ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/301464069

Building a Large Dataset for Model-based QoE Prediction in the Mobile Environment

Conference Paper · November 2015

DOI: 10.1145/2811587.2811631

CITATIONS
2

READS
22

4 authors:

Lamine Amour
Université Paris-Est Créteil Val de Marne - Univ…
7 PUBLICATIONS **12** CITATIONS

SEE PROFILE

Sami Souihi
Université Paris-Est Créteil Val de Marne - Univ…
23 PUBLICATIONS **31** CITATIONS

SEE PROFILE

Said Hoceini
Université Paris-Est Créteil Val de Marne - Univ…
68 PUBLICATIONS **211** CITATIONS

SEE PROFILE

Abdelhamid Mellouk
Université Paris-Est Créteil Val de Marne - Univ…
262 PUBLICATIONS **1,289** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Project

Belief functions theory in telecommunications and network technologies View project

Project

Implementation of a new adaptive and scalable model for the real QoE estimation : Application to the mobile services View project

All content following this page was uploaded by Lamine Amour on 21 February 2017.

The user has requested enhancement of the downloaded file.

Building a Large Dataset for Model-based QoE Prediction in the Mobile Environment

Lamine Amour, Souihi Sami, Said Hoceini and Abdelhamid Mellouk
University of Paris-Est Créteil Val de Marne (UPEC)
Image, Signal and Intelligent Systems Lab-LiSSi and Network & Telecoms Dept, IUT CV
122 rue Paul Armangot, 94400 Vitry sur Seine, France
(lamine.amour, sami.souihi, hoceini, mellouk) @u-pec.fr

ABSTRACT

The tremendous growth in video services, specially in the context of mobile usage, creates new challenges for network service providers : How to enhance the user's Quality of Experience (QoE) in dynamic wireless networks (UMTS, HSPA, LTE/LTE-A). The network operators use different methods to predict the user's QoE. Generally to predict the user's QoE, methods are based on collecting subjective QoE scores given by users. Basically, these approaches need a large dataset to predict a good perceived quality of the service. In this paper, we setup an experimental test based on crowdsourcing approach and we build a large dataset in order to predict the user's QoE in mobile environment in term of Mean Opinion Score (MOS). The main objective of this study is to measure the individual/global impact of QoE Influence Factors (QoE IFs) in a real environment. Based on the collective dataset, we perform 5 testing scenarios to compare 2 estimation methods (SVM and ANFIS) to study the impact of the number of the considered parameters on the estimation. It became clear that using more parameters without any weighing mechanisms can produce bad results.

Keywords

Quality of Experience (QoE); Mobile environment; Crowdsourcing; Mean opinion score; Smartphone; Video.

1. INTRODUCTION

A large growth in Internet based devices (e.g. Smart phone, Tablet, etc.) causes the emergence of multimedia service that changed our daily lives. Our life is increasingly be made of continuous interaction with multimedia services : Email consultation, ticket booking, live games, etc. In this context the use of traditional monitoring networks based only on Quality of Service (QoS) optimization are not sufficient to ensure user's requirements. That is why, system actors (service provider, network operator,...etc.) are investigating a new concept called Quality of Experience (QoE) or simply User Quality. This new concept affected several areas such as multimedia services and the medical field to evaluate the real quality perceived by users. To pinpoint the problem of user perception, many works were proposed and several community were created such as *Qualinet* in Europe. Although, this concept is still hard to estimate. One reason for this difficulty is the large number of parameters, which overall impact has not been evaluated yet. All these parameters or metrics are called Quality of Experience Influence Factors (QoE IFs)[1]. To try to deal with the QoE IFs impact on the user's estimation, we propose this work to present some recent works (frameworks) using the crowd-sourcing approach to study the QoE issue. The main objective of our proposed testbed is to subjectively evaluate the user's QoE using a video application. In our experimentation, the participants use android devices in mobile environments and evaluate the quality using the Mean Opinion Score (MOS). In our approach, we collect and consider the impact of several factors (QoE IFs). This paper is structured as follows : In section 2, we discuss the state of the art of QoE and some recent frameworks using QoE for multimedia services. In section 3, we present our mobile test campaign (conditions, setup, procedure). Then, we present the dataset collected and evaluate the importance of the impact of each QoE IFs on the user perception in section 4. Finally, we conclude our work by giving some perspectives.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.
MSWiM'15 , November 02-06, 2015, Cancun, Mexico
Copyright 2015 ACM. ISBN 978-1-4503-3762-5/15/11 ...$15.00
DOI: http://dx.doi.org/10.1145/2811587.2811631.

