

Aarhus University

Research & Development Project

PERCEPTUAL IMAGE EVALUATION

Author:

Line Aggerbo Johansen
201302451

Supervisor:

Christian Fischer Pedersen

January 7, 2016

TABLE OF CONTENTS

1	Introduction	1
2	Assessment Techniques	2
2.1	Absolute Category Rating	2
2.2	Degradation Category Rating	3
2.3	Subjective Assessment Methodology for Video Quality	4
2.4	Discussion	5
3	Design	6
3.1	Section heading	6
4	Implementation	7
4.1	Section heading	7
5	Concluding remarks	8

INTRODUCTION

This report describes the process of a Research and Development project carried out at Aarhus University. The project is developed as a subproject of a Master Thesis which is carried out in in spring 2016. The purpose of the Master Thesis is to conduct a statistical comparison of subjective image quality evaluation with objective image quality metrics, e.g. MSE, PSNR, and SSIM.

The purpose of this Research and Development project is to analyze multiple subjective image quality evaluation assessment techniques. From this analysis one assessment technique is chosen to be carried out. This project leads to an implemented application which will be used in the Master Thesis described above.

All design and implementation choices made throughout this project is reasoned and described in this report together with screen shots from the final implementation of the application.

ASSESSMENT TECHNIQUES

When collecting subjective image quality evaluations numerous of techniques exist. Three of the existing techniques are described in this chapter. The benefits and drawbacks of using each assessment technique in this project is also discussed in this chapter.

In a subjective image quality evaluation assessment scenario an observer are presented for one or more stimulus (video or image) at a time and must evaluate the presented stimulus.

2.1 Absolute Category Rating

Absolute category rating (ACR) is a discrete rating system most commonly used for video quality testing [1] but the usage of it for image quality testing is also seen [3]. The rating scale has five categories; bad, poor, fair, good, and excellent, and these labels are used for all the ACR rating scales.

The rating scale used in ACR can be a five-grade rating scale from 1 (bad) to 5 (excellent). This is referred to as the ACR-5. It can also consist of a higher number of grades like the ACR-9 or ACR-11 which are presented together with the ACR-5 in both [1] and [3].

The process of the ACR system is to present each stimulus to the observer one at a time and only once. For stimuli being images each image is presented to the observer for approximately 10 seconds followed by a gray screen from where the observer must rate the presented stimulus. The time for the observer to do the rating depends on the rating scale but should not exceed 10 seconds. The process is illustrated in Figure 2.1.

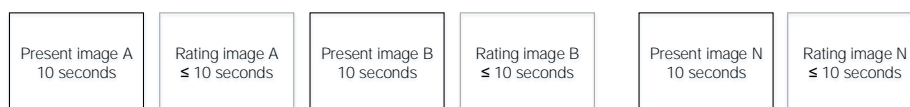


Figure 2.1: The process of the ACR assessment technique. An image is presented for 10 seconds followed by a gray screen from where the observer can evaluate the image. This process is repeated until all images has been evaluated.

At the end of the process the observer should have assigned a rating score to all presented images. When all observers have assigned a rating score for each stimulus a mean opinion score (MOS) is found for each stimulus by averaging all rating scores.

2.2 Degradation Category Rating

Degradation category rating (DCR) is similar to ACR a discrete rating system. This system is most commonly used when quality evaluation of stimulus compared with some original stimulus is wanted [1]. Due to the comparison the labels for the DCR system is a little different; very annoying, annoying, slightly annoying, perceptible but not annoying, and imperceptible. Besides the labels the rating scale is the same as in ACR where 1 is 'very annoying' and 5 is 'imperceptible'.

The process of the DCR system is much similar to the ACR system. The difference consists in two instead of one stimulus (the original stimulus and the stimulus for evaluation) being presented to the observer before the rating is being made. Between the two images a short break of two seconds is placed and the observer sees a gray screen. As with ACR the time for the stimulus to be presented is approximately 10 seconds when stimulus is being images. The process is illustrated in Figure 2.2. As with ACR a MOS can be found for each image.



Figure 2.2: The process of the DCR assessment technique. Two stimulus are presented to the observer with a small break in between. The observer rates the second stimulus (modified) compared with the first stimulus (original).

