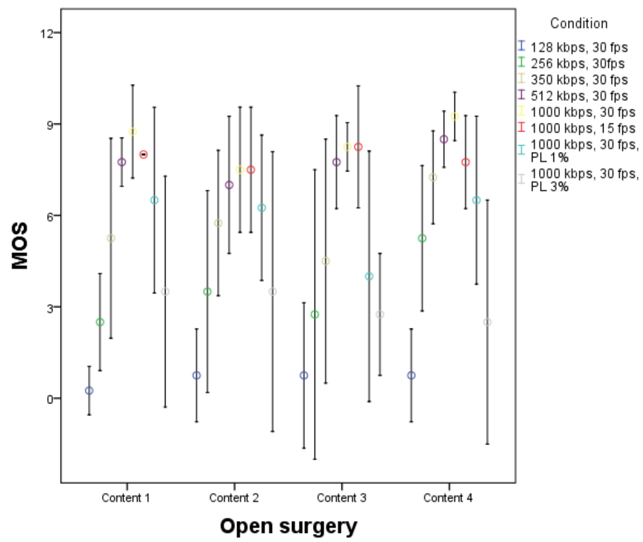


FIGURE 5. **Illustration of the mean opinion score (MOS) averaged across subjects for each distorted video for open surgery. "Content" refers to a source video. Error bars indicate a 95% confidence interval.**

TABLE 5. **Results of the ANOVA to evaluate the effect of "Bit rate" and "Content" on the perceived quality for open surgery.**

Factor	df	F	p-value
Bit rate	4	89.445	<0.001
Content	3	3.255	0.074
Bit rate * Content	12	2.230	0.032

TABLE 6. **Results of the ANOVA to evaluate the effect of "Frame rate" and "Content" on the perceived quality for open surgery.**

Factor	df	F	p-value
Frame rate	1	4.765	0.117
Content	3	1.809	0.216
Frame rate * Content	3	2.168	0.162

interactions of the fixed independent variables are included in the analysis. The results are summarised in Table 5, where the F-statistic (i.e., F) and its associated degrees of freedom (i.e., df) and significance (i.e., p-value) are included; and show that bit rate has a significant effect on perceived quality. The post-hoc test reveals the following order in quality (note that commonly underlined entries are not significantly different from each other):

128 kbps (<MOS> = 0.63) < 256 kbps (<MOS> = 3.50) < 350 kbps (<MOS> = 5.69) < 512 kbps (<MOS> = 7.75) < 1 Mbps (<MOS> = 8.44) (also see the comparison of MOS in Fig. 6).

Also, the interaction between video content and bit rate is significant, which implies that the difference in quality between different bit rates is not the same for the four source videos.

The results of the ANOVA to evaluate the effect of frame rate on videos are summarised in Table 6, and show that the effects are not statistically significant. The post-hoc test reveals the following order in quality (note that commonly underlined entries are not significantly different from each other):

15 fps (<MOS> = 7.88) < 30 fps (<MOS> = 8.44) (also see the comparison of MOS in Fig. 7).

The results of the ANOVA to evaluate the effect of packet loss on videos are summarised in Table 7, and show that packet loss rate has a significant effect on perceived quality. The post-hoc test reveals the following order in quality:

3% (<MOS> = 3.063) < 1% (<MOS> = 5.81) < 0% (<MOS> = 8.44) (also see the comparison of MOS in Fig. 8).

