VARIUM

Visual ARtifacts Interference Understanding and Modeling

Attachment 1 - Project Plan

Program: NUFFIC/CAPES, call 3.2011   Project Area: Engineering and Computer Science

Project Proponents:

Institute	Coordinator	Contact
Delft University of technology	Dr. Judith A. Redi	j.a.redi@tudelft.nl
University of Brasilia	Prof. Mylène C.Q. Farias	Mylene@unb.br

TABLE OF CONTENTS

1 PROJECT SUMMARY AND AIMS

Background: Visual material is nowadays mainly produced and distributed in a digital format. Together with many advantages (ease of processing and storage), the digitization of media content comes with the main disadvantage that its quality can easily decay during every phase of its life (capture, transmission, storage, and/or display), as well as due to processing applied along the way, such as compression. To prevent and correct degradation of visual quality, the scientific community has developed quality control tools (i.e. objective quality metrics), that allow estimating the visual quality of media in an automatic way, yet consistent with human perception. Although many video quality models have been proposed, they usually address single artifacts in quality decrease, such as sharpness or noisiness. Little work has been done on studying and characterizing combinations and interferences of individual artifacts (spatial and temporal), which are most likely to happen.

Research goal and methodology: In this project, we are interested in understanding the impact on perceived quality of combinations of relevant artifacts, and their relationship with content. We aim at designing an **objective metric for overall video quality** that takes into account **specific spatial and temporal artifacts, their mutual impact and their mutual importance** for a broad range of image contents. As a further element to strengthen our research, we aim at including visual attention in the developed quality metrics. To ensure consistency with visual perception, we first plan to perform a series of psychophysical experiments on visual quality perception. These experiments will determine the impact on quality first of a set of spatial and temporal artifacts by themselves, and later of their combinations, taking into account different content with varying spatial and temporal characteristics. The set of artifacts to be studied will be chosen among those perceptually most relevant for digital video applications (e.g. blockiness, blurriness, packet-loss, jerkiness, etc.). We intend to use the knowledge acquired in the psychophysical experiments to design an objective quality metric that is adequate for the digital video scenario.

Partnership and educational goals: This project aims at building a strong collaboration among three major universities (**Universidade de Brasília, Universidade Estadual de Campinas** and **Delft University of Technology**), which excel in the field of video processing and visual perception understanding. The experience of the researchers involved in the project is complementary; nevertheless, they share a common background in the development of visual media technologies. Therefore, the project is based on an ease of collaboration and will stimulate the exchange of know-how between high level research groups in the field of visual quality assessment. Furthermore, the proposed plan envisions the exchange of young researchers (MSc and PhD students), who will benefit of the educational services of both proponent universities. Both these aspects meet the focus of the Nuffic / CAPES program, which is to promote scientific exchange between Dutch and Brazilian institutions and to form high level human resources in both countries.

2 INTRODUCTION AND MOTIVATION

The digital transformation of images and video offers many advantages over the existing analog methods. Compared to analogue image material, digital content can be more effectively stored or transmitted through a broadcasting channel, different algorithms can be used to reduce defects or enhance the quality of the image material, and additional information can be added to the content to enlarge options for segmentation and indexing of the content. As a consequence of these advantages, digital image content nowadays is used in different applications covering communication, entertainment, medical information representation, and security. For each of these application areas, the advantages mentioned above are used in a different way and to a different extent, and in some cases even dynamically adaptive in order to optimize the experience for the user of the content.

The advantages of digital visual material, however, do not come without some disadvantages as well. Similar as with analog content, the quality of digital content may decrease when impairments are introduced during capture, transmission, storage, and/or display, as well as by any signal processing algorithm that may be applied along the way (e.g., compression, etc.). Impairments are defined as visible defects (flaws) and can be decomposed into a set of perceptual features called artifacts [1-2]. The artifacts found in digital image applications only have a *spatial* component, while the artifacts found in digital video applications can be classified as *spatial* or *temporal* artifacts. Spatial artifacts are characterized by the presence of degradations that vary (mainly) within one frame or along the spatial domain. Examples of spatial artifacts include blocking, blurring, ringing, noise, mosaic patterns, etc. Temporal artifacts are degradations that vary across the frames of the video or along the temporal domain. Examples of temporal artifacts include motion compensation mismatches, mosquito effects, ghosting, smearing, jerkiness, and so on. Some artifacts may incorporate both spatial and temporal characteristics, like for example the packet losses artifact that is caused by network losses.

In most applications, the transmitted images or video are destined for human consumption, and as a consequence, humans will decide on the quality and visual experience of the material. Therefore, human perception should be taken into account when optimizing the quality and experience of the material. For example, human perception comes into play when trying to establish the degree to which a video can be compressed, to decide on which visual enhancements should be applied, and on how these visual enhancements should be optimized. The most accurate way to determine the quality of a video is by measuring it using psychophysical experiments with human subjects [3]. Unfortunately, psychophysical experiments are very expensive, time-consuming and hard to incorporate into a design process or an automatic quality of service control. Therefore, the ability to measure video quality accurately and efficiently, without using human observers, is highly desirable in practical applications. To this end, the research community develops automatic algorithms that predict the quality of image material as perceived by human observers. These automatic algorithms are hereafter referred to as *objective image/video quality metrics.* When sufficiently reliable (i.e. when sufficiently reproducing the quality experienced by human observers), these objective quality metrics can be employed to monitor the quality of image material along the whole broadcasting chain, to compare the performance of image/video processing systems and algorithms, and to optimize the algorithms and parameter settings for an image/video processing system. In addition, to quantify the performance of an image or video communication system in a real-time application, it is important that objective quality metrics are also sufficiently fast and computationally efficient, and that they give an estimate of the quality of the image material (or the quality change) at each stage of the communication system (acquisition, compression, transmission, or display).

Objective video quality metrics can be classified as *data metrics*, which measure the fidelity of the signal without considering its content, or *picture metrics*, which estimate quality considering the visual information contained in the data. Customarily, objective quality metrics in the area of image processing have been largely limited to a few data metrics, such as the mean absolute error (MAE), the mean square error (MSE), and the peak signal-to-noise ratio (PSNR). Although over the years data metrics have been widely criticized for not correlating well with perceived quality, it has been shown that such metrics can predict subjective quality ratings with reasonable accuracy as long as the comparisons are made with the same content, the same technique, or the same type of distortions [4-5]. One of the major reasons why data metrics do not generally perform as desired is because they do not incorporate any human visual system (HVS) features in their computation. It has been discovered that in the primary visual cortex of mammals an image is not represented in the pixel domain, but in a rather different manner [5]. Unfortunately, the measurements produced by commonly used metrics like MSE or PSNR are simply based on a pixel to pixel comparison of the data, without considering what is the content and the relationships among pixels in an image (or frames). They also do not consider how the spatial and frequency content of the impairments are perceived by human observers. In the past few years, a big effort in the scientific community has been devoted to the development of new objective metrics, which correlate better with the human perception of quality [6-14]. Quality metrics with best performances are metrics that analyze visible differences (*error sensitivity metrics*) between a test and a reference signal, taking into account aspects of the HVS considered relevant to quality, such as color perception, contrast sensitivity, and pattern masking [13-14]. Most of these metrics have the disadvantages of being computationally complex and, frequently, requiring the reference signal at the point of measurement. Because of this, it is difficult to use this approach in many real-time applications. One possible solution to this problem is to use the *feature extraction* approach, which looks for higher-level features or attributes of the image or video (e.g. sharpness/blur, contrast, fluidity, artifacts, etc.) that are considered relevant to quality. These algorithms measure the magnitude of these high level features in order to estimate quality.