2.3 Subjective Assessment Methodology for Video Quality

Subjective assessment methodology for video quality (SAMVIQ) is another assessment technique which differs from the other assessment techniques in multiple ways even though the labels; bad, poor, fair, good, and excellent are the same as the ones used in ACR. As the name says this technique is developed for use in video quality evaluation but the use of it for image quality evaluation is also seen [3].

The rating scale in SAMVIQ differs hence it goes from 0 to 100 where the label 'bad' is aligned with the score 10 and the label 'excellent' is aligned with the score 90. This much larger scale has shown to be unnecessary large and the observers ratings might be grouped in 5 or 9 groups over the rating scale range [3].

The process for the system also differs from earlier described systems. In SAMVIQ the process is split into a number of scenes where each scene contains a number of different variations of the same stimulus. Within a scene the observer can browse between the different variations of the stimulus. Also the observer is able to evaluate the image while observing it. When all stimulus has been rated another scene might be presented and the process repeats itself. Notice that the order of the different variations of a stimulus changes for every scene. The process is illustrated in Figure 2.3.

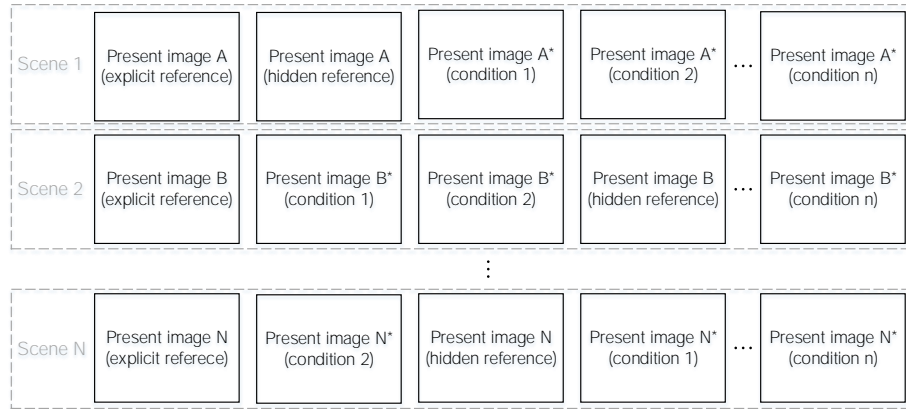


Figure 2.3: SAMVIQ method

The illustration in Figure 2.3 shows the use of a hidden reference. In SAMVIQ the use of a hidden reference is mandatory [2].

2.4 Discussion

The right choice of assessment technique depends highly on the application which it is used for. In this project it should be used to evaluations of images which lack of quality is imposed by some kind of lossy compression technique. The image quality will differ in range of very good quality to very poor quality. The evaluation assessment technique therefore must work for both cases. The hardest thing to evaluate is the smaller errors which can not be easily seen with the eye. Therefore the greatest constraint for the assessment technique is that it should give the observer the best opportunity to see and evaluate these type of errors.

The greatest benefit of the ACR method is that it is easy to understand and implement. The drawback with respect to this project is that the ACR does not have an explicit reference which is recommended in evaluations when the impairments of the images are small [1]. Both DCR and SAMVIQ has the ability of using an explicit reference.

One difference of DCR and SAMVIQ is the rating scale. As mentioned before experiments show that the rather large rating scale in SAMVIQ is superfluous [3] and the 5, 9 or 11 grade rating scale in DCR would be enough.

Another difference is the time an observer has to rate each image. In DCR the time for evaluation is limited and the observer only get one chance to see each image where SAMVIQ has no time constraints and the observer gets to see the image as many times and for as long time as wanted.

DESIGN

3.1 Section heading

IMPLEMENTATION

4.1 Section heading

CONCLUDING REMARKS

Bibliography

- [1] ITU-T Recommendation P.910. Subjective video quality assessment methods for multimedia applications. Technical report, 1999.
- [2] Franc Kozamernik, Paola Sunna, E Wyckens, and DI Pettersen. Subjective quality of internet video codecs - Phase 2 evaluations using SAMVIQ. *EBU Technical Review*, (January), 2005.
- [3] David M Rouse, Romuald Pépion, Patrick Le Callet, and Sheila S Hemami. Trade-offs in subjective testing methods for image and video quality assessment. In Bernice E. Rogowitz and Thrasyvoulos N. Pappas, editors, *Proc. SPIE*, volume 7527, pages 75270F–75270F–11, feb 2010.