2. RELATED WORKS

Lately we hear the word Quality of experience on the lips of many people. Some of them say it is an objective measure, others say it strongly related to the user, and another category sees it is interdisciplinary domain between social science, psychology science, cognitive science and economics science. So what do people expect from the QoE ? and how was it shown in the multimedia area ?.

Quality of Experience (QoE), or simply User Quality is a measure of the experience of a customer with a service (Web browsing, phone call, broadcast TV, ...etc.). QoE presents a multidisciplinary emerging field based on several areas (social psychology, cognitive science, economics and engineering sciences). The QoE concept has become very important in several areas such as multimedia services, the medical field and marketing. The International Telecommunication Union (*ITU*) defined in 2007, the QoE in [7] as a human subjective experience. This experience is represented as the overall acceptability of an application or service, perceived subjectively by the end user.

Over time, QoE became a topic of interest in recent years.

We will try bellow to expose some of QoE works, in order to present their results and contributions.

In [4], authors present Android application which is able to evaluate and analyze the perceived Quality of Experience (QoE) for YouTube service in wireless terminals ($UMTS$ and $WIFI$). The application has been tested over 17 Android terminals in one month. The added value of this tool is informing the user about potential causes that lead to a low Mean Opinion Score (MOS) as well as provides some hints to improve it. After each YouTube session, the users may optionally qualify the session through an online opinion survey. The main finding of this work are : (i) The experience has shown that the theoretical model (taken from the literature) provides slightly more pessimistic results compared to user feedback. (ii) The use of the heuristic measurement quantification proposed in [8] increases the MOS from the opinion survey in about 20% compared to the theoretical model, which was obtained from wired scenarios.

In [10], the authors conduct two experiments and simulating two different usage contexts. Each experiment was conducted as completely randomized. The authors used two kinds of variables : (i) Independent variables : category (static and dynamic), resolution, and frame rate. (ii) Dependent variables : picture quality, continuity, and overall satisfaction. Smartphone owners can watch videos while they are sitting, walking, or standing in various environments. Diverse settings of encoding elements for digital videos were compared in static and dynamic situation. This research shows a lot of results including a low resolution and present enough continuity for dynamic videos in a sitting condition. Low frame rate and resolution can be used to encode a static video if it is shown in a walking context. A dynamic video would deliver a worse quality than a static video in the same condition.

Finally, Hoβfeld et al[6] elaborate QoE management requirements for two complementary network scenarios (wireless mesh Internet access networks vs. global Internet delivery). The authors provide also a QoE model for YouTube taking into account impairments like initial and stalling delay. They present two YouTube QoE monitoring approaches operating on the network and the end user level. Finally, they demonstrate how QoE can be dynamically optimized in both network scenarios with two exemplary concepts, AquareYoum and FoG, respectively. This study shows many results including : (i) The highly non-linear relationship between technical impairment level (QoE IFs) and quality perception. (ii) The stalling has strong QoE impact and should be avoided by all means, e.g. by increasing initial delay to fill the video buffer. Finally, this study lets us understand how QoE management can truly improve the user experience while at the same time increase the efficiency of network resource allocation, and give an exhaustive list of key Influence Factors on YouTube QoE (QoE IFs).

3. MOBILE TEST CAMPAIGN

Generally, the user's Quality of Experience (QoE) for multimedia services is evaluated by using the two methods : subjective method and the objective method. Subjective method is proposed by the International Telecommunication Union (ITU) Rec. $P.800$ and the Video Quality Expert Group (VQEG). It consists of a group of people watching distinct video sequences under a specific controlled environment, and rate their quality. The Mean Opinion Score (MOS) is an example of a subjective measurement method in which users rate the video quality by giving five different point scores from 5 to 1, where 5 is the best and 1 is the worst quality. The second quality evaluation method is the Objective method which uses different models of human expectations and tries to estimate the performance of the video streaming service in an automated manner, without involving humans.

The main objective of our proposed testbed is to evaluate subjectively the user's QoE using an OTT video application, by considering the impact of several factors (QoE IFs). QoE IFs are classified into different categories : network, application, devices, user feedback, etc. The objective of the study is to measure the individual/global impact of each IF category on QoE in order to build a solid correlation QoE IFs/QoE function. The testbed uses various cellular communication networks (HSPA, 3G(UMTS), 4G(LTE)), where the influence of different parameters is examined in the real time environment.

