Universidad de Málaga

Escuela Técnica Superior de Ingeniería de Telecomunicación

TESIS DOCTORAL

Video Quality Assessment in Underwater Acoustic Networks

Autor:

José Miguel Moreno Roldán

Directores:

Javier Poncela González

Pablo Otero Roth

Málaga 2018

UNIVERSIDAD
DE MÁLAGA

AUTOR: José Miguel Moreno Roldán

http://orcid.org/0000-0001-5023-2350

EDITA: Publicaciones y Divulgación Científica. Universidad de Málaga

AUTORIZACIÓN PARA LA LECTURA DE LA TESIS

D. Javier Poncela González y D. Pablo Otero Roth, profesores doctores del Departamento de Ingeniería de Comunicaciones de la Universidad de Málaga,

CERTIFICAN

que D. José Miguel Moreno Roldán, Ingeniero de Telecomunicación, ha realizado en el Departamento de Ingeniería de Comunicaciones de la Universidad de Málaga, bajo su dirección el trabajo de investigación correspondiente a su TESIS DOCTORAL titulada:

"Video Quality Assessment in Underwater Acoustic Networks"

En dicho trabajo, se han propuesto aportaciones en el problema de la evaluación de calidad de vídeo en redes acústicas subacuáticas. Esto ha dado lugar a varias publicaciones científicas, superando el requisito de 1 punto ANECA del programa de doctorado regulado por el Real Decreto 99/2011.

Por todo ello, los directores de la tesis AUTORIZAN su presentación.

Málaga 18 de diciembre de 2017

Los directores:

Fdo: Javier Poncela González

Fdo: Pablo Otero Roth

Abstract

Underwater imagery is increasingly drawing attention from the scientific community since pictures and videos are invaluable tools in the study of the vastly unknown oceanic environment that covers 90% of the biosphere in our planet. However, Underwater Sensor Networks must cope with the harsh channel that sea water constitutes. Medium range communication is only possible with acoustic modems which feature very limited transmission capabilities: peak bitrates of a few dozens of kbps. When transmitting video information, the reduced bitrates force heavy compression, yielding much higher levels of distortion than in other existing video services. Furthermore, underwater video users are ocean researchers or other types of specialist and, therefore, their quality perception is also different from the perception of a general group of users. The peculiarities described call for a dedicated study on video quality assessment for underwater networks.

This doctoral thesis tackles the video quality assessment problem and presents contributions in the two main areas of quality assessment: subjective assessment and objective assessment. The reference for quality perception in any service is human opinion and thus, an analysis of subjective quality is the first step in this work. The experimental design and the results of a test planned according standardized psychometric methods are presented. The subjects involved in the quality assessment test were ocean scientists. Video sequences were recorded in actual exploration expeditions and were processed to simulate the conditions found in Underwater Communications. The presented experimental results show how videos are considered to be useful for scientific purposes even in very low bitrate conditions.

Objective video quality assessment methods are algorithms designed to automatically deliver quality scores. They have become essential in network planning and service operation stages, since subjective experiments are considered expensive and unfeasible for some tasks. This dissertation presents three specialized models for objective VQA, designed to match the special requirements of UWNs and based on machine learning techniques. The first method focuses on simplicity and computes a quality estimation from two application parameters. The second model uses ordinal logistic regression to estimate a full distribution of scores for a video from a combination of application parameters and image metrics. The third method is founded on the perception mechanism of the human visual system and predicts quality from distortion assessment performed through video processing. All the models have been trained with actual user data gathered from subjective tests. The performance analysis shows how the estimated quality presents a very good correlation with human scores and how the proposed methods outperform other existing methods that have not been designed for underwater video.

Content

List of Figures

List of Tables

Acronyms

ACR – Absolute Category Rating

ANOVA – Analysis of Variance

AUV – Autonomous Underwater Vehicle

DCR – Degradation Category Rating

GOB – Good or better

GOF – Goodness of Fit

ITU – International Telecommunication Union

IQA – Image Quality Assessment

FR – Full Reference

GGD – Generalized Gaussian Distribution

LCC – Linear Correlation Coefficient

LSA – Least Squares Approximation

MOS – Mean Opinion Score

NLR – Non-Linear Regression

NR – No Reference

NSS – Natural Scene Statistics

NVS – Natural Video Statistics

OLR – Ordinal Logistic Regression

PC – Pair Comparison

POW – Poor or worse

QoE – Quality of Experience

QoS – Quality of Service

QVGA – Quarter Video Graphics Array

QQVGA – Quarter-QVGA

RMSE – Root Mean Square Error

ROV – Remote Operated Vehicle

RR – Reduced Reference

SI – Spatial Perceptual Information

SROCC – Spearman Rank Order Correlation Coefficient

SSE – Sum of Squares due to Error

SVM – Support Vector Machine

TI – Temporal Perceptual Information

UWSN – Underwater Sensor Network

VQA – Video Quality Assessment

Chapter 1

Introduction

This introductory chapter presents the motivation and goals of the research work in this doctoral thesis. The importance of underwater video services is explained in Section 1.1. A brief introduction to the specificities of underwater communications is given in Section 1.2. The aims of this work are described in Section 1.3. Section 1.4 provides an overview of the contents of this dissertation.

1.1 The necessity of UW video services

Oceans are a driving force in our planet. They influence weather, regulate temperature and ultimately support all living organisms. Humans have been inexorably linked to the ocean throughout history using their waters for transport, commerce and nourishment of both body and soul. They cover almost three quarters of the surface of the earth and yet most of them remain unexplored. It is estimated that only 5% has been seen by human eyes [1].

Underwater imagery is an essential tool in the research and study of the oceans since it provides invaluable information about the mostly unknown contents of the seabed and the water column. However, the gathering of scientific underwater pictures and video footage is currently very expensive. It involves exploration expeditions with vessels in which divers (shallow waters) or robots are submerged for a number of recording sessions.

Engineering underwater works also need live images to be carried out, in places were human life is not possible without artificial means and, in consequence, Remote Operated Vehicles (ROV) are used instead. Underwater robots can do many complicated tasks, they can even tie knots in a rope, but under the remote control of a human operator [2], [3]. Examples of ROV applications are diver observation, platform, pipeline and submarine

cables inspection, drilling and construction support, debris and garbage removal, telecommunications support and object location and recovery [4].

When an electromechanical cable can be deployed that is not a major issue. Nevertheless, in some cases either cable free or wireless systems are required. Typically, a cable free ROV is designated as Autonomous Underwater Vehicle (AUV). Underwater wireless video transmission at practical distances is only possible with acoustic carriers, where bandwidth is scarce.

The costs of ROVs and the staff required to handle them are very high. Oceanographers, have also developed alternative cheaper systems, but they lack the precision of their more expensive counterparts and can lead to poor quality recordings with non-aimed content and a considerable amount of dust. For instance, the group of Marine Geology of the Spanish Institute of Oceanography (IEO-GEMAR) [5] developed its own ROV, that they designated as VOR, equipped with digital single-lens-reflex and video cameras [6].

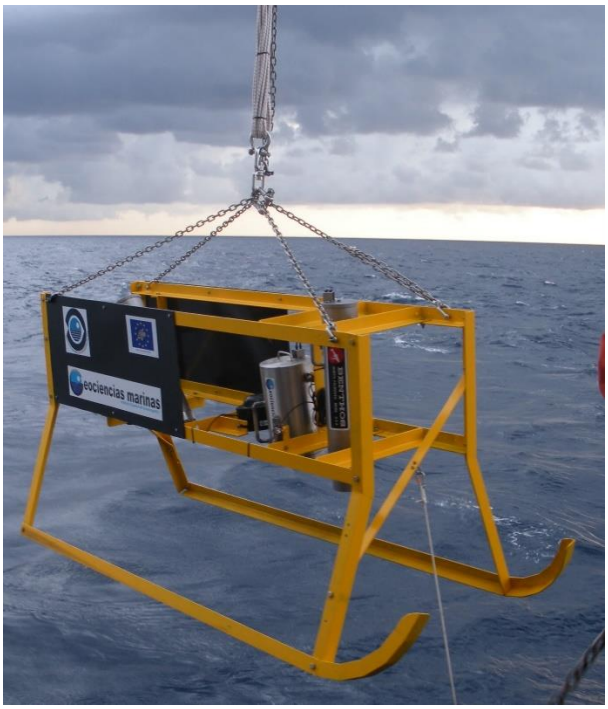

Fig. 1.1. Low cost ROV developed by IEO-GEMAR [6] (with permission of the authors).

ROVs and AUVs are not the only platforms for video recording and transmission. The deployment of wireless sensor networks capable of video capturing and transmission would be a major technological breakthrough allowing for continuous monitoring of underwater environments or cooperative exploration with autonomous underwater vehicles (AUVs). A video service with enough quality could lead to decisions about the re-planning of AUVs path even if instant remote controlling were not possible due to transmission delays.

Beyond the scientific and engineering utilities, other applications in the areas of defense, tourism, fishery and education can also be envisioned for underwater video services, opening a broad field of application for this technology.

1.2 The challenges of UW communications

Wireless communications have experienced a rocketing evolution in the last 50 years. Smartphones are the epitome of this development: an almost ubiquitous device that includes a range of wireless technologies that allow users to connect nearby devices (Bluetooth), local area networks (Wi-Fi), the global telephone network (GPRS/3G/4G) and receive positioning information from satellites (GPS/Galileo). However, these communications take place in the atmosphere or the space, where electromagnetic propagation can be efficiently used as an energy transportation mechanism.

Everything changes when we dive into the underwater environment. Electromagnetic waves suffer from large attenuation and can only be used for very short ranges. Optical communications achieve data rates of a few Megabits per second, but their operating range is limited to some meters and a proper alignment between transmitter and receiver is required [7], [8], [9]. Acoustic waves are the only feasible alternative for mid-range communications (up to a few kilometers) and yet, some negative characteristics of the propagation must be dealt with: the attenuation for high frequencies limits the available bandwidth [10]; the propagation speed is quite low (around 1500 m/s) and dependent on salinity, temperature and depth; multipath propagation occurs in the surface and the seafloor; and colored, non-gaussian noise. State-of-the art acoustic modems offer peak data rates between 32.2 and 64.2 kilobits per second (in ranges from 1 kilometer to 300 meters). These data rates are reduced by network operation (shared access, error correction, protocol overhead) leaving a very low net bitrate to the applications. Due to this constraint, applications have been traditionally restricted to telemetry and other low volume data services.

Beyond the difficulties linked to acoustic communications, there is one additional consideration to be made about the logistics of an underwater network. In most of the networks deployed on earth, a technician can travel to a node location to perform maintenance tasks or to retrieve local information (i.e. perform local monitoring during a test). In underwater networks, the location of the nodes is considered virtually unreachable once they have been deployed. Recovering a node might be possible, but very expensive. The consequences of this inaccessibility are twofold. Firstly, no information from the original signals gathered by the sensor is available (the specific impact of this on video quality assessment will be explained in chapter 2). Secondly, batteries are expensive to be replaced and efficient underwater recharging mechanisms are yet to be developed [11]. Therefore, the node operation should be as energy efficient as possible to extend battery life as long as possible.

1.3 Aims of this work

A natural question arises when the importance of underwater imagery is jointly considered with the strongly limited capabilities of underwater communication. Is it possible to offer a video service within the conditions of underwater acoustic networking?

Or if we tweak the question from an engineering point of view: what is the quality we can expect from a video service offered under the constraints of underwater acoustic networks? Video Quality Assessment has been a topic of interest for the scientific community for decades, from television broadcasting to Youtube playback in smartphones. However, as many other aspects of underwater communications, underwater video service differs from equivalent existing services in other types of networks.

The purpose of this work is to delve deeper into underwater video quality, studying quality assessment techniques, analyzing existing quality data and quality models and their suitability to the underwater scenario and gathering new data, proposing new models whenever the existing research fails to provide a satisfactory result.

1.4 Overview

This dissertation is organized as follow. Chapter 1 describes the motivation and goals of this research work. Chapter 2 serves as an introduction to video quality assessment and other key topics that will help give a better understanding of the contributions presented here. Chapter 3 contains the first contribution of this thesis, the execution of an experimental procedure to determine the quality perception of scientific underwater video and the video encoding parameters that can be used to achieve a quality that is useful for scientific purposes. Chapter 4 focuses on objective quality assessment and deals with the suitability analysis of the only standardized model for this task and proposes two parametric models that outperform the standard as the second contribution of this work. Chapter 5 is also dedicated to objective quality assessment but from an image-analysis point of view. A novel quality prediction model is presented that also outperforms other state-of-the-art models that were not developed considering the specifics of underwater environments. Finally, Chapter 6 contains the conclusions and the suggested future work that could further improve the current technology for video quality assessment in underwater networks.

Chapter 2

Fundamentals of Video Quality Assessment

Quality of Experience studies are a key aspect of the performance evaluation of every telecommunication service that is offered to human users. It focuses on measuring the quality perceived by the user of a given service as a whole. This is essential for providing any service in an efficient and resource-optimized way while maximizing the user satisfaction. As an example, a network link will not offer the best user experience with the highest possible data rate transmission. It depends on the service: users of mobile videogames will prefer a lower response time, video streaming users will be more satisfied with a steadier connection and the interaction with an ATM machine will be limited by other factors once the data rate has reached a certain threshold. Video Quality Assessment has been a topic of interest for network engineers since the beginning of television broadcasting. In this chapter, the two main approaches to VQA are presented in Section 2.1. The peculiarities of VQA for Underwater Acoustic Networks are explained in Section 2.2. Finally, some background on the mathematical tools utilized in this thesis for VQA is provided in Section 2.3.

2.1 Types of Video Quality Assessment

According to existing standards and literature, quality for a network service can be assessed with two different kinds of approaches: subjective methods and objective methods. Both approaches usually aim to produce a standard quality metric known as Mean Opinion Score (MOS) which represents observed (for subjective methods) or estimated (for objective methods) quality in a numeric scale. The standard range is [1, 5] (higher is better) although some other scales can be used for particular applications ([1, 10] interval is used for a higher discriminant capability). Despite its simplicity, MOS has become almost ubiquitous as QoE measure.

2.1.1 Subjective Quality Assessment

The first group of techniques aims to get quality values directly from human evaluators. A group of viewers are presented a sequence of *stimuli* (video signals) which they are asked to score on a quality scale. These scores are statistically processed to compute MOS values for different conditions of service provisioning. The video signals in the test present several degrees or kinds of impairments with respect to the original signal such as blocking, blurring, lack of smoothness, etc. This degradation is produced by different factors that can be grouped in two categories: compression impairments (codec, compression bitrate, framerate…) and transmission impairments (packet loss, burst duration…). The statistical analysis of subjective quality data is able to find relationships between the variation in the factors and the variation in the quality perception.

The International Telecommunication Union (ITU) has standardized several methodologies for subjective quality tests. The process specified in BT.500 [12] recommendation has been used for decades although it is intended for television pictures. A more recent standard for "Subjective video quality assessment methods for multimedia applications" is described in P.910 [13] and has been used as reference recommendation for this work. The procedure presented in P.910 includes recommendations about different aspects of the quality assessment. An overview and brief description of all of them is presented below. The intended purpose is not to reproduce the contents of the recommendation here, but to provide some insight on what it is included within it.

A. **Source signal** recommendations describe how to record, store and select the video signals that will be used for the quality test.
    i. Illumination of the recording environment should be typical for each particular environment.
    ii. Recording system: camera hardware and video format must be reported and ensure a minimum quality.
    iii. Scene characteristics: spatial perceptual information and temporal perceptual information metrics for video sequences are defined. These parameters play a crucial role in the relationship between quality and compression. Therefore, the set of scenes should span the full range of spatial and temporal information of interest to users of the service under test. Also, the number of scenes should be enough to avoid boring the viewers and to achieve a minimum reliability of the results.
B. **Test methods** recommendations describe different procedures of presenting the scenes to the viewers and the way they should score them.
    i. Absolute category rating (ACR). The signals are presented sequentially. A grey screen is shown for a few seconds between every two videos to provide the users some time to score. A five-level categorical scale is used for rating overall quality (a single category is assigned to the whole signal). Categories with their associated numerical value (in brackets) are "bad" (1), "poor" (2), "fair" (3), "good" (4) and "excellent" (5). This method is fast to implement, easy for the users and the presentation is similar to the common use of the video system.

ii. Absolute category rating with hidden reference (ACR-HR). This test method is a variation of the ACR where the original unimpaired video signal for each scene in the test is shown as any other test *stimulus*. The same categorical scale is used, but now a differential score is computed with the equation D(sequence) = Q(sequence) – Q(reference) + 5.

iii. Degradation category rating (DCR). Video signals are presented in pairs (either simultaneously or with a short time lapse between them). The first element of the pair is the original unimpaired scene, the second element of the pair is a degraded version of the scene. Viewers are then asked to rate the degree of impairment in the second element in a categorical scale: "very annoying" (1), "annoying" (2), "slightly annoying" (3), "perceptible, but not annoying" (4), "imperceptible (5)". DCR should provide an accurate evaluation of the fidelity to the original signal, but it is also less intuitive for users.

iv. Pair comparison (PC). Video signals are presented in pairs of different impaired signals. All possible combinations should be used. After each pair, users are asked to choose the preferred element in the context of the test scenario. A PC test is not aimed to obtain quality scores. Instead, it is used to discriminate between factors. A typical example is the comparison of two video codecs which perform similarly regarding their numerical impairment assessment. Even in this situation, viewers could perceive one of them as more pleasant.

C. **Evaluation procedures** recommendations describe several conditions of the testing environment that ensure repeatability and statistical consistency on the results.

i. **Viewing conditions** should be kept constant to some specified values during the tests. These include viewing distance, luminance parameters of the screen where the videos are shown and illumination of the test room.

ii. **Processing and playback systems** can be used in real-time to display the images while they are transmitted through the system producing the impairments. As an alternative procedure, pre-processing of the source signal to create a new set of impaired videos for the test is also suggested.

iii. **Viewers** should not be directly involved in picture quality evaluation as part of their work. Although it is allowed that a small group of experts carry out some preliminary testing before a larger test is performed. Regarding the number of subjects, it is recommended that between 4 and 40 viewers with normal or corrected-to-normal visual acuity should take part in the test.

iv. **Instructions to viewers** must be given prior to the experiment and it is recommended that they undertake a short **training session** with a few videos that will not be part of the results of the test. The signals in the training stage should give an idea of the range and type of impairments. However, the users must be informed that the worst and best quality in the training videos do not have to correspond with the lowest and highest categories in the scale.

D. **Statistical analysis and reporting of results**. A result report for a subjective VQA experiment should include:
   i. The details of the experimental set-up.
   ii. The quality data, along with the method used to assess data consistency.
   iii. An analysis of variance with classical techniques to evaluate the significance of the results.
   iv. Optionally, a cumulative distribution of scores table.

These standardized recommendations have been used as a reference for the subjective quality assessment procedures in the present work. The specific details of the conducted study can be found in chapter 3.

2.1.2 Objective Quality Assessment

Subjective methods have a clear disadvantage in terms of the amount of resources needed. Since a considerable amount of time and a group of people are required, subjective studies are a slow and expensive approach to quality assessment. The second group of methods for quality assessment pursues the estimation of quality values from mathematical models. These models might require a database of subjective quality scores to be built, but then, they are able to compute the quality estimation automatically. Therefore, they solve the main issues of the subjective approach and have received attention from the scientific community in the last decades.

Several classifications can be found for objective quality assessment methods. The ITU group them in three categories according to the way the quality estimation is carried out [14]:

1. Intrusive monitoring. Both transmitted and received signals are required as inputs of the quality model. To capture the signals and compute the estimation, some modifications of the communication system are necessary. This might affect the system performance even if it is only because the users are aware of the monitoring activities.

2. Non-intrusive monitoring: only the received signal is used as input.

3. Network planning: no signal needed, estimation is based on network parameters for the transmission.

Another classification of the ITU standardization activities in quality assessment [15] is shown in Table 2.1. This categorization is based on the input information required for the model and the primary application of the assessment.

Table 2.1. Classification of the standardization activities of the ITU on Quality Assessment

Input information	Media signal	Packet header information	Quality design parameters	Packet header and pay-load information	Combination of any
Primary application	Quality bench-marking	In-service nonintrusive monitoring	Network planning, terminal/application designing	In-service nonintrusive monitoring	In-service nonintrusive monitoring
	Media-layer model	**Parametric packet-layer model**	**Parametric planning model**	**Bitstream layer model**	**Hybrid model**

The third and last classification presented here is based on the presence of a "reference signal", the unimpaired original signal. It is frequently found in the literature (see section 2.3) since the availability of this information usually defines the approach of the assessment algorithm.

1. Full Reference (FR) methods. The complete original signal is required to perform the estimation. The quality score is computed comparing the impaired signal with the original. A basic FR method uses the peak signal-to-noise ratio (PSNR) in equation 2.1 of the video frames to estimate quality.

$$PSNR = 10\log_{10}\left(\frac{MAX_I^2}{MSE}\right) \tag{2.1}$$

where MAX_I is the maximum possible pixel value of the image and MSE for a pair of images is defined in equation 2.2

$$MSE = \frac{1}{MN}\sum_{i=0}^{M-1}\sum_{j=0}^{N-1}[I_o(i,j) - I_i(i,j)]^2 \tag{2.2}$$

where M, N are the width and height of the images, I_o is the original image and I_i is the impaired version of the image.

The PSNR is fast and easy to compute, but it is a too simplistic way to assess distortion. Two pairs of pictures with a similar PSNR could have a completely different MOS since different distortions are perceived differently by the human visual system. A more advanced method is the Perceptual Evaluation of Video Quality (PEVQ), which was standardized on the ITU-T Recommendation J.247 [16]. The Structural Similarity Index (SSIM) proposed in [17] has been proved to outperform PEVQ and is also a reference.

2. Reduced Reference (RR) methods. Some information extracted from the original video is used for the quality evaluation along with the impaired signal. The ITU-T Recommendation J.249 "Perceptual video quality measurement techniques for digital cable television in the presence of a reduced reference" [18] describes several RR methods.

3. No Reference (NR) methods. No information from the original signal is required in these methods. Usually the impaired signal is utilized for the quality score computation, but network planning methods, which use networks parameters can also be considered part of this category.

Regardless of the convenience of objective VQA methods, subjective data is considered the ultimate reference in quality assessment and performance of any method is normally checked against human scores. Typical performance metrics of objective methods are based on how well the automatically computed scores correlate with subjective scores.

2.2 Challenges in Underwater VQA

As commented in the introductory chapter, underwater networks constitute a particularly challenging environment for video transmissions. This statement can be extended to video quality assessment and some specific considerations should be made.

2.2.1 Considerations for underwater subjective VQA

Video services in UWSNs are meant to be used by a very specific public: oceanographic scientists, operators of companies managing oceanic resources, safety and security specialist among others. Not only it is more difficult to find an appropriate group of evaluators, but also each of these professionals have a different perception of quality depending on the tasks they usually perform with video information. As an example, a research group could be satisfied with an extremely low framerate video from a network with stationary nodes because it allows them to estimate the flow of individuals from a species under study. On the contrary, a service with the same parameters could be completely useless as visual feedback from an autonomous vehicle used for exploration.

To circumvent some of the obstacles in subjective VQA, the authors in [19] propose crowdsourcing as an alternative technique for QoE assessment. Crowdsourcing uses the Internet as a "virtual laboratory" providing access to a larger pool of users with a larger diversity and a faster turnover of test campaigns, thus reducing costs. Nevertheless, there are also strong limitations on the testing scenarios since it is not possible to control any of the environmental variables detailed in the ITU recommendation (viewing distance, room illumination, display size and brightness). These variables can widely change for one user to the next in the most common video services. YouTube clips are watched in a smartphone under direct sunlight, but also in a 50'' TV panel in a living-room. Additionally, it is not easy to fully monitor internet connections for participants and consequently it would not be possible to separate the effects of different variables under study, e.g. if a low score was produced by a poor-quality video coding or a burst in the

packet loss due to a problem in the user's Wi-Fi connection. All these problems make it arguable if test results can be compared and question the validity of the statistical analysis of results. However, users of underwater video applications would generally work in environments with very similar conditions: a desktop computer in a research facility. An on-board laboratory in a research vessel can be also considered as an alternate scenario, but, again, all of them will be similar. In this low variation context, an uncontrolled open test could deliver better results without some of the disadvantages of crowdsourcing.

2.2.2 Considerations for underwater objective VQA

Objective VQA was proposed as a solution to the hindrances of subjective testing. Some user produced quality data can be needed to build a model, but once it has been verified, MOS values can be estimated without expensive subjective experiments. Other QoE prediction models do not even require an initial dataset of subjective data, since they are based on modelling the human visual system (HVS) or the perception of image impairments. This kind of assessment in UWSNs must, yet, consider the following aspects:

- Network location. Due to the nature of the underwater environment, the nodes are virtually unreachable once they have been deployed. A battery replacement could be scheduled on the long-term, but recovering a node is an expensive task and is not usually planned for any other purpose.

- Energy saving. Since node retrieval should be delayed for as long as possible, network design must keep energy consumption at a low level. Intensive processing task or additional transmissions for quality measuring should be limited.