A popular type of feature extraction metrics is the *artifact metrics*, which estimate the strength of the artifacts (or attributes) being perceptually most relevant, and combine these values in order to obtain a quality estimate [11-12]. The artifact metrics have the advantages of being simple and not necessarily requiring the reference signal at the measurement point. Another advantage of artifact metrics is that they can be particularly useful for post-processing and enhancement algorithms, providing information about which artifacts or attributes need to be mitigated or enhanced. Nevertheless, their design requires a good understanding of the characteristics of the various individual artifacts (and attributes), their mutual interference, and their

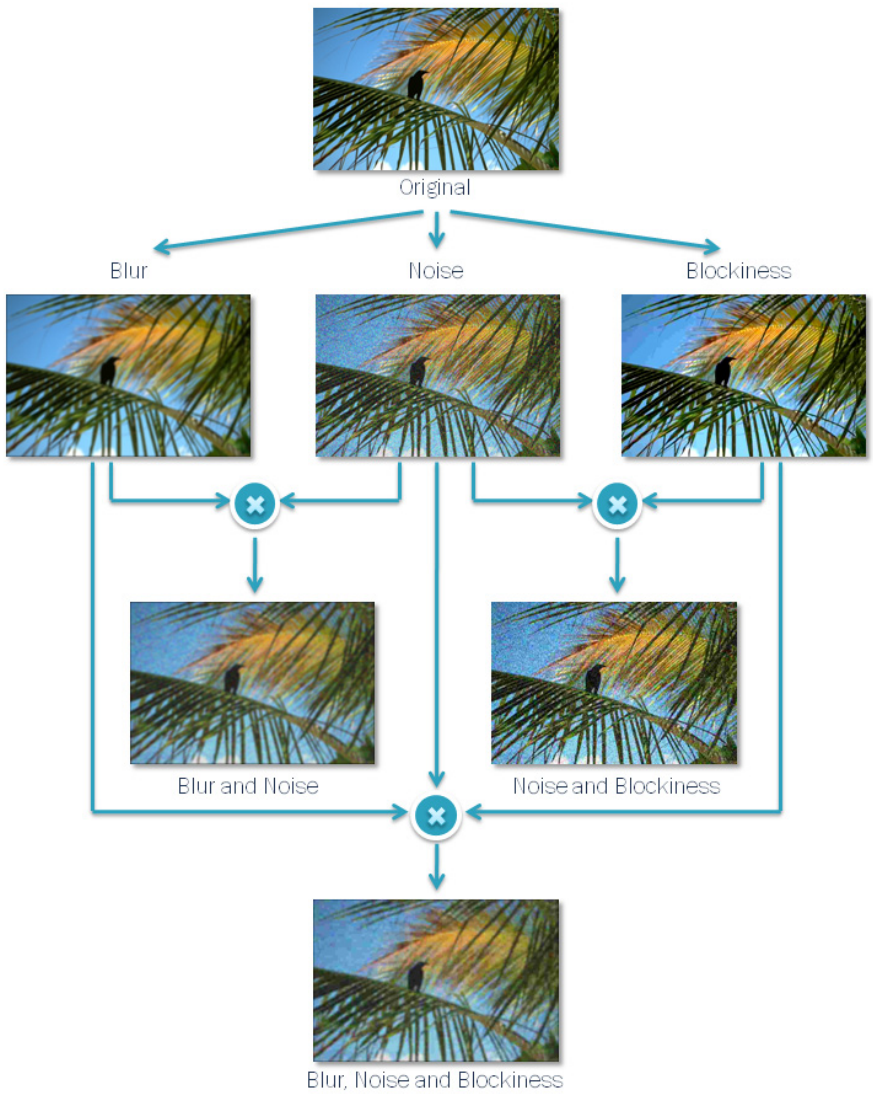


Figure 1 – Examples of single artifacts and combinations of artifacts.

interference with the content of the image material. In fact, most digital visual material is affected by multiple artifacts, visually interfering with each other (see, e.g. Figure 1). Examples are the co-presence of acquisition noise and blurring, or again of blurring and blockiness as produced by compression algorithms. In both cases, the effect on perceived quality of the combination of artifacts can hardly be predicted by the linear combination of the annoyance estimated for the single artifact. Masking and other interaction effects can indeed occur (e.g., blur attenuating noise), what makes the prediction strategy more complex and very dependent on the artifacts involved.

Little work has been done on studying and characterizing combinations and interferences of individual artifacts, as pointed out by A. Moorthy and A. Bovik in a very recent paper [14]. Some research on the mutual impact of blockiness, ringing and blur as a consequence of compression, and on the related effect of the content of the image material is reported in literature [15-16]. So far, however, this is limited to the combination of spatial artifacts. In this project, we are interested in further extending these first attempts by understanding the visibility and annoyance of additional relevant artifacts and their relationship with content. With this extension, we finally aim at an objective metric for overall video quality that takes into account specific spatial and temporal artifacts, their mutual impact and their mutual importance for a broad range of image content. To obtain this goal, first we plan to perform a set of psychophysical experiments to measure the visibility and annoyance of a set of spatial and temporal artifacts by themselves, taking into account different content with varying spatial and temporal characteristics [17]. The set of artifacts to be studied will be chosen among the most perceptually relevant artifacts for digital video applications (e.g. blockiness, blurriness, packet-loss, jerkiness, etc.). Second, in order to study their interactions, we plan to measure the visibility and annoyance of a set of artifacts in presence of another set of artifacts. For example, we plan to study the interactions among different spatial artifacts (e.g. blurring versus blocking), among different temporal artifacts (e.g. jitter versus packet loss), and across temporal and spatial artifacts (e.g. blurring versus jerkiness). Finally, we plan to use the knowledge acquired to design a perceptual quality metric that is adequate for the digital video scenario.

3 PROPOSED WORK

Perceived quality has often been considered as a multidimensional quantity [18-21], where each dimension represents a psychological continuum corresponding to the perceptual strength of a single artifact. Although real life multimedia content typically contains the coexistence of multiple spatial and temporal artifacts, humans express a single judgment when asked to compare the quality of such content, independent on the number of artifacts affecting them. The average Internet viewer will prefer the quality of a video based on its overall appearance, and not on the (conscious) independent evaluation of the annoyance of various artifacts, and their possible interactions. Hence, in this project the human ability to scale images presenting different artifacts (even overlapping) on a single psychological continuum is assumed, and the underlying dimension is called "overall quality".