The testbed experiment mainly consists of the following elements :

- A dedicated mobile application has been developed for experimentation.
- The evaluation was performed at different locations.
- Users were trained to perform the test.
- Several types of terminals were used (e.g. smartphones (with different CPU capacities), tablet, TV, HD screen, laptop,... etc.).
- Several types of videos were used (e.g. sport, movie trailer, documentary, news, music,... etc.).

3.1 Testbed overall design

In the experimental setup, users watched videos on different devices such as smartphones, tablets and android TV using different networks HSPA, UMTS and LTE. The experimental setup is shown in Figure 1, where the user uses different devices and networks to watch the desired video contents.

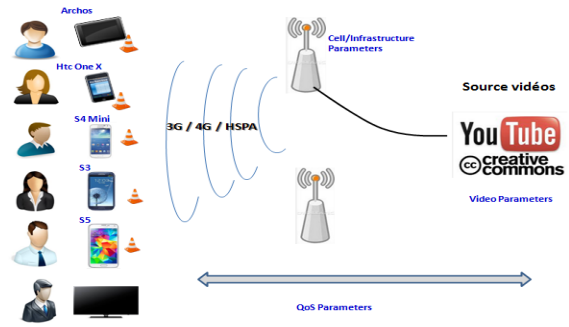

Figure 1: Testbed setup.

Testbed experiment takes place at different locations in an urban cellular environment which is based on crowdsourcing approach. The experimentation was held in the last week of January 2015 in different places in the LiSSi Laboratory and Networks & Telecommunications Department (NTD) (122, Rue Paul Armangot, University of Paris-Est Créteil, 94400, Vitry sur Seine, France). In this testbed, several locations were selected (un-controlled environment). Each one of these locations is characterized by a different RSS (Received Signal Strength). When the video session (a set

of videos seen by each user) end, then client provides its quality experience feedback in term of MOS (Mean Opinion Score), which is stored in the remote database.

3.2 Tested conditions

The key goal for crowdsourcing approach is to had a good exploitable data sample. To achieve this objective, we used and collected several QoE IFs (Table 1 gives an overall tested condition used in our testbed). For the tests we selected 40 different video sequences of 240p and 360p resolution (4 videos for each one of the 10 video categories). The experimentation was made in a totally un-controlled environment, and users give their MOS at the end of each video. In fact, different locations of LiSSI Laboratory and Networks & Telecommunications Department (NTD) were used to study the influence of network coverage in different scenarios.

Parameters	Description
Video	**-10 Types** : News, Cartons,etc. **-Duration** : 60 secondes **-Resolution** : 244p and 360p
Devices	**Samsung 5** : V. 4.2.2, SDK=19 **Samsung 4 Mini** : V. 4.2.2, SDK=19 **Samsung 3** : V. 4.1.2, SDK=16 **HTC X** : V. 4.1.1, SDK=16 **Archos android TV** : V. 4.0.4, SDK=16 **Archos Tablet** : V. 4.1.1, SDK=15
Operator	Orange, Free and Bouygues

Table 1: Overall tested conditions.

We have selected 10 video types' (Figure 2 shows screen shots of some video types used). In each type, we choose 4 videos in YouTube (under Creative Common license).

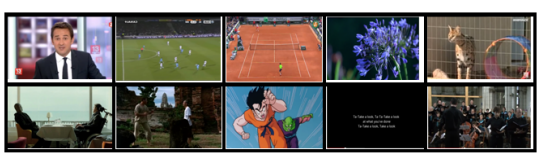

Figure 2: Screenshots of used videos.

3.3 Tested procedure

In this experimental testbed, each user has tested a set of videos (in one session). All members were students or researchers from different disciplines aged from 17 to 40 years old with a little or no experience of this kind of evaluation. In addition, the participants use smart devices (phone/tablet) with the installed application that starts the experimental video session for the current participant. Each video session provides opportunity to the participant for selecting the desired video content type, and provides the feedback of video's quality in terms of MOS. In fact, the different usage scenarios are considered in order to observe the influence of network performance at different locations (the laboratory, the Department of Networks & Telecoms,...etc.). At the end of each video, the participant provides its quality of perception about the video quality, and additionally answers a few questions that are stored in the database. The questions used in the testbed consist of starting video time, the image and audio synchronization, the image quality, the sound quality and the MOS (1 : very bad/ 5 : very good).