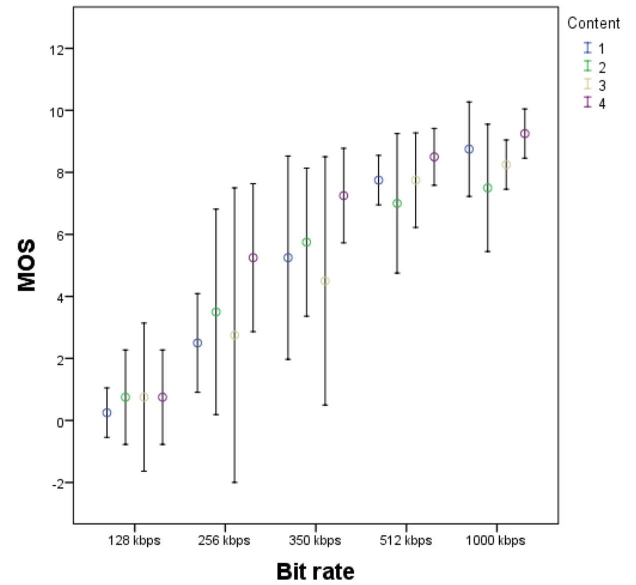


FIGURE 6. **Illustration of the effect of bit rate on open surgery videos. Error bars indicate a 95% confidence interval.**

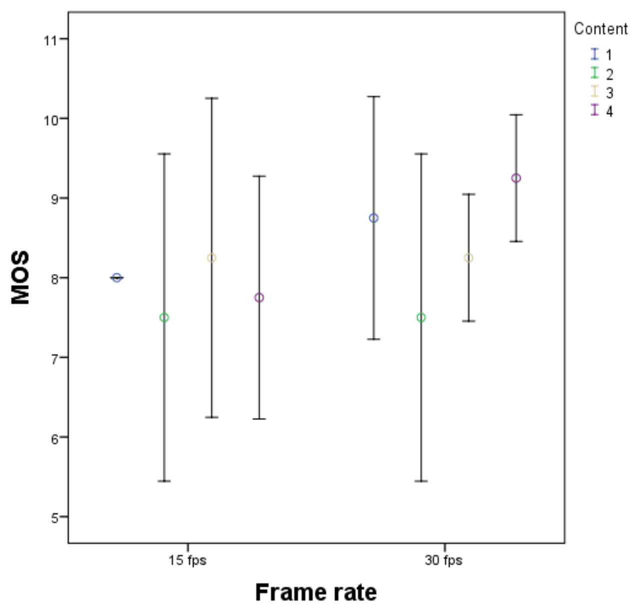


FIGURE 7. **Illustration of the effect of frame rate on open surgery videos. Error bars indicate a 95% confidence interval.**

TABLE 7. Results of the ANOVA to evaluate the effect of "Packet loss rate" and "Content" on the perceived quality for open surgery.

Factor	df	F	p-value
Packet loss rate	2	17.858	0.003
Content	3	2.673	0.082
Packet loss rate * Content	6	2.538	0.059

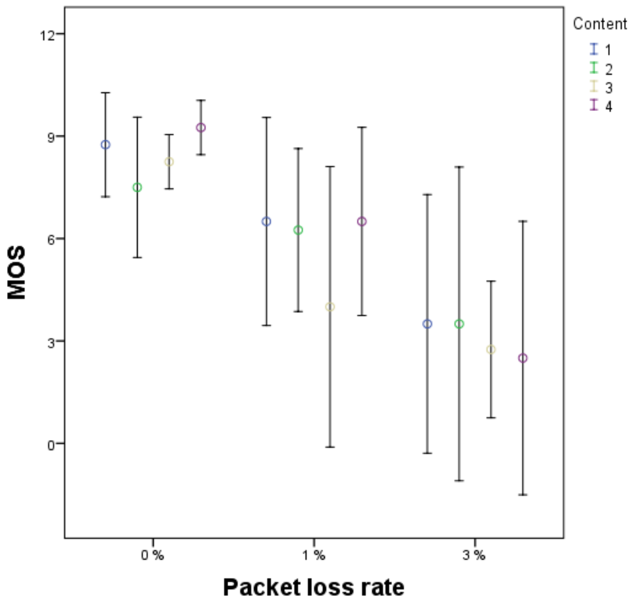


FIGURE 8. Illustration of the effect of packet loss on open surgery videos. Error bars indicate a 95% confidence interval.

TABLE 8. Results of the ANOVA to evaluate the effect of "Bit rate" and "Content" on the perceived quality for laparoscopic surgery.