However, to deeply understand how artifacts combine from a quality perception point of view, we first need to acquire knowledge on their standalone impact. In order to study the individual artifacts, their mutual interactions and their interaction with the content, we will perform a set of psychophysical experiments [3]. Experiments are a valuable research tool for better understanding how humans judge image or video quality and perceive impairments. Also, the subjective ratings/responses obtained from these experiments are considered a benchmark for the development of the objective quality metrics. Therefore, the goal of the experiments is to gather information about the *visibility*, *annoyance*, *description*, and *perceptual strengths* of the artifacts, and to determine the *relative importance* of the artifacts to overall perceived quality.

The *visibility* of an artifact refers to whether the artifact is noticed within an image content. It is defined related to a visibility threshold, which corresponds to the distorting signal strength (total squared error) that allows 50% of the observers to notice the artifact. It is important to know these thresholds for single artifacts independently, but, in our particular case, it is also necessary to study visibility thresholds of artifacts when combined with others. This will help in better modeling artifacts interactions. The *annoyance* of an artifact is a measure of the degradation of the visual content, and it's dependent on the visibility and on the distorting signal strength. To measure it, impairment scores, rather than quality scores, can be used to avoid double-ended scales. Impairment scores are a subjective measure of the degradation of a video and are usually expressed in relation to a reference (unimpaired video). Observers can also be trained to recognize specific artifacts when combined (*description*) and estimate their *perceptual strength*. This way, we can have an idea of how the visibility and annoyance are being affected by the video content and the presence of other artifacts. The final step consists in connecting visibility, annoyance, description and perceptual strength information to the overall appearance of the combined of artifacts. This can be achieved by measuring overall quality scores of videos impaired with different combinations of artifacts (with different strengths as well). This approach allows connecting perceptual responses of single artifacts to their combined impact. A video quality ruler can be used for measuring overall quality scores. The quality ruler is a psychometric tool which enables measuring overall quality scores with a high degree of reliability and reproducibility over various experiments [22].

As a further element to strengthen our research, we aim at including visual attention in the developed quality metrics. When observing a scene, the human eye typically filters the large amount of visual information available by focusing on selected (salient) regions [23, 24]. This selection process is actively controlled through oculomotor mechanisms. These mechanisms allow the gaze of attention to hold on a particular location (fixation) or to shift to a preferred location when sufficient information has been collected from the current one (saccades). Fixations are instinctively concentrated on highly informative areas; as a consequence, the amount of data to be further processed by the brain is minimized, yet maximizing the quantity of useful information. The selection of the fixation locations, also known as bottom-up attention, is fully driven by the intrinsic visual properties of the scene observed.

It has been hypothesized that visual distortions appearing in less salient areas might be less visible and therefore less annoying. A partial confirmation of this hypothesis was already provided in [25], where the authors show how blocking artifacts in the background influence the final quality judgment to a lesser extent as compared to artifacts located in highly salient areas. As a consequence, researchers have started lately to incorporate saliency information in objective metrics, for example through visual importance pooling [26-28]. We aim at incorporating the analysis of visual attentive paths in our study as well. To do so, we will record eye-movements throughout all the planned subjective experiments. These recordings will be performed by means of an eye-tracker, available at Delft University of Technology.

4 RESEARCH PLAN AND DIVISION OF TASKS

We envision our work to be composed of four work packages (WP): a first one (WP1) dedicated to the study of perceptual properties (visibility, annoyance, description and strength) of single and combined artifacts, a second one (WP2) focused on studying the impact on overall quality of artifacts and their combinations, a third one (WP3) which is focused on the study of visual attention in relation to visual quality perception, and a fourth one (WP4) aiming at producing an objective quality metric from the integration of the data collected from the previous work packages. Figure 2 shows a diagram that depicts the four work projects (WP) and their interconnections.

4.1 WP1: STUDYING THE PERCEPTUAL PROPERTIES OF SINGLE AND COMBINED ARTIFACTS.

This WP aims at studying the perceptual impact of single artifacts related to the image content. We plan to use four different tasks to gather the necessary data. The tasks are: (1) detecting an artifact, (2) judging its annoyance, (3) judging its strength, and (4) describing its appearance. The four tasks are described below.

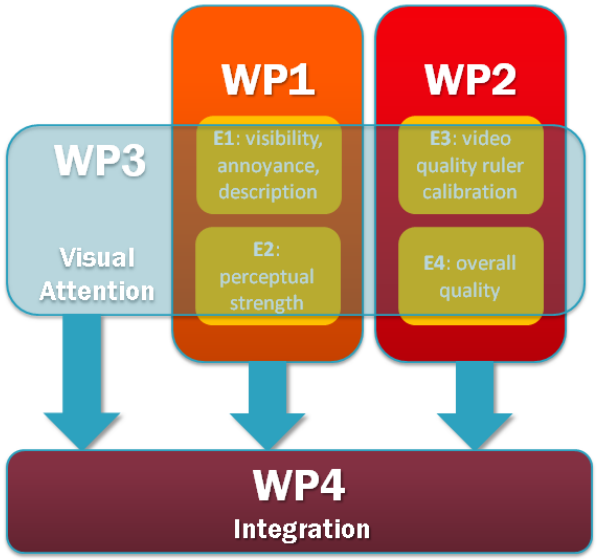
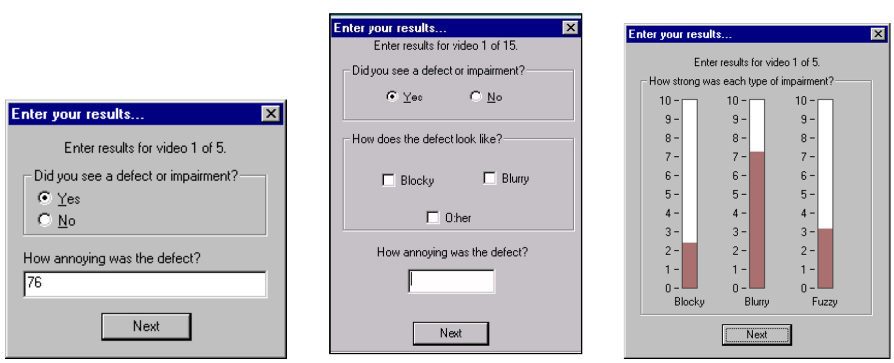


Figure 2 - Schematic representation of the planned work and division of the tasks

The *detection task* consists of detecting a spatially and/or temporally localized impairment in an image or video sequence. The test subjects are instructed to search each stimulus (i.e. image or video) for impairments. After each stimulus is shown, the subjects are asked to answer the following question: "Did you see a defect or impairment?" The subject is supposed to choose a 'yes' or 'no' answer.