These questions allow us to evaluate the user feedback for each video measure. The answer of each question is between 1 and 5 where 1 indicates that the quality is not acceptable or very bad (time to start very long, lag between picture and audio very high,...etc.) and 5 indicates that the quality is very good (time to start is very fast, no lag between picture and audio,...etc.).

3.4 Results

A total of 63 subjects, 45 men and 18 women, participated in the subjective assessment to construct the dataset. All members were researchers or students from different disciplines aged from 17 to 40 years old with little or no experience of this kind of evaluation. All of them were non-experts in assessing the video quality. The experiment is conducted using the different wireless Internet connections, such as 3G, 4G, HSPA. All the subjects spent at least 5 minutes on watching a session, and 18 of them watched at least plus than 10 videos on one session. Therefore, according to the users' answers, it is reasonable to assume that they are familiar with video-watching applications.

The collected dataset contains 646 samples with 33 several parameters divided on different classes : network, application, devices, user feedback, etc. Figure 3 shows the overall distribution of the MOS by devices and by operators.

MOS/Device	Android TV	Tablet	S3	S4 Mini	S5	HTC	Total
1	1	0	1	4	4	39	49
2	4	2	5	11	14	10	46
3	11	16	18	47	30	21	143
4	1	8	22	141	52	49	273
5	0	1	15	59	34	26	135
Total	17	27	61	262	134	145	646

QoE IFs	MF_1	MF_2
FR (Frame rate)	18.0003	28.08163
frame_video_loss	10.79586	19.24571
Bitrate	1008.7092	1152.2351
Mean_Buf	5166.5503	20245.1714
Input_bytes	5352.5415	5680.91859
lost_A	2.2792	6.8665

Figure 3: Distribution of the MOS.

4. ANALYSIS

The objective of this section is to conduct a study on application factors that impact the QoE. We start by studying the correlation between these factors in pairs wise. Then, we compare 2 QoE estimation methods to evaluate the impact of number of the considered factors on the estimation.

4.1 Dataset QoE IFs chosen

A large number of volunteers participating in our testbed, and we gather the impact of many QoE IFs. In this paper, we focus on the parameters that have significant impact on the user's QoE. We consider application layer factors (QoA) which we qualified important because : (i) they proved already they influence the user perception. (ii) They want to predict just user perception with application QoA IFs. (iii) They are collected in a non intrusive way (just in the end-user level).

To identify the relationship between these factors (QoA factors) on YouTube QoE, we study the intensity of the connection which may exist between these factors and the user's MOS. To attempt this end, we use the *rcorr* method in *Hmisc* Package *R*[2] to plot Correlation matrix for the QoE IFs chosen (Figure 4).

The main aim of the correlation matrix is to study the intensity of the connection which may exist between these factors and the MOS.

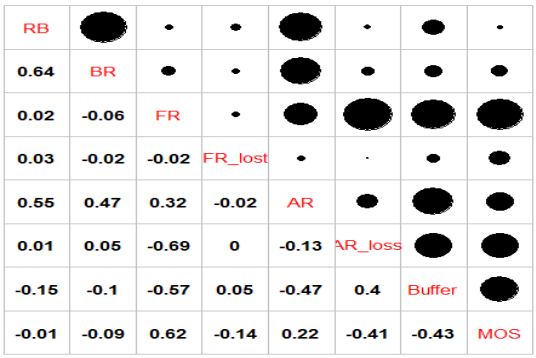

Figure 4: Correlation matrix result.

Table 2 below presents the QoE IFs description.

QoE IF	N	Description
FR (Frame rate)	$f1$	Frames number that are projected or displayed per second.
Buffer	$f2$	The time for one buffering. In fact it's the result of dividing the total buffering duration on the bufferization number.
AR_lost	$f3$	Number of audio bytes that are lost.
AR (Audio rate)	$f4$	Number of audio bytes that are received per second.
FR_loss	$f5$	Number of frames or images that are lost.
BR (Bitrate)	$f6$	Bits number that are conveyed or processed per second.
RB (Read bytes)	$f7$	Total number of video bytes read.

Table 2: Factors description.

The x-axis represents the first value in the compared pair, while the y-axis depicts the second one.