Factor	df	F	p-value
Bit rate	4	63.676	<0.001
Content	3	3.395	0.067
Bit rate * Content	12	1.970	0.058

C. RESULTS FOR LAPAROSCOPIC SURGERY

We take the same statistical approach as used above to analyse the subjective data collected for laparoscopic surgery.

As a result of the outlier removal and subject exclusion procedure, none of the scores given by the surgeons was detected as an outlier, and therefore, no subject was excluded. The Pearson coefficient is 89% between overall quality and "colour", and is 94%, 96% and 96% for "contrast", "relief" and "texture" respectively. Fig. 9 illustrates the mean opinion score (MOS) averaged over all surgeons for each distorted video for laparoscopic surgery.

An ANOVA was performed to evaluate the effect of bit rate on videos. The results are summarised in Table 8, and show that bit rate has a significant effect on perceived quality. The post-hoc test reveals the following order in quality (note that commonly underlined entries are not significantly different from each other):

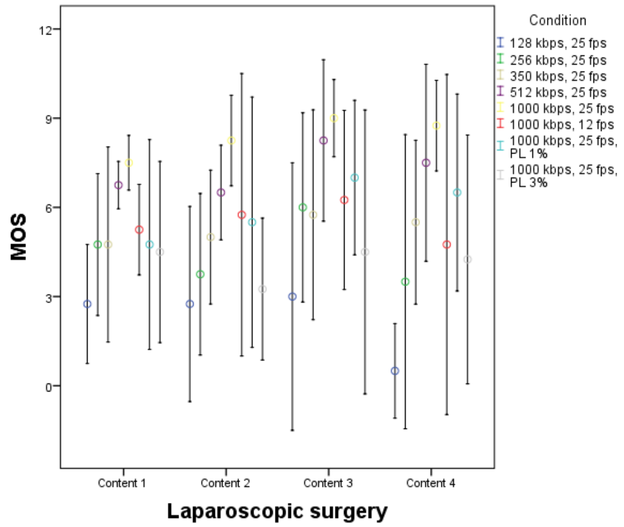


FIGURE 9. Illustration of the mean opinion score (MOS) averaged across subjects for each distorted video for laparoscopic surgery. "Content" refers to a source video. Error bars indicate a 95% confidence interval.

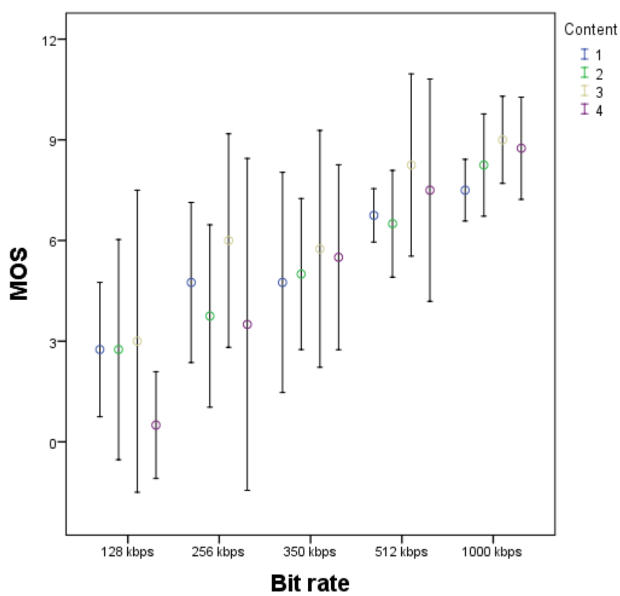


FIGURE 10. Illustration of the effect of bit rate on laparoscopic surgery videos. Error bars indicate a 95% confidence interval.

128 kbps (<MOS> = 2.25) < 256 kbps (<MOS> = 4.5) < 350 kbps (<MOS> = 5.25) < 512 kbps (<MOS> = 7.25) < 1 Mbps (<MOS> = 8.13) (also see the comparison of MOS in Fig. 10).