The *annoyance task* consists of giving a numerical judgment of how annoying/bad the detected impairment is. Examples of original and highly impaired videos are shown in advance in order to make the participants of the experiment acquainted with the level of impairment they can expect during the experiment, and in order to give them a first impression on how to assess the perceived impairments on a scoring scale. The annoyance task



(a) (b) (c)

Figure 3. Dialog boxes used for subjective data entry in: (a) detection and annoyance task, (b) description task, and (c) perceptual strength task.

can be performed together with the detection task, so that the participants first indicate whether they see an artifact, and if so, how annoying this artifact is. Both tasks can be handled throughout a graphical user interface such as the one displayed in Figure 3(a).

The *description task* consists of judging the appearance of the detected impairment. To this purpose, subjects are asked to provide a lexical characterization of the appearance of the artifacts. Thus, the subject is requested to answer the following question: "How would you describe the defect?" To answer this question the subject is presented with a number of descriptors or classifiers, depending on the artifacts used in the experiment. Figure 3(b) displays an example of a dialog box for the description task, as used in one of our previous experiments [28]. In that particular experiment, observers were given three descriptors to describe the impairment - 'blocky', 'blurry' and 'other'. The observer could pick as many descriptors as needed to describe the impairment.

The *perceptual strength task* consists of asking the subjects for an estimate of how strong or visible a set of artifacts are in the detected impairment. This type of task requires that subjects be taught how each artifact looks like. Therefore, in the training stage subjects are shown a set of stimuli illustrating the set of artifacts being measured. In the trials, after each stimulus is played, the subject is asked to enter a judgment on a continuous interval scale. If no impairments are seen, subjects are instructed not to enter a judgment. Figure 3(c) displays an example of a dialog box for the strength task, as used in our previous experiments [28].

We plan to complete this first part of the study by performing two experiments. The first experiment (E1) will gather the detection, annoyance, and description tasks. From the data collected, psychometric and annoyance

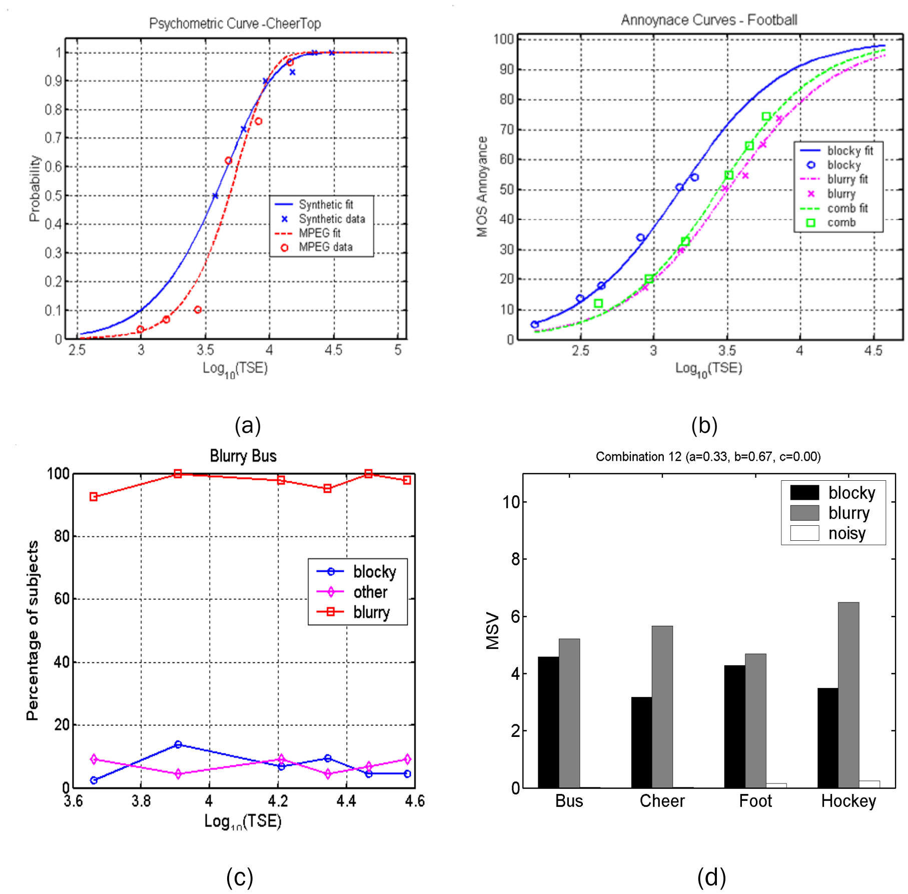


Figure 4. Sample data gathered from: (a) detection, (b) annoyance task, (c) description task, and (d) perceptual strength task.

functions of individual and combined artifacts will be determined. Sample psychometric and annoyance curves obtained from previous experiments are shown in Figure 4(a) and 4(b), respectively. The second experiment (E2) will involve only strength tasks. The data gathered from this experiment will serve to examine the relationship between the perceptual strength of individual artifacts. Comparing the data of the two experiments (sample results are shown in Figures 4(c) and 4(d)), we will study how the overall annoyance is related to the perceptual strength of individual artifacts. This procedure was already successfully exploited in the past [28], where the data on detection threshold, annoyance, and perceptual strength for a set of videos containing blockiness, blurriness, ringing, and noise were combined by means of a Minkowski metric.

4.2 WP2: STUDYING THE IMPACT OF ARTIFACTS AND THEIR COMBINATION ON OVERALL QUALITY

Although promising and proven valuable, the approach described in WP1 uses single-ended scales, which don't take into account the fact that, when combined, some artifacts may mask other artifacts, and as such may increase the perceived quality. In addition, a Minkowski metric might not be sufficient to mimic interactions among the artifacts at a perceived level. As a consequence, it is also important to understand the impact of the perceptual strengths of the individual artifacts on the overall quality perception. To this purpose, we aim at collecting data on the impact of single artifacts and their combination on overall quality by evaluating the same image material involved in E1 and E2 of WP1 using the quality ruler method as a psychometric tool.

Brian Keelan's *quality ruler* method [19, 30, 31] is able to provide precise and standardized measures of perceived quality for large sets of stimuli. It is based on the use of a set of reference images of known quality that are evenly distributed along a pre-calibrated quality scale (the Standard Quality Scale, SQS), as depicted in Figure 5. The pre-calibration can be arranged so that the ruler spans JNDs of overall quality [19]. Observers are asked to position the stimuli along the SQS, by matching them with the reference images. The quality score assessed for each stimulus then corresponds to its position along the SQS, providing a JND-based measure of

Figure 5. Schematic representation of the Quality Ruler method. Observers are asked to position test stimuli between two reference samples by visual matching. The SQS is calibrated in overall quality JNDs.

overall visual quality. The use of the SQS seems to free the methodology from range effects [22, 32], allowing the production of absolute scores comparable over experiments performed at different times and places. This feature is desirable when large amounts of data need to be collected and is particularly suitable when envisioning collaboration between institutions that are far apart from each other, such as TUD and University of Brasilia. In its current version, the quality ruler is implemented to measure overall quality of images.

