MSU Perceptual Video Quality tool 1.0

Tool for subjective video quality evaluation

Project, ideas: Dmitriy Vatolin

Implementation: Oleg Petrov

Table of contents

Introduction to "MSU Perceptual Video Quality" Tool

"MSU Perceptual Video Quality tool" is a tool for subjective video quality evaluation. It can be used to evaluate performance of video processing system using experts' opinion. This tool consists of two parts.

First part is "MSU Perceptual Video Quality - task manager". It can be used to create task file and to collect expert's results. You just add video files (.avi, .avs) that you want to compare, set amount of test that you want expert to go through, give name to task and save task file.

Second part is "MSU Perceptual Video Quality - player". It can play video files from task in pairs and record experts' reaction.

## Subjective Video Quality Evaluation

Types of video quality evaluation

There are two ways to evaluate video quality: by objective or subjective measures. Objective measures are automatically extracted from video file (or from pair of files), for example by computer program [3]. But video processing systems are created to give best quality for users, not for programs, and objective measures cannot fully characterize reaction of those who might use such system. Subjective measures are extracted from marks given by people, and can help to evaluate users' opinion.

Methods of subjective quality evaluation

There are a lot of subjective comparison methods, each with its advantages and disadvantages. Most of them can be found in [2]. Often expert is asked to evaluate quality after sequence playback has finished. This scheme is not good when you want to compare videos whose quality is scene-dependent or time-varying (as for compressed video files); in such cases it's better to record experts' opinion during playback. Our "MSU Perceptual Video Quality" tool implements both of these approaches and contains six types of subjective comparison.

Double Stimulus Impairment Scale (DSIS)
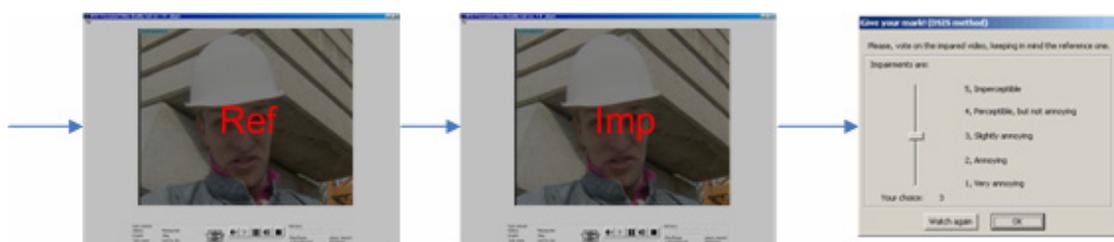
In this method two sequences are played simultaneously, first one is the reference, second is an impaired one. And after playback expert is asked to give his opinion using comparison scale:
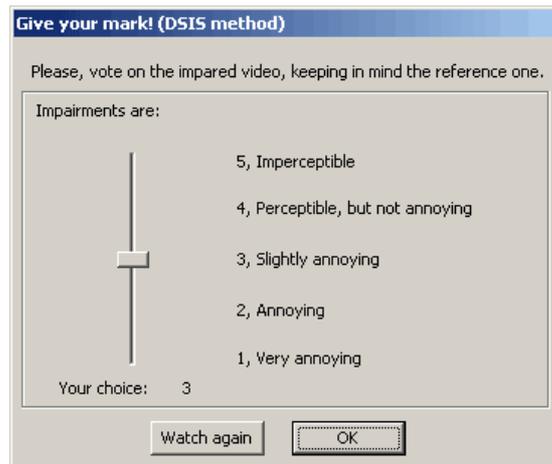
Only impaired video is evaluated, it gets mark from 1 to 5, the higher the better. In results, this mark is scaled to the range 0..10.