- Terminal vs node assessment. An efficient objective VQA method could be used as built-in network intelligence for real-time optimization of video services. A complete VQA can be only done in the user end of the communication since that is the only point where the signal has gone through both encoding and transmission impairments. However, decisions on adjustment of video or network parameters must be forwarded to the nodes consuming a portion of the limited bitrate and suffering a non-negligible transmission delay. A partial assessment for encoding impairments can be done on every node. In this case a low energy technique should be used to avoid a waste of the scarce node energy.

These considerations demand a detailed analysis of the classes of objective VQA (see Section 2.1), their suitability and their advantages and disadvantages.

A. Full Reference assessment

FR methods have been extensively used in VQA. However, FR methods require the original signal and UWSNs bitrates make unfeasible the transmission of the unimpaired video. Storage and deferred retrieval could be an option for sensor networks deployed on the surface but underwater nodes are meant to be recovered only in exceptional situations if ever. Additionally, these methods usually involve heavy image processing calculations which would be expensive in terms of energy for underwater nodes. The usefulness of these methods is hence reduced to laboratory tests.

B. Reduced Reference assessment

RR methods only require some features of the source signal to perform the quality estimation. This could partially solve the drawback of FR methods since the amount of information to be transmitted along with the video signal is greatly reduced. Still, the binary rate for the extra quality assessment data must be kept a small fraction of the video bitrate or it will burden video quality. The methods described in ITU J.249 [18] are focused on TV services in the order of Mbits/s and the imposed overhead ranges from 15 to 256 kbits/s, which make it unsuitable for the application bitrates available in UWSNs. Recent research in [20] proposed a method with a 0.875 kbits/s bandwidth for image features which stands for about 10% of an 8 kbits/s video flow. Although this bitrate could be reduced for lower fps video, RR feature extraction usually involve some intensive image processing which might be energy-unaffordable for an underwater node.

C. No Reference assessment

NR methods compute the quality estimation without other information than the received signal. This can be done with a pixel based analysis (evaluating image impairments), a bit-stream analysis (a study of encoding parameters without actually decoding the video) or a network parameter analysis (accounting for statistics such as the packet loss rate or the burst loss length). For VQA performed in the terminal, it seems an efficient method, since no extra node energy must be used and additional network resources are limited to forwarding adaptation information to underwater nodes. Good performing NR methods can be found in the literature [21], [22], although none of them has been tested with underwater video.

NR techniques also present a valuable alternative for local on-node assessment. Specifically, methods based on network parameters would save image processing and therefore a waste of energy for VQA purpose.

2.3 Statistical tools for VQA

As in every aspect of engineering, Mathematics play a crucial role in Video Quality Assessment. Most of the tools used for this thesis belong to the field of statistics and probability. Some of them can be frequently found in engineering (probability distributions, distribution fitting) while others are usually associated with life sciences or psychology (statistical tests and classical analysis of variance) where experiments with subjects are an essential part of research in these areas. Machine learning is also a branch of statistics that has become popular and of wide application in many different disciplines. This section aims to provide a brief description on three different topics. Although they could seem unrelated to each other, they constitute key concepts in the mathematical tools for video quality assessment.

2.3.1 Statistical testing and analysis of variance

Subjective quality assessment has already been described as the process of obtaining quality scores from human users. Unlike other aspects of engineering, video viewers

themselves cannot be described with equations or simulated. We can, of course, build models for user behavior, but then we are moving into the area of objective assessment. While we stay within the subjective study we must use statistical inference to design experiments and analyze the results. The purpose of these tools is drawing conclusions on experimental data. As an example, we could imagine the simple case a single video content is shown to two different groups of people changing the compression codec, asking them to rate the quality from 1 to 10. The results of this experiment would most likely be different for each group. There are several reasons for these differences that are usually broken down into main groups [23]:

- Measurement variability. It is caused by the measuring mechanism or instrumentation.
- Environmental variability. It is introduced by changes in the "external conditions" of the experiment. In our example, we could have shown the video clips in different lightning conditions, such as direct sunlight and living-room illumination.
- Treatment application variability. It is caused by the changes in a variable or factor the experimental design is controlling. In our example, the compression codec is changed so the two codecs would be our treatments.
- Subject-to-subject variability. It is caused by the fact that every subject will have a different quality perception and this, will issue a different quality score.

Statistical analysis considers what types of results we would get if specific conditions are met and if we were to repeat an experiment many times, and then to compare the observed result to these hypothetical results and characterize how typical the observed result is [23]. In our example experiment, we could minimize measurement and environmental variability as much as possible (using reliable measuring instruments and identical environments) and assume certain conditions so it can be stated whether the differences between the quality scores are caused by typical subject variation or by the change in the treatment (video codec).

The usual steps for statistical analysis are [23]:

1. Choose a model that is a reasonable match for the data from the experiment. The model is expressed in terms of the population from which the subjects and the outcome variable were drawn. Define parameters of interest.
2. Using the parameters, define a null hypothesis and an alternative hypothesis which correspond to a question of interest.
3. Choose a statistic which has different distribution under the null and alternative hypothesis.
4. Calculate the null sampling distribution of the statistic.
5. Compare the observed statistic to the null sampling distribution of the statistic to calculate a p-value for a specific null-hypothesis.
6. Perform assumption checks to validate the degree of appropriateness of the model assumptions.
7. Use expert judgment to interpret the statistical inference in terms of the underlying science.

The hypotheses are statements about the population parameters that express different characterizations of the population which correspond to different scientific hypotheses. In a two-treatment-group case (like our simple example), the usual null hypothesis is that the two population means are equal (average perceived quality for the population is the same) and the usual alternative hypothesis is that means are unequal (average perceived quality for the population is not the same). Once we have chosen a statistic and computed its null sampling distribution we can evaluate if its value is "typical", i.e. if there is a high probability we find this value and, thus, we do not have grounds to reject the null hypothesis. Formally, a p-value is the probability that any given experiment will produce a value of the chosen statistic equal to the observed value in our actual experiment or something more extreme (in the sense of less compatible with the null hypothesis), when the null hypothesis is true and the model assumptions are correct. The usual convention is to reject the null hypothesis if the p-value is less than or equal to 0.05 and retain it otherwise. This cutoff value (usually denoted by α) is called the significance level of a test.

The analysis of variance or ANOVA [24] is a statistical inference procedure that defines all the mentioned steps. Different versions of the tool are available with different models and assumptions. The version used in this work is the multiway within-subjects ANOVA. It is used to check if there are significant differences in the mean across several factors with several levels on each factor using the same group of subjects. The statistic used by ANOVA is the F-statistic [24].

2.3.2 Machine Learning

An abstract definition of machine learning [25] says that it is "the programming of a digital computer to behave in a way which, if done by human beings or animals, would be described as involving the process of learning". In general, we could say that the behavior of an algorithm does not only depend on some "rules" but also on a set of "examples". There is a close relationship between machine learning and statistics. In fact, we could say that a simple linear regression algorithm is a form of machine learning: some rules for computing the model parameters are given, but then, the algorithm will "learn" to draw a line from a dataset.

There are two features that make machine learning techniques a powerful tool to tackle many different problems. The first one is the offloading of detailed programming instructions. In image processing, it is easy to write a simple computer program to count white pixels from a black and white thresholded image if it is known that they determine the size of a given object. However, it is difficult even to imagine the instructions that will find out if there is a face or not in a grayscale image. Machine learning is an alternative for those problems, since it is easy to feed an appropriate algorithm with a set of positive (images with a face) and negative (images without a face) examples. The second important feature of these techniques is the possibility of retraining. Once an algorithm has been trained, it is possible to add examples to the data set and improve the results. Moreover, once an algorithm has been proven successful for an application, it can be retrained with a different dataset for a similar application with good chances of also providing a satisfactory solution that can be later adjusted.

A frequent classification of machine learning techniques refers to the learning approach and the example dataset:

- Supervised learning. The elements of the input dataset are associated with an expected output. This is the case of the independent variable value for dependent variable point in linear regression or the labels "yes", "no" in the aforementioned example of identifying if a picture has a face on it or not.
- Unsupervised learning. The elements of the dataset have no information about the expected output. The algorithm should be able to find patterns from the input data.
- Reinforcement learning. The system learns from being exposed to the data and given "rewards" when the output behavior is the right one (or "punishments" otherwise). An algorithm that can learn to play a game by playing and receiving feedback on its performance (victories, good and bad moves, etc.).

There are also several broad categories for machine learning algorithms according to the purpose and outputs of the system:

- Regression. The system computes the value of a continuous magnitude.
- Classification. The system assigns a category (from a finite set) to the inputs.
- Clustering. The system divides the inputs into groups. The difference with classification is that the groups are not known in advance.
- Dimensionality reduction. The inputs are mapped into a space with a lower dimensionality.

Machine learning models are commonly called predictive models, particularly for certain applications. Although the word "prediction" could sound esoteric, in this context it is just used as a synonym for "estimation". In the field of video quality assessment, we will say that a model is able to predict quality in the sense that is able to estimate how human users would score the video quality. The models used in this work are supervised regression models since they will be trained with some videos of known quality (supervised learning) and they will compute the quality values (or the quality distribution) as continuous magnitudes.

2.3.3 Natural Scene Statistics and the Human Visual System

A very powerful analogy between the visual quality assessment problem and a communication system is proposed in [26]. Our visual world is described as a "transmitter". The physical properties of matter and light generate a signal that can be captured by sensors, digitalized, processed, stored or transmitted and displayed in screens. This system introduces distortion and is regarded as the communication "channel" in this analogy. Finally, the "receiver" is the human visual system that forms the perceptual image signal in the human brain. As in any other communication system, the modelling of the different elements helps us understand how the system works.

Natural Scene Statistics (NSS) is a theory about the transmitter model. It states that images originated from the capture of our world (natural scenes) exhibit statistical regularities. These regularities are not present in other images that do not resemble our world. Computer-generated images generally lack these regularities, but also natural

images that have suffered distortions caused by the "channel" lose these properties. A useful NSS model [27], assumes that if the lowest spatial frequencies are removed from a natural image, the pixel values follow a Gaussian Scale Mixture distribution. The importance of this model is better understood when contrasted to the findings on how the human visual system (HVS) works. Some studies [28], [29] show that the architecture of neurons involved in early visual processing is generally regarded as having evolved to efficiently encode and analyze images from the real world, i.e., images that exhibit the statistical properties proposed by NSS models.

Integrating the proposals of these theories, we could define quality as fidelity to the real world and thus, identify good quality with images that match NSS properties. Subsequently, the perception of bad quality (distortion) is linked to the departure of the statistical regularity that natural images present. This concept is the foundation of the objective quality assessment method presented on chapter 5.

Chapter 3

Subjective Quality Assessment for Underwater Video

As explained in chapter 2, Subjective VQA is based on experiments with human viewers. The goal is obtaining data about the perceived quality by directly asking participants to rate video clips. Further statistical processing is used to gain insightful information. This chapter presents a literature review for the topic in Section 3.1. Section 3.2 thoroughly describes the experiment performed for this research work. Section 3.3 contains the results of the experiment, the statistical analysis and the discussion of them.

3.1 Literature review

Multimedia data acquisition is currently difficult in wireless underwater networks due to the low data rates available. Proposals in existing literature try to circumvent this obstacle with different solutions. In [30], pictures captured by sensor nodes are gathered by an autonomous underwater vehicle (AUV). This AUV travels to the position of each node and downloads the information through an optical link. This method suffers from delays due to the time required to complete a round trip through all deployed sensors. [31] suggests a different setting with underwater sensors wired to a buoy equipped with a communication unit transmitting over the air with an 802.11b modem. This kind of dual node requires heavy anchoring, is only suitable for shallow waters and is more vulnerable to potential damages. Other existing studies include image quality considerations. [32] proposes three classes of quality of service (QoS) to optimize the network performance and acknowledges that mechanisms to meet application level QoS requirements and to map them into network-layer metrics have not been primary concerns in mainstream research on UWSN. Although the authors make a significant contribution with their cross-layer protocol, assessing the quality remains an unaddressed task. [33] does obtain a quality measure of the studied service but it only considers still images and the quality

is assessed through computing the peak signal to noise ratio on standard test images (not even related to the underwater context).

The first step of this thesis work is motivated by the need to have subjective QoS data for video services under the constraints of the underwater medium. The importance of subjective quality assessment is supported by similar studies already performed such as [34] for general purpose video, [35] for mobile video or [36] for HTTP based streaming. As a result of these three papers, a remarkable database with subjective quality information is publicly available [37]. However, all of these works use videos with much higher bitrates than those achievable in UWSNs. Some previous works highlight the effect of network parameters such as the packet loss rate in [38]. Other works highlight environmental viewing conditions as the relation between the viewing distance and the image resolution [39]. Studies on how a specific service and their users can affect the quality assessment process can also be found in [40] for telemedicine multimedia applications. In the latter, the conclusions emphasize the differences in quality perception when the video application is being used by a medical expert and how this is highly dependent on the specialty area.

Other papers on this topic focus on general parametric models for opinion-based quality estimation. A thorough compilation and comparison can be found in [41] along with the authors' own proposal. The reference model is [42], which is part of the International Telecommunication Union (ITU) standard "G.1070 Opinion model for video-telephony applications" [14]. Although targeted for a very particular application, it has been used for other multimedia services due to the lack of a more appropriate standard for video quality assessment. Other models cited in [41] differ from Yamagishi's proposal in the video and network parameters that they take as input variables (e.g., bitrate, frame rate, packet loss, video content). Conclusions of [41] show that the best estimations are computed with their model and the G.1070 model, depending on the video impairment type, but the performance analysis is made with bitrates that cannot be attained in UWSNs. Nevertheless, all the analyzed models share the main limitation that a set of coefficients derived from subjective data is needed to complete the model equations. These coefficients are linked to certain input variable ranges and some additional settings like video size, resolution or compression codec. The models only provide sets of coefficients for a small group of settings. Outside these configurations they cannot be used unless new coefficients are computed. For example, three out of five sets of coefficients included in G.1070 require video bitrates over 256 kbps, a figure far from the achievable rate with current acoustic technology. The above-mentioned drawbacks of parametric models make subjective quality assessment essential in the feasibility analysis of multimedia services for UWSNs.

3.2 Experimental UW VQA

Subjective quality assessment in video services requires a careful planning accounting for all the different aspects of the experiment. The specific features of the target service in terms of video configuration parameters are detailed in Section 3.2.1. The following

experimental set-up information, as described in P.910 recommendation, is provided: details about the recording environment and equipment producing the source signal (Section 3.2.2); scene creation and selection criteria (Section 3.2.3); psychometric method used in the quality test (Section 3.2.4); equipment and settings used for performing this test (Section 3.2.5); full description of the viewing conditions (Section 3.2.6).

3.2.1 Target service

Video services studied here could be provided in wireless sensor networks with anchored nodes for monitoring underwater ecosystems or with autonomous vehicles for visual exploration of seafloors. Current underwater networks are in an early stage of development and state of the art acoustic modems reach a peak data rate of 62.5 kbps with a 300 m operating range [43]. Bitrates available in the application layer are highly limited and the quality is seriously burdened by this constraint. Nevertheless, underwater video as considered in this thesis is not a service to be provided to a great number of heterogeneous users like other video services such as IPTV, videoconferencing or video-sharing streaming. Instead, it is considered as a tool for a very specific public and a particular use: scientific exploration and monitoring of areas that are otherwise very difficult and expensive to reach. We also need to consider the specific features of state-of-the-art differential video encoders which leverage intra-frame and inter-frame similarities. Because of this, a direct relation between parameters cannot be found, i.e., a 10 fps video is not twice the size of the 5 fps equivalent. Taking this into account, a sensible choice of video parameters (the quality of which will be assessed in the experiment) has been based on a previous study [44]. In this preliminary analysis, a simplified quality test was conducted among a reduced group of viewers (not related to ocean science) leading to the following choice for the current experiment:

Bitrates: 8, 14, 20 kbps.
Frame rates: 1, 5, 10 fps.
Resolution: 320 x 240, 160 x 120 pixels.
Color depth: RGB (3 x 8 bits) and Grayscale (8 bits) video.

In this previous work, the bitrate was found to be the main limiting factor. Below 8 kbps, video contents were difficult to distinguish, even in low resolutions. An effective data rate higher than 20 kbps is unrealistic due to protocol overheads and competition among several nodes. In these bitrate conditions, a standard 25 fps rate produced fuzzy images. For equivalent conditions, viewers preferred videos with lower smoothness, but enhanced frame sharpness. This fact suggests that, below a bitrate threshold, lower frame rates offer better quality. However, for two MOS vs bitrate curves plotted for different frame rates, the bitrate value for the crossing point depends on the particular frame rates being compared. Because of this, the selection of frame rates for the current experiment ranges from 1 to 10 fps. Concerning image size, only low resolutions (QVGA, QQVGA) are suitable for these conditions since higher resolution pictures would appear hazy and suffer heavily from distortion artifacts. Finally, the impact of color/grayscale streams in quality will also be studied.

3.2.2 Recording environment and source signal

Video sources for this experiment have been provided by the Spanish Institute of Oceanography (Instituto Español de Oceanografía, IEO) from real exploration expeditions in the scope of the project "Life+Indemares-Chimeneas de Cádiz" [45]. Images were captured with the underwater vehicle VOR APHIA 2012, a prototype developed by the GEMAR research group (IEO). It includes its own illumination system consisting of two high power LED spotlights of 19,000 lumens in a 60-deg. angle and a Canon Legria HF R106 as camera/recording system. This camcorder features a 1/5.5 type CMOS sensor and AVCHD as video encoding format with 4:2:0 color sampling scheme. Recording settings are 1440 x 1080 pixels resolution, 25 frames per second, automatic white-balance and automatic focus. AVCHD is a compressed format using H.264. The quality provided is considered to be high enough so that there are no visible compression artifacts, particularly considering that the test sequences will be downscaled by a 4.5 linear factor.

3.2.3 Scene selection

Image contents range from almost static shoots of plain sea bottom to fast navigation of areas with complex layouts of irregular rocks and different kinds of underwater flora and fauna. As a representative sample of these conditions, a number of 56 scenes with a 12-s duration were chosen to be used as potential test sequences. All the video manipulations needed for the test have been made with Avidemux 2.6 software. This includes clipping, resampling, downscaling and re-encoding. H.264 differential compression produces a variable bitrate video. The average bitrate of a time window within the scene does not have to match the average bitrate for the whole scene. The bitrate for a set of frames depends on the previous frames, their bitrate and the adjusting bitrate algorithm. Every 12-s sequence is encoded at a random starting position within a longer 120-s clip. This way, we randomize the unwanted effects due to the variable bitrate coding. The 120-s segments have been processed to generate test sequences with all possible combinations of parameters mentioned in Section 3.2.1. Two important attributes for scenes are the spatial and temporal perceptual information because high variation scenes will be further impaired when encoded. To measure the characteristics of the scenes in these two dimensions we have generated sequences with the largest evaluation resolution (320 × 240 pixels) and a very high bitrate (3000 kbps). This bitrate guarantees that there are no impairments in the clips aside from those due to the reduced resolution. The metrics for spatial perceptual information (SI) and temporal perceptual information (TI) are specified in (3.1) and (3.2) [13]. These metrics, as defined in the recommendation, are only applied to the luminance plane of the images (F_n is the n-th frame represented as a pixel matrix containing this luminance information). The Sobel operator in (3.1) is a convolutional kernel operator used for edge detection [46], the result of applying this operator over a frame is also a pixel matrix. The std_{space} operator computes standard deviation of luminance values within a single pixel matrix. The max_{time} operator selects the maximum value of the argument (spatial standard deviation for a pixel matrix in both cases) over the set of all processed video frames in the clip:

$$\mathrm{SI} = \max_{time}\{\mathrm{std}_{space}[\mathrm{Sobel}(F_n)]\} \tag{3.1}$$

$$\mathrm{TI} = \max_{time}\{\mathrm{std}_{space}[M_n(i,j)]\}, \text{ with } M_n(i,j) = F_n(i,j) - F_{n-1}(i,j) \tag{3.2}$$

Figure 3.2 shows the (SI, TI) plane for all sequences. It can be observed that no samples but one has a high value for one dimension and a low value for the other dimension. The Pearson correlation coefficient for both dimensions is ρ = 0.8416, which shows a substantial linear dependence. Scenes have been divided in two groups according to their content variation features. The threshold is the median in the variable with the largest scattering, SI.

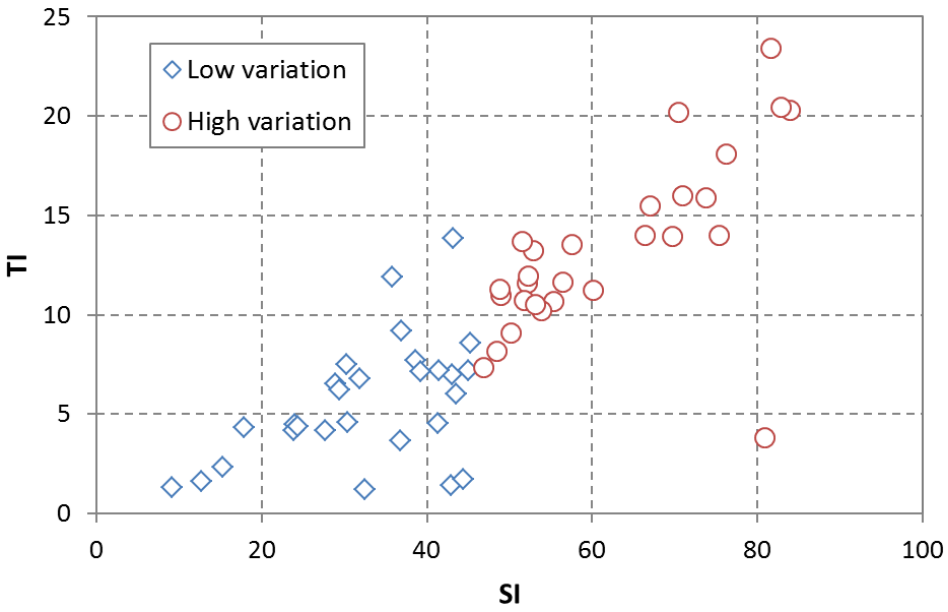

Figure 3.2. Scatter diagram for Spatial and Perceptual Information in test scenes.

Scenes in the same group are considered equivalent in terms of content variability. Instead of choosing a few representative scenes, repeated with every configuration, each evaluation condition is applied to a different video content. Thereby we avoid two important side effects observed in the previous study [44]:

1. Learning effect: viewers can recognize objects from a low-quality scene if they have seen them previously in a better quality. Thus, their opinion could be biased.
2. Boredom effect: even in short sessions (less than fifteen minutes) volunteer viewers get bored if the same contents are displayed repeatedly. This may cause a loss of interest and introduce unwanted factors in the test.

Initially, contents from the 56-sample scene database were intended to be randomly assigned to each viewing condition according to content variation required (see Section 3.2.6). However, a close inspection of the clip file sizes showed that the average bitrate defined in the compression settings was sometimes ignored. Compressed scenes with a

deviation higher than ±10% of the target size were removed from the database and only the remaining clips were employed in the random assignment.

3.2.4 Test method

The absolute category rating (ACR) method described in P.910 has been used to assess video quality. It employs a standard five-level scale: bad, poor, fair, good and excellent. This method requires a short explanation time and the single stimulus presentation is the most similar one to the typical use of the video sequences.

Some viewers can feel unsure on how to use the scale and change their evaluation criteria after some scenes. Taking this into account, some "dummy" scenes have been introduced at the beginning of the test to stabilize viewers' ratings. Figure 3.3 shows the time pattern for the presentation: 12 s for scene visualization and 8 s for voting.

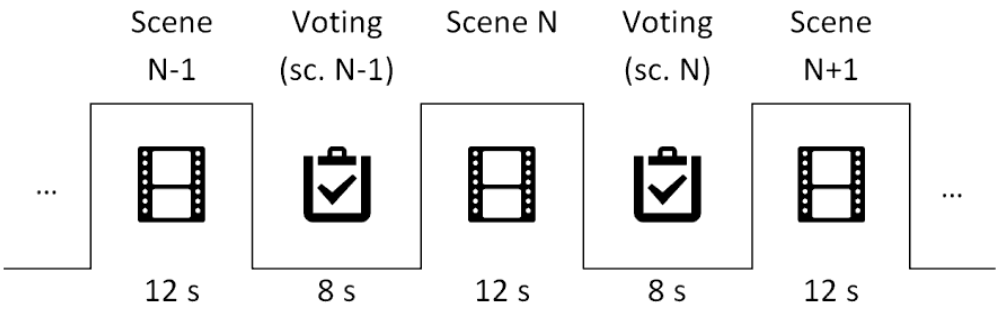

Figure 3.3. Scene-voting sequence time pattern.

Once the viewers had ended the scene scoring, they were asked to rate the scientific utility of the quality categories they had just used in the test. They were provided a five rank (ACR-like) scale with the following values: useless, barely useful, moderately useful, quite useful and very useful. This question was designed to go beyond the perceived quality as an abstract concept, linking it to another magnitude, also subjective but with a more specific meaning. After delivering the test form viewers went through a short interview in which they were asked to give a qualitative opinion about the images they had just seen.