It will be part of WP2 to produce a new version of the quality ruler that is targeted at videos (hereafter referred to as E3). We will follow the procedure advised by Keelan [19] and perform a large paired comparison experiment including video material impaired with all the artifacts under study in the project. This will allow extracting the overall quality JNDs on which the video SQS will be calibrated.

Once the quality ruler is implemented, it will be used to assess the overall quality resulting of individual artifacts and of combination of artifacts (hereafter referred to as experiment E4). Measuring overall quality for the individual artifacts can be considered as an alternative for the annoyance scoring. Performing both the annoyance task and the quality scaling of the artifact will illustrate: (a) the accuracy that can be obtained with each of the methodologies, and (b) the relation between annoyance and quality for individual artifacts. The measurement of the quality of combined artifacts on the SQS will provide useful information on the interaction of artifacts on a quality level.

4.3 WP3: STUDYING VISUAL ATTENTION DEPLOYMENT WHILE QUALITY SCORING

Throughout every experiment an eye-tracker will be used to register the movements of the participants' eyes. An eye-tracker identifies the location of the pupil from the reflections of infrared light on the retina. The infrared light source is mounted above the lens of an infrared camera. Since infrared falls outside the human visual sensitivity spectrum, the viewer is not distracted by the light emitted from the system. Eye-tracker output data can be processed in order to obtain a description of the eye movements in terms of fixations and saccades; from that saliency maps [33] can be extracted. Saliency maps are a visual representation of the probability that a location of the scene is attended by the average observer. By computing the average saliency map across observer per frame, we obtain a map of the most attended locations in the scene at a precise moment in time. Thus, we can determine whether artifacts were actually foveated during the observation, and relate this to the subjective data collected in the experiment. Further useful information can be derived from the comparison of saliency maps over different types of tasks and artifacts.

We expect that the experimental task affects the viewing strategy to some extent [34]. The same can be hypothesized for different artifacts and quality levels of the videos [35]. Differences in viewing strategies will point out which visual features are important to the perception and quality degradation of artifacts, and as such will provide essential information for design of a perceptual model. More specifically, temporal and spatial saliency data will allow us to discover if and under which circumstances the artifact visibility is such that it distracts the attention away from natural saliency, thus impacting on the final quality judgment. Furthermore, these data will give us more insight in perceptual differences between temporal and spatial artifacts, and in the impact of single versus combined artifacts (and the related masking).

4.4 WP4: INTEGRATION OF THE OUTCOMES OF WP1, WP2 AND WP3 IN AN OBJECTIVE MODEL OF VISUAL QUALITY PERCEPTION

The last phase of the project will be dedicated to the fusion of the outcomes of the other three work packages (see Figure 2). Results from WP3 on artifact visibility and visual strategies will be used to make the Minkowski-based model delivered in WP1 more robust (e.g. by adjusting the metric parameters according to the saliency of the artifacts). In a second stage, the ability of the enriched Minkowski-based model to predict overall quality will be tested. The overall quality scores will serve as ground truth to re-consider the impact of each artifact in the combination and will help in modeling masking effects. In case the Minkowski model cannot accurately reproduce these interactions, different modeling functions will be explored.

5 PLANNING OF THE PROJECT ACTIVITIES

The project aims at building a strong collaboration among three excellent universities (Universidade de Brasília (UnB), Universidade Estadual de Campinas (Unicamp), and Delft University of Technology (TUD)) and at stimulating the exchange of know-how between high level research groups in the field of image quality assessment. The Brazilian team will contribute to the project by bringing their expertise in video quality enhancement and assessment, and especially in the assessment of complex impairments. The Dutch team, instead, is experienced in overall quality assessment and in the impact of visual attention on quality perception.

5.1 COMPOSITION OF THE RESEARCH TEAM

Tables 1 and 2 show the members of the Brazilian and Dutch research teams, respectively.

Table 1– Composition of the Brazilian team

Name	Institution	Function
1. Prof. Mylène Christine Queiroz de Farias, Ph.D. (Univ. of California Santa Barbara, 2004)	Universidade de Brasília (UnB)	Researcher, Coordinator
2. Prof. Ricardo Lopes de Queiroz, Ph.D. (University of Texas at Arlington, 1994)	Universidade de Brasília (UnB)	Researcher
3. Prof. Bruno Luiggi Macchiavello, Ph.D. (Universidade de Brasília, 2009)	Universidade de Brasília (UnB)	Researcher
4. Prof. Max Henrique Machado Costa, Ph.D. (Stanford University, 1983)	Universidade Estadual de Campinas (Unicamp)	Researcher
5. Rafael Sarres de Almeida, M.Sc. (Universidade de Brasilia, 2004)	Universidade de Brasília (UnB)	Ph.D. student
6. Alexandre Fieno da Silva, M.Sc. (Universidade Federal de Uberlândia, 2006)	Universidade de Brasília (UnB)	Ph.D. student
7. Pedro Garcia Freitas	Universidade de Brasília (UnB)	M.Sc. student
8. Tainá Borges Andrade Garrido	Universidade de Brasília (UnB)	M.Sc. student
9. Renan Utida Ferreira	Universidade de Brasília (UnB)	Ph.D. student

Table 2– Composition of the Dutch team

Name	Institution	Function
1. Dr. Judith Alice Redi, Ph.D. (University of Genoa, 2010)	Delft University of Technology	Researcher, Coordinator
2. Prof. Ingrid Heynderickx, Ph.D. (University of Antwerp, 1986)	Delft University of Technology	Researcher
3. Prof. Huib De Ridder (Technical University of Eindhoven, 1987)	Delft University of Technology	Researcher
4. Dr.Hantao Liu	Delft University of Technology	Researcher
5. Hani Alers	Delft University of Technology	Ph.D. student
6. Dr. Harold Nefs	Delft University of Technology	Researcher

The **Brazilian team** in Universidade de Brasilia (UnB) and Universidade Estadual de Campinas (Unicamp) have a large experience in the area of video quality, digital television, and video coding. The Brazilian coordinator, Professor Mylène Farias, is a specialist in the area of video quality, having worked specifically on the subjective assessment of video quality and on the related development of no-reference video quality metrics. Professors Max Costa, Bruno Macchiavello, and Ricardo de Queiroz have both been part of the Brazilian research team that implemented digital television in Brazil. The research facilities available to the Brazilian team include the Laboratório de Imagens, Sinais e Áudio (LISA – Laboratory of Images, Signals and Audio) at UnB and the Laboratório de Comunicações (Communications Laboratory) at Unicamp. Both laboratories are equipped with state of the art computers, multimedia servers, and high quality displays.