Double Stimulus Continuous Quality Scale (DSCQS) type I
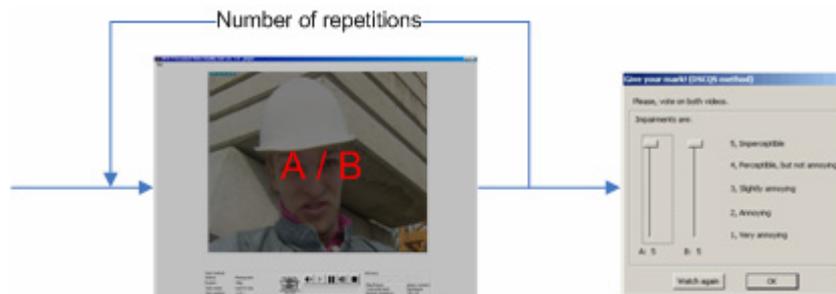
In type I videos are played in pairs in one playback window. Each pair is repeated a given amount of times ("repetition" parameter in the Task Manager). During playback expert is free to switch between two videos, named "Sequence A" and "Sequence B" by pressing Tab button. One of videos is the reference one, but expert is not informed about it. After playback expert is asked to give his opinion about each video sequence:
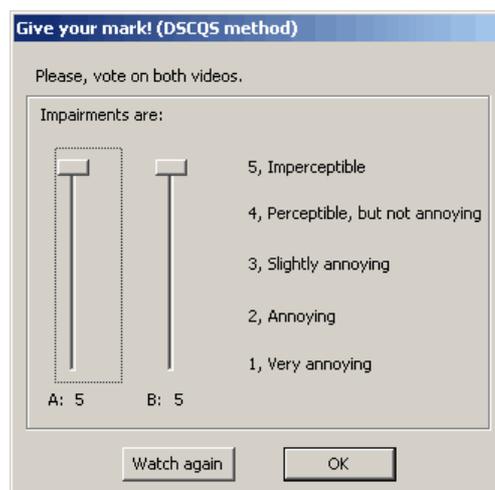
Both sequences are evaluated. Mark of each of them is counted like 5.0 + (sequence_mark – another_sequence_mark) * 1.25, thus it is scaled to 0..10. Note that mark for a video is treated as a difference of scores.

Double Stimulus Continuous Quality Scale (DSCQS) type II
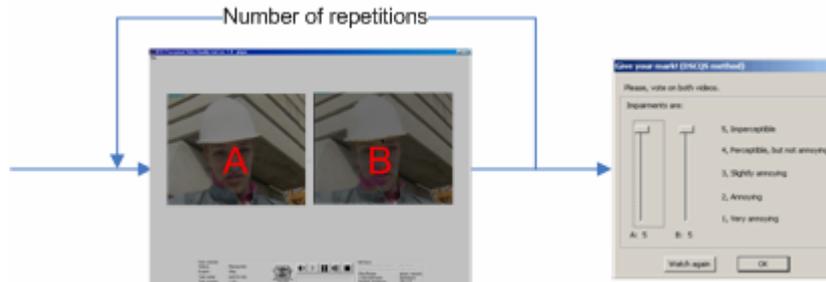

In type II (which is used more often) videos are played in pairs, and both videos are shown simultaneously. Each pair is repeated a given amount of times ("repetition" parameter in the Task Manager). As in type I, one of videos is the reference one, but expert is not informed about it. Impairment scale and resulting mark are the same as in type I.

Stimulus Comparison Adjectival Categorical Judgment (SCACJ)
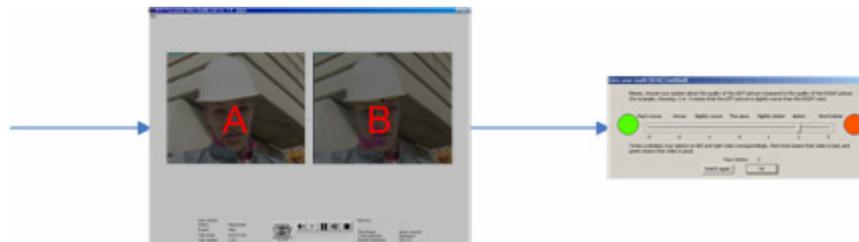
In this method two sequences are played simultaneously, and after playback expert is asked to give his opinion using comparison scale:
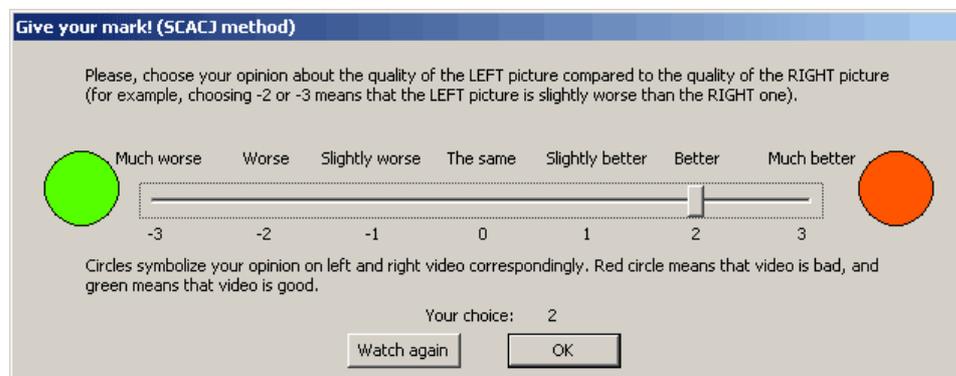
Original score is a value from -3 to 3. Mark for the left video is counted as this score + 3 and for the right one as 3 - this score. In pair of sequences, if the first one got m1 and second m2, m1 + m2 is always 6 (because there are six point on the comparison scale). In results this mark is scaled to 0..10.