3.2.5 Evaluation procedures

The playback system was an HTML5 application developed specifically for this purpose. The application starts with an instruction screen with a start button. Once the button is pressed, clips are presented in sequence (see Figure 3.3) and no further interaction is required from the user. Viewers were informed about the test procedures and given a paper form to write down the score for each scene. The application was displayed in a 14'' screen configured with WXGA (1366 × 768 pixels) resolution. Scenes were centered in the screen with visualization size 320 × 240 pixels (diagonal length 3.57'') for both resolutions, with the background color set to 50% grey.

Illumination conditions were measured using a photometer (Sekonik L-758DR [47]) for both the screen and the room where the test was conducted. Table 3.1 collects the illumination requirements given in recommendation P.910 alongside with the measured values in the test. Chromaticity was not measured since required D65 illuminant corresponds with daylight and the only source of light in the room was natural light shaded by adjustable panels. Illumination conditions were kept constant during the entire test. The viewing distance was approximately 50 cm or 8 H using the picture height as reference (as defined in P.910). This matches the conditions in which the images would be used in a real service (an application on a lab computer). A minimum of four viewers are required for statistical processing, although P.910 suggests at least 15 viewers should participate in the experiment. They should not be directly involved in picture quality evaluation. A total of 21 viewers took part in the test, all of them ocean scientists, geologists and biologists, from the Oceanographic Málaga Center of the IEO. The group features a wide variety of research interests such as sedimentology, submarine morphology, plankton, taxonomy of small species, benthos, and fishery. Some of them were acquainted with the use of images in their everyday work although none of them had been involved in video quality assessment before. Therefore, this sample of viewers met the requirements.

Table 3.1. Illumination conditions.

Parameter	**Requirement**	**Measured Value**
Peak luminance of the screen.	100–200 cd/m²	111.4 cd/m²
Ratio of luminance of inactive screen to peak luminance.	≤0.05	0.001
Ratio of luminance of the screen when displaying only black level in a completely dark room to that corresponding peak white.	≤0.1	0.004
Ratio of luminance of background behind picture monitor to peak luminance of picture.	≤0.2	0.006
Background room illumination	≤20 lux	2.5 lux

3.2.6 Test conditions

Each evaluation condition is a combination of test variables which characterizes a scene to be scored by viewers. The test consisted of 31 evaluation conditions arranged in five blocks as follows:

- **Block 1**: dummy conditions for score stabilization (see Section 3.2.4). The opinions issued for these conditions were discarded and not included in the test results.

- **Block 2**: conditions for measuring the impact of the three levels of bitrate and frame rate with low variation content. Resolution and color settings for this block are QVGA and RGB.
- **Block 3**: conditions for measuring the impact of the two levels of resolution and color with low variation content. Bitrate and frame rate settings for this block are 20 kbps and 5 fps.
- **Block 4**: conditions for measuring the impact of the three levels of bitrate and frame rate with high variation content. Resolution and color settings for this block are QVGA and RGB.
- **Block 5**: conditions for measuring the impact of the two levels of resolution and color with high variation content. Bitrate and frame rate settings for this block are 20 kbps, 5 fps.

In this block set up, only two parameters are changing within a block. This configuration allows for a better statistical analysis of results (see Section 3) which would be otherwise difficult to interpret.

Blocks are presented in the same order to all viewers but scenes within a block are randomly reordered for each test. This approach reduces the negative effects of specific ordering of scenes.

A full detailed list of evaluation conditions is provided in Table 3.2. Values in combined cells are set for the whole block. Four representative frames are provided in Figure 3.4 as a reference for the kind of content and quality being assessed in the test.

Table 3.2. Evaluation conditions.

Block	ID	Br[a]	Fr[b]	Resolution	Color	CV[c]
1	D1	8	10	QVGA	RGB	Low
	D2	14	5	QVGA	Grayscale	Low
	D3	20	1	QVGA	Grayscale	Low
	D4	14	1	QQVGA	RGB	High
	D5	8	5	QVGA	RGB	High
2	01	8	1	QVGA	RGB	Low
	02	8	5			
	03	8	10			
	04	14	1			
	05	14	5			
	06	14	10			
	07	20	1			
	08	20	5			
	09	20	10			
3	10	20	5	QQVGA	RGB	Low
	11			QQVGA	Grayscale	
	12			QVGA	RGB	
	13			QVGA	Grayscale	
4	14	8	1	QVGA	RGB	High
	15	8	5			
	16	8	10			
	17	14	1			
	18	14	5			
	19	14	10			
	20	20	1			
	21	20	5			
	22	20	10			
5	23	20	5	QQVGA	RGB	High
	24			QQVGA	Grayscale	
	25			QVGA	RGB	
	26			QVGA	Grayscale	

[a]Bitrate; [b]Frame rate; [c]Content Variation.

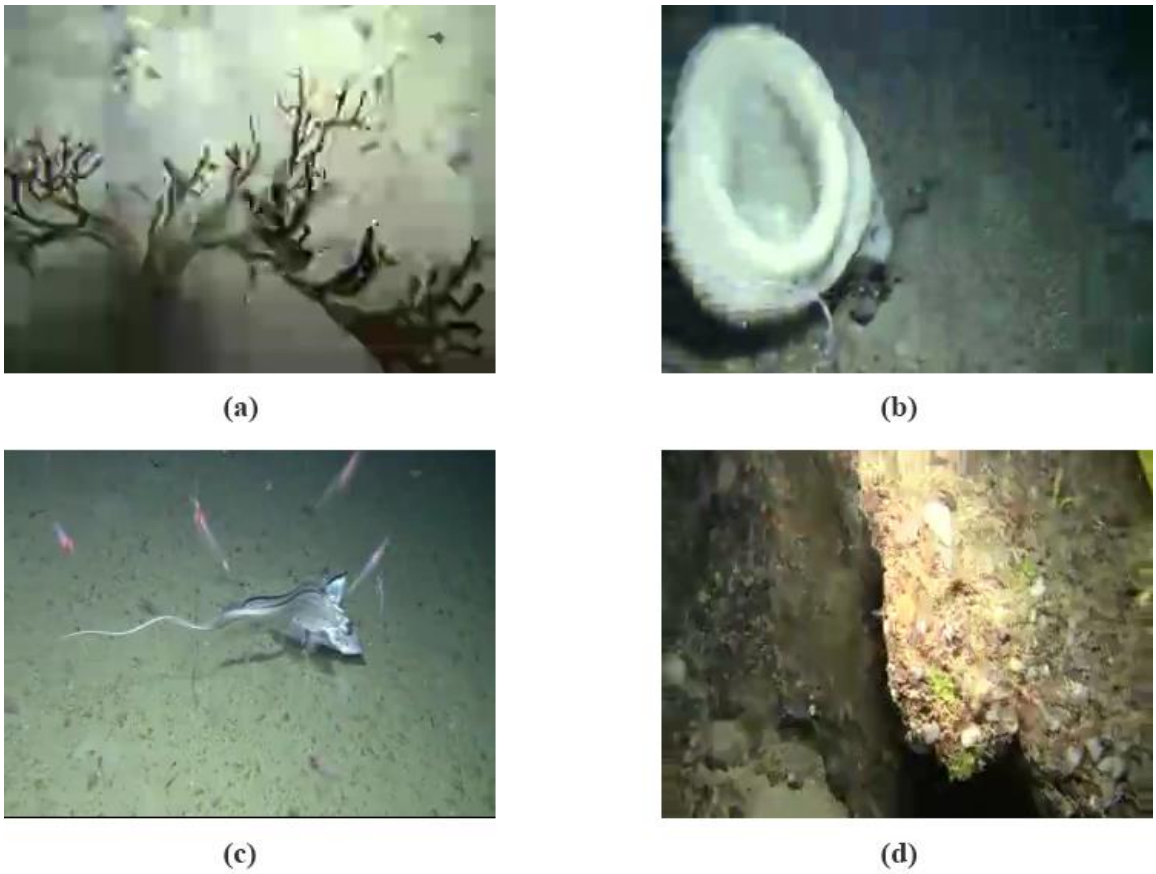

Figure 3.4. Sample frames (QVGA, RGB color). (**a**) 8 kbps–5 fps—high variation content; (**b**) 14 kbps–5fps—high variation content; (**c**) 20 kbps–5 fps – low variation content, (**d**) 14 kbps–1 fps – high variation content.

3.3 Results and discussion

This section describes the statistical processing of the data acquired in the experiment according to the guidelines given in P.910. The main statistic is the MOS, but other meaningful indicators have been also computed. Additionally, analysis of variance (ANOVA) tests have been performed to check the significance of the MOS differences across different blocks.

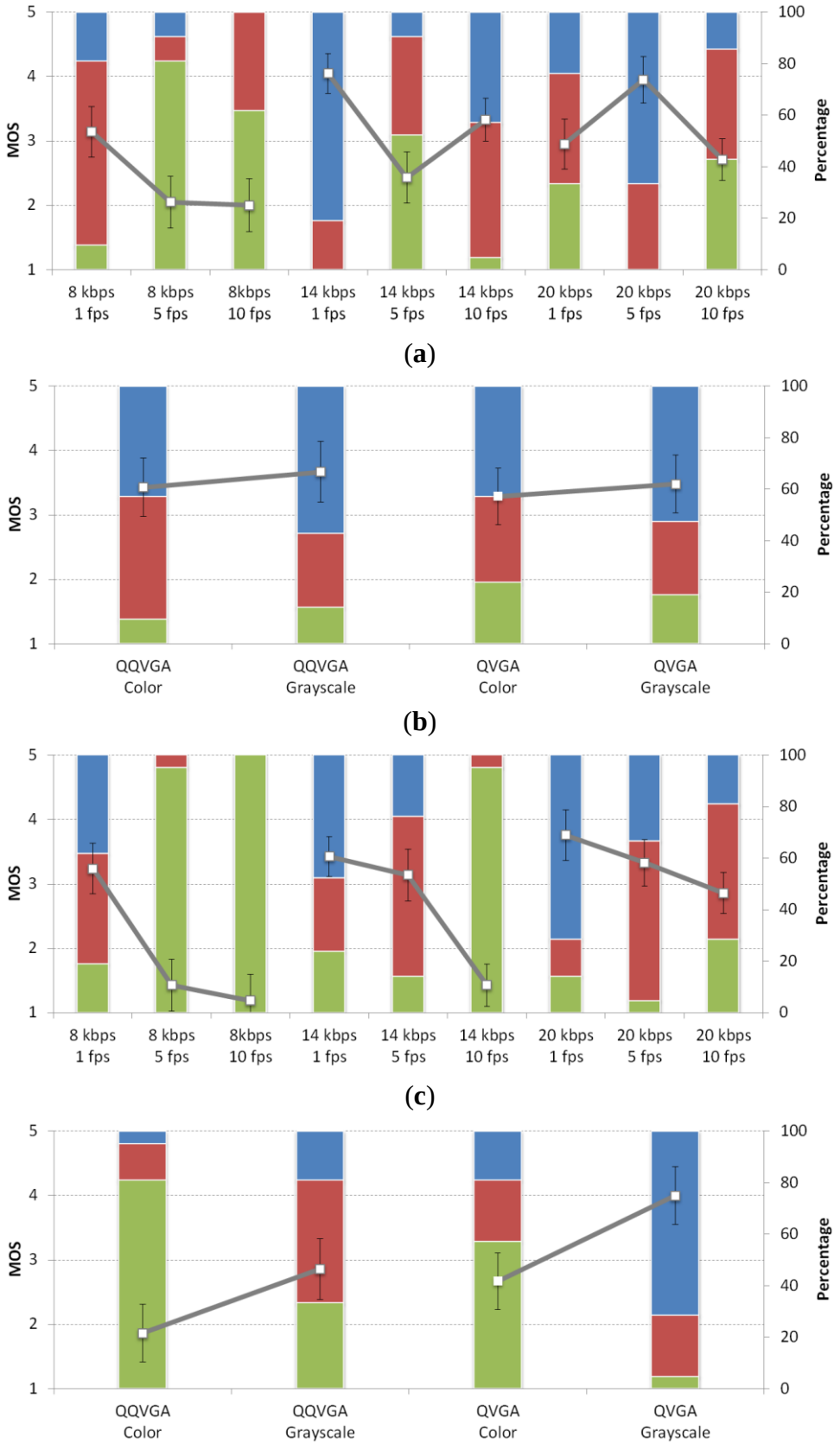

Figure 3.5. MOS values and cumulative distribution of scores as the percentage value of "good or better" (GOB-blue), fair (FAIR-red) and "poor or worse" (POW-green) scores. (**a**) Block 2; (**b**) Block 3; (**c**) Block 4; (**d**) Block 5.

Figure 3.5 shows, for every condition in the test, the MOS value with the 95% confidence interval (error bars). The stacked columns chart shows the cumulative distribution of quality scores in three groups of categories (each of them with a different color): the percentage of good or better (GOB) samples in the blue (upper) column, the percentage of fair (FAIR) samples in the red (middle) column and the percentage of poor or worse (POW) samples in the green (lower) column.

Figures 3.5a and c plot the MOS values for low and high variation samples in the bitrate/frame rate groups (blocks 2 and 4). Apart from some exceptions it can be said that, for a given bitrate, lower frame rates achieve better quality, while for a given frame rate, higher bitrates are scored better. It can also be seen that the scenes with low variation contents get better scores than the high variation ones.

Figure 3.5b and d show the MOS values for resolution/color samples (blocks 3 and 5). Regardless of the content variation, grayscale scenes (second and fourth columns) score better than color scenes (first and third columns) for a given resolution. If color is the given parameter, higher resolutions (third and fourth columns) get a higher MOS for high variation contents while the opposite happens for low variation scenes, although in this case MOS differences are very small. As before, most of the conditions in the low variation block have better average scores than the equivalent conditions in the high variation block.

To test the statistical significance of the MOS values obtained in the tests we have used analysis of variance techniques (ANOVA) [24]. A two-way within-subjects test has been performed for each block. Using this test, we can accept or reject the hypothesis of equal means in ANOVA for each block, i.e., we can attribute the differences between MOS values to changes in parameters (if we reject the hypothesis) or to other random effects in the sampling process (if we accept it). This acceptance or rejection is based on one of the results of the ANOVA test, the p-value, which represents a probability. A high p-value means the hypothesis under test may be accepted, while a low p-value means we should reject the hypothesis. In this experiment, two factors for each block have been defined (see Section 3.2.6). Three levels per factor are tested for bitrate (8, 14, 20 kbps) and frame rate (1, 5, 10 fps), while two levels are used for resolution (160 × 120, 320 × 240 pixels) and color (RGB, Grayscale). The software IBM SPSS Statistics 22 has been used [24]. We have used a significance $\alpha = 0.05$ and the repeated measures option since every subject has been used to evaluate every condition in the test and thus the answers for the same subject are not independent. This fact also provides improved statistical power.

The repeated measures test requires the assumption of sphericity, which is checked with the Mauchly's test [49]. Sphericity assumption will be rejected if the result (p-value) of the Mauchly's test is below 0.05. In this case, corrected p-values (Lower-bound, Greenhouse-Geisser [50] and Huynh-Fedt [51]) for ANOVA should be calculated. Multivariate tests do not require sphericity and are also a common tool to complete the comparison.

Table 3.3. ANOVA results.

Block 2 — Maulchly's Test of Sphericity					
Within subjects effect	Maulchly's W				Sig.
Bitrate	0.832				0.173
Frame rate	0.944				0.947
B*Fr	0.282				0.006
Block 2 — Test of within subjects effects					
Source		df	MS	F	Sig.
Bitrate	S.A.[a]	2	14.926	28.380	0.000
Frame rate	S.A.	2	8.720	19.988	0.000
B*Fr	G-G[b]	2.705	16.967	23.405	0.000
Block 2 — Multivariate tests					
Effect		Value		F	Sig.
B*Fr	Pillai's T.	0.771		14.320	0.000
	Hotelling's T.	3.370		13.320	0.000
Block 4 — Maulchly's Test of Sphericity					
Within subjects effect	Maulchly's W				Sig.
Bitrate	0.941				0.562
Frame rate	0.853				0.221
B*Fr	0.739				0.782
Block 4 — Test of within subjects effects					
Source		df	MS	F	Sig.
Bitrate	S.A.	2	29.370	69.858	0.000
Frame rate	S.A.	2	42.926	68.580	0.000
B*Fr	S.A.	4	6.140	21.158	0.000
Block 3 — Test of within subjects effects					
Source		df	MS	F	Sig.
Color	S.A.	1	0.583	1.429	0.246
Resolution	S.A.	1	0.964	2.477	1.131
R*C	S.A.	1	0.012	0.041	0.841

Block 5 — Test of within subjects effects					
Source		df	MS	F	Sig.
Color	S.A.	1	20.012	64.160	0.000
Resolution	S.A.	1	28.583	65.962	0.000
R*C	S.A.	1	0.583	1.522	0.232

[a]Sphericity Assumed; [b]Greenhouse-Geisser.

A summarized version of the full analysis output is shown in Table 3.3. The table contains the main results of the analysis: sum of squares, degrees of freedom, mean squares, F ratio and p-value (under the "Sig." column). For blocks testing bitrate and frame rate, the Mauchly's sphericity test results are shown first. The hypothesis of sphericity should only be rejected for the interaction effect in block 2 ($p = 0.006$). The corrected value for significance, calculated using the Greenhouse-Geisser method, is shown in the within subjects effects table. The results of two multivariate tests (Pillai's and Hotelling's Traces) are also included to complete the comparison. Other multivariate tests offered by SPSS had consistent values. The computed significances allow rejecting the hypothesis of equal means for blocks 2 and 4 ($p < 0.001$) and state that changes in the MOS are due to changes in parameters. We can also say that changes for different levels of bitrate depend on the frame rate and vice versa ($p < 0.001$).

The blocks analyzing resolution and color have only two levels per factor, so sphericity checking is not applicable. In this case, the null hypothesis can be rejected for single effects in the high content variation group ($p < 0.001$), but not for the interaction ($p = 0.232$). The hypothesis of equal means cannot be rejected for any of the effects in the low variation group either ($p = 0.246$ for color, $p = 1.131$ for resolution and $p = 0.841$ for interaction).

Summarizing, the analysis of variance verifies that there is enough statistical significance to consider that the differences in the MOS values are due to differences in the bitrates and frame rates, but a similar conclusion for color/resolution can only be drawn for the high variation content block. For the color/resolution block with low variation content the results are inconclusive.

The data gathered about scientific utility (SU) has been used to compute a linear regression model between this measure and the MOS. Utility is mapped to integer values from 0 (useless) to 4 (very useful). The mean utilities for each value in the MOS scale have been computed and used as a dependent variable. The estimated regression line shown in Equation (3) has been obtained as a result:

$$\mathrm{SU} = 0.8583 * \mathrm{MOS} - 0.2409 \qquad 1 < \mathrm{MOS} < 5 \tag{3.3}$$

Figure 3.6 plots the measured points, the regression line and equation $y = x - 1$. This equation represents the mapping function if there were an identity correspondence between MOS values and utility. It can be seen that though both lines are very close, the

regression line is always above the identity line. This means that evaluators perceived a better utility than quality for all samples. This gap is wider, up to half a point, for low MOS and it decreases as the MOS increases.

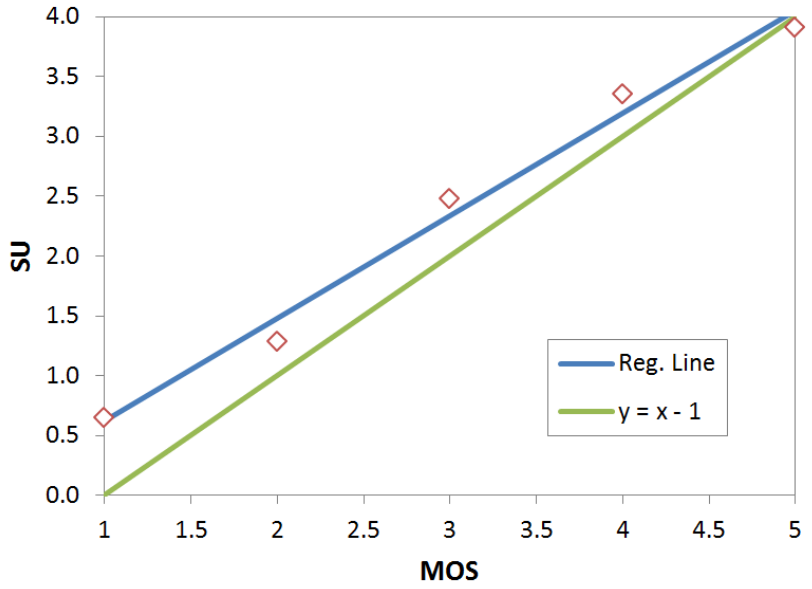

Figure 3.6. Scatter diagram for MOS *versus* scientific utility and estimated regression line.

The qualitative opinions gathered in the interviews cannot be analyzed through statistical methods, but are very consistent with the quantitative results. Most viewers stated that they preferred sharp images even if the video smoothness was not high. Another frequent remark was that grayscale images were as useful as the RGB videos for typical tasks. Also, participants often mentioned that, in spite of their low definition, images are still useful for common processes such as identification and counting of species or identification of seafloor morphology.

The MOS results provided in this section show how the analyzed video services achieve moderate quality scores. Even with the strong limitations in the encoding parameters, viewers in the test scored a good number of evaluation conditions in the fair category with some of them actually reaching the good category. These results notably differ from the G.1070 estimations for similar conditions (see Section 4.3). They allow planning of underwater video services with bitrates as low as 8 kbps at the application layer with an expected perceived quality of 3 out of 5 in the MOS scale and GOB between 20%–40%. When asked about the scientific utility of the samples, most of the scores fell in the moderately useful and quite useful categories. Finally, the viewers' qualitative comments corroborate the quantitative analysis. This agreement supports the idea that MOS is being shifted because of the special access conditions.

Chapter 4

Parametric Objective Quality Assessment for Underwater video

If repeated measures of quality are required, subjective VQA becomes a cumbersome task as it can be understood from the experimental setup in chapter 3. Objective VQA aims to obtain quality information in an automated way, without the intervention of human viewers. The convenience of no reference methods for underwater networks has already been discussed (see Section 2.2.2). Parametric methods are especially useful for network planning (see Section 2.1.2). This chapter starts with a literature review on the topic (Section 4.1) and the description of the subjective dataset used as a basis in this chapter (Section 4.2). A suitability study for the standardized parametric method in ITU-T Recommendation G.1070 is presented next, concluding that it is not appropriate for underwater video (Section 4.3). Then, two parametric methods with different applications are presented (Sections 4.4 and 4.5).

4.1 Literature Review

Most of the existing bibliography on objective VQA focuses on mean opinion score (MOS) estimation. The MOS statistic stems from subjective quality tests. In these studies, the users issue a score for every video sample in a categorical quality scale. A five-class scale (bad, poor, fair, good and excellent) is often employed and classes are mapped to numerical values (1–5) for easier processing.

The MOS is the average of these values across all the users for each sample. Objective VQA methods estimate this value because it is an intuitive and easy to use quality metric. There are several pixel-based and bitstream-based good performing NR methods available [21], [52], [22], [53]. Nonetheless, all of them involve a considerable amount of image or feature extraction processing. This processing load can be considered reasonable for typical computing capabilities, even for inexpensive equipment. However,

energy saving is a priority in UWSNs and intensive processing tasks should be avoided, as mentioned above.

Some parametric network planning methods can also be found in the literature [42], [54]-[62]. These techniques are lighter in processing since they only require the evaluation of a function to compute the quality estimation. A performance comparison between all of them was conducted in [41], concluding that the best results for encoding impairments are obtained with [62] and the best results for transmission impairments are achieved with the ITU standard G.1070 [14]. However, the procedure proposed in [62] is not strictly a parametric method since the video content is introduced in the model with the average sum of absolute differences (SAD) per pixel and, therefore, actual video signals should be used to compute the quality estimation.

Machine learning techniques have also been applied successfully to the problem of VQA. The work in [63] describes a RR method using a convolutional neural network which is usually regarded as one of the machine learning procedures with a higher computational cost [64]. Another study [65] proposes a NR support vector machine (SVM) regression but, again, a moderate amount of processing is required to extract the eighteen different features necessary for the estimation. Similar problems can be found in [52], where the number of features increases up to fifty-four. A decision tree is trained in [66] to develop a NR bitstream-based method but the work focuses on a subjective dataset with a very high coding bitrate to resolution ratio, which greatly differs from our environment.

MOS has already been found an insufficient metric unable to provide information about user diversity; however, a number of investigations have tried to overcome this limitation. A very interesting approach to QoE research is offered in [67], where a model with additional statistics is provided; it shows how the MOS hides relevant information. Nevertheless, QoE is addressed generally and video services are only included as a use case. Moreover, the authors state that these particular services do not fully fit the binomial distribution of scores proposed in the paper. Another noteworthy effort to overcome the MOS limitations has been done in [68]. The authors use machine learning techniques to build a prediction model for the proposed metrics: the degree of general acceptability and the degree of pleasant acceptability. Yet, it is based on a non-standard subjective data experiment which requires a complex procedure. None of the works mentioned in this section take into account the scarcity of resources we have described for UWSNs nor do they consider the specificities in user perception which have been observed in scientific applications (see Chapter 3). In this chapter, we present two machine learning models for quality estimation designed for underwater video. The first of them is a NR parametric planning model based on surface fitting regression techniques. This model is able to provide a computationally fast and lightweight estimation of quality with only two service parameters (bitrate and framerate).