The **Dutch team** (IQLab) is based in the faculties of Electrical Engineering, Mathematics and Computer Science (EEMCS) and Industrial Design (IDE) of Delft University of Technology (TUD). The IQLab joins a group of researchers with expertise in methodologies for subjective quality assessment and in objective modeling of perceived quality (including visual attention). These researchers can rely on the well-equipped Experience Lab of the Man Machine Interaction group of TUD (http://mmi.tudelft.nl/experiencelab/). The Experience Lab features high quality monitors, a stereoscope, calibration equipment, an advanced lighting system in the ceiling, and an eye-tracking device that will be used in the proposed research (see previous section). The coordinator of the Dutch Research team, Dr. Judith Alice Redi is Assistant Professor at TUD and is an expert in visual quality perception modeling via computational intelligence tools and in visual attention implications in quality assessment tasks. Professor Ingrid Heynderickx has a long-term experience in perceptive evaluation of display and lighting systems. Besides being the head of the Delft Image Quality Laboratory (http://mmi.tudelft.nl/iqlab/index.html), she is also part of the Group of Visual Experiences at Philips Research Laboratories. Prof. Huib de Ridder is an experienced researcher in the field of visual perception and user quality of experience, with a strong emphasis on methodologies for subjective quality assessment. Hani Alers and Hantao Liu are both experienced researchers in the field of image and video quality perception. Dr. Harold Nefs is a senior researcher at TUD with a strong background in visual perception and its modeling. As well as the other researchers of the Dutch team, he is member of the perceptual intelligence lab (π-Lab, http://pi-lab.nde-lab.nl/aboutus.html).

It is worth pointing out that the researchers of the Dutch and Brazilian research teams have previously worked together in a joint project between University of California Santa Barbara and Philips Research Laboratory, exploring perceptual aspects involved in video quality.

5.2 PLAN OF ACTIVITIES

The project is planned for a period of four years. The first three years will be used to perform the subjective experiments and collect the data useful for the formulation of the final quality model. The fourth year will be used to integrate this information and design the model of the impact of combined artifacts on visual quality. As mentioned earlier, TUD has available state-of-the-art facilities for visual perception studies. To use all facilities to their full advantage, it is crucial that perceptual experiments are performed in the Delft Experience Lab of TUD. Furthermore, since more students are involved in Brazil, they should have the ability to benefit from the experience of the Dutch team in performing perceptual experiments, favoring in this way the exchange of knowledge between the two countries.

Annual meetings to coordinate the project development are expected to be held in Brazil, since that requires traveling of a smaller team.

To conclude a full cycle of one subjective experiment, we plan to perform 4 steps:

1) Experiment planning and test material preparation: Depending on the stage at which the experiment is performed, this phase could take between 3 and 6 months. The tasks involved in this phase can be performed at a distance, at the same time in Brazil and the Netherlands. Nonetheless, to ensure coordination a yearly meeting is to be envisioned.

2) Experiment running: We expect that this phase will take approximately one month. The experiments should be performed in the Netherlands, as TUD will provide the facilities and the equipment needed.
3) Data analysis: This phase can be carried out separately in the Netherlands and in Brazil, provided an agreement on the strategy to be adopted is settled in advance (3 to 6 months).
4) Wrap-up and publication of the outcomes: The outcomes of each experiment should be discussed in an end-of-experiment meeting, during which a draft of a publication should also be produced. The writing of the publication can then be carried out separately in the Netherlands and in Brazil (3 months).

As discussed in the “Research Plan and Division of Tasks” session, four types of experiments are planned: E1-E4 (see Figure 1). From experiments WP1-E1 and WP1-E2, visibility, annoyance, and description data will be collected. In parallel to this, eye-tracking data (WP3) will also be collected and analyzed. The collection will coincide with the timing envisioned for running the experiments. The data analysis will be performed in the following 6 months. Experiment WP2-E3 will serve to set up the *quality ruler* adapted to video quality assessment and, therefore, is an exception to the described division. Given the strong expertise of the Dutch team in the field, all the four phases should be performed at Delft University of Technology. Finally, the integration phase will take approximately 9 months. The activities involved in this phase should be performed consistently in one of the two locations, as coordinating activities at a distance would be unfeasible.

To favor the exchange of scientific knowledge and to consolidate the educational aspects of this proposal, we plan to perform 3 student exchanges throughout the project. The students will be hosted at TUD facilities without further charge for tuition fees and operational costs, besides those related to the use of the experimental facilities (see annex financial plan). A first internship is envisioned at the very beginning of the project, aiming at educating one PhD student on empirical research methods targeted to perceptual studies. The visiting student will take care of the set-up of experiment WP1-E1, including the arrangement of the experimental environment and the eye-tracker calibration. After collecting experimental data, the student will