In our crowdsourcing campaign, we focused on quantifying the impact of QoA factors on the QoE perceived on Youtube service using just the factors presented above. In fact,the key influence factors are identified by Pearson's correlation coefficients as described in the correlation matrix (Figure 4). This figure clearly shows that we can categorize the impact of different factors into 3 classes :

- The key factor class : In this class we have the frame rate ($f1$), mean_buffer ($f2$) and loss_audio ($f3$) respectively : 0.62, -0.43 and -0.41 of correlation rate (Pearson correlation). These factors are the most impacting because it was proven by some works in the QoE area that separately they impact on the user perception. For example for the FR, [8] and [9] proved that it plays an essential role in the prediction of the MOS user. Concerning, Mean_buffer, we can cite the work of [12], who measures the QoE of HTTP video streaming.
- The modest factor class : In this class we have : AR Audio byte rate ($f4$), frame lost video($f5$) and input bitrate($f6$) which have respectively 0.22, -0.14 and -0.09 of correlation rate (Pearson correlation). That is can be explained by subjectivity of our dataset. These factors participate to predict QoE, however, in our dataset, their correlation with the user's MOS is limited compared to the first class.
- The no correlation class : In this class we find also the input bytes read. This factor is not on correlation with th user's MOS (r=0.01). This value can be explained by the fact that the read byte number is not very important because it depends on the video compression format (H264, MP4...etc.) and the frame rate[8].

4.2 Methods used

In order to select the best method, we analyze the impact of various parameters on the perceived user's QoE in the mobile video environment. We have designed a comparison of two prediction models (based on the classification), which is implemented on R software [3]. These methods are : SVM Support Vector Machines (SVM)[5] and Adaptive-network-based fuzzy inference system (ANFIS)[11]. For our experimentation in R software [3], we used respectively yhe "$e1071$" package to test the SVM method and the "$FRBS$" package for ANFIS method.

4.3 Experimentation

In our experience, we use 6 variables (several scenarios) as inputs from the dataset, which is described in the section 6.1. Further, we perform 5 test scenarios using the SVM and ANFIS to calculate the RMSE (Root Mean System Error).

$$RMSE = \sqrt{\frac{\sum_1^n (f_i - y_i)^2}{n}} \quad (1)$$

where : f_i is the prediction of MOS, y_i is the true value of the MOS and n is the total number of the considered samples.

In fact, in this experimentation there is 5 scenario types, which are differentiate by the number of QoE IFs, as presented in the next table.

Testbed	QoE IFs
Scenario 1	$\{f1, f2\}$
Scenario 2	$\{f1, f2, f3\}$
Scenario 3	$\{f1, f2, f3, f4\}$
Scenario 4	$\{f1, f2, f3, f4, f5\}$
Scenario 5	$\{f1, f2, f3, f4, f5, f6\}$

Table 3: QoE IFs for each scenario.

In one scenario the procedure works as follows : the data set is divided into 3 sub-samples. The two first parts of dataset are used for training and the third one use for testing. For each method, we use 10-fold cross-validation. In which each fold, the RMSE is calculated. Then, we calculate the scenario Average RMSE.

To run each scenario, we use the R software tool[3]. This tool has produced many outputs, which provide the information about the model prediction using two methods (SVM, ANFIS). Initially, we focus on the RMSE. We compare the error rate between the different models that engender the different QoE prediction model. The results are illustrated in Figure 5. The x-axis represents the different considered scenarios (see the table 3) and the y-axis represents the RMSE prediction performance.

4.3.1 Discussion

In our experimentation, two learning models are trained to calculate the RMSE error rate based on MOS prediction.

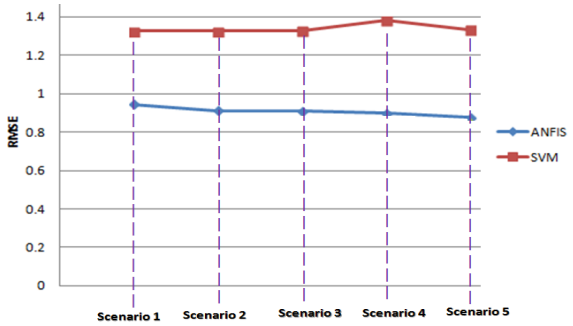

Figure 5: Comparison of RMSE results in of the SVM and ANFIS prediction model.

The ANFIS [11] based model has performed better with an average of 0.85 for the 5 scenarios. For the SVM model [5], it has less prediction performance with an average of 1.35 for the same scenarios (See figure 5).