The second model goes beyond MOS and computes estimations of full score distributions from the same service parameters and two video content features and, thus, it is a RR hybrid model. Ordinal logistic regression serves as a machine learning foundation algorithm for this model, which also produces quality predictions with lightweight processing.

4.2 Subjective dataset

The database of videos and subjective quality scores used in this chapter is a subset of the information gathered in the experiment described in chapter 3. The features of the video clips selected for building the parametric models are shown in Table 4.1. Other video features kept constant for all the clips were the compression format H.264, RGB (24 bits) color and QVGA (320 × 240) resolution. Clips are grouped for comparison purposes in two blocks according to their content variation: a high variation content (HVC) block and a low variation content (LVC) block. Additionally, an alternative, reduced, low variation content block (rLVC) will be considered. This block contains every point in the LVC block except rows with ID 06 and 07 (see Table 4.1). In chapter 3, the relation between quality to usefulness for the specific application of scientific underwater video was discussed. The MOS values for clips 06 and 07 break the trend of the whole dataset and it is possible that the particular content of these clips shifted the opinion of the viewers since some starfish can be seen in clip 06, while clip 07 contains plain seafloor with a few scattered small cavities. Although further subjective tests should be conducted to assess this hypothesis, it is useful to consider the rLVC block as a tool to avoid overfitting.

Table 4.1. Video features for model fitting and machine learning algorithms.

Block	ID	B_r (kbps)	F_r (fps)	SI	TI
LVC	01	8	1	23.95	4.46
	02	8	5	23.87	4.19
	03	8	10	24.35	4.38
	04	14	1	30.35	4.58
	05	14	5	27.66	4.16
	06	14	10	29.35	6.24
	07	20	1	41.46	7.18
	08	20	5	36.87	9.20
	09	20	10	39.23	7.13
HVC	10	8	1	67.13	15.42
	11	8	5	75.43	13.96
	12	8	10	57.69	13.46
	13	14	1	71.11	15.92
	14	14	5	66.52	13.92
	15	14	10	76.33	18.05
	16	20	1	71.11	15.92
	17	20	5	60.21	11.20
	18	20	10	53.95	10.15

4.3 Suitability study of ITU-T G.1070

The VQA model proposed in [42] is the only parametric model standardized by ITU as part of the ITU-T G.1070 recommendation: "Opinion model for video-telephony applications" [14]. Although it was designed for this specific application some authors consider it a general reference for parametric models [41]. The model computes a MOS value from a group of equations which take as input parameters the bitrate (Br), the frame rate (Fr) and the packet-loss rate. The model coefficients must be selected according to some other service variables: the compression codec, the video resolution and the physical display size. The recommendation provides five "provisional" coefficient sets in Appendix I (not an integral part of the recommendation), which can be used under some restrictions regarding bitrate and packet loss. Due to these restrictions, only coefficients in sets #1 and #2 could be used in underwater video. Figure 4.1 shows the MOS values predicted by this model for a 5–25 kbps bitrate interval. Different curves are plotted for common frame rates and the coefficient group corresponds to MPEG4, QVGA and 4.2'' sized images. The highest MOS is predicted for 25 kbps and 1 fps with a value of 1.301 in a one to five scale, which does not agree with the subjective data in chapter 3. Therefore, the first step in the search for a parametric model for underwater VQA is calculating a new set of coefficients for the G.1070 model.

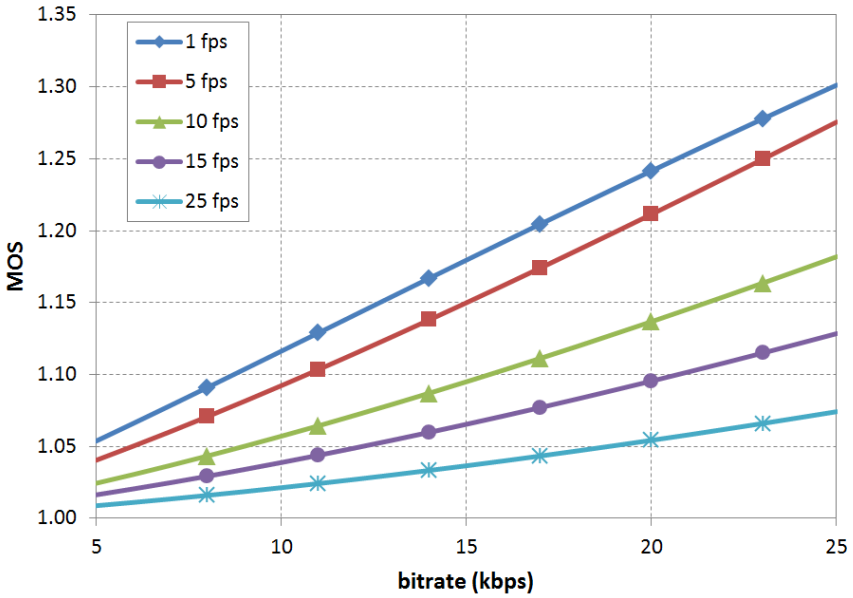

Figure 4.1. MOS values predicted by G.1070 for MPEG4, QVGA and 4.2'' videos.

According to the available information in the available experimental data, the G.1070 model can be simplified as shown in Equations (4.1)–(4.3), where MOS is the quality prediction in the usual 1–5 MOS scale, *Ofr* is the optimal frame rate for a given bitrate, I_{Ofr} is the maximum video quality for a given bitrate and D_{Fr} is the degree of robustness

due to the frame rate. This model does not take into account content variation and therefore SI, TI information must be discarded:

$$MOS = 1 + \mathrm{I}_{Ofr}\exp\left(\frac{[\ln(Fr) - \ln(Ofr)]^2}{2D_{Fr}^2}\right) \tag{4.1}$$

$$Ofr = v_1 + v_2 B, \qquad 1 \le Ofr \le 3 \tag{4.2}$$

$$I_{Ofr} = v_3 - \frac{v_3}{1 + \left(\frac{Br}{v_4}\right)^{v_5}}, \qquad 1 \le I_{Ofr} \le 4 \tag{4.3}$$

$$D_{Fr} = v_6 + v_7 Br, \qquad 0 < D_{Fr} \tag{4.4}$$

Annex A in the recommendation specifies the methodology for deriving the coefficients from a subjective quality dataset. The procedure is based on successive least square approximations (LSA). The first step obtains, for each bitrate, estimations of intermediate parameters *Ofr*, I_{Ofr} and D_{Fr}, based on frame rate values. However, the LSA for our subjective data cannot be solved in the real domain as shown in Table 4.2. For the high variation content, the imaginary part of the intermediate parameters could be considered negligible since it is nine orders of magnitude smaller than the real part and thus the coefficients can be calculated with another LSA approximation. Table 4.3 contains these results along with the goodness of fit (GOF) standard measures: the sum of squares due to error (SSE), the R square (R^2) and the root mean squared error (RMSE). The values of R^2 and RMSE indicate a very poor fit quality: R^2 is negative, showing that even a simple linear regression (plane) would be more appropriate for the data; and the RMSE is of the same order of magnitude as the data. The poor performance of this model could be attributed to the fact that it was designed for a very specific application (video telephony), which greatly differs from underwater video services in several important aspects, such as video content and features, purpose of the video service and user expectancies. These differences can considerably change user perception of quality.

Table 4.2. Intermediate parameter estimation for deriving the coefficients of the G.1070 model.

Br (kbps)		8	14	20
LVC	**Ofr**	1.013×10^{-7}	$-30.7385 - 7.813 \times 10^{-8}$ i	$2.969 - 1.138 \times 10^{-13}$ i
	I_{Ofr}	31.826	2.219 − 0.07 i	$3.336 - 1.206 \times 10^{-13}$ i
	D_{Fr}	6.906	14.01 + 2.155 i	$1.05 - 8.857 \times 10^{-14}$ i
HVC	**Ofr**	1.878	1.101	0.682
	I_{Ofr}	4.955	2.204	0.577
	D_{Fr}	$2.43 + 1.066 \times 10^{-9}$ i	$1.688 - 5.507 \times 10^{-9}$ i	$1.811 + 6.434 \times 10^{-9}$ i

Table 4.3. HVC coefficients for the G.1070 model and GOF statistics.

v_1	v_2	v_3	v_4	v_5	v_6	v_7
2.445	0.0459	1.946	7.935	32.431	– 0.294	0.094
SSE			**R^2**			**RMSE [1]**
36.9130			−0.0561			2.5906

[1] RMSE averaged over the difference between the number of samples and the number of parameters in the model.

4.4 No-Reference parametric model

4.4.1 Model development

In the previous section, we have shown that the ITU reference model does not seem suitable for the experimental data. As an approximation to an appropriate model, we have used the thin plate spline interpolation method [69] to find a surface for each of the content variation subsets. The thin plate spline is defined as the unique minimizer of the energy function defined in (4.5) for the two-dimensional case. This method provides a perfect fit ($R^2 = 1$) for the given control points. The minimization constraint produces a smooth surface (minimally "bended") which matches the assumption of no great variations in quality values between the studied input variables. The surface can be defined as in (4.6), a weighted sum of the radial basis function in (4.7) where $x^{(i)}$ are the control points, K is the number of points and a_i, w_i are the optimization parameters. In this case, the control points are the samples from our subjective dataset, considering the bitrate as our first dimension or feature (x_1) and the framerate as the second feature (x_2). The resulting MOS for a given sample is $y^{(i)}$. This interpolation technique produces a representative surface but not the practical model we aim for, since the complexity of the resulting equation makes it difficult to interpret the coefficients. Figure 4.2 shows three surface plots of thin plate splines fitting the subjective dataset. Figure 4.2 a is obtained from points in the high variation content (HVC) block as control points, while points in the low variation content (LVC) and reduced low variation content (rLVC) blocks are for Figure 4.2 b,c, respectively. The shapes of the HVC and rLVC surfaces are very similar. Even the LVC surface could be regarded as reasonably similar, except for the bending forced by the anomalies already mentioned in Section 4.2. The model proposal in Section 4.4.2 is motivated by the resemblance between this geometrical profile and a sigmoid function.

$$E_{tps} = \sum_{i=1}^{K} \left| f\left(x^{(i)}\right) - y^{(i)} \right|^2 + \lambda \iint \left[\left(\frac{\partial f^2}{\partial x_1^2} \right)^2 + 2 \left(\frac{\partial f^2}{\partial x_1 \partial x_2} \right)^2 + \left(\frac{\partial f^2}{\partial x_2^2} \right)^2 \right] dx_1 dx_2 \quad (4.5)$$

where

$$f(x_1, x_2) = a_0 + a_1 x_1 + a_2 x_2 \sum_{i=1}^{K} w_i \varphi\left(\left| (x_1, x_2) - x^{(i)} \right| \right) \quad (4.6)$$

and

$$\varphi(r) = r^2 ln(r). \tag{4.7}$$

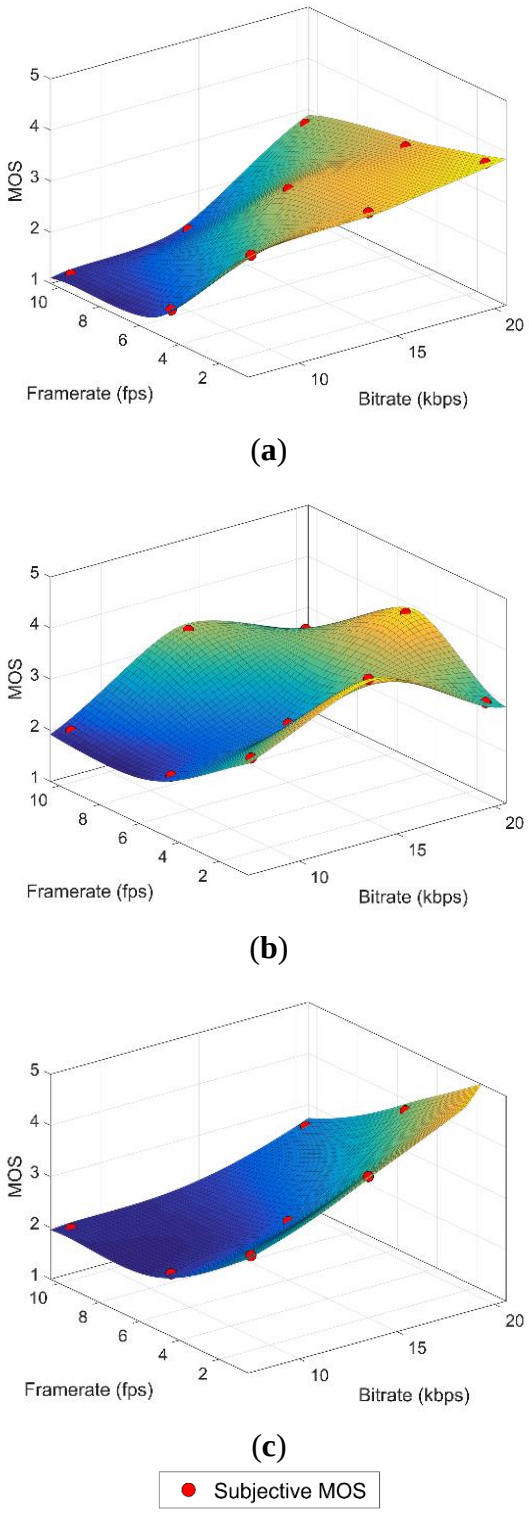

Figure 4.2. Thin plate spline surfaces. (**a**) HVC block, (**b**) LVC block, (**c**) rLVC block.

4.4.2 Model equations

The first proposal is a regression model based on the generalized logistic function (4.8) [70]. The coefficients in the logistic function are relatively easy to relate to the function behavior and thus an interpretation of their values can be extracted. The proposed model extends the generalized logistic function for two dimensions and includes a linear function with a nonzero y-intercept term for the exponential in the denominator to improve the fitting performance. Two different variations of the model according to different optimization objectives are given:

i. Generalization—Non-linear regression model (NLR.G).

Equation (4.9) achieves a more consistent behavior of the model outside the range of the subjective dataset. The asymptotes of the surface are set to the limits of the quality scoring scale (1–5).

ii. Accuracy—Non-linear regression model (NLR.A).

Equation (4.10) achieves a better fitting for the points in the subjective dataset (higher R^2):

$$f(x) = L + \frac{U - L}{(A + Be^{-Cx})^{1/v}}, \tag{4.8}$$

$$f(x) = 1 + \frac{4(A)^{1/v}}{(A + Be^{-(c_0 + c_1 x_1 + c_2 x_2)})^{1/v}}, \tag{4.9}$$

$$f(x) = L + \frac{K}{(A + Be^{-(c_0 + c_1 x_1 + c_2 x_2)})^{1/v}}. \tag{4.10}$$

Parameters L, U, K, A, B, c_i, υ are optimized with the non-linear least squares method applied to our subjective dataset. The coefficients computed for every block can be found in Table 4.4 for the NLR.G model and in Table 4.5 for the NLR.A model. The corresponding goodness-of-fit statistics (SSE, R^2, RMSE) are shown in Tables 4.6 and 4.7. These values are used as a performance metric of the model. Figure 4.3 contains plots for each model surface: NLR.G surfaces are in the left column while the right column shows the NLR.A surfaces.

Table 4.4. Coefficients for the NLR.G model.

Block	A	c_0	c_1	c_2	v
HVC	6.994	5.569	0.0977	–0.1512	3.623×10^{-4}
LVC	487.1	–1.008	0.05259	−0.05686	5.195×10^{-3}
rLVC	23.33	–15.31	0.7495	–1.224	10.37

Table 4.5. Coefficients for the NLR.A model.

Block	**L**	**K**	**A**	**B**
HVC	1.291	3.518	1.539	2.411
LVC	2.505	7.83	3.864	11.11
rLVC	1.933	2.264	1.362	4.158
Block	c_0	c_1	c_2	**v**
HVC	−1.952	0.6349	−0.9421	1.013
LVC	−16.62	3.128	−6.671	0.7034
rLVC	−9.609	1.063	−1.906	5.672

Table 4.6. GOF statistics for the NLR.G model.

Block	**SSE**	**R^2**	**RMSE** *
HVC	0.959	0.8809	0.4896
LVC	2.687	0.3945	0.9186
rLVC	0.3916	0.9084	0.4425

* RMSE averaged over the difference between the number of samples and the number of parameters in the model.

Table 4.7. GOF statistics for the NLR.A model.

Block	**SSE**	**R^2**	**RMSE** *
HVC	0.116	0.9856	0.34
LVC	1.936	0.5637	1.391
rLVC	0.2609	0.939	–

* RMSE averaged over the difference between the number of samples and the number of parameters in the model.

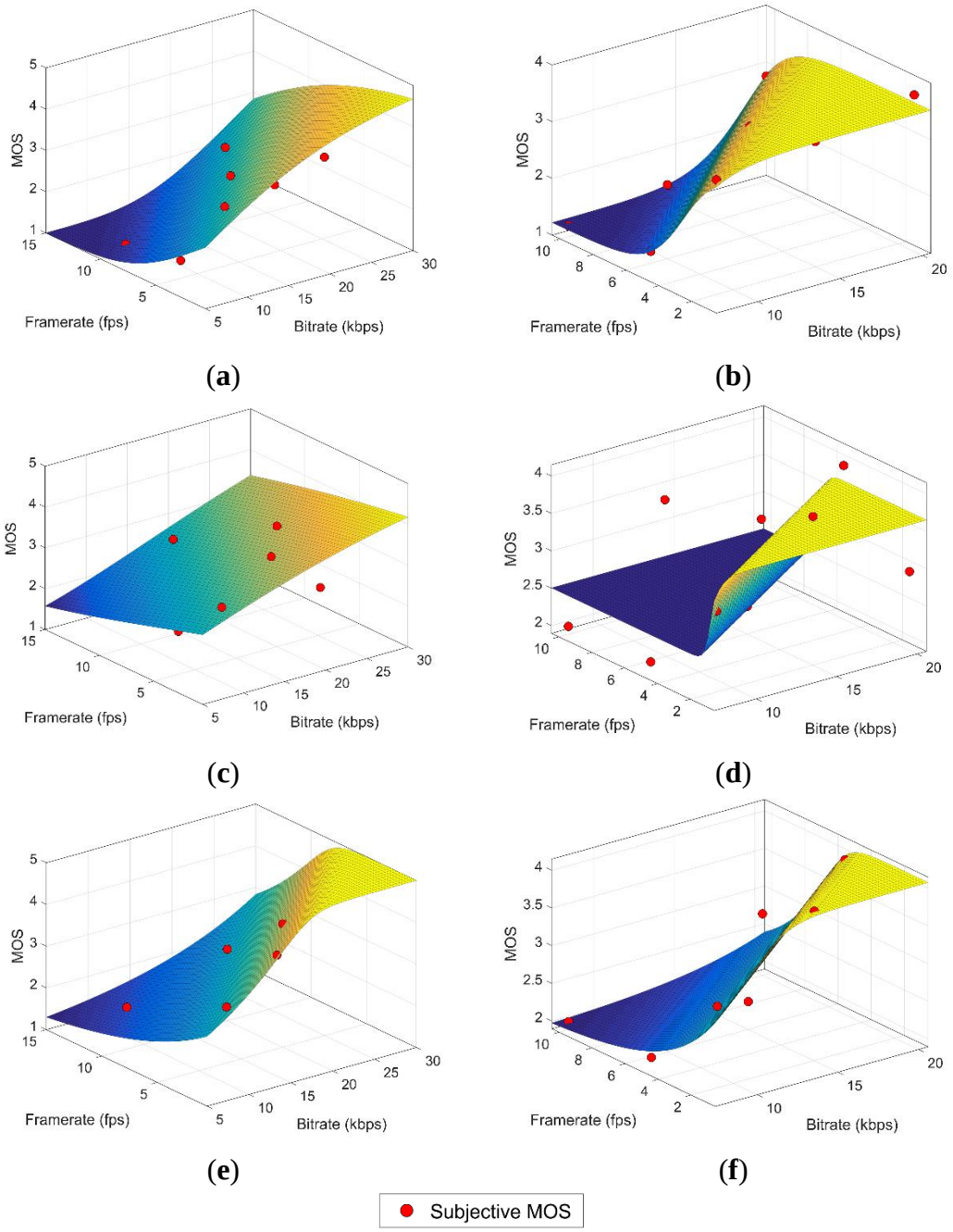

Figure 4.3. NR model surfaces. (**a**) NLR.G–HVC, (**b**) NLR.A–HVC, (**c**) NLR.G–LVC, (**d**) NLR.A–LVC, (**e**) NLR.G–rLVC, (**f**) NLR.A–rLVC. Note that the bitrate axis in (**a**,**c**,**e**) has been extended to show the generalization behavior.

The model fit for high variation content video is very good for the NLR.G model ($R^2 \approx 0.88$) and excellent for the NLR.A model ($R^2 \approx 0.98$). The performance of the model is poor ($R^2 \leq 0.6$) for the low variation content videos because the model cannot fit the non-sigmoid shape of the cloud of points. However, the performance dramatically rises to the levels of high variation content for the reduced low variation content dataset, with an excellent performance of both NLR.G ($R^2 \approx 0.91$) and NLR.A ($R^2 \approx 0.94$). The RMSE value is also considerably low for both blocks (RMSE ≤ 0.5), taking into account that it

is not being averaged over the total number of samples but over the difference between the number of samples and the number of parameters in the model.

Beyond goodness-of-fit considerations, NLR models offer an easily computable approach to objective quality assessment. An underwater node could obtain an estimation of the MOS in a fast and energy-efficient way since it only requires the video coding bitrate and framerate as input variables and no further calculation is needed.

4.5 Reduced-Reference hybrid model

In spite of the advantages of the NLR model, for some applications it could be regarded as too simplistic. Firstly, video content is only considered in a coarse way, as two big blocks of high and low variation content. Secondly, relevant data is lost when computing the MOS because the information about the distribution of scores is discarded when they are transformed into a single averaged value.

The ordinal logistic regression (OLR) [71], [72] is a classification method for multiclass problems with a natural order among the response categories. Thus, it is perfectly suitable for the quality assessment experiment in which users issue scores within an ordered categorical scale (bad, poor, fair, good and excellent). In OLR a sample or observation x is a group of values of the input variables associated to a distribution of scores for the outcome variable. All the video features described in this chapter in Section 2 will be used as inputs of the model. Therefore, each observation is a four-component vector including as features the bitrate (x_1), the framerate (x_2), the SI (x_3) and the TI (x_4). We call $\pi_i(x)$ the probability of the observation x to be in the i-th category. For k categories of the outcome variable, the method computes the $k-1$ logarithms of the odd ratios or logits, i.e. the logarithms of the probability of being in a given category or any category below (γ_j) divided by the probability of being in any superior category.

The model is based on the proportional odds assumption, which states that these logits can be represented by a linear model with a different intercept term θ_j for every logit but the same coefficients β for all the predictors (4.11). The $\pi_i(x)$ probabilities are obtained from the model as in (4.12). An estimator for the MOS is proposed in (4.13). Even though the original target was departing from the MOS simplistic approach to QoE assessment, the MOS estimator can still be useful for comparing with other models:

$$log\left(\frac{\gamma_j(x)}{1-\gamma_j(x)}\right) = \theta_j + \beta^T x, \;\; where \;\; \gamma_j(x) = \sum_{i=1}^{j} \pi_i(x) \tag{4.11}$$

$$\pi_i(x) = \gamma_j(x) - \gamma_{j-1}(x) = \frac{\exp(\theta_j+\beta^T x)}{1+\exp(\theta_j+\beta^T x)} - \frac{\exp(\theta_{j-1}+\beta^T x)}{1+\exp(\theta_{j-1}+\beta^T x)}, \tag{4.12}$$

$$\text{with } \gamma_5 = 1, \gamma_0 = 0 \; \forall x$$

$$MOS_{OLR} = \sum_{i=1}^{5} i \times \pi_i(x) \tag{4.13}$$

The IBM SPSS Statistics [48] software has been used for the model fitting through maximum-likelihood estimation and for the analysis of results. Since it has been already shown that there are some interactions between the model inputs (i.e. the video features; see Section 4.2), we have followed an iterative procedure to build the model. This procedure creates a model with as many interaction terms as possible. It, then, discards the non-significant interactions based on their *p*-value. A high *p*-value means the interaction is non-significant and it can be discarded; otherwise, the interaction term is retained. The procedure can be described with the following two-step loop:

For *i* = *num_features* to *i* = 1

1. Compute a model including every possible interaction except the ones that have been discarded in a previous iteration.
2. Check the *p*-value of the coefficient for every interaction term of *i-th* order (or main effect if *i* = 1). If $p > 0.05$, the interaction is considered non-significant and thus removed from subsequent iterations.