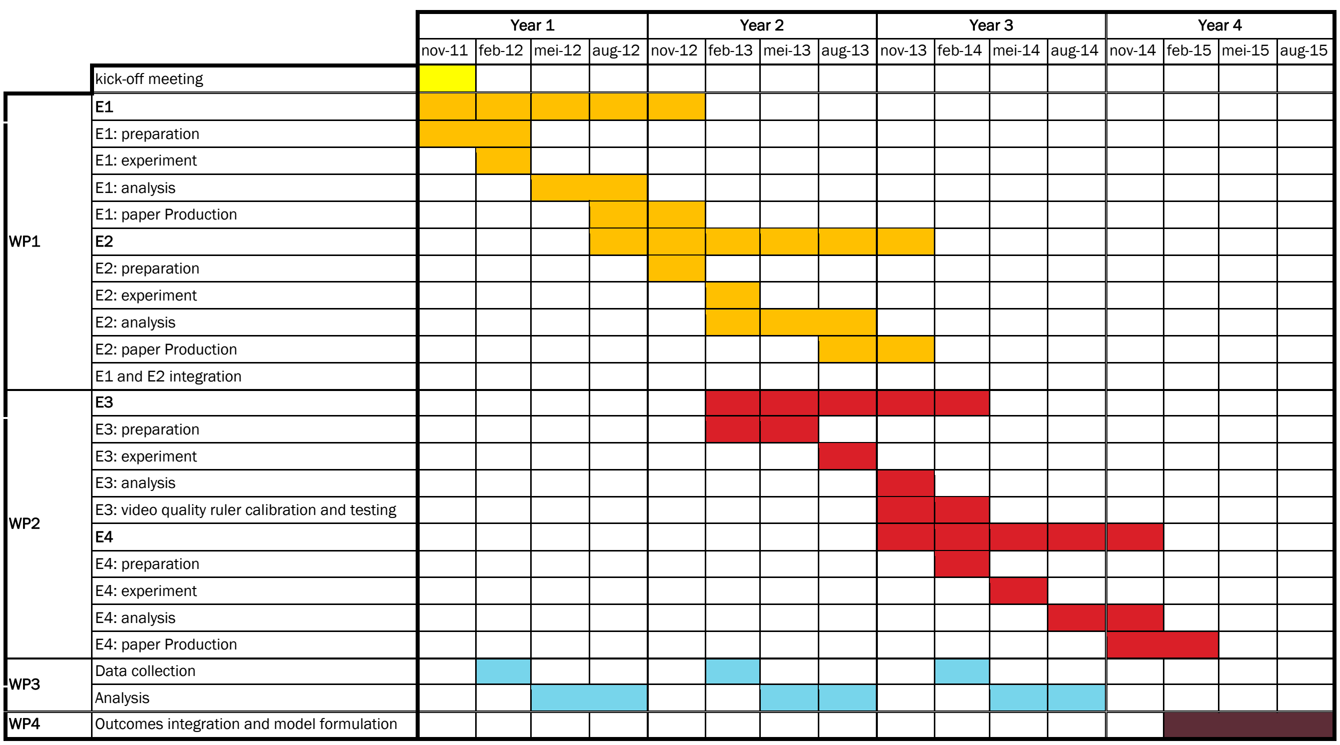


Figure 6 – Timeline of the project activities

also learn how to process and analyze the information collected. These activities will last for a period of 6 months. For experiment WP2-E3, we envision a 9 months internship for a Brazilian PhD student at TUD. In this case, the student will deal with a very specific problem, i.e. the implementation of a psychometric tool. Several experimental sessions will be needed to define a well calibrated SQS; the student will learn how to deal with problems arising from data collection and analysis such as individual differences, inversions in scaling and inapplicability of standard models to the data. Furthermore, the tool will have to be evaluated and an experimental protocol for the video quality ruler will have to be defined. Finally, for the integration phase, we envision an additional 9 months internship for a Brazilian PhD student at TUD. The student will benefit from the expertise of the Dutch team in the development of an objective metric that includes saliency information. Machine learning techniques may be investigated to model the mechanisms observed in the experimental phase.

The outcomes of the internship and of the whole project will eventually be discussed in a plenary meeting in Brazil. The timeline of the project activities is shown in Figure 6.

5.3 MISSION PLAN FOR THE PERIOD OF 4 YEARS (NOVEMBER 2011– OCTOBER 2015):

- November 2011 – Dutch mission to Brazil (Brasília and/or Campinas).
    - Goal: Kick-off meeting. Elaboration on details of the WP1-E1 experiment with the goal of estimating visibility, annoyance, and description of a set of spatial and temporal artifacts. An eye-tracker (WP3) will also be used in the experiment to register the movements of the participants' eyes.
    - Dutch researchers involved: Ingrid Heynderickx (15 days), Judith Redi and Huib de Ridder (15 days).
- January 2012 – Brazilian mission to The Netherlands (Delft).
    - Goal: First experimental phase (6 months internship). Learning outcomes: acquisition of basic skills in psychophysics and empirical research methods. Activities: experimental setup for WP1-E1, including eye-tracker set-up and calibration. Execution of the experiment and data analysis.
    - Brazilian researchers involved: 1 PhD Student (internship of 6 months).
- February 2012 – Brazilian mission to The Netherlands (Delft).
    - Goal: Preparation and execution of WP1-E1 experiment. Activities: Definition of statistical analysis to be performed on the data collected in the WP1-E1 experiment (i.e., visibility, annoyance, description, and attention data (WP3), collected with the eye tracker).
    - Brazilian researchers involved: Mylène Farias (40 days) and Ricardo de Queiroz (15 days).
- November 2012 – Dutch mission to Brazil (Brasília and/or Campinas).
    - Goals: Discussion of the results of WP1-E1 and WP3 (related to first experiment). Set up of a scientific paper to report these results. Elaboration on the details for experiment WP1-E2 with the goal of measuring the perceptual strength of spatial and temporal artifacts.
    - Dutch researchers involved: Ingrid Heynderickx and/or Judith Redi and/or Huib de Ridder (15 days).
- February 2013 – Brazilian mission to The Netherlands (Delft).
    - Goal: Preparation and execution of WP1-E2 experiment. Definition of statistical analysis to be performed on the collected data from the WP1-E2 experiment (i.e., perceptual strength and attention data (WP3), collected with an eye tracker).

    - Brazilian researchers involved: Mylène Farias (40 days).
- March 2013 – Brazilian mission to The Netherlands (Delft).
    - Goal: Implementation of the Video Quality Ruler (9 months internship). Learning outcomes: experience in implementing psychometric tools and methodologies. Activities: Preparation and execution of WP2–E3 experiment. Calibration and testing of the Video Quality Ruler.
    - Brazilian researchers involved: 1 Ph.D. Student (internship of 9 months).
- November 2013 – Dutch mission to Brazil (Brasília and/or Campinas).
    - Goal: Discussion of the results of WP1-E2 and WP3 (related to second experiment). Start preparation of a scientific paper to report these results. Update on WP2 – E3. Elaboration on the details for the experiment WP2-E4, with the goal of measuring overall quality. An eye-tracker (WP3) will also be used in the experiment to register the movements of the participants' eyes.
    - Dutch researchers involved: Ingrid Heynderickx and/or Judith Redi and/or Huib de Ridder (15 days).
- February 2014 – Brazilian mission to The Netherlands (Delft).
    - Goal: Discussion of the results of WP2-E3. Preparation and execution of WP2–E4. Definition of the statistical analysis to be performed on the collected data. An eye-tracker (WP3) will also be used in the experiment to register the movements of the participants' eyes.
    - Brazilian researchers involved: Mylène Farias (40 days) and Max Costa (15 days).
- November 2014 – Dutch mission to Brazil (Brasília).
    - Goal: Discussion of the results of WP2–E4 and WP3. Preparation of a scientific paper to report these results.
    - Dutch researchers involved: Ingrid Heynderickx and/or Judith Redi and/or Huib de Ridder (15 days).
- February 2015 – Brazilian mission to The Netherlands (Delft).
    - Goal: Overview of all data collected throughout WP1, WP2, and WP3, and discussion on the approach of how to integrate all these data into one objective model for perceived overall quality (WP4)
    - Brazilian researchers involved: Mylène Farias (40 days).
- February 2015 – Brazilian mission to The Netherlands (Delft).
    - Goal: Design of an objective quality metric for the prediction of the annoyance of combined artifacts. Learning outcomes: objective quality metrics design, data fusion, machine learning for perception modeling. Activities: analysis and integration of data collected throughout WP1, WP2, and WP3. Development of a perceptual metric based on the results and conclusions gathered. Validation of the model (WP4).
    - Brazilian researchers involved: 1 Ph.D. Student (internship of 9 months).
- October 2015 - Dutch mission to Brazil (Brasília).
    - Goal: Conclusive meeting. Discussion on the outcomes of WP4. Preparation of a scientific paper with the final results.
    - Dutch researchers involved: Ingrid Heynderickx, Judith Redi and Huib de Ridder (15 days).

6 REFERENCES

[1] International Telecommunication Union, “ITU Recommendation P.930: Principles of a reference impairment system for video," 1996.

[2] M. Yuen and H. R. Wu, "A survey of hybrid MC/DPCM/DCT video coding distortions," Signal Processing, vol. 70, pp. 247 - 278, October 1998.

[3] International Telecommunication Union, “ITU-T Recommendation BT.500-8: Methodology for the subjective assessment of the quality of television pictures,” 1998. E.G Eskicioglu and P.S. Fisher, “Image quality measures and their performance,” IEEE Trans. Image Processing, 43(12), pp. 2959–2965 (1995).