The results of the ANOVA to evaluate the effect of frame rate on videos are summarised in Table 9, and show that the effects are statistically different. The post-hoc test reveals the following order in quality:

12.5 fps (<MOS> = 5.50) < 25 fps (<MOS> = 8.13) (also see the comparison of MOS in Fig. 11).

The results of the ANOVA to evaluate the effect of packet loss rate on videos are summarised in Table 10, and show that

TABLE 9. Results of the ANOVA to evaluate the effect of "Frame rate" and "Content" on the perceived quality for laparoscopic surgery.

Factor	df	F	p-value
Frame rate	1	10.099	0.050
Content	3	1.538	0.271
Frame rate * Content	3	1.112	0.394

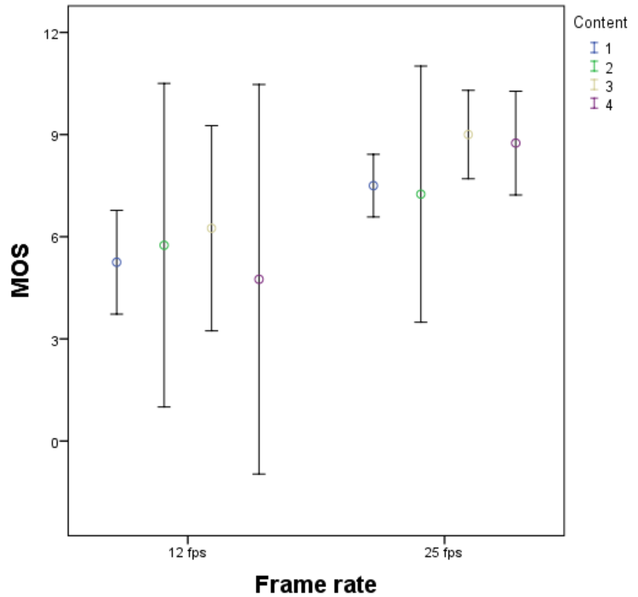


FIGURE 11. Illustration of the effect of frame rate on laparoscopic surgery videos. Error bars indicate a 95% confidence interval.

TABLE 10. Results of the ANOVA to evaluate the effect of "Packet loss rate" and "Content" on the perceived quality for laparoscopic surgery.

Factor	df	F	p-value
Packet loss rate	2	15.029	0.005
Content	3	6.187	0.094
Packet loss rate * Content	6	0.795	0.586

packet loss has a significant effect on perceived quality. The post-hoc test reveals the following order in quality:

3% (<MOS> = 4.13) < 1% (<MOS> = 5.88) < 0% (<MOS> = 8.13) (also see the comparison of MOS in Fig. 12).

IV. DISCUSSION

The results of our study provide insights into how variations in video distortion can affect the quality perception of surgeons in the practice of telesurgery. For both surgical procedures, compression artifacts have statistically significant effect on perceived quality. For open surgery, the way the video quality changes with the bit rate depends on the video content. However, for laparoscopic surgery, the impact of the different bit rates have on video quality is the same for all video scenes. The impact of video content on the

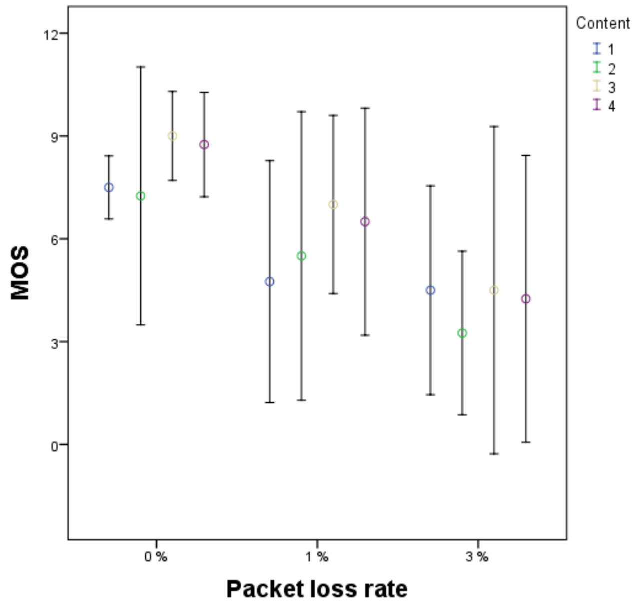


FIGURE 12. Illustration of the effect of packet loss on laparoscopic surgery videos. Error bars indicate a 95% confidence interval.