After this iterative procedure, the remaining main effects and interactions as well as the computed coefficients are shown in Table 4.8 along with the intercept term for each category. Table 4.9 collects the result of two χ^2 tests. The "model fitting test" is a Likelihood Ratio χ^2 test between a model with only the intercept term and the final model. The *p*-value for the model is significant ($p = 0.005$) and indicates that the final model fits the dataset better than a model with constant odds based on the marginal probabilities of each outcome category. The "parallel lines test" is an analogy between the final model and a multinomial model where no natural ordering is considered between categories and therefore different β coefficients are obtained for every logit estimator. The *p*-value is non-significant ($p = 1.00$), so there is no evidence to reject the assumption of proportional odds and there is a single set of β coefficients. Several R^2 values are provided in Table 4.10.

Table 4.8. Coefficients for the OLR model.

Category/Logit	Coefficient	Value
$\frac{\text{bad}}{\text{poor or better}}$	θ_1	6.839
$\frac{\text{poor or worse}}{\text{fair or better}}$	θ_2	8.891
$\frac{\text{fair or worse}}{\text{good or better}}$	θ_3	11.066
$\frac{\text{good or worse}}{\text{excellent}}$	θ_4	13.097
Effect/Interaction		
Framerate	β_1	0.333
SI	β_2	−0.871
TI	β_3	0.607
Bitrate*Framerate	β_4	−0.083
Bitrate*SI	β_5	0.024
Framerate*SI	β_6	0.090
Framerate*TI	β_7	−0.318
SI*TI	β_8	0.037
Bitrate*SI*TI	β_9	−0.002

Table 4.9. Chi-Squared tests for the OLR model.

Test		-2 Log Likelihood	χ^2	df *	p
Model fitting	Intercept only	485.514	–	–	–
	Final **	235.726	249.788	9	<0.005
Parallel lines	Null hypotesis **	235.726	–	–	–
	General	232.135	3.591	27	1.000

* degrees of freedom, ** fitted OLR model.

Table 4.10. Pseudo-R^2 and R^2 statistics for the OLR model.

p-R^2 – C&S *	p-R^2 – N **	p-R^2 – M ***	R^2 - MOS_{OLR}
0.484	0.509	0.220	0.90

* Cox and Snell, ** Nagelkerke, *** McFadden.

We have computed an R^2 value for the MOS_{OLR} estimator and the subjective MOS values in the dataset, resulting in a 90% of the variance explained by the proposed estimator. This result is similar to the R^2 obtained with the NR models. Pseudo-R^2 values are also included in the results as Cox and Snell [73], Nagelkerke [74] and McFadden [75]. These pseudo-R^2 values, as discussed in [76], cannot be interpreted like a classic R^2 in a least squares regression since they do not provide a comparison between the predicted values and those in the dataset, but between the fitted model and the only-intercept model described above. However, they serve to compare different models. To provide a graphical approach for the goodness-of-fit, Figure 4.4 plots the category probability distribution P_i for every observation in the subjective dataset as estimated by the OLR model against the proportions of scores computed from the subjective data π_i. It can be observed how the model provides a very good fit for most cases, with an excellent performance for some of the observations (IDs 05 and 15) and only a small amount of higher errors (category 3 in IDs 01 and 03, and category 2 in ID 09). In particular, 71.1% of the π_i estimations show a deviation smaller than 0.1.

We could also consider the mode of the distribution (the value that occurs most frequently) of scores to be the best categorical guess for the quality. In this case, if we select the categories with $\max(\pi_i(x))$ and $\max(P_i(x))$ as the classification decision, the accuracy of the classification method is 83.3%.

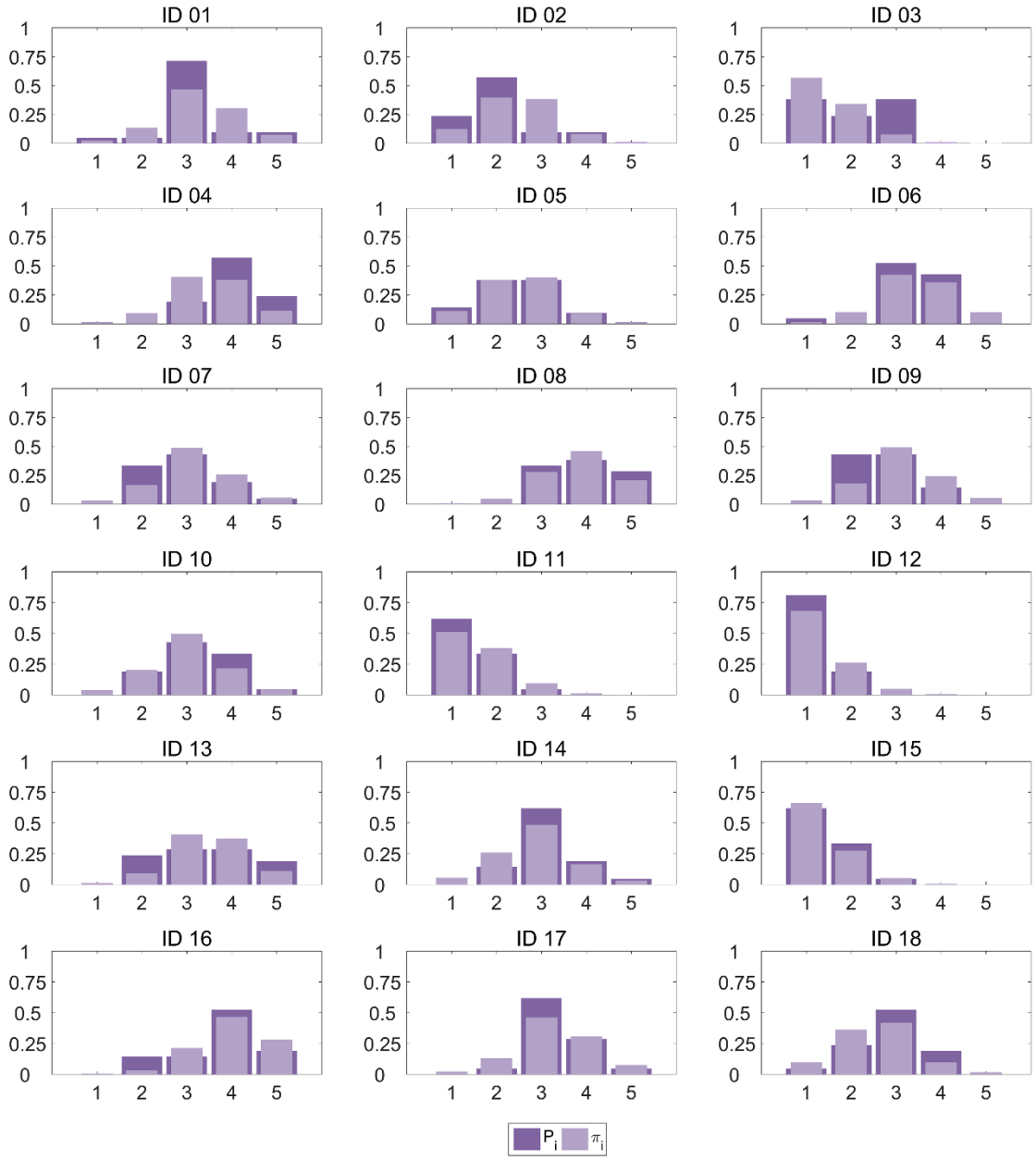

Figure 4.4. Proportions of scores from subjective data and estimated probabilities from OLR model.

Chapter 5

Pixel-based Objective Quality Assessment for Underwater Video

If a continuous monitoring of video quality is required, pixel-based methods might be a better choice for the quality assessment task. These methods use the received video frames and apply signal processing techniques to compute the quality estimation. The approach used in this chapter utilizes machine learning techniques to learn from actual human scores. In a way, the method presented here acts as a computerized viewer. As in the two previous chapters, a literature review is first presented (Section 5.1). Then, the model is described (Section 5.2) and a performance analysis and comparison with other methods is presented (Section 5.3).

5.1 Literature Review

Over the last few years there have been extensive improvements in the development of video services. A recent forecast indicates that by 2021, 82 percent of all consumer Internet traffic will be IP video [77]. Widespread broadband connections both landline and mobile are key technologies in this growing trend. In this context, algorithms for automatic evaluation of video quality are an important research area in science and industry. This is also true in underwater video transmission, where acoustic modems are used. However, due to channel characteristics, these modems feature strongly restricted bitrates, with peak bitrates of 64 kbps in the physical layer [43]. Therefore, until recently, only limited attention has been paid to video transmission over these networks. Specific encoding schemes for underwater video have been proposed focusing in low bitrate [78] and error resiliency [79]. However, the performance was evaluated with the Peak Signal to Noise Ratio (PSNR) an image metric that does not reflect the human perception of quality and subjective opinions are considered the ultimate reference for quality assessment. The acoustic modem design in [80] features video transmission capabilities, but in a short communication range under 20 m. The subjective quality assessment study in chapter 3 shows how video generated with the heavy constraints imposed by current

acoustic equipment can reach a relatively high perceptual quality, thereby becoming useful as considered by a group of experienced ocean scientists.

The natural next step is the application of an objective video quality assessment (VQA) method, i.e. the estimation of the average quality of a distorted video perceived by a group of users. No Reference (NR) algorithms only make use of the distorted video to estimate the quality, while other techniques require some partial information of the original signal (Reduced Reference) or even the undistorted, pristine, video (Full Reference). In UWSNs, quality information can be used to improve the management of the network resources and thus, it is desirable that the VQA method can be used on the service provisioning stage. NR methods are, then, quite appropriate for this task, since the recovery of the original signal is unfeasible due to the virtual inaccessibility of the nodes and the transmission of additional information about the original signal can be regarded as a burden for an already narrow channel. The topic of image quality assessment has been addressed for underwater networks in [81] and [82]. These studies are nevertheless focused on color, contrast and blur distortions due to underwater light absorption in raw images. In underwater video transmission, the amount of distortion caused by the encoding is much higher than the color distortion and the image quality metrics proposed in [81] and [82] are not useful as a starting point for our video quality assessment analysis.

The subjective study in chapter 3 suggests that perceptual quality in underwater video is shifted, due to the comparative advantage that an underwater network would mean over current video capture methodologies (which involve expensive submergible robots in limited expeditions). An effective objective VQA method for UWSNs should correlate well with human judgements. One successful pixel-based NR-VQA algorithm with this characteristic is V-BLINDS [83]. It was trained with the widely used LIVE VQA database [34] and its performance was evaluated on the same database and, also, on the EPFL-Polimi database [84]. V-BLINDS should be generalizable through re-training, but its performance is not satisfactory for the contents and levels of distortion of underwater video as we will show later (see section 5.3.2). VIIDEO [85] is another recent NR-VQA method which tries to overcome the dependence on human scores or any other knowledge about the distortion, but it also fails to provide good correlations against experimental human scores (see section 5.3.2). Other NR-VQA methods can be found in the literature and have already been extensively surveyed [83], [85]. However, they are either designed to evaluate other, specific kinds of distortion, or do not include information about subjective opinions or involve extracting a high number of features. As a process intensive task, feature extraction should be reduced as much as possible in underwater nodes due to power consumption issues. The method proposed here presents both a good correlation with human scores, and does not require a large number of features.

5.2 No-reference VQA method for underwater video

The model is based on the theory of Natural Scene and Video Statistics (NSS/NVS). This theory states that undistorted images and videos exhibit certain statistical properties that are lost when the same content is exposed to distortion. It also assumes that the human visual system has adapted through evolution to those characteristics and thus, they are relevant in visual perception. The algorithm proposed here is built upon a machine learning prediction model that utilizes six NVS features extracted from the video. The model is trained and validated with an underwater video database (see section 5.3.1).

5.2.1 Model Foundation

The spatial NSS model in [27] relies on the statistical properties of a transformation of the image by mean subtraction and divisive normalization:

$$\hat{I}(i,j) = \frac{I(i,j) - \mu(i,j)}{\sigma(i,j) + 1} \tag{5.1}$$

where $i \in \{1,2,\dots M\}$, $j \in \{1,2,\dots N\}$ are spatial indices with M and N the image dimensions and

$$\mu(i,j) = \sum_{k=-K}^{K} \sum_{l=-L}^{L} w_{k,l}\, I(i+k, j+l) \tag{5.2}$$

$$\sigma(i,j) = \sqrt{\sum_{k=-K}^{K} \sum_{l=-L}^{L} w_{k,l} \big(I(i,j) - \mu(i,j)\big)^2} \tag{5.3}$$

estimate the local mean and contrast, respectively, where $w_{k,l}$ is a 2D circularly-symmetric Gaussian weighting function sampled out to 3 standard deviations ($K = L = 3$) and rescaled to unit volume. The coefficients (5.1) are known to follow a Gaussian distribution when I is a natural image [27], but this behavior is disrupted when the images have been distorted. This separation from Gaussianity can be modeled by the generalized Gaussian distribution (GGD) with zero mean given by:

$$f(x;\alpha,\sigma^2) = \frac{\alpha}{2\beta\Gamma(1/x)} \exp\left(-\left(\frac{|x|^2}{\beta}\right)^{\alpha}\right) \tag{5.4}$$

where

$$\beta = \sigma \sqrt{\frac{\Gamma(1/\alpha)}{\Gamma(1/\alpha)}} \tag{5.5}$$

and $\Gamma(x)$ is the gamma function:

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt \qquad a > 0. \tag{5.6}$$

In a GGD distribution, the parameter α defines the shape of the distribution with $\alpha = 2$ corresponding to a Gaussian distribution. Other values of α imply departure from Gaussianity and, thus, distortion. The GGD distribution parameters can be estimated using the procedure in [89]. This model has been used in several successful image quality assessment methods [86], [87], [88].

The proposed NVS model is based on the statistics of frame differences, which have previously been used as a reliable tool to measure distortion in the temporal domain [83], [85]. We define a frame difference as the difference between two consecutive frames in a video sequence with *M* frames as:

$$\Delta F^t = F^{t+1} - F^t \;\; \forall t \epsilon \{0,1,2 \ldots, M-1\} \tag{5.7}$$

It has been shown [85] that frames differences from pristine videos also exhibit a Gaussian distribution when processed by the transformation in (5.1) (with $I(i,j) = \Delta F^t$) and that this Gaussianity is lost when they are distorted. The proposed model leverages these statistical regularities to extract a set of six features $(f_1 \ldots f_6)$ from a video that are used to estimate the quality.

5.2.2 Full Frame Difference Features

The first two features (f_1, f_2) are computed from the complete frame differences in the sequence. We denote $\widehat{\Delta F^t}$ to be the frame differences transformed by (5.1). For each $\widehat{\Delta F^t}$, the shape parameter α of the corresponding GGD distribution is estimated. Then, the first feature is computed as the average (over time) of the shape parameter:

$$f_1 = \frac{1}{M-1} \sum_{t=1}^{M-1} \alpha(\widehat{\Delta F^t}) \tag{5.8}$$

Pictures usually contain multiscale information, and the effect of distortion has an impact across different scales. Hence, we also compute the shape parameter for a reduced resolution version of the frame difference (using bicubic interpolation) by a factor of 2:

$$f_2 = \frac{1}{M-1}\sum_{t=1}^{M-1} \alpha\left(\widehat{\Delta F_2^t}\right) \tag{5.9}$$

5.2.3 Patched Frame Difference Features

A very important component of distortion happens locally. Therefore, features (f_3, f_4, f_5) are obtained from rectangular patches of the transformed frames differences $\widehat{\Delta F^t}$. Note that these transformations have already been computed for the feature f_1 and, thus, time and energy can be saved if the system has enough memory or the processing order is optimized. For every patch in each $\widehat{\Delta F^t}$, the shape parameter α of the associated GGD distribution is computed. We call A_p the set with the resulting α values. Then, the elements of A_p are grouped into three levels l_1, l_2, l_3 determined by the thresholds u_1, u_2. Finally, the features (f_3, f_4, f_5) are calculated as the number of α's per frame difference in each level:

$$f_3 = \frac{n\left(\mathrm{A}_{l_1}\right)}{M-1} \; with \; \mathrm{A}_{l_1} = \left\{\alpha \epsilon A_p \middle|\, \alpha < u_1\right\} \tag{5.10}$$

$$f_4 = \frac{n(\mathrm{A}_{l_2})}{M-1} \; with \; \mathrm{A}_{l_2} = \{\alpha \epsilon A_p |\, u_1 \leq \alpha \leq u_2\} \tag{5.11}$$

$$f_5 = \frac{n(\mathrm{A}_{l_3})}{M-1} \; with \; \mathrm{A}_{l_3} = \left\{\alpha \epsilon A_p \middle|\, \alpha > u_2\right\} \tag{5.12}$$

where $n(S)$ is the cardinality (number of elements) of the set S.

5.2.4 Single Frame Feature

Spatial distortion in single frames is also a significant contributor to the perceived video quality. In addition to the previously described features based on frame differences, we add a feature (f_6) related to the quality of individual pictures. The NIQE [87] algorithm is used for this purpose in V-BLIINDS, however NIQE has no knowledge of human opinions, which are particularly important in our case, as we have already mentioned. The BRISQUE [86] IQA algorithm is more robust, as it does utilize perceptual information gathered from users to assess picture quality. Hence, we use it to compute f_6 as the time average of the quality score for each frame:

$$f_6 = \frac{1}{M}\sum_{t=1}^{M} BRISQUE(F^t) \tag{5.13}$$

5.2.5 Prediction Model

Several regression techniques can be used to learn a mapping between the defined NVS features and the human judgements in the underwater video database. Here, we use a Support Vector Machine Regressor (SVR) with a Gaussian (or radial basis function) kernel. SVRs have been extensively and successfully used for image and video quality assessment [83], [86], [91].

5.3 Performance Evaluation

5.3.1 Underwater Video Database

The main purpose of the proposed VQA algorithm is the evaluation of underwater video quality. Although underwater imagery has received wide attention by the scientific community, available databases are focused on still pictures and object detection [90], [92]. The dataset in chapter 3 is the only underwater video database which includes subjective scores and is suitable for quality assessment tasks. It contains 31 different scenes spanning a wide range of temporal and spatial variation. These contents are compressed in H.264 with several target average bit rates (between 8 and 20 kbps), frame rates (between 1 to 10 fps), color depth (8-bit greyscale and 24-bit RGB) and resolutions (QVGA and QQVGA upscaled to QVGA). These parameters were chosen considering the limited transmission capabilities of the underwater acoustic channel. Figure 5.1 shows some sample frames which illustrate the contents of the database. The original uncompressed (but resized) frame is also included to provide a better understanding of the compression and the level of distortion.

A total of 20 viewers took part on the subjective quality experiment. All of them were scientists from the Spanish Oceanographic Institute using underwater imagery in their research, but were unfamiliar with quality assessment tasks. A value of Mean Opinion Score (MOS) for each video was computed from their judgements and stored in the database. A detailed description of the experiment is given in Chapter 3.

(**a**)

(**b**)

Figure 5.1. Sample frames from the underwater video database: pristine **(a)** and distorted **(b)**.

5.3.2 Prediction Performance

The procedure used to evaluate the performance of the proposed algorithm is based on the correlation of the predictions with human scores. Since the performance metric for V-BLIIND and VIIDEO is also founded on this correlation, the comparison with these algorithms is direct. We split the video database into two non-overlapping sections: 70% of the elements were used for training and 30% of the elements were used for testing. During the training stage, the features $(f_1..f_6)$ and the MOS of the videos in the training set were fed to the SVR. The thresholds selected for features based on patched frame differences were set to u_1 = 1.8, u_2 = 2.2. This phase also includes a grid search optimization for two internal parameters of the SVR model: the kernel coefficient and the box constraint. Then, predictions for the test set (whose samples are unknown to the trained model) were computed, and the linear correlation coefficient (LCC) and the Spearman rank order correlation coefficient (SROCC) between the predictions and the subjective MOS values were calculated. This procedure was repeated 1000 times with random 70/30 splits of the database as cross-validation, and the median of the correlation coefficients was taken as the performance metric.

To illustrate the procedure, Figure 5.2 shows the scatter plot for the subjective quality scores against the predicted quality scores computed in 10 runs of the test phase (i.e. for

unknown samples). An approximate linear trend between the two variables can be observed.

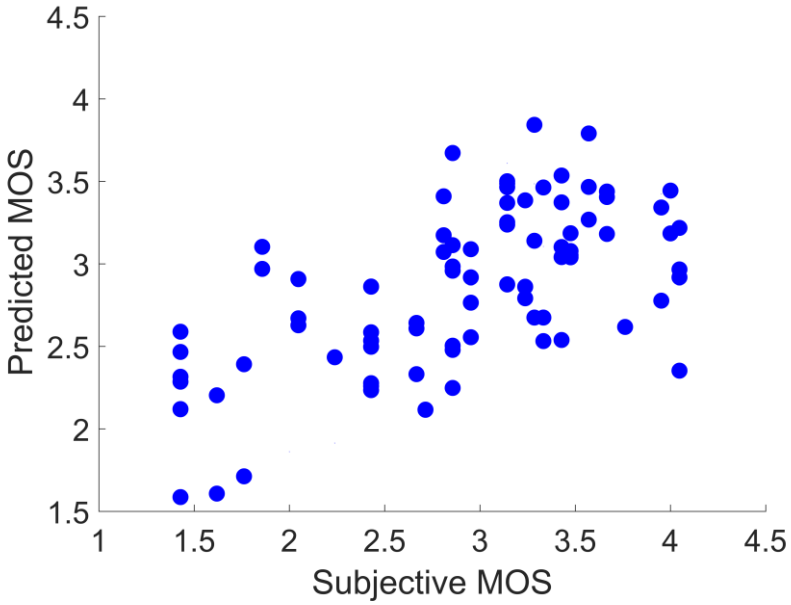

Figure 5.2. Scatter plot for the subjective quality scores against the predicted quality scores computed in 10 runs of the test phase.

The results of the performance evaluation are shown in Table 5.1. It also contains the correlation coefficients between the subjective scores and the predictions of the algorithms V-BLIINDS (trained on the LIVE database) and VIIDEO. In addition, since V-BLIINDS can be adapted to a new set of subjective scores, the median correlation coefficients for a retrained version of V-BLIINDS are also included. The procedure for this retraining was identical to the one already described for the proposed algorithm. It can be seen how VIIDEO and V-BLIINDS perform poorly, this was an expected result for V-BLIINDS since subjective perception in the underwater dataset is very different from perception on the LIVE database. The retrained V-BLIINDS (rV-BLIINDS in Table 5.1) shows an improved performance with a LCC and a SROCC of about 0.50, but it is still unsatisfactory. The proposed algorithm outperforms the three alternatives with larger correlation coefficients (LCC=0.81, SROCC=0.76).

Table 5.1. Linear and Spearman Correlation coefficients for subjective and predicted scores in the underwater video database.

Algorithm	**LCC**	**SORCC**
Proposed algorithm	0.81	0.76
V-BLIINDS	0.34	0.34
rV-BLIINDS [a]	0.50	0.49
VIIDEO	0.12	0.11

[a] retrained V-BLIINDS.

A parallel procedure was conducted to assess the weight of each group of features in the performance of the model. Three separate 1000-repetition training/testing procedures were executed using only one of the defined group of features (full frame differences, patched frame differences and single frames) in the prediction model. Table 5.2 contains the resulting median correlation coefficients for each single group of features. All of the groups performed better than the V-BLIINDS, rV-BLIINDS and VIIDEO algorithms (LCC=SORCC=0.60 in the worst case), and no individual set of features reaches the prediction power of the combination (in Table 5.1).

The results provided in this section show how the proposed algorithm can be successfully used for pixel-based video quality estimation with better performance than other existing machine learning approaches.

Table 5.2. Linear and Spearman Correlation coefficients for subjective and predicted scores using only one group of features (1000 repetitions of the training/testing procedure).

Features	**LCC**	**SORCC**
Full frame differences	0.64	0.60
Patched frame differences	0.71	0.62
Single frames	0.60	0.56

Chapter 6

Conclusions and future work

6.1 Conclusions

Quality assessment is an essential aspect of video service provisioning, but it turns out to be critical in highly constrained environments. It allows identifying the configuration parameters which make the difference between a useless service and a valuable one. This is the case for prospective underwater networks with a low available data rate, but also with the promising possibility of dramatically reducing costs of collecting images for scientific purposes. This thesis work tackles the whole video quality assessment problem addressing its two main aspects: subjective quality studies and objective quality estimation.

6.1.1 Subjective Quality Assessment

The first contribution presented focuses on subjective quality assessment. An experiment to gather quality data was designed and performed under the guidelines of the ITU P.910 recommendation. The Spanish Institute of Oceanography provided video sources from real underwater footage and a group of ocean scientists as evaluators.

The statistical processing of the collected data shows how the potential users of the video service rate conditions under test with grades between poor and good. A good number of conditions fall around the fair quality category, which supports the utility of this kind of video transmissions. Viewers often preferred 1 fps, grayscale sequences which obtained MOS > 3 (fair) with only one exception.