[4] Zhou Wang and Alan C. Bovik, “Mean Squared Error: Love It or Leave It? A new look at signal fidelity measures”, IEEE Signal Processing Magazine, 26(1), pp. 98-117 (2009).

[5] B. Girod, “What’s wrong with mean-squared error?,” in Digital images and human vision, edited by Andrew B. Watson, MIT Press, Cambridge, Massachusetts, pp. 207–220 (1993).

[6] S. Daly, “The visible differences predictor: an algorithm for the assessment of image fidelity,” in Digital Images and Human Vision, edited by Andrew B. Watson, MIT Press, Cambridge, Massachusetts, pp. 179–206 (1993).

[7] Zhou Wang, Alan C. Bovik, Hamid R. Sheikh, and Eero P. Simoncelli, “Image quality assessment: From error visibility to structural similarity,” IEEE Transactions on Image Processing, 13, pp. 600–612 (2004).

[8] VQEG, “Final report from the video quality experts group on the validation of objective models of video quality assessment - Phase II,” Tech. Report (2003).

[9] S. Winkler and P. Mohandas, “The evolution of video quality measurement: From PSNR to hybrid metrics,” IEEE Transactions on Broadcasting, 54(3), pp. 660–668 (2008).

[10] M.H. Pinson and S. Wolf, “A new standardized method for objectively measuring video quality,” IEEE Transactions on Broadcasting, 50(3), pp. 312–322 (2004).

[11] J. Caviedes and J. Jung, "No-Reference Metric for a Video Quality Control Loop," Proc. 5th World Multiconference on Systemics, cybernetics, and Informatics, July 2001, vol. 13, part 2, pp. 290-5.

[12] FARIAS, Mylene Christine Queiroz de ; MITRA, Sanjit K . No-Reference Video Quality Metric based on Artifact Measurements. In: IEEE International Conference on Image Processing, 2005, Genova. Proc. of IEEE International Conference on Image Processing, 2005. ICIP 2005., 2005. v. 3. p. 141-144.

[13] W. Lin, C.-C. Jay Kuo, Perceptual Visual Quality Metrics: A Survey. J. Vis. Commun. (2011).

[14] A.K. Moorthy and A.C. Bovik, “Visual Quality Assessment Algorithms : What Does the Future Hold? ”International Journal of Multimedia Tools and Applications, Special Issue on Survey Papers in Multimedia by World Experts, Vol: 51 No: 2, February 2011, Page(s): 675-696

[15] FARIAS, Mylene Christine Queiroz de, FOLEY, John M, MITRA, Sanjit Koumar; Detectability and Annoyance of Synthetic Blocky, Blurry, Noisy, and Ringing Artifacts. IEEE Transactions on Signal Processing, v. 55, p. 2954 - 2964, 2007.

[16] M. S. Moore, J. M. Foley, and S. K. Mitra, "Defect visibility and content importance: Effects on perceived impairment," Image Communication, vol. 19, pp.185-203,February 2004.

[17] Quan Huynh-Thu and Mohammed Ghanbari, “Modelling of spatio–temporal interaction for video quality assessment ”, Signal Processing: Image Communication 25 (2010) 535–546 .

[18] Jr. Ahumada, A. J. and C.H. Null, “Image quality: a multimensional problem,” Digital Images and Human Vision, pp. 141–148, 1993.

[19] Keelan, B., “Handbook of image quality: characterization and prediction,” Marcel Dekker, Inc., New York, 2002

[20] S. Bech, R. Hamberg, M. Nijenhuis, C. Teunissen, H. Looren de Jong, P. Houben and S.K. Prainanik, “The rapid perceptual image description method (RaPID)”, SPIE 2657, 1996.

[21] H. de Ridder and G. M. M. Majoor, “Numerical category scaling: an efficient method for assessing digital image coding impairments,” in Proc. SPIE Human Vision and Electronic Imaging: Models, Methods and Applications, Santa Clara, CA, USA, 1990, vol. 1249, pp. 65–77.

[22] Redi, JA, Liu H, Alers H, Zunino R, Heynderickx I. 2010. Comparing subjective image quality measurement methods for the creation of public databases. IS&T/SPIE Electronic Imaging 2010 and Image Quality and System Performance VII

[23] Yarbus, A., “Eye movements and vision,” New York: Plenum Press, 1967.R. Desimone & J. Duncan, “Neural mechanisms of selective visual attention,” Annual Review of Neuroscience no. 18, 1995
[24] R. Desimone & J. Duncan, “Neural mechanisms of selective visual attention,” Annual Review of Neuroscience no. 18, 1995
[25] H. Alers, H. Liu, J.A. Redi and I. Heynderickx “Studying the risks of optimizing the image quality in saliency regions at the expense of background content,” in proc. IS&T/SPIE Electronic Imaging - Image Quality and System Performance VII, 2010
[26] A.K. Moorthy and A.C. Bovik, “Visual importance pooling for image quality assessment”, IEEE Journal on Special Topics in Signal Processing, Special Issue on Visual Media Quality Assessment, Vol: 3 No: 2, April 2009, 193-201
[27] H. Liu and I. Heynderickx, “The added value of visual attention in objective quality metrics”, in Proc. IEEE Int. Conf. ICIP 2009, 2009.
[28] J. Redi, H. Liu, P. Gastaldo, R. Zunino and I. Heynderickx, “How to apply spatial saliency into objective metrics for JPEG compressed images?”, In. Proc. IEEE ICIP 2009, 2009.
[29] Mylene C.Q. Farias, No-Reference and Reduced Reference Video Quality Metrics: New Contributions, Ph.D Dissertation, Department of Electrical and Computer Engineering, University of California, Santa Barbara, CA, USA, September 2004.
[30] Keelan, B., and Urabe, H., “ISO 20462, A psychophysical image quality measurement standard,” Proc. SPIE 5294, 2004.
[31] Jin, E.W., Keelan, B.W., Chen, J., Phillips, J.B., and Chen, Y., “Softcopy quality ruler method: implementation and validation”, *Proc. SPIE 7424*, 2009.
[32] Parducci, A., and Perrett, L.F., “Category rating scales: Effects of relative spacing and frequency of stimulus values”, Journal of Experimental Psychology, no. 89, 1971.
[33] C. Koch and S. Ullman, “Shifts in Selection in Visual Attention: Toward the Underlying Neural Circuitry,” Human Neurobiology, vol. 4, no. 4, pp. 219-27, 1985
[34] Ninassi, O. Le Meur, P. L. Callet, D. Barba, and A. Tirel, “Task impact on the visual attention in subjective image quality assessment,” Proc. EUSIPCO-06, 2006.
[35] Redi JA, Liu H, Zunino R, Heynderickx I. “Interactions of visual attention and quality perception”, In: IS&T/SPIE Electronic Imaging 2011 and Human Vision and Electronic Imaging XVI., Vol. 7865. 2011.