former case is probably due to that content itself may induce an intrinsic difference in sensitivity to distortion and thus, in the annoyance of distortion. We also found that at high bit rate (i.e., 1 Mbps), transmission errors (i.e., packet loss) can significantly reduce the perceived video quality for both surgical procedures in the similar way. The quality of videos for open surgery is not sensitive to the change of frame rate; but for laparoscopic surgery, videos with a lower frame rate are scored lower in quality than videos with a higher frame rate.

It should, however, be noted that our study with a subjective experiment is intrinsically time-consuming and thus limited with respect to the amount of test stimuli, the stimulus variability and the number of participants. Adding more experimental data to the evaluation would be highly beneficial, especially in terms of adding confidence to the generalizability of the conclusions.

V. CONCLUSIONS

In this paper, we investigated the assessment of the quality of videos for telesurgery, via both qualitative and quantitative research. First, we performed semi-structured interviews with surgeons to obtain reliable qualitative data in terms of the quality attributes of telesurgery videos, the contextual/environmental aspects for telesurgery practice, and the quality assessment behaviour of experts. Second, based on the findings of the qualitative study, we designed and conducted a perception experiment, where videos of different surgical procedures, and for each procedure videos of diverse content distorted with various levels of degradation were assessed by surgeons. Statistical evaluation showed that the perceived quality tends to depend on the specific procedure of telesurgery; within each surgical procedure of interest, the perceived quality of videos tends to be affected by content, frame rate, compression ratio and amount of transmission errors. Our study provides new insights into the perception of video quality in the context of telesurgery, and the findings can be used to optimise the perception of telesurgery practitioners.

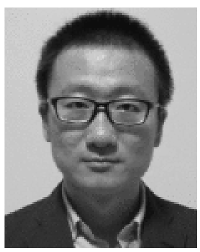
ACKNOWLEDGMENT

Many thanks go to the surgeons who were involved in their experiments. The study would not have been possible without their expertise.