Although considered higher bitrates also tended to produce better opinion values, these differences can be considered a marginal enhancement depending on the frame rates being compared. This can be seen for the high variation content samples with 1 fps, where there is an improvement of only 16% when bitrate is increased by 150%. Furthermore, MOS values are considerably higher than those predicted by the G.1070 parametric model for similar video conditions.

The results lead to a better understating of video quality perception in underwater environments and could help to harness the existing technology to provide an effective instrument for oceanic research. Furthermore, the subjective quality information in this contribution allows for comparison of objective methods with actual human scores. This is a key piece for the objective quality estimation studies that constitute the second and third contribution in this work.

6.1.2 Objective Quality Assessment. Parametric models

The second contribution of this thesis work studies objective quality estimation. Not every quality estimation method is suitable for this particular problem, due to the peculiarities of underwater communications: nodes are difficult to reach once deployed, the limited bandwidth does not allow for an extra communication channel for measuring purposes, and energy saving restrictions prevent the use of intensive processing tasks.

This work exposes the unsuitability of the standardized ITU parametric method for underwater video quality estimation and presents two alternative models based on machine learning algorithms. These models are able to successfully accommodate the specific perception of quality revealed by subjective tests while taking into account the aforementioned special conditions. The first model is a parametric no reference estimation method and, therefore, only the evaluation of the model equation is required to predict MOS values. It shows a very good fit to the subjective data ($R^2 \approx 0.9$) and can be used for network planning applications but also to obtain a fast, lightweight processing estimation of the quality for real-time adaptation. The second model is a reduced reference method with a similar performance in terms of MOS prediction ($R^2 \approx 0.9$) but it further explores the concept of quality estimation. This technique, built upon ordinal logistic regression, is capable of predicting the distribution of user scores and thus provides a full characterization of quality beyond the simplistic common MOS statistic. This approach has not been previously applied to video quality assessment and delivers a more reliable way to assess user satisfaction and quality of experience.

6.1.3 Objective Quality Assessment. Pixel-based models

The third contribution is also focused on objective quality estimation, but, in this case, video processing tools are used to extract information from the video pixels. A new no-reference algorithm is described. The algorithm is based on Natural Video Statistics and focused on underwater video transmitted through acoustic networks. The number of features extracted for the evaluation is small and belong to a compact feature space for greater computational simplicity, which is a requirement for energy constrained underwater nodes. The quality estimation power of the proposed method shows good correlation against human scores (Linear Correlation ≈ 0.80, Spearman Rank Correlation ≈ 0.75), performing better on underwater video than state-of-the-art algorithms.

6.2 Future work

Underwater video services are just starting to be considered as a possibility for underwater networks. This dissertation can be considered as a point of departure for the vast research field that opens with the development of new technology and the growing interest for underwater exploration. Some suggestions for continuing this research are given below:

- Building extended underwater video quality databases is essential for the understanding of quality perception and the development and validation of quality estimation algorithms. Several of these databases are available for other applications (TV broadcasting, videoconferencing). The future work with quality databases is twofold:
    - Extended analysis of underwater video content. The new databases should contain contents for environments different than exploration expeditions, such as surveillance, monitoring of installations, flora or fauna, etc.
    - Extended analysis of underwater video perception. Additional subjective tests should be conducted to improve the availability of human scores. Different applications should also be considered and, thus, other professionals than ocean scientist should take part in the tests.
- Study of alternatives for underwater video capturing and compression. Current video technology has been developed for a recording environment where the air is the medium the light traverses between the objects and the sensor. Water has different properties and further research is needed to assess if that can be leveraged for a more efficient processing. Quality perception can be affected by potential changes in the capturing and compression mechanisms.
- Deep learning techniques are being applied to a wide variety of problems. Convolutional neural networks are starting to be applied to video quality assessment problems and underwater networks could also benefit from the developments on this field.

Appendix A

Resumen (Summary in Spanish)

A continuación, se presenta un resumen en español del contenido de esta tesis doctoral. La estructura de secciones de este resumen es idéntica la estructura de capítulos de la tesis. Las figuras y tablas no se reproducen aquí, pues, en general, su contenido no necesita traducción. Cuando sea necesario, se referenciará la imagen o tabla correspondiente por su numeración.

1. Introducción

Los océanos son una fuerza motriz en nuestro planeta. Influyen en el tiempo meteorológico, regulan la temperatura y son el soporte último de todos los organismos vivientes. La humanidad ha estado ligada a los océanos a lo largo de toda la historia. Ha utilizado sus aguas para transporte, comercio y alimento físico e intelectual. Los océanos cubren tres cuartas partes de la superficie de nuestro planeta y, sin embargo, se estima que sólo un 5% ha sido explorado.

Fotografías y vídeos son herramientas esenciales en el estudio del vasto y desconocido ecosistema oceánico que cubre el 90% de la biosfera en nuestro planeta. Sin embargo, la adquisición de imágenes subacuáticas conlleva en la actualidad un coste muy alto. En oceanografía son necesarias campañas de exploración con buques en las que buceadores o robots se sumergen en número limitado de sesiones de grabación. El caso de la ingeniería oceánica es similar y se requieren vehículos operados remotamente (ROVs) o autónomos (AUVs) para tareas como inspección de instalaciones, apoyo en la construcción de instalaciones, retirada de residuos y localización y recuperación de objetos.

El coste de este equipamiento y del personal necesario para gestionarlo es muy elevado. El despliegue de redes inalámbricas de sensores con capacidad para grabar y transmitir vídeo constituiría un avance tecnológico significativo a la vez que reduciría los costes de obtención de este tipo de información. Las aplicaciones de los servicios de vídeo subacuáticos podrían extenderse, más allá de la ciencia y la ingeniería, a áreas como la defensa, el turismo, la pesquería y la educación.

Las comunicaciones inalámbricas han experimentado una rapidísima evolución en las últimas décadas y los *smartphones* se han convertido en el paradigma de esa revolución. Sin embargo, todo ese desarrollo se ha hecho para comunicaciones basadas en el electromagnetismo como medio de propagación. Todo cambia cuando nos sumergimos en el agua. Las ondas electromagnéticas sufren una gran atenuación que las hacen útiles únicamente en corto alcance (pocos metros). Las ondas acústicas son la única alternativa viable para comunicaciones de medio rango (hasta unos pocos kilómetros) y aun así hay que tener en cuenta algunas características negativas de la propagación: la atenuación a altas frecuencias limita el ancho de banda disponible; la velocidad de propagación es lenta (en torno a 1500 m/s) y depende de la salinidad, la temperatura y la profundidad; la componente multicamino debida a reflexiones en la superficie y el fondo marino; y un ruido coloreado y no gaussiano. Los modems acústicos más recientes ofrecen tasas binarias de pico entre 32.2 y 64.2 kilobits por segundo (en rangos de 300 m a 1 km). Estas tasas de datos quedan reducidas por la operación de la red (acceso compartido, corrección de errores, sobrecarga de protocolos) dejando una tasa disponible para aplicación muy baja. Debido a esta limitación, las aplicaciones han estado restringidas tradicionalmente a telemetría y otros servicios de bajo volumen de datos.

Más allá de las dificultades ligadas a las comunicaciones acústicas, hay que considerar la logística de una red subacuática. En este tipo de redes, la localización de los nodos es virtualmente inalcanzable tras el despliegue. La recuperación de un nodo es posible, pero de un coste muy elevado. El impacto de esta situación es doble. En primer lugar, no es posible recolectar información de la señal original (que pudiera estar almacenada en un nodo). En segundo, la duración de las baterías debe extenderse el máximo posible puesto que el coste de reemplazarlas es muy alto.

Cuando se considera la importancia de las imágenes submarinas con las fuertes limitaciones de las comunicaciones en este entorno, surge una pregunta natural. ¿Es posible ofrecer servicios de vídeos en las condiciones de las redes acústicas subacuáticas? O, si enfocamos la pregunta desde el punto de vista de la ingeniería, ¿qué calidad podemos esperar de un servicio de vídeo bajo las restricciones impuestas por las redes acústicas subacuáticas?

La evaluación de calidad de vídeo (en inglés, *Video Quality Assessment* o VQA) ha sido una disciplina de interés durante décadas, desde las primeras difusiones de televisión hasta la reproducción de *clips* de Youtube en *smartphones*. El propósito de esta tesis doctoral es profundizar en la calidad de vídeo subacuático, estudiando las técnicas de evaluación, analizando datos y modelos de calidad existentes, su adecuación al escenario subacuático. Asimismo, se propone la obtención de nuevos datos y la construcción de nuevos modelos cuando los existentes no produzcan resultados satisfactorios.

2. Fundamentos de Evaluación de Calidad de Vídeo.

Los estudios de calidad de experiencia son un aspecto clave de la evaluación de rendimiento de todo servicio de telecomunicación que tiene a un usuario en el extremo de la comunicación. Es esencial para ofrecer un servicio de manera eficiente y optimizando los recursos a la vez que se maximiza la satisfacción del usuario. En esta sección se describen los dos enfoques a la evaluación de calidad (2.1), las peculiaridades

de la VQA para redes acústicas subacuáticas (2.2) y se proporciona información sobre las herramientas matemáticas utilizadas en este trabajo (2.3).

2.1 Tipos de Evaluación de Calidad de Vídeo

De acuerdo a la literatura existente, la calidad para un servicio de red puede evaluarse con dos enfoques: utilizando métodos subjetivos y utilizando métodos objetivos. Las dos técnicas tienen por objetivo producir una métrica estándar de calidad conocida como Puntuación de Opinión Media (en inglés, *Mean Opinion Score* o MOS). Esta métrica representa una calidad promedio observada, en el caso de los métodos subjetivos, o estimada, en el caso de los métodos objetivos. El rango habitual es numérico en el intervalo [1, 5] aunque pueden utilizarse otras escalas para aplicaciones específicas. A pesar de su simplicidad, la MOS se ha transformado en la medida de calidad de experiencia más extendida.

2.1.1 Evaluación de calidad subjetiva

El primer grupo de técnicas tiene por objetivo obtener valores de calidad directamente de evaluadores humanos. En VQA, una secuencia de estímulos (señales de vídeo) es presentada a un grupo de espectadores a los que se pide que puntúen los estímulos en una escala de calidad. Estas puntuaciones se procesan estadísticamente para calcular valor de MOS para diferentes condiciones de provisión de servicio. Las señales de vídeo en el experimento presentan diferentes tipos o niveles de degradación con respecto a la señal original tales como artefactos de compresión, desenfoque, pérdida de sensación de continuidad, etc. Esta degradación es producida por diferentes factores que pueden agruparse en dos categorías: degradación por compresión (debidos al códec, la tasa binaria de compresión, la tasa de cuadro, etc.) y degradación por transmisión (debidos a la tasa de pérdida de paquetes, la duración de las ráfagas, etc.). El análisis estadístico de los datos de calidad subjetivos puede encontrar relaciones entre la variación de los factores y la variación en la percepción de calidad.

La Unión Internacional de Telecomunicaciones (UIT) ha estandarizado varias metodologías para la realización de experimentos de calidad subjetiva. El procedimiento especificado en la BT.500 [12] ha sido utilizado desde hace décadas para diferentes propósitos, aunque fue diseñado para imágenes de televisión. Un estándar más reciente, aunque de contenido similar, es el descrito en la recomendación P.910 [13] para aplicaciones multimedia y ha sido utilizado como referencia en esta tesis. El estándar describe todas las partes del procedimiento de recogida de datos subjetivos incluyendo:

A. Grabación, almacenamiento y selección de la **señal fuente**.
B. **Métodos de test** para presentación de imágenes a espectadores y puntuación por parte de estos.
C. **Procedimientos de evaluación** referentes a las condiciones de visualización, los sistemas de procesado y reproducción, los espectadores y las instrucciones para los espectadores.
D. **Análisis estadístico e informe de resultados.**

Los detalles específicos del estudio de calidad subjetiva realizado en este trabajo de investigación se encuentran en la sección 3.

2.1.2 Evaluación de calidad objetiva

Los métodos subjetivos tienen una clara desventaja en términos de la cantidad de recursos necesarios. Debido a que requieren un grupo de personas y un tiempo considerable, los métodos subjetivos son una opción lenta y con un coste elevado. El segundo grupo de técnicas para evaluación de calidad persigue la estimación de calidad por medio de modelos matemáticos. Estos modelos pueden necesitar una base de datos de calidad subjetiva para ser construidos, pero después son capaces de estimar la calidad automáticamente, sin necesidad de espectadores. Por tanto, resuelven los principales problemas de los métodos subjetivos y han recibido atención por parte de la comunidad científica.

La clasificación más utilizada en la literatura para los métodos objetivos tiene en cuenta la información de entrada de la estimación y la presencia de una referencia, esto es, la señal original sin distorsiones. Tiene tres categorías:

A. Métodos de referencia completa (Full Reference o FR). Requieren de la señal original para calcular la estimación de calidad. Normalmente, ésta se calcula mediante una comparación entre la señal sin distorsiones y la degradada.
B. Métodos de referencia reducida (Reduced Reference o RR). No requieren de la señal original, pero sí de algunas características extraídas de ella. Habitualmente, el volumen de esta información es mucho menor y en ocasiones puede ser transmitida a través de la red (consumiendo parte del ancho de banda disponible).
C. Métodos sin referencia (No Reference o NR). Estos métodos no requieren de ninguna información de la señal original. Los métodos NR pueden hacer uso de la señal degrada completa (los basados en pixeles) o parte de ella como las cabeceras del flujo de vídeo (basados en la capa de paquetes) o simplemente algunos parámetros de la transmisión (conocidos como métodos de planificación, pues suelen usarse en la fase de diseño).

2.2 Retos en la evaluación de calidad de vídeo subacuático

Como se ha comentado en la sección 1, las redes submarinas constituyen un entorno particularmente desafiante para las transmisiones de vídeo. Esto incluye la evaluación de calidad y deben hacerse algunas consideraciones específicas.

2.2.1 Retos en la VQA subjetiva

Los servicios de vídeo en redes submarinas tienen un público muy específico: oceanógrafos, operadores de empresas que gestionan recursos oceánicos y especialistas en seguridad entre otros. No sólo es más difícil encontrar un grupo apropiado de evaluadores, también ocurre que cada profesional tendrá una percepción diferente de la calidad dependiendo de las tareas que realiza habitualmente con el vídeo.

Los autores de [19] proponen una metodología subjetiva mediante crowdsourcing, es decir, utilizar Internet como un "laboratorio virtual" en el que es más fácil la participación. Sin embargo, la metodología estándar requiere un control de las

condiciones de realización del experimento de evaluación que no pueden conseguirse si los espectadores acceden al test a través de internet. Aún está por estudiar la influencia de esta variabilidad de condiciones de entorno en los valores de calidad percibida.

2.2.2 Retos en la VQA objetiva

En el caso de la evaluación objetiva, hay algunos aspectos importantes de las redes acústicas subacuáticas a tener en cuenta:

- Ubicación de los nodos. No es posible acceder a los nodos para recuperar la señal original.
- Tasa binaria muy limitada.
- Ahorro de energía. Dado que las baterías son virtualmente irreemplazables, la evaluación de calidad debe consumir la menor cantidad de recursos energéticos posible.

Los métodos de referencia completa no son adecuados puesto que necesitan la señal original, cuya recuperación no es generalmente factible. Los métodos de referencia reducida pueden ser útiles, siempre que la tasa de bit necesaria para enviar la información adicional para evaluación de calidad sea suficientemente pequeña. Esta consideración, válida para cualquier tipo de red si se quiere que el método interfiera mínimamente con la prestación del servicio, es crítica en redes subacuáticas donde la tasa de transmisión es muy reducida. Por ejemplo, los métodos descritos en la recomendación J.249 de la UIT [18] imponen una sobrecarga de entre 15 y 256 kbits/s lo que los hace inviables para esta aplicación. Los métodos sin referencia no necesitan información de la señal original sin distorsión lo que los hace especialmente interesantes para la evaluación de calidad de vídeo en redes subacuáticas.

2.3 Herramientas matemáticas para la evaluación de calidad de vídeo

Esta sección proporciona una breve introducción a las herramientas matemáticas utilizadas en esta tesis doctoral en el contexto de la evaluación de calidad de vídeo. Su objetivo es facilitar la comprensión de algunos términos y la manera en que son utilizados en el desarrollo de las contribuciones de este trabajo.

2.3.1 Tests estadísticos y análisis de varianza

La evaluación de calidad subjetiva se ha descrito como el proceso de obtener información de calidad directamente de usuarios humanos. El estudio de este comportamiento no puede ser descrito con ecuaciones o simulado (en tal caso estaríamos hablando de evaluación calidad objetiva). Por tanto, es un ámbito intrínsecamente experimental en el que se debe utilizar la inferencia estadística para diseñar experimentos y analizar los resultados. La inferencia nos permite extraer conclusiones significativas de los resultados aceptando (o no) con un alto margen de confianza que la variabilidad en los datos obtenidos se debe a los factores bajo estudio. En concreto, el Análisis de Varianza clásico (ANOVA) [24] nos permite aceptar si la diferencia entre las medias de una medida en varios grupos debe considerarse estadísticamente significativa o si por el contrario no hay evidencia suficiente para considerar que las medias son diferentes. En este trabajo, se

hace uso de la versión intra-sujetos con múltiples vías del ANOVA para estudiar las medias de la puntuación de calidad (MOS) variando factores tales como la tasa de transmisión o la tasa de cuadro.

2.3.2 Aprendizaje Máquina

Se conoce como aprendizaje máquina a "la programación de un ordenador para que se comporte de un modo que, si fuese realizado por humanos o animales, sería descrito como un proceso de aprendizaje" (traducido de [25]). Podríamos decir que el comportamiento del algoritmo programado no sólo depende una serie de "reglas", sino también de un conjunto de "ejemplos". Esta disciplina está estrechamente relacionada con la estadística: las técnicas de regresión, en las que se construye un modelo (comportamiento) a partir de una nube puntos (ejemplos), son consideradas pertenecientes al ámbito del aprendizaje máquina. Estos modelos reciben el calificativo de predictivos, dado que se utilizan con frecuencia para estimar el comportamiento de un sistema. En este contexto, la palabra "predicción" se utiliza de manera intercambiable con el término "estimación". En evaluación de calidad objetiva, diremos que un modelo es capaz de predecir la calidad en el sentido de que es capaz de estimar cómo los usuarios del servicio lo puntuarían.

2.3.3 Estadísticos de escenas naturales y el sistema visual humano

El trabajo publicado en [26] propone una analogía muy expresiva entre un sistema de comunicaciones y el problema de la evaluación de calidad. Nuestro mundo visual se describe como un "transmisor". Las propiedades físicas de la materia y de la luz generan una señal que puede ser capturada por sensores, digitalizada, procesada, almacenada o transmitida y mostrada en una pantalla. Este sistema introduce distorsiones y por tanto equivale al "canal". Finalmente, el "receptor" es el sistema visual humano que forma la imagen percibida en el cerebro. Como en cualquier otro sistema de comunicaciones, el modelado de estos elementos nos ayuda a comprenderlo mejor.

La teoría de estadísticos de escenas naturales (en inglés, *Natural Scene Statistics* o NSS) afirma que las imágenes que se originan de la captura de nuestro mundo (escenas naturales) poseen regularidades estadísticas. Estas regularidades no están presentes en otras imágenes que se alejan de las naturales tales como los gráficos generados por computador. También las imágenes distorsionadas por el canal de nuestra analogía sufren esta desviación de los estadísticos con respecto a las escenas naturales. Un modelo NSS interesante [27] asume que, si las frecuencias espaciales más bajas son eliminadas de una imagen natural, los valores de los píxeles siguen una distribución de mezcla de gaussianas. La importancia de este modelo se entiende mejor cuando se contrasta con los descubrimientos sobre el funcionamiento del sistema visual humano. Algunos estudios [28], [29] consideran que la arquitectura de las neuronas involucradas en el sistema visual temprano ha evolucionado para codificar y analizar de manera eficiente imágenes del mundo real, es decir, imágenes que exhiben las propiedades estadísticas propuestas por los modelos NSS. Integrando estas propuestas, podemos definir la calidad como fidelidad al mundo real y, por tanto, identificar buena calidad con imágenes que preservan las propiedades de NSS. De la misma manera, la percepción de una mala calidad está relacionada con la desviación de la regularidad estadística que presentan las imágenes naturales.

3. Evaluación de calidad subjetiva para vídeo subacuático

La evaluación subjetiva de la calidad está basada en experimentos con espectadores humanos. La finalidad es obtener información sobre la calidad percibida pidiendo a los participantes que puntúen cortes de vídeo. Tras el experimento, el procesado estadístico de estos datos permite obtener información relevante sobre la percepción de calidad. En esta sección se presenta una revisión de la literatura existente en el área (3.1), la descripción del experimento realizado en este trabajo de investigación (3.2) y los resultados del experimento junto con su análisis estadístico (3.3).

3.1 Revisión de la literatura

La importancia de la evaluación de calidad de vídeo subjetiva está respaldada por numerosos estudios realizados para diversas aplicaciones como vídeo de propósito general [34], vídeo para móviles [35] o "streaming" basado en HTTP [36]. Algunos de los estudios existentes se centran en el efecto sobre la calidad de factores concretos como la tasa de pérdida de paquetes [38], la distancia a la pantalla del espectador y la resolución [39] o el hecho de que se trate de una aplicación profesional tal y como la telemedicina [40]. Sin embargo, todos los trabajos citados utilizan tasas de transmisión mucho más altas que las alcanzables en redes subacuáticas y, por tanto, parten de un servicio de vídeo notablemente diferente. Esto justifica la primera contribución de esta tesis, un experimento para la obtención de datos de calidad subjetiva en el contexto de prestación de servicio de una red subacuática.

3.2 Experimento para la evaluación de calidad

3.2.1 Servicio de referencia

Los servicios de vídeo que se estudian en este experimento pueden corresponder a redes inalámbricas de sensores con nodos anclados para monitorización de ecosistemas submarinos o con robots autónomos para la exploración del fondo marino. Los modems acústicos más recientes alcanzan una tasa de datos pico de 62.5 kbps en un rango de 300 m [43]. Considerando u estudio previo [44], la selección de parámetros de vídeo para el experimento son:

Tasas binarias: 8, 14, 20 kbps.
Tasas de cuadro: 1, 5, 10 kbps.
Resoluciones: 320 x 240 píxeles (QVGA) y 160 x 120 píxeles (QVGA).
Profundidad de color: RGB (3 canales de 8 bits por canal) y escala de grises (8 bits).

En el estudio previo se observó que los contenidos se hacen difíciles de distinguir por debajo de 8 kbps, incluso con resoluciones bajas. Una tasa de transmisión efectiva por encima de 20 kbps es poco realista debido a la sobrecarga de los protocolos de red y los mecanismos de acceso compartido entre varios nodos. En ese intervalo de tasas de bit, una tasa de cuadro estándar de 25 fps produce imágenes con una distorsión demasiado

alta. Por el mismo motivo, las resoluciones de los vídeos en el experimento son también relativamente bajas (QVGA, QQVGA).

3.2.2 Entorno de grabación y señal fuente.

El metraje de vídeo para el experimento fue proporcionado por el Instituto Español de Oceanografía (IEO). Los archivos fueron capturados en campañas de exploración en el marco del proyecto "Life+Indemares-Chimeneas de Cádiz" [45] con el vehículo subacuático VOR APHIA 2012, un prototipo desarrollado por el grupo de investigación GEMAR (IEO). El vehículo incluye su propio sistema de iluminación y una cámara de vídeo digital Canon Legria HF R106. Las imágenes grabadas por este sistema son de calidad suficiente como para considerarlas sin distorsión a efectos del experimento de calidad subjetiva.

3.2.3 Selección de escenas

Un total de 56 escenas de 12 segundos de duración fueron elegidas como potenciales secuencias en el test de calidad. Su contenido varía desde tomas casi estáticas de un fondo marino prácticamente plano a otras con rápida navegación de áreas con rocas irregulares y diferentes especies de fauna y flora submarina. Estas escenas son clasificadas en dos grupos de acuerdo con las métricas de variación de información perceptual definidas en la recomendación de la UIT (ecuaciones 3.1 y 3.2). La figura 3.2 muestra el plano de variación espacial (SI) y temporal (TI) de las muestras de vídeo. El umbral de separación es la mediana de la variable con mayor dispersión (SI), obteniendo de esta manera dos grupos: uno de contenido de alta variación y otro de baja variación.

3.2.4 Método de test

El método utilizado para evaluación de calidad de vídeo subjetiva es el de calificación en categorías absolutas (en inglés, *Absolute Category Rating* o ACR) descrito en la recomendación UIT P.910. Dicho método emplea una escala categórica en cinco niveles de calidad: mala, pobre, moderada, buena y excelente. Las escenas se muestran en un orden aleatorio (diferente para cada participante) en una secuencia como la de la figura 3.3. Los usuarios disponen de unos segundos entre escena y escena para calificar la que acaban de ver. Además, una vez finalizado el test, se pidió a los usuarios que calificaran la utilidad de cada categoría en la escala de calidad en una escala, también categórica, de cinco niveles: sin utilidad, poco útiles, moderadamente útiles, bastante útiles y muy útiles. Esta pregunta (no incluida en el procedimiento ACR estándar) se diseñó para ir más allá de la calidad como concepto abstracto y relacionarla con otra medida, también subjetiva, pero más específica.