Both SVM and ANFIS models confirm that using a large number of QoE IFs produces a better user's perception estimation than using a few number. However, due to the interaction between parameters, estimation results may be worse than expected as given in [1]. For example, SVM model performs better using 4 factors ($RMSE = 1.32$) as compared to use 6 factors, which results RMSE equal to 1.385. Concerning the impact of different QoE IFs (QoA factors) and according to the correlation matrix given in section 4.1, we select 3 classes of QoA factors/user's MOS. The first one contains the key factors presented by : Frame rate, time for one buffering and the number of audio bytes lost. The second class contains audio byte rate, frame lost video and input bitrate. Finally, we confirm that the number of bytes is not important because it depends on the video compression format (H264, MP4...etc.) and the frame rate as explained in [8].

5. CONCLUSIONS

In recent years, tremendous growth of video mobile traffic has created a new challenge for network service provider : How to maximize user's Quality of Experience (QoE) in the mobile environment. To deal with this challenge, a subjective evaluation of the user's QoE is required. In this context, we propose a framework based on an android system and Youtube platform to study the impact of several QoE IFs on the perceived quality. The aim of this framework is to build a large dataset for QoE prediction in the dynamic wireless networks (UMTS, HSPA, LTE/LTE-A). Based on the built dataset, we perform an experimentation to study the relationship between QoE IFs and to highlight the most correlated factors. Based on these factors, we compare two estimation methods (SVM and ANFIS) and we highlight the relationship between the number of the considered factors and the estimation accuracy. In fact, our experimentation showed that using a large number of QoE IFs produces a better user's perception estimation than using a small number. However, we must not overlook the interaction between factors. Finally, we find that, ANFIS performance is better than the SVM model. As perspective, this framework can be extended with the introduction of new factors categories such as : QoS (Strength signal, cell load rate,...etc.) or QoA (CPU, battery,...etc.). In addition, we will continue to work on improving our dataset in order to build a bigger and more consistent dataset by performing tests in different cells and including more users with different profiles and other QoE IFs.

6. ACKNOWLEDGMENTS

This work has been funded by the LiSSi laboratory from the UPEC university in the framework of the French cooperative project PoQEMoN, *Pôle de Compétitivité Systematic* (FUI 16).

7. REFERENCES

[1] L. Amour, S. Souihi, S. Hoceini, and A. Mellouk. A hierarchical classification model of qoe influence factors. *13th International Conference on Wired and Wireless Internet Communications*, May 25-27, 2015.

[2] H. F.E. and C. Dupont. Package 'hmisc'. available at:. `http://cran.r-project.org/web/packages/Hmisc/Hmisc.pdf"`, April, 2015.

[3] I. for Statistics and M. of WU. R software. `http://www.r-project.org/`.

[4] G. Gómez, L. Hortiguela, Q. Pérez, J. Lorca, R. García, and M. C. A. Torres. Youtube qoe evaluation tool for android wireless terminals. *CoRR journal*, 1405-4709, 02 Jun, 2014.

[5] S. Gupta, R. Kambli, S. Wagh, and F. Kazi. Support-vector-machine-based proactive cascade prediction in smart grid using probabilistic framework. *Industrial Electronics, IEEE Transactions on*, 62(4):2478–2486, April 2015.

[6] T. Hoßfeld, F. Liers, R. Schatz, B. Staehle, D. Staehle, T. Volkert, and F. Wamser. *Quality of Experience Management for YouTube: Clouds, FoG and the AquareYoum*, volume 35. 2012.

[7] ITU-T. Report 2007. definition of quality of experience (qoe). *International Telecommunication Union, Liaison Statement*, January 2007.

[8] A. Khan, L. Sun, and E. Ifeachor. Content-based video quality prediction for mpeg4 video streaming over wireless networks. *Journal of Multimedia*, 4(4), 2009.

[9] Menkovski, V., Exarchakos, G., Liotta, A., and A. Sánchez. Measuring quality of experience on a commercial mobile tv platform. In *Advances in Multimedia (MMEDIA), 2010 Second International Conferences on*, June 2010.

[10] H.-I. C. Murat Copcu. The quality of contextual experience of multimedia on the smartphone. *International Journal of Emerging Science and Engineering (IJESE)*, Volume-3 Issue-3:30–33, January 2015.

[11] L. S. Riza, C. Bergmeir, F. Herrera, and J. M. Benitez. Learning from data using the r package" frbs". *Fuzzy Systems (FUZZ-IEEE), 2014 IEEE International Conference on*, pages 2149–2155, 2014.

[12] H. T., S. M., M. Hirth, T. Zinner, P. Tran-Gia, and R. Schatz. Quantification of youtube qoe via crowdsourcing. In *Multimedia (ISM), 2011 IEEE International Symposium on*, pages 494–499, Dec 2011.

View publication stats