Subjective Assessment Method for Video Quality Evaluation (SAMVIQ)
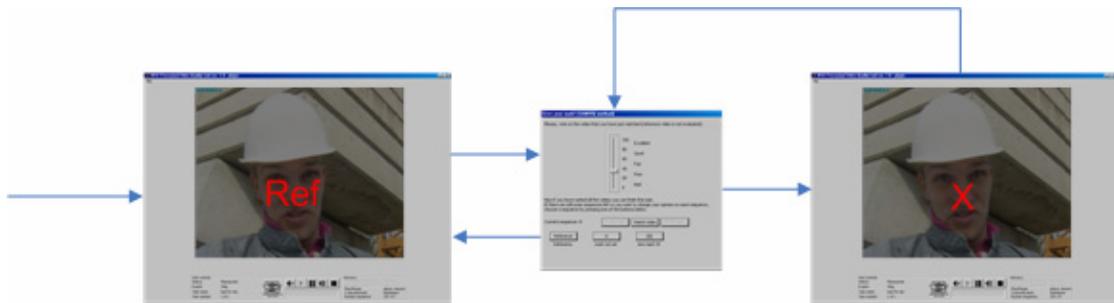
This new method was created by EBU and was recently sent for standardization.

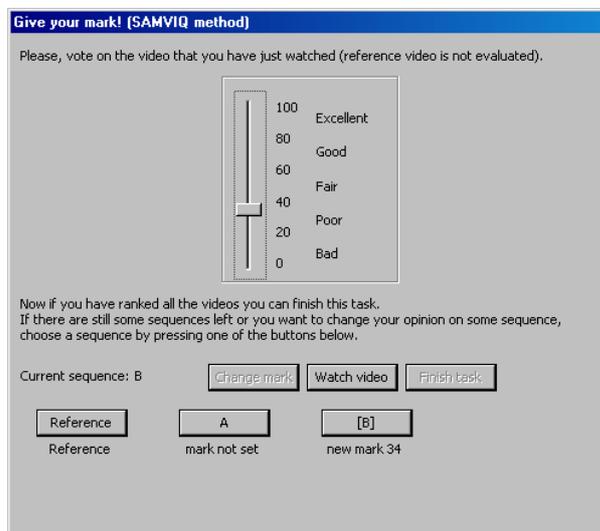During test procedure expert is able to play any sequence from the test set and give it mark, plus he is able to play reference video. Mark is in the range from 0 to 100, in results it is scaled to the range 0..10.

MSU Continuous Quality Evaluation
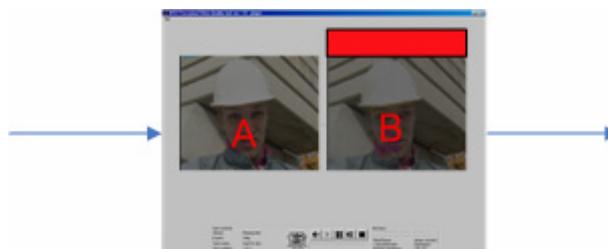
This method implements idea of continuous evaluation: two sequences are played simultaneously, and if during playback expert thinks that one sequence is worse than another, he should click left or right arrow button on his keyboard, depending on a position of the sequence he dislikes. Red mark arises on top of the video an expert clicked against.

At the end of playback, if an expert clicked m1 times against the left sequence and m2 times against the right one, then mark for the first is 10 * m2/(m1 + m2) and mark for the second is 10 * m1/(m1 + m2). If no clicks were made, both videos get 5.0.