REFERENCES

[1] N. R. Council, *Telemedicine: A Guide to Assessing Telecommunications for Health Care*. Washington, DC, USA: The National Academies Press, 1996.
[2] D. Holt, A. Zaidi, J. Abramson, R. Somogyi, "Telesurgery: Advances and trends," *Univ. Toronto Med. J.*, vol. 82, 2004.
[3] C. Cavaro-Ménard and A. Naït-Ali, *Compression of Biomedical Imaging and Signals*. Hoboken, NJ, USA: Wiley, 2008, p. 288.
[4] M. Razaak and M. Martini, "Medical image and video quality assessment in e-health applications and services," in *Proc. IEEE 15th Int. Conf. e-Health Netw. Appl. Services*, Sep. 2013, pp. 6–10.
[5] H. Liu, J. Koonen, M. Fuderer, and I. Heynderickx, "The relative impact of ghosting and noise on the perceived quality of MR images," *IEEE Trans. Image Process.*, vol. 25, no. 7, pp. 3087–3098, Jul. 2016.
[6] A. Chaabouni, Y. Gaudeau, J. Lambert, J.-M. Moureaux, and P. Gallet, "Subjective and objective quality assessment for H 264 compressed medical video sequences," in *Proc. 4th Int. Conf. Image Process. Theory, Tools Appl.*, Paris, France, 2014, pp. 1–5.
[7] A. Chaabouni, Y. Gaudeau, J. Lambert, J.-M. Moureaux, and P. Gallet, "H.264 medical video compression for telemedicine: A performance analysis," in *IRBM*, vol. 37. Amsterdam, The Netherlands: Elsevier, 2015, pp. 40–48.
[8] A. E. Kumcu *et al.*, "Visual quality assessment of H.264/AVC compressed laparoscopic video," *Proc. SPIE*, vol. 9037, pp. 1–12, Mar. 2014.
[9] A. Kumcu *et al.*, "Selecting stimuli parameters for video quality assessment studies based on perceptual similarity distances," *Proc. SPIE*, vol. 9399, pp. 1–10, Sep. 2015.
[10] M. Martini, C. Hewage, M. Nasralla, R. Smith, I. Jourdan, and T. Rockall, "3-D robotic tele-surgery and training over next generation wireless networks," in *Proc. 35th Annu. Int. Conf. IEEE EMBS*, Osaka, Japan, Apr. 2013, pp. 6244–6247.
[11] N. Nouri, D. Abraham, J.-M. Moureaux, M. Dufaut, J. Hubert, and M. Perez, "Subjective MPEG2 compressed video quality assessment: Application to tele-surgery," in *Proc. 7th IEEE Int. Symp. Biomed. Imag.*, Rotterdam, The Netherlands, May 2010, pp. 764–767.
[12] Y. Shima, A. Suwa, Y. Gomi, H. Nogawa, H. Nagata, and H. Tanaka, "Qualitative and quantitative assessment of video transmitted by DVTS (digital video transport system) in surgical telemedicine," *J. Telemed. Telecare*, vol. 13, no. 3, pp. 148–153, 2007.
[13] *Methodology for the Subjective Assessment of the Quality of Television Pictures*, document ITU-R, BT.500-11, 2002.
[14] *Subjective Video Quality Assessment Methods for Multimedia Applications*, document ITU-T. 910, 2008.
[15] R. Weiss, *Learningfrom Strangers: The Art and Method of Qualitative Interview Studies*, New York, NY, USA: Free Press, 1995, p. 256.
[16] H. Bernard, *Research Methods in Anthropology: Qualitative and Quantitative Approaches*. USA: AltaMira Press, 1988, p. 520.
[17] S. Jamshed, "Qualitative research method-interviewing and observation," *J. Basical Clin. Pharmacy*, vol. 5, pp. 87–88, May 2014.
[18] J.-L. Ermine, *Challenges and Approaches for Knowledge Management*. New Delhi, India: CEA, 2000.
[19] R. E. Clark and D. F. Feldon, "Cognitive task analysis," *Handbook of Research on Educational Communications and Technology*. USA: Lawrence Erlbaum Associates, 2006.
[20] J. Annett and N. Stanton, *Task Analysis*. Boca Raton, FL, USA: CRC Press, 2000, p. 248.
[21] A. Strauss and J. Corbin, *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*. Thousand Oaks, CA, USA: SAGE Publications, 1998, p. 312.
[22] T. Wiegand, G. J. Sullivan, G. Bjøntegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 560–576, Jul. 2003.
[23] G. Hasslinger and O. Hohlfeld, "The Gilbert–Elliott model for packet loss in real time services on the Internet," in *Proc. 14th GI/ITG Conf. Meas. Modelling Eval. Comput. Commun. Syst.*, 2008, pp. 1–15.
[24] K. J. Robson and C. J. Kotre, "Pilot study into optimisation of viewing conditions for electronically displayed images," *Radiat. Protection Dosimetry*, vol. 117, nos. 1–3, pp. 298–303, 2005.
[25] E. Krupinski, "Diagnostic accuracy and visual search efficiency: Single 8 MP vs. dual 5 MP displays," *J. Digit. Imag.*, vol. 30, pp. 144–147, Sep. 2017.
[26] M. Razaak and M. Martini, "Rate-distortion and rate-quality performance analysis of HEVC compression of medical ultrasound videos," in *Proc. 4th Int. Conf. Sel. Topics Mobile Wireless Netw.*, Rome, Italy, 2014, pp. 230–236.
[27] Z. Wang and A. Bovik, *Modern Image Quality Assessment: Synthesis Lectures on Image, Video and Multimedia Processing*. USA: Morgan & Claypool, 2006, p. 156.
[28] H. R. Sheikh, M. F. Sabir, and A. C. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Trans. Image Process.*, vol. 15, no. 11, pp. 3440–3451, Nov. 2006.