3.2.5 Procedimiento de evaluación

El sistema de reproducción utilizado fue una aplicación HTML5 desarrollada específicamente para este propósito. La pantalla de visualización era de 14'' con una resolución WXGA. La ventana de visualización de vídeo tenía un tamaño constante de 320x240 píxeles (para ambas resoluciones en el test), es decir 3.57'' de diagonal. Las condiciones de iluminación fueron controladas con paneles orientables y mantenidas constantes a lo largo del test mediante mediciones con un fotómetro (tabla 3.1). Un total

de 21 espectadores participaron en el test, todos ellos oceanógrafos de diferentes especialidades del Centro Oceanográfico de Málaga (IEO).

3.2.6 Condiciones en el test

Cada condición de evaluación es una escena de vídeo con una combinación de las variables del test. El test consiste en 31 condiciones de evaluación organizadas en cinco bloques como sigue:

- Bloque 1: condiciones de estabilización. Tienen por objetivo familiarizar a los usuarios con el procedimiento y no son tenidas en cuenta en los resultados.
- Bloque 2: condiciones para medir el impacto de los tres niveles de tasa de bit y los tres niveles de tasa de cuadro con contenido de baja variación. La resolución y el color se fijan en QVGA y RGB, respectivamente.
- Bloque 3: condiciones para la medida del impacto de los dos niveles de resolución y los dos niveles de color con contenido de baja variación. La tasa de bit y la tasa de cuadro se fijan en 20 kbps y 5 fps, respectivamente.
- Bloque 4: condiciones para medir el impacto de los tres niveles de tasa de bit y los tres niveles de tasa de cuadro con contenido de alta variación. La resolución y el color se fijan en QVGA y RGB, respectivamente.
- Bloque 5: condiciones para la medida del impacto de los dos niveles de resolución y los dos niveles de color con contenido de baja variación. La tasa de bit y la tasa de cuadro se fijan en 20 kbps y 5 fps, respectivamente.

La tabla 3.2 detalla la configuración de las condiciones de evaluación. La figura 3.4 contiene algunos cuadros de muestra para ilustrar el contenido de los vídeos en el test.

3.3 Resultados y discusión

La puntuación media (MOS) para cada condición es el principal estadístico para la representación de la calidad. También se han calculado otros estadísticos, y se ha realizado un análisis de varianza (ANOVA) para comprobar la significancia de las diferencias entre las medias producidas por los diferentes factores. La figura 3.5 presenta la MOS para todas las condiciones con un intervalo de confianza del 95%. Las barras asociadas muestran la distribución acumulada de puntuaciones en tres grupos: porcentaje de calificaciones como buena o excelente (azul), porcentaje de calificaciones como moderada (rojo) y porcentaje de puntuaciones como pobre o mala (verde). En general, dada una tasa de bit fija, la calidad es más alta con tasas de cuadro más bajas. Lo contrario ocurre para tasas de cuadro fijas, la calidad aumenta con la tasa de bit. El análisis de varianza revela estas diferencias como significativas. En cuanto a la resolución y el color, se obtienen mejores resultados con imágenes en escala de grises y mayor resolución, aunque las diferencias son menores y además, en este caso, el análisis de varianza no permite concluir que sus diferencias sean significativas. Los resultados detallados del ANOVA se encuentran en la tabla 3.3.

Los datos de utilidad se utilizaron para calcular una regresión linear con la MOS. La ecuación 3.3 y la figura 3.6 muestran como existe una tendencia fuertemente lineal entre

estas dos variables. Puede decirse entonces que en el caso del vídeo subacuático para uso científico hay una identificación entre utilidad y calidad.

Los resultados de este experimento muestran cómo los servicios de vídeo analizados alcanzan una calidad que los hace viables. Incluso con las importantes limitaciones en los parámetros de configuración, los espectadores calificaron un buen número de condiciones de evaluación como moderadas, algunas de ellas incluso alcanzando una valoración de buena calidad. Los resultados permiten la planificación de servicios con tasas binarias tan bajas como 8 kbps en la capa de aplicación y una calidad esperada de 3 sobre 5 en la escala de MOS. Por otra parte, la mayoría de puntuaciones de utilidad entran en las categorías de "utilidad moderada" y "muy útil". Aunque es algo por demostrar, es posible que el origen de estas puntuaciones relativamente altas en un vídeo que para otras aplicaciones no se consideraría esté en la ventaja que una red submarina supone sobre los métodos actuales de obtención de imágenes en este entorno.

4. Evaluación de calidad objetiva para vídeo subacuático. Método paramétrico

Cuando se requieren medidas repetidas de calidad, el análisis de la calidad subjetiva es una tarea pesada y costosa. La evaluación de calidad objetiva pretende obtener información de calidad de una manera automatizada, sin la intervención de espectadores humanos. En esta sección se presentan dos modelos paramétricos para estimación de calidad de vídeo subacuático. El primero está basado en técnicas de ajuste de superficies y genera una estimación computacionalmente muy ligera basada en la tasa de bit y de cuadro. El segundo modelo va más allá de la MOS como medida de calidad y calcula una distribución completa de puntuaciones utilizando, además de los parámetros del modelo anterior, dos características del contenido del vídeo.

4.1 Revisión de la literatura

La mayoría de los métodos de VQA objetiva se centran en estimar la MOS. Algunos de estos métodos pueden encontrarse en la literatura [42], [54]-[62]. El estudio publicado en [41] realiza un análisis comparativo de todos ellos concluyendo que el mejor método para degradación por transmisión es el propuesto en el estándar G.1070 de la UIT y el mejor para degradación por compresión es el propuesto en [62]. Sin embargo, estos métodos están diseñados para vídeo que no contempla las fuertes restricciones de parámetros del contexto de las transmisiones submarinas.

Por otro lado, algunos estudios ya han considerado la MOS insuficiente como medida de calidad. Un enfoque interesante se ofrece en [67] donde se muestra como la MOS esconde información relevante y se propone un modelo con estadísticos adicionales. A pesar de ello, este estudio se refiere a la calidad de experiencia como un concepto más amplio y sólo incluye alguna información sobre servicios de vídeo como caso de uso que, además, no se adecúa a las restricciones del modelo. Otro estudio digno de mención sobre medidas de calidad más allá de la MOS ha sido publicado en [68]. Los autores usan técnicas de

aprendizaje máquina para construir un modelo de predicción de las nuevas métricas propuestas: el grado de aceptación general y el grado de aceptación como agradable. En los estudios mencionados, no se tienen en cuenta los reducidos recursos de las redes submarinas ni las diferencias de percepción propias del vídeo científico resultantes del estudio presentado en la sección 3.

4.2 Base de datos de calidad subjetiva

La base de datos utilizada en esta sección es un subconjunto de la presentada en el experimento de la sección 3. El resumen de las características de los cortes de vídeo seleccionados para los modelos paramétricos se encuentra en la tabla 4.1. Están agrupados en dos bloques: uno de contenido de baja variación espacio-temporal (en inglés, *Low Variation Content* o LVC) y uno de contenido de alta variación espacio-temporal (en inglés, *High Variation Content* o HVC). Además, se considera un bloque alternativo de baja variación (rLVC) en el que se han eliminado los contenidos con identificador 06 y 07 partiendo de la hipótesis de que su contenido particular puede haber desviado su MOS.

4.3 Estudio de viabilidad del modelo ITU-T G.1070

El modelo para evaluación de calidad de vídeo propuesto en [42] calcula la MOS con un conjunto de ecuaciones que toman como parámetro la tasa de bit, la tasa de cuadro y la tasa de pérdida de paquetes. Los coeficientes del modelo deben ajustarse basándose en otras variables de servicio: el códec de compresión, la resolución y el tamaño de visualización. La figura 4.1 muestra los valores de MOS estimados para unas condiciones de visualización similares a las de la base de datos de calidad subjetiva. Los valores de calidad se encuentran en el intervalo entre 1 y 1.3 lo que difiere considerablemente de los datos de calidad subjetiva obtenidos. El anexo A de la recomendación especifica cómo obtener un conjunto de coeficientes a partir de datos subjetivos. Los resultados se muestran en las tablas 4.2 y 4.3. El modelo no permite un ajuste consistente pues se obtienen coeficientes complejos de parte imaginaria no despreciable (LVC) o unos parámetros inaceptables de la bondad del ajuste (en inglés, *Goodness of Fit* o GOF).

4.4 Modelo sin referencia (NR) paramétrico

4.4.1 Desarrollo del modelo

Para encontrar un modelo apropiado a los datos subjetivos se ha utilizado la interpolación "thin plate spline". Este método minimiza la función energía (ecuación 4.5) y proporciona un ajuste perfecto para los puntos de control. Al mismo tiempo produce una superficie suave (minimimante "doblada") que encaja con la hipótesis de que los valores de calidad no varían bruscamente en el dominio considerado para las variables de entrada: la tasa de bit y la tasa de cuadro. La superficie se define en la ecuación 4.6 como una suma ponderada de funciones de base radial (ecuación 4.7). La figura 4.2 muestra tres gráficas de superficie que ajustan la base de datos subjetiva. Las sub-figuras a, b y c corresponden

respectivamente con los bloques HVC, LVC y rLVC. Se observa cómo las superficies de los bloques HVC y rLVC son muy similares entre sí, compartiendo la forma sigmoide que motiva el modelo propuesto a continuación.

4.4.2 Ecuaciones del modelo

La primera propuesta es un modelo de regresión no lineal (en inglés, *Non Linear Regression* o NLR) basado en la función logística generalizada (ecuación 4.8). Los coeficientes de esta función son relativamente fáciles de interpretar geométricamente. El modelo extiende la función logística generalizada a dos dimensiones y se proporcionan dos variantes de acuerdo con diferentes objetivos de optimización:

- Variante de Generalización (NLR.G). Ecuación 4.9. Las asíntotas de la superficie se fijan a los límites de la escala de calidad (1, 5) por lo que este modelo es más útil si se pretende utilizar fuera de los límites de los datos subjetivos.
- Variante de precisión (NLR.A). Ecuación 4.10. Alcanza una mayor bondad de ajuste.

Los parámetros se optimizan con el método de los mínimos cuadrados (versión no lineal) y los resultados se muestran en las tablas 4.4 y 4.5 para cada variante. Los estadísticos de bondad de ajuste se muestran en las tablas 4.6 y 4.7. La figura 4.3 contiene gráficas de superficie para cada variante y grupo de contenido. Utilizando el parámetro R^2 para medir la bondad del ajuste (donde $R^2=1$ indica un ajuste perfecto), puede observarse un ajuste muy bueno para los grupos HVC y rLVC con $R^2 > 0.9$ en casi todas las combinaciones.

4.5 Modelo híbrido de referencia reducida

A pesar de las ventajas del modelo NLR, éste puede resultado demasiado simplista según la aplicación. Como alternativa, se propone un segundo modelo paramétrico basado en regresión logística ordinal (en inglés, *Ordinal Logistic Regression* o OLR). Este método de clasificación es adecuado para problemas multiclase en los que existe una ordenación natural entre las clases como es el caso de la evaluación de calidad de vídeo. En este caso, no se realiza la división en bloques según la cantidad de variación y la estimación se realiza en función de la tasa de bit, la tasa de cuadro, la variación espacial y la variación temporal. El método OLR se basa en la asunción de las cuotas proporcionales expresadas en la ecuación 4.11. Las estimaciones calculadas con este modelo no producen un único valor de opinión media (MOS) sino una distribución completa de proporciones de puntuación para cada categoría, es decir, la proporción de usuarios que calificarían la calidad del vídeo como excelente, buena, moderada, pobre o mala. A efectos de comparación, la MOS puede calcularse fácilmente a partir de esta información, como se indica en la ecuación 4.13.

El modelo ha sido ajustado con estimación de máxima verosimilitud con el software IBM SPSS Statistics [48] que también se utiliza para el análisis de resultados. La construcción del modelo sigue un procedimiento iterativo con el objetivo de capturar todas las

interacciones significativas entre las variables de entrada. Dicho procedimiento puede describirse con el siguiente bucle de dos pasos:

Para i = número_de_características a i = 1:

1. Calcular un modelo incluyendo todas las interacciones posibles excepto las que han sido descartadas en una iteración previa.
2. Comprobar el p-*value* de todos los términos de interacción de orden i-ésimo (o los efectos principales si i = 1). Si $p > 0.05$, la interacción se considera no significativa y el término se descarta.

La tabla 4.8 muestra el resultado de aplicar este procedimiento al conjunto de datos subjetivo. La tabla 4.9 recoge los resultados de dos tests que verifican la adecuación del modelo. La tabla 4.10 contiene los parámetros bondad de ajuste que verifican la calidad del ajuste tanto para la distribución de puntuaciones como para la MOS calculada a partir de ella. La figura 4.4 muestra la distribución real de puntuaciones en el conjunto de datos subjetivo junto a las estimaciones calculadas por el modelo. Se observa como un 71.1% de las proporciones estimadas tiene una desviación menor de 0.1 con respecto a las observadas. Además, si se selecciona la categoría de calidad con mayor proporción de calificaciones como la mejor estimación categórica de calidad, la precisión de la clasificación es del 83.3%.

5. Evaluación de calidad objetiva para vídeo subacuático. Método basado en píxeles.

Cuando se requiere una monitorización continua de la calidad, los métodos basados en píxeles y procesado de imagen y vídeo pueden proporcionar una mejor estimación de la calidad. El método propuesto en esta sección utiliza técnicas de aprendizaje máquina para aprender de las características de las imágenes en un vídeo y relacionarlas con las puntuaciones emitidas por humanos.

5.1 Revisión de la literatura

Aunque existen abundantes métodos sin referencia basados en procesado de imagen para la evaluación de calidad de imágenes estáticas, no existe la misma profusión para la evaluación de calidad de vídeo. El algoritmo V-BLINDS [83] muestra un buen rendimiento en términos de correlación con puntuaciones subjetivas y ha sido probado con dos bases de datos diferentes la “LIVE VQA” [34] y la EPFL-Polimi [84]. Además, V-BLINDS puede reentrenarse para otras bases de datos. Otro método recientemente publicado es VIIDEO [85] que intenta superar la dependencia de puntuaciones subjetivas. Sin embargo, estos métodos están diseñados para tipos específicos de distorsión o no incluyen información sobre opiniones de espectadores humanos o involucran la extracción de un gran número de características.

5.2 Método sin referencia para la evaluación de calidad de vídeo subacuático

El modelo propuesto está basado en la teoría de estadísticos de escenas y vídeos naturales (en inglés, *Natural Scene Statistics* o NSS). Esta teoría afirma que las imágenes y vídeos sin distorsión exhiben ciertas propiedades estadísticas que se pierden cuando el contenido sufre distorsión. También asume que el sistema visual humano ha evolucionado de acuerdo a esas características y, por tanto, son relevantes en la percepción visual.

5.2.1 Fundamentos del modelo

El modelo espacial de NSS en el que se basa el método de evaluación de calidad propuesto se describe en las ecuaciones 5.1 a 5.6. En resumen, los coeficientes de la operación de normalización divisiva sobre una imagen siguen una distribución gaussiana cuando la imagen es natural [27]. Cuando la imagen sufre distorsión, la separación de la gaussianidad puede modelarse con la distribución gaussiana generalizada con media cero. El parámetro α de la distribución define su forma, un valor de 2 corresponde con una distribución gaussiana. Otros valores del parámetro indican separación de la gaussianidad y, por tanto, distorsión. Este modelo ha sido utilizado con éxito para evaluar la calidad de imágenes estáticas [86], [87], [88]. Se ha demostrado que la diferencia de cuadro (ecuación 5.7) tiene propiedades similares a las descritas para las imágenes naturales y ha sido utilizada para medir la distorsión en el dominio temporal [83], [85].

5.2.2 Características y modelo de predicción

El modelo que se propone a continuación aprovecha las propiedades comentadas y extrae seis características (en inglés, *features*) de los cuadros del vídeo que se utilizan para estimar la calidad. Las ecuaciones 5.8 a 5.13 describen cómo obtener las características. La primera de ellas (f_1) se obtiene calculando la media temporal del parámetro α calculado sobre todas las diferencias de cuadro normalizadas. Debido a que las imágenes contienen información en múltiples escalas, la segunda característica (f_2) es idéntica a la primera, pero se calcula sobre una versión diezmada por un factor de 2 de la diferencia de cuadro normalizada. Una componente importante de la distorsión ocurre de manera local, por lo que las características tercera a quinta (f_3 a f_5) utilizan subregiones de las diferencias de cuadro normalizadas. Para cada subregión se calcula el ajuste a la distribución gaussiana generalizada y el parámetro α. Los valores obtenidos se agrupan en tres conjuntos de acuerdo a dos umbrales y las características se calculan como el cardinal de cada conjunto normalizado por el número de diferencias de cuadro. La última característica (f_6) tiene en cuenta que la distorsión espacial en cuadros individuales es otra componente importante de la percepción de la distorsión, por tanto, se utiliza el algoritmo BRISQUE [86] para calcular el promedio de calidad para todos los cuadros individuales en el vídeo. Como modelo de predicción se utiliza una regresión con máquina de vectores de soporte (en inglés, *Support Vector Machine Regressor* o SVR) con un núcleo de función de base radial. El entrenamiento y evaluación se describen en la sección 5.3.

5.2.3 Evaluación de rendimiento

Para la evaluación del rendimiento del algoritmo se utilizan todas las escenas incluidas en la base de datos de puntuaciones subjetivas descrita en la sección 3. El procedimiento

utilizado se basa en la correlación con las puntuaciones subjetivas. La base de datos se divide en dos partes, el 70% de los elementos se utilizan para el entrenamiento del modelo SVR y el 30% restante para las pruebas de rendimiento. En el entrenamiento el SVR se alimenta con las seis características descritas (entradas) calculadas para cada vídeo del conjunto de entrenamiento y la MOS correspondiente (salida). Para la evaluación, se calculan las predicciones de calidad para los vídeos en el conjunto de test (que son desconocidos para el modelo) y se calculan el coeficiente de correlación de Pearson y el coeficiente de correlación de Spearman entre las estimaciones y las puntuaciones subjetivas. Como técnica de validación cruzada, este procedimiento se repite 1000 veces con particiones 70/30 aleatorias. La mediana de cada coeficiente de correlación se utiliza como métrica de rendimiento. Para ilustrar el procedimiento, la figura 5.2 muestra una gráfica de dispersión para las puntuaciones estimadas frente a las subjetivas. Por claridad, sólo se muestran 10 repeticiones, pero ya puede observarse una tendencia aproximadamente lineal.

Los resultados de la evaluación de rendimiento se encuentran en la tabla 5.1. Además de las medianas de los coeficientes de correlación del algoritmo propuesto se han calculado las métricas análogas para los algoritmos VIIDEO, V-BLINDS y una versión de este último reentrenada con la base de datos subjetiva de vídeo subacuático que se utiliza en este trabajo. Puede verse cómo el algoritmo propuesto supera las tres alternativas con coeficientes de correlación más altos (CC Pearson = 0.81, CC Spearman = 0.76).

Para evaluar el peso de cada grupo de características se llevó a cabo un procedimiento similar. Se ejecutaron tres procesos de 1000 repeticiones utilizando sólo uno de los grupos de características. La tabla 5.2 contiene los resultados de esta comparación. Se observa que todos los grupos mejoran la predicción de los modelos alternativos, aunque ninguno de ellos alcanza el poder de predicción de la combinación.

6. Conclusiones y líneas futuras

La evaluación de calidad es un aspecto esencial de la provisión de servicios de vídeo. Es especialmente crítica en entornos con capacidades de transmisión muy restringidas puesto que permite identificar los parámetros de configuración que suponen la diferencia entre un servicio inútil y uno de valor. Este es el caso de las redes acústicas submarinas con una tasa de transmisión disponible muy baja, pero también con la posibilidad de reducir en gran medida los costes de la recolección de imágenes. Este trabajo de tesis aborda el problema completo de la evaluación de calidad de vídeo en sus dos principales áreas: los estudios de calidad subjetiva y la estimación objetiva de calidad.

La primera contribución presentada en este trabajo se centra en la percepción subjetiva de la calidad. Se diseña y ejecuta un experimento siguiendo la recomendación UIT P.910. El Instituto Español de Oceanografía proporcionó los archivos fuentes de vídeo y un grupo de científicos oceanógrafos como evaluadores. El procesado estadístico de estos datos muestra cómo los potenciales usuarios del vídeo valoran un buen número de las condiciones de prueba en torno a la categoría de calidad moderada, lo que apoya el uso de este tipo de transmisiones. Los espectadores prefieren escenas con un cuadro por

segundo en escala de grises. Este tipo de escenas obtienen puntuaciones de calidad media por encima de 3 en una escala del 1 al 5 con una única excepción. Aunque mayores tasas binarias producen valores de calidad más altos, esta mejora puede considerarse marginal dependiendo de las tasas de cuadro que se comparen. Estos resultados conducen a una mejor comprensión de la calidad de vídeo en entornos subacuáticos y ayudan al mejor aprovechamiento de la tecnología existente como un instrumento efectivo en la investigación oceánica. Además, la información de calidad subjetiva de esta contribución es fundamental para la estimación objetiva de la calidad que constituye el núcleo de las siguientes contribuciones en este trabajo.

No todos los métodos de estimación de calidad son adecuados para el problema que plantean las redes subacuáticas: nodos difíciles de alcanzar una vez desplegados, ancho de banda limitada y restricciones energéticas. La segunda contribución de esta tesis expone la falta de adecuación del método paramétrico estandarizado por la UIT para estimación de calidad de vídeo y presenta dos alternativas basadas en algoritmos de aprendizaje máquina. Estos modelos son capaces de adaptarse con éxito a la percepción subjetiva a la vez que tienen en cuenta las condiciones especiales antes mencionadas para redes subacuáticas. El primer modelo no requiere referencia y muestra un ajuste muy bueno a los datos subjetivos ($R^2 \approx 0.9$). El segundo modelo utiliza una referencia reducida y tiene un rendimiento similar en términos de MOS. Sin embargo, el método va más allá de la estimación de la puntuación media y es capaz de predecir la distribución de puntuaciones. Este enfoque no ha sido aplicado antes a la evaluación de calidad de vídeo y proporciona un modo más fiable de evaluar la calidad de experiencia y la satisfacción del usuario.

La tercera contribución también se encuadra en la estimación de la calidad objetiva, pero, en este caso, se utilizan técnicas de procesado de vídeo para extraer información de los píxeles. Se describe un algoritmo que no requiere referencia basado en la estadística de escenas naturales. El número de características extraídas de las imágenes es suficientemente pequeño como para no suponer una carga de procesado y el consiguiente coste energético. Asímismo, pertenece a un espacio de características compacto. La potencia de estimación muestra una buena correlación con las puntuaciones subjetivas (Correlación lineal ≈ 0.80, Correlación de Spearman ≈ 0.75) superando a los algoritmos más recientes en la evaluación de vídeo subacuático.

Esta tesis doctoral puede considerarse un punto de partida para el vasto campo de investigación que se abre con el desarrollo de nuevas tecnologías para transmisiones submarinas y el creciente interés en la exploración oceánica. A continuación, se proporcionan algunas sugerencias para la continuación de esta investigación:

- Construcción de bases de datos de calidad subjetiva extendidas. Este trabajo puede realizarse en dos ramas:
    - Análisis extendido del contenido del vídeo submarino.
    - Análisis extendido de la percepción de vídeo submarino.
- Estudio de alternativas para la captura y compresión de vídeo submarino.
- Utilización de técnicas de aprendizaje profundo para estimación de calidad objetiva.

Appendix B

Curriculum Vitae

Experience

- Associate Scientist for Digital Phenotyping February 2018 – Currently
  Bayer Vegetable Seeds
- Visiting Researcher February 2017 – May 2017
  Laboratory for Image and Video Engineering – The University of Texas at Austin
- Research Engineer December 2012 – January 2017
  Department of Communications Engineering - University of Málaga
  Research and Development Engineer for Industrial Projects.
- Teaching Assistant First semester 2016 – 2017
  Department of Communications Engineering - University of Málaga
  Course: Telecommunications Networks and Systems. Degrees: Telematics Engineering and Electronic Engineering.

Education

- Telecommunications Engineering (Bachelor's+Master's Degree), University of Málaga.

Journal papers

- J.M. Moreno-Roldán, J. Poncela, P. Otero, A.C. Bovik, "No-Reference Video Quality Assessment for Underwater Acoustic Networks". *Submitted to the IEEE Journal of Oceanic Engineering*.
- M.Á. Luque-Nieto, J.-M. Moreno-Roldán, P.Otero, J. Poncela. "Optimal scheduling and fair service policy for STDMA in underwater networks with

acoustic communications". *Sensors*, vol. 18, Art. 612, 2018. DOI: 10.3390/s18020612.