LUCIE LÉVÊQUE received the B.Eng. and M.Eng. degrees in cognitics from the Ecole Nationale Supérieure de Cognitique, Bordeaux, France, in 2013 and 2015, respectively, and the M.Sc. degree in biomedical imaging from the University of Angers, Angers, France, in 2015. She is currently pursuing the Ph.D. degree with the School of Computer Science and Engineering, Cardiff University, Cardiff, U.K. Her research interests include visual media quality assessment and medical image perception.



WEI ZHANG (S'14) received the B.Sc. and M.Sc. degrees from Xidian University, Xi'an, China, in 2011 and 2013, respectively. He is currently pursuing the Ph.D. degree with the School of Computer Science and Informatics, Cardiff University, Cardiff, U.K. His research interests include image analysis, video processing, and human visual perception.

CHRISTINE CAVARO-MÉNARD received the M.Sc. and Ph.D. degrees from the Ecole doctorale Santé, Sciences et Technologies, Tours, France. She is currently a Lecturer with the University of Angers, Angers, France.

Her research interests include medical image and video compression, engineering, and medicine.

PATRICK LE CALLET (SM'14) received the M.Sc. and Ph.D. degrees in image processing from the Ecole Polytechnique de l'Université de Nantes, Nantes, France. He was a Student with the Ecole Normale Supérieure de Cachan, where he sat the Aggrégation (credentialing exam) in electronics of the French National Education. He was an Assistant Professor from 1997 to 1999 and a Full-Time Lecturer from 1999 to 2003 with the Department of Electrical Engineering of Technical Institute of the University of Nantes. Since 2003, he has been involved in teaching with the Electrical Engineering and the Computer Science Departments, Engineering School, Ecole Polytechnique de l'Université de Nantes, where is currently a Full Professor. He led for ten years (2006–2016) the Image and Video Communication Laboratory, CNRS IRCCyN. Since 2015, he has been the Scientific Director of the cluster Ouest Industries Créatives gathering over ten institutions, including three universities to strengthen the Research, Education and Innovation of the Région Pays de Loire in the field of creative industries. He is mostly engaged in research dealing with the application of human vision modeling in image and video processing. His current centers of interest are quality of experience assessment, visual attention modeling and applications, perceptual video coding and immersive media processing. He has co-authored over 250 publications and communications and co-invented 16 international patents on these topics. He was one of the five members of the Board of Governor of CNRS IRCCyN (250 researchers) from 2013 to 2016. Since January 2017, he has been one of the seven members of the Board of Governors of the CNRS LS2N Laboratory (450 researchers), as representative of Polytech Nantes. He serves or has been served as Associate Editor or Guest Editor for several journals such as the IEEE TIP, the IEEE STSP, the IEEE TCSVT, the *SPRINGER EURASIP Journal on Image and Video Processing*, and the SPIE JEI. He is currently serving the IEEE IVMSP-TC and the IEEE MMSP-TC. He is also one the Founding Member of the EURASIP Special Areas Team on *Image and Video Processing*.

HANTAO LIU (S'07–M'11) received the Ph.D. degree from the Delft University of Technology, Delft, The Netherlands, in 2011. He is currently an Assistant Professor with the School of Computer Science and Informatics, Cardiff University, Cardiff, U.K. His research interests include visual media quality assessment, visual attention modeling and applications, visual scene understanding, and medical image perception. He is currently serving the IEEE MMTC as the Chair of the Interest Group on Quality of Experience for Multimedia Communications. He is also an Associate Editor of the IEEE Transactions on Human–Machine Systems and the IEEE Transactions on Multimedia.