- J.M. Moreno-Roldán, M.Á. LuqueNieto, J. Poncela, P. Otero, "Objective Video Quality Assessment for Underwater Scientific Applications based on Machine Learning". *Sensors*, vol. 17, no. 4, 2017. DOI 10.3390/s17040664.
- M.Á. Luque-Nieto, J.M. Moreno-Roldán, J. Poncela, P. Otero, "Optimal fair scheduling in S-TDMA sensor networks for monitoring river plumes". *Journal of Sensors*, vol. 2016, Article ID 8671516, 2016. DOI 10.1155/2016/8671516.
- J.M. Moreno-Roldán, M.Á. Luque-Nieto, J. Poncela, V. Díaz-del-Río, P. Otero, "Subjective Quality Assessment of Underwater Video for Scientific Applications". *Sensors*, vol. 15, no. 12, pp. 31723-31737, Dec. 2015. DOI 10.3390/s151229882.
- J. Poncela, J.M. Moreno, M. Aamir, "M2M challenges and opportunities in 4G", *Wireless Personal Communications*, Volume 85, Issue 2, pp 463-481, Nov 2015. DOI 10.1007/s11277-015-2746-y.
- N. Perez, J. Poncela, J. M. Moreno-Roldan, M. Memon, "IntelCity, Multiplatform Development of Information Access Platform for Smart Cities", *Wireless Personal Communications*, Volume 85, Issue 2, pp 407-420, Nov 2015. DOI 10.1007/s11277-015-2749-8.

Conference papers

- M.A. Luque-Nieto, J.M. Moreno-Roldán, J. Poncela, P. Otero, "Operación Multihop eficiente en redes Acústicas Submarinas de Sensores con planificación TDMA". *Actas del V Congreso Nacional de i+d en Defensa y Seguridad (DESEi+d 2017),* Toledo (Spain), 22-24 November 2017.
- J.M. Moreno-Roldán, M.Á. Luque-Nieto, P. Otero y J. Poncela, "Challenges in video quality assessment for underwater wireless sensor networks", in *Proc. 4th Int. Conf. on Computing for Sustainable Global Development (INDIACom)*, New Delhi (India), 1–3 March 2017.
- M.A. Luque-Nieto, J.M. Moreno-Roldán, J. Poncela y P. Otero, "Análisis de planificaciones para redes acústicas submarinas de sensores con topología lineal", *Actas del IV Congreso Nacional de i+d en Defensa y Seguridad (DESEi+d 2016)*, San Javier (Murcia), 16-18 November 2016, pp. 779-786.
- J.M. Moreno-Roldán, M.A. Luque-Nieto, J. Poncela, P. Otero, V. Díaz-del-Río, "Calidad de transmisión de vídeo en redes de comunicaciones submarinas". *Actas del III Congreso Nacional de i+d en Defensa y Seguridad (DESEi+d)*, Marín (Spain), 19-20 November 2015, pp. 689-695.
- M.A. Luque-Nieto, J.M. Moreno-Roldán, J. Poncela, P. Otero, L.M. Fernández-Salas, V. Díaz-del-Río, "Planificación S-TDMA en redes submarinas de comunicaciones inalámbricas". *Actas del III Congreso Nacional de i+d en Defensa y Seguridad (DESEi+d)*, Marín (Spain), 19-20 November 2015, pp. 11-18.

- J.M. Moreno-Roldán, M.A. Luque-Nieto, P. Otero, J. Poncela, L.M. Fernández-Salas, V. Díaz-del-Río, "Real-time monitoring of underwater environments (Monitorización de entornos submarinos en tiempo real)". *Proccedings of the VIII International Symposium on the Iberian Atlantic Margin (MIA*), Málaga (Spain), 21-23 September 2015, pp. 49-52.
- J.M. Moreno-Roldán, M.A. Luque-Nieto, V. Díaz-del-Río, P. Otero y J. Poncela, "Parametric quality assessment in underwater video", in *Proc. 2nd Int. Conf. on Computing for Sustainable Global Development (INDIACom)*, New Delhi (India), 11–13 March 2015 (ISBN: 978-93-80544-14-4), pp. 1270-1274.
- M.A. Luque-Nieto, J.M. Moreno-Roldán, J. Poncela y P. Otero, "Reliable Transmissions in Fair STDMA Underwater Sensor Networks" in *Proc. 2nd Int. Conf. on Computing for Sustainable Global Development (INDIACom)*, New Delhi (India), 11–13 March 2015 (ISBN: 978-93-80544-14-4). pp. 1285-1289.

Management and Organization Experience

- Web and Publicity chair for the Global Conference on Wireless and Optical Communications 2016 (GCWOC'16), Málaga 5-7 September 2016.

Industrial Projects Experience

- Design and development of multiplatform (web and SmartTV) user interfaces for a Smart Cities application. Collaboration project: University of Málaga and Abaccus S. L. Principal Investigator: Prof. Javier Poncela.
- Design and development of an interactive digital signage system integrating Natural User Interfaces and Profile detection based on Microsoft Kinect. Collaboration project: University of Málaga and Ingenia S. A. Principal Investigator: Prof. Javier Poncela.

Other related experience

- Internship on software validation of a network analysis tool (6 months). AT4 Wireless, S. A. U.

Bibliography

[1] National Oceanic and Atmospheric Administration. “How much of the ocean have we explored?”, National Ocean Service website, https://oceanservice.noaa.gov/facts/exploration.html, accessed on 10/10/2017.

[2] Marine Technology Society, What is a ROV?, http://www.rov.org/rov_overview.cfm, accessed on 17/11/2017

[3] R.D. Christ and R.L. Wernli. “The ROV Manual: a User Guide for Observation Class Remotely Operated Vehicles”, 2nd ed., Elsevier Science&Technology, 2013, ISBN 9780080982885

[4] Marine Technology Society, ROV Applications - What ROVs can do, http://www.rov.org/rov_applications.cfm#, accessed on 17/11/2017

[5] IEO, Geociencias Marinas, http://geologiamarina.blogspot.com.es/, accessed on 17/11/2017

[6] IEO, Infraestructura y equipos, http://geologiamarina.blogspot.com.es/2009/03/infraestructura-y-equipos.html, accessed on 17/11/2017

[7] H. Kaushal and G. Kaddoum, “Underwater optical wireless communication”, *IEEE Access* (ISSN: 2169-3536), Vol. 4, pp. 1518-1547, 11 Apr. 2016, DOI: 10.1109/ACCESS.2016.2552538

[8] Bluecom, underwater optical communication system. https://www.sonardyne.com/product/bluecomm-underwater-optical-communication-system/, accessed on 17/11/2017

[9] Aquamodem Op1, underwater optical communication system. http://www.aquatecgroup.com/aquamodem/aquamodem-op1, accessed on 17/11/2017

[10] W.H. Thorp and D.G. Browning. “Attenuation of low frequency sound in the ocean”, *Journal of Sound and Vibration* (Elsevier), Vol. 26, pp. 576-578, 1973

[11] S.M.A. Shakar, G. Griffiths, and A.T. Webb, “Power sources for unmanned underwater vehicles”, chapter 2 in G. Girffiths (ed.), *Technology and Applications of Autonomous Underwater Vehicles*, Taylor & Francis, 2003, pp. 19-36, ISBN: 0-203-54794-8.

[12] Methodology for the Subjective Assessment of the Quality of Television Pictures. ITU-R Recommendation BT.500–13. 2012. Available online: https://www.itu.int/rec/R-REC-BT.500 (accessed on 14 December 2015).

[13] Subjective Video Quality Assessment Methods for Multimedia Applications. ITU-T Recommendation P.910. 2008. Available online: https://www.itu.int/rec/T-REC-P.910 (accessed on 14 December 2015).

[14] Opinion Model for Video-Telephony Applications. ITU-T Recommendation G.1070. 2012. Available online: https://www.itu.int/rec/T-REC-G.1070 (accessed on 14 December 2015).

[15] A. Takahashi, D. Hands and V. Barriac, "Standardization activities in the ITU for a QoE assessment of IPTV", *IEEE Commun. Mag.* 2008, 46, 78–84.

[16] Objective perceptual multimedia video quality measurement in the presence of a full reference. ITU-T Recommendation J.247, 2008. Available online: https://www.itu.int/rec/T-REC-J.247 (accessed on 14 September 2016).

[17] Z. Wang, A. C. Bovik, H. R. Sheikh and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity", *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600-612, April 2004

[18] Perceptual video quality measurement techniques for digital cable television in the presence of a reduced reference. ITU-T Recommendation J.249, 2010. Available online: https://www.itu.int/rec/T-REC-J.249 (accessed on 14 September 2016).

[19] T. Hoßfeld et al., "Best Practices for QoE Crowdtesting: QoE Assessment with Crowdsourcing", *IEEE Transactions on Multimedia*, vol. 16, no. 2, pp. 541-558, Feb. 2014.

[20] Lin Ma, K. N. Ngan and Long Xu, "Reduced reference video quality assessment based on spatial HVS mutual masking and temporal motion estimation", *2013 IEEE International Conference on Multimedia and Expo (ICME)*, San Jose, CA, 2013, pp. 1-6.

[21] M. A. Saad, A. C. Bovik and C. Charrier, "Blind Image Quality Assessment: A Natural Scene Statistics Approach in the DCT Domain", *IEEE Transactions on Image Processing*, vol. 21, no. 8, pp. 3339-3352, Aug. 2012.

[22] Y. Kawayoke and Y. Horita, "NR objective continuous video quality assessment model based on frame quality measure," *2008 15th IEEE International Conference on Image Processing*, San Diego, CA, 2008, pp. 385-388.

[23] H. J. Seltman, Experimental design and analysis. Available online: http://www.stat.cmu.edu/~hseltman/309/Book/Book.pdf. Accessed on: 12/11/2017

[24] D.C. Montgomery, Design and Analysis of Experiments; John Wiley and Sons: New York, NY, USA, 2013; pp. 65–130.

[25] A. L. Samuel, "Some Studies in Machine Learning Using the Game of Checkers", *IBM Journal of Research and Development*, vol. 3, no. 3, pp. 210-229, July 1959. DOI: 10.1147/rd.33.0210. URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5392560&isnumber=5392559

[26] A.C. Bovik, "Automatic Prediction of Perceptual Image and Video Quality", *Proceedings of the IEEE*, Vol. 101, No. 9, September 2013.

[27] D. L. Ruderman, "The statistics of natural images," *Netw. Comput. Neural Syst.*, vol. 5, no. 4, pp. 517–548, 1994.

[28] D. J. Field, "Relations between the statistics of natural images and the response properties of cortical cells", *J. Opt. Soc. Amer.* A, vol. 4, pp. 2379–2394, 1987.

[29] B. A. Olshausen and D. J. Field, "Sparse coding with an overcomplete basis set: A strategy employed by V1?", *Vis. Res.*, vol. 37, no. 23, pp. 331–3325, 1997.

[30] I. Vasilescu, K. Kotay, D. Rus, M. Dundabin and P. Corke, "Data Collection, Storage and Retrieval with an Underwater Sensor Network", *Proceedings of the SenSys'05, San Diego*, CA, USA, 2–4 November 2005; pp. 154–165.

[31] T. Suzuki, K. Kato, E.Makihara, T. Kobayashi, H. Kono, K. Sawai, K. Kawabata, F. Takemura, N. Isomura and H. Yamashiro, "Development of Underwater Monitoring Wireless Sensor Network to Support Coral Reef Observation", *Int. J. Distrib. Sens. Netw*, 2014.

[32] D. Pompili and I.F. Akyildiz, "A multimedia cross-layer protocol for underwater acoustic sensor networks", *IEEE Trans. Wirel. Commun*, 2010, 9, 2924–2933.

[33] P. Sarisaray-Boluk, V.C. Gungor, S. Baydere and A.E. Harmanci, "Quality aware image transmission over underwater multimedia sensor networks", *Ad Hoc Netw.* 2011, 9, 1287–1301.

[34] K. Seshadrinathan, R. Soundararajan, A.C. Bovik and L.K. Cormack, "Study of Subjective and Objective Quality Assessment of Video", *IEEE Trans. Image Process.* 2010, 19, 1427–1441.

[35] A.K. Moorthy, L.K. Choi, A.C. Bovik and G. de Veciana, "Video Quality Assessment on Mobile Devices: Subjective, Behavioral and Objective Studies", *IEEE J. Sel. Top. Signal Process.* 2012, 6, 652–671.

[36] C. Chen, L.K. Choi, G. de Veciana, C. Caramanis, R.W. Heath and A.C. Bovik, "Modeling the Time-Varying Subjective Quality of HTTP Video Streams With Rate Adaptations", *IEEE Trans. Image Process.* 2014, 23, 2206–2221.

[37] H.R. Sheikh, Z. Wang, L. Cormack and A.C. Bovik, LIVE Image Quality Assessment Database Release 2. Available online: http://live.ece.utexas.edu/research/quality (accessed on 4 November 2015).

[38] F. DeSimone, M. Naccari, M. Tagliasacchi, F. Dufaux, S. Tubaro and T. Ebrahimi, "Subjective Quality Assessment of H.264/AVC Video Streaming with Packet Losses", *EURASIP J. Image Video Proc.* 2011.

[39] K. Gu, M. Liu, G. Zhai, X. Yang and W. Zhang, "Quality Assessment Considering Viewing Distance and Image Resolution", *IEEE Trans. Broadcast.* 2015, 61, 520–531.

[40] B. Tulu and S. Chatterjee, “Internet-based telemedicine: an empirical investigation of objective and subjective video quality”, *Decis. Support Syst.* 2008, 45, 681–696.

[41] J. Joskowicz, R. Sotelo and J.C.L Ardao, “Towards a general parametric model for perceptual video quality estimation”, *IEEE Trans. Broadcast.* 2013, 59, 569–579.

[42] K. Yamagishi and T. Hayashi, “Video-Quality Planning Model for Videophone Services”, *ITE* 2008, 62, 1050–1058.

[43] Evologics GmbH. S2C M HS Modem Product Information. Available online: http://www.evologics.de/en/products/acoustics/s2cm_hs.html (accessed on 4 November 2015).

[44] J.M. Moreno-Roldán, M. A. Luque-Nieto, V. Díaz-del-Río, P. Otero and J. Poncela, “Parametric Quality Assessment in Underwater Video”, *Proceedings of the 2nd International Conference on Computing for Sustainable Global Development (INDIACom)*, New Delhi, India, 11–13 March 2015; pp. 1270–1274.

[45] V. Díaz del Río, G. Bruque, L.M. Fernández Salas, J.L. Rueda, E. González, N. Lopez-Gonzalez, D. Palomino, F. José López, C. Farias, R.F.L. Sánchez, et al. “Volcanes de Fango del Golfo de Cádiz”, Proyecto LIFE+INDEMARES; Alimentación y Medio Ambiente: Madrid, Spain, 2014; p. 130.

[46] R.O. Duda and P. Hart, Representation and initial simplifications. In Pattern Classification and Scene Analysis; John Wiley and Sons: New York, NY, USA, 1973; pp. 271–272.

[47] Sekonik Corp. L-758DR Digital Lightmeter Product Information. Available online: http://www.sekonic.com/products/l-758dr/overview.aspx (accessed on 4 November 2015).

[48] IBM Corp. IBM SPSS Statistics Base 22. Software—Manual. Available online: ftp://public.dhe.ibm.com/software/analytics/spss/documentation/statistics/22.0/en/client/Manuals/ (accessed on 14 December 2015).

[49] J.W. Mauchly, Significance Test for Sphericity of a Normal n-Variate Distribution. Ann. Math. Stat. 1940, 11, 204–209.

[50] S.W. Greenhouse and S. Geisser, On the methods in the analysis of profile data. Pyschometrika 1959, 24, 95–112.

[51] H. Huynh and L.S. Feldt, “Estimation of the Box Correction for Degrees of Freedom from Sample Data in Randomized Block and Split-Plot Designs”, *J. Educ. Stat.* 1976, 1, 69–82.

[52] J. Sogaard, S. Forchhammer, J. Korhonen, “Video quality assessment and machine learning: Performance and interpretability”, *Proceedings of the 2015 Seventh InternationalWorkshop on Quality of Multimedia Experience (QoMEX)*, Pylos-Nestoras, Greece, 26–29 May 2015; pp. 1–6.

[53] L. Anegekuh, L. Sun, E. Jammeh, I.H. Mkwawa and E. Ifeachor, "Content-Based Video Quality Prediction for HEVC Encoded Videos Streamed Over Packet Networks", *IEEE Trans. Multimedia* 2015, 17, 1323–1334.

[54] F. You, W. Zhang and J. Xiao, "Packet Loss Pattern and Parametric Video Quality Model for IPTV", *Proceedings of the Eighth IEEE/ACIS International Conference on Computer and Information Science (ICIS)*, Shanghai, China, 1–3 June 2009; pp. 824–828.

[55] A. Raake, M.N. Garcia, S. Moller, J. Berger, F. Kling, P. List, J. Johann and C. Heidemann, "T-V-model: Parameter-based prediction of IPTV quality", *Proceedings of the 2008 IEEE International Conference on Acoustics, Speech and Signal Processing*, Las Vegas, NV, USA, 31 March–4 April 2008; pp. 1149–1152.

[56] H. Koumaras, A. Kourtis, D. Martakos and J. Lauterjung, "Quantified PQoS assessment based on fast estimation of the spatial and temporal activity level", *Multimedia Tools Appl.* 2007, 34, 355–374.

[57] M. Ries, C. Crespi, O. Nemethova and M. Rupp, "Content based video quality estimation for H.264/AVC video streaming", *Proceedings of the 2007 IEEEWireless Communications and Networking Conference*, Kowloon, China, 11–15 March 2007; pp. 2668–2673.

[58] J. Gustafsson, G. Heikkila and M. Pettersson, "Measuring multimedia quality in mobile networks with an objective parametric model", *Proceedings of the 2008 15th IEEE International Conference on Image Processing*, San Diego, CA, USA, 12–15 October 2008; pp. 405–408.

[59] A. Khan, L. Sun and E. Ifeachor, "Content-based video quality prediction for MPEG4 video streaming over wireless networks" *J. Multimedia* 2009, 4, 228–239.

[60] Q. Huynh-Thu and M. Ghanbari, "Temporal Aspect of Perceived Quality in Mobile Video Broadcasting", *IEEE Trans. Broadcast.* 2008, 54, 641–651.

[61] Y.F. Ou, Z. Ma, T. Liu, Y. Wang, "Perceptual Quality Assessment of Video Considering Both Frame Rate and Quantization Artifacts", *IEEE Trans. Circuits Syst. Video Technol.* 2011, 21, 286–298.

[62] J. Joskowicz and J.C.L. Ardao, "Combining the effects of frame rate, bit rate, display size and video content in a parametric video quality model", *Proceedings of the 6th Latin American Networking Conference*, Quito, Ecuador, 12–13 October 2011; pp. 4–11.

[63] P. Le Callet, C. Viard-Gaudin and D. Barba, "A Convolutional Neural Network Approach for Objective Video Quality Assessment" *IEEE Trans. Neural Netw.* 2006, 17, 1316–1327.

[64] C. Edwards, "Growing pains for deep learning", *Commun. ACM* 2015, 58, 14–16.

[65] M. Shahid, A. Rossholm and B. Lövström, "A no-reference machine learning based video quality predictor", *Proceedings of the 2013 Fifth InternationalWorkshop on*

Quality of Multimedia Experience (QoMEX), Klagenfurt am Wörthersee, Austria, 3–5 July 2013; pp. 176–181.

[66] A. Hameed, R. Dai and B. Balas, "A Decision-Tree-Based Perceptual Video Quality Prediction Model and its Application in FEC for Wireless Multimedia Communications", *IEEE Trans. Multimedia* 2016, 18, 764–774.

[67] T. Hoßfeld, P.E. Heegaard and M. Varela, "QoE beyond the MOS: Added value using quantiles and distributions", *Proceedings of the 2015 Seventh International Workshop on Quality of Multimedia Experience (QoMEX)*, Pylos-Nestoras, Greece, 26–29 May 2015; pp. 1–6.

[68] W. Song and D.W. Tjondronegoro, "Acceptability-Based QoE Models for Mobile Video", *IEEE Trans. Multimedia* 2014, 16, 738–750.

[69] F.L. Bookstein, "Principal warps: Thin-plate splines and the decomposition of deformations", *IEEE Trans. Pattern Anal. Mach. Intell.* 1989, 11, 567–585.

[70] F.J. Richards, "A Flexible Growth Function for Empirical Use", *J. Exp. Bot.* 1959, 10, 290–300.

[71] P. McCullagh, "Regression Models for Ordinal Data", *J. R. Stat. Soc. Ser. B (Methodol.)* 1980, 42, 109–142.

[72] P. McCullagh, J.A. Nelder, "Models for polytomous data" in *Generalized Linear Models*; Chapman & Hall: London, UK, 1989; pp. 151–155.

[73] D.R. Cox and E.J. Snell, Analysis of Binary Data, 2nd ed.; Chapman & Hall: London, UK, 1989.

[74] N.J.D. Nagelkerke, "A note on a general definition of the coefficient of determination", *Biometrika* 1991, 78, 691–692.

[75] D. McFadden, "Conditional logit analysis of qualitative choice behavior", in *Frontiers in Econometrics*; Zarembka, P., Ed.; Academic Press: San Diego, CA, USA, 1974; pp. 105–142.

[76] W. Hosmer, S. Lemeshow and R.X. Sturdivant, "Applied Logistic Regression", John Wiley and Sons: Hooboken, NJ, USA, 2013.

[77] Cisco Visual Networking Index: Forecast and Methodology, 2016–2021. Available online: http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/complete-white-paper-c11-481360.pdf.

[78] [Hoag-97] D. F. Hoag, V. K. Ingle and R. J. Gaudette, "Low-Bit-Rate Coding of Underwater Video Using Wavelet-Based Compression Algorithm", IEEE Journal of Oceanic Engineering, vol. 22, no. 2, pp. 393-400, April 1997.

[79] [Zhang-16] Y. Zhang, S. Negahdaripour and Q. Li, "Error-resilient coding for underwater video transmission", OCEANS 2016 MTS/IEEE Monterey, Monterey CA, 2016, pp. 1-7.

[80] [Martins 2015] M. S. Martins, J. Cabral, G. Lopes and F. Ribeiro, "Underwater acoustic modem with streaming video capabilities", OCEANS 2015 – Genova, Genoa 2015, pp. 1-7.

[81] [Yang-2015] M. Yang and A. Sowmya, "An Underwater Color Image Quality Evaluation Metric", IEEE Transacations on Image Processing, vol. 24, no. 12, pp. 6062-6071, Dec. 2015.

[82] [Panetta-2016] K. Panetta, C. Gao and S. Agaian, "Human-Visual-System Inspired Underwater Image Quality Measures", IEEE Journal of Oceanic Engineering, vol. 41, no.3, pp. 541-551, July 2016.

[83] M. A. Saad, A. C. Bovik, and C. Charrier, "Blind prediction of natural video quality," *IEEE Trans. Image Process.*, vol. 23, no. 3, pp. 1352–1365, Mar. 2014.

[84] F. De Simone, M. Tagliasacchi, M. Naccari, S. Tubaro and T. Ebrahimi, "A H.264/AVC video database for the evaluation of quality metrics," *2010 IEEE International Conference on Acoustics, Speech and Signal Processing*, Dallas, TX, 2010, pp. 2430-2433.

[85] A. Mittal, M. A. Saad and A. C. Bovik, "A Completely Blind Video Integrity Oracle," *IEEE Trans. on Image Process.*, vol. 25, no. 1, pp. 289-300, Jan. 2016.

[86] A. Mittal, A. K. Moorthy and A. C. Bovik, "No-Reference Image Quality Assessment in the Spatial Domain", *IEEE Trans. Image Process.*, vol. 21, no. 12, pp. 4695-4708, Dec. 2012.

[87] A. Mittal, R. Soundararajan and A. C. Bovik, "Making a "Completely Blind" Image Quality Analyzer", *IEEE Signal Processing Letters*, vol. 20, no. 3, pp. 209-212, March 2013.

[88] L. Zhang, L. Zhang and A. C. Bovik, "A Feature-Enriched Completely Blind Image Quality Evaluator", *IEEE Trans. Image Process.*, vol. 24, no. 8, pp. 2579-2591, Aug. 2015.

[89] K. Sharifi and A. Leon-Garcia, "Estimation of shape parameter for generalized Gaussian distributions in subband decompositions of video," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 5, no. 1, pp. 52–56, Feb. 1995.

[90] F. Oleari, F. Kallasi, D. Lodi Rizzini, J. Aleotti, S. Caselli, "An Underwater Stereo Vision System: from Design to Deployment and Dataset Acquisition", *Proc. of the IEEE/MTS OCEANS*, Pages 1-6, May 19-21, 2015.

[91] M. Narwaria and W. Lin, "Objective image quality assessment based on support vector regression," *IEEE Trans. Neural Netw.*, vol. 21, no. 3, pp. 515–519, Mar. 2010.

[92] C. Spampinato, G. Nadarajan, J. Chen-Burger and R. Fisher, "Detecting, tracking and counting fish in low quality unconstrained underwater videos", *Proc. 3rd Int. Conf. on Computer Vision Theory and Applications (VISAPP)*, Funchal, Vol.2, pp 514-520, 2008.