Tool description

Step-by-step process of making subjective test

With "MSU Perceptual Video Quality Tool" it is very easy to make subjective quality evaluation. Just go through following steps:

1. Prepare task files using "MSU Perceptual Video Quality - task manager".

2. Group task files and set color blindness test if needed.

2. Experts go through tasks using "MSU Perceptual Video Quality - player".

3. Calculate results using "MSU Perceptual Video Quality - task manager".

All these steps are explained below.

Preparing task file

Choose files for task, task name and amount of tests

This tool supports **all variety of .avi files and .avs (AviSynth files)**. Add only files that can be opened on experts' computer! It is assumed that all files are different versions of one sequence (for example, compressed with different codecs). To compare a number of different sequences, create task file for each.

If you use .avs file, your script should return a value, i.e. have string "return clip" in the end. **Task name** is very important: When expert finishes task, subfolder with name of task is created in folder that contains task file. All results of experts are stored there. Results for particular expert are stored in file "expert_name".csv. **Tests amount** defines number of times when pair of sequences is shown to the expert during the task. Add extra tests if you want to check experts' reliability (using MSU averaging method described below).

Choose task type

With this tool you can make six types of subjective comparisons. All of them methods are described before.

Additional task options

- Task coverage
  Choose 'all possible pairs' if you want to have overall comparison of all sequences in a task (each to each). If you want to compare one sequence to all others, choose 'task target to all other films'. For DSCQS, DSIS and SAMVIQ a reference video is required by a test procedure, so this option is disabled.

- Swap frames

This option can be set only for MSUCQE method. With this enabled, left and right sequence will exchange their places during playback. You need to enter numbers of frames on which it should happen (or this numbers can be filled in automatically based on scene change detection algorithm).

- Average framerate
  With this option enabled, two sequences with different framerate will be played with averaged framerate. Otherwise, each will be played with its own framerate.

- Enable pause
  Sometimes quality of video seems good during playback and very bad on still frames, and vice versa. You can allow expert to pause sequence by checking this checkbox.

- Enable rewind
  You can allow user to rewind video by checking this checkbox.

Grouping tasks and color blindness test

Expert can run "MSU PVQ - player" and open any task on his or her choice.
It is possible to group any number of tasks, so expert doesn't need to open them. To do this, create .bat file with following contents:

```
"MSU perceptual video quality player.exe" "c:\tasks\task1.tsk"
"c:\tasks\task2.tsk" "c:\tasks\task3.tsk"
```

When expert runs this file, he will go through task1.tsk, task2.tsk and task3.tsk consequently.

Persons with color blindness may give unstable results, so you may want to exclude their marks from final MOS. To test an expert for the color blindness, create following .bat file:

```
"MSU perceptual video quality player.exe" -c
```

Results of this test will be added to each file with results of this expert. Test can be combined with a group of task:

```
"MSU perceptual video quality player.exe" -c "c:\tasks\task1.tsk"
"c:\tasks\task2.tsk" "c:\tasks\task3.tsk"
```

In this case expert will go through the color blindness test, and then through task1.tsk to task3.tsk.

Expert testing

Expert starts "MSU Perceptual Video Quality - player". On startup expert should enter his or her name. Two experts can't have the same name - in this case, expert will be offered to change his name during results' saving. If set, expert goes through color blindness test.

Expert goes through tests according to the method of assessments. When all tests in the task are finished, program saves expert result and another task (if there are any) begins. Results are stored in subfolder with name of task in folder that contains task file. Results for particular expert are stored there in file "expert_name".csv. Format of results is as follows:

task type,	[string with type of task]	
average framerate,	[0|1]	//is "average framerate" option enabled
pause allowed,	[0|1]	
rewind allowed,	[0|1]	
one to each,	[0|1]	
number of tests,	[number of tests]	
number of videos,	[number of videos]	
reference video,	[reference video name]	//only if task is "one to each"
video,	mark	
[video name],	[mark]	//meaning of mark depends on test methodology
...,	...	
screen resolution,	[resolution values]	
video,	decompressor	
[video name],	[decompressor name]	
...,	...	
time of assessment,	[time]	
Results of color blindness test		

Calculating results

Open task (*.tsk) for which you want to average result and click "Count results".

There are three ways of calculating final mark: **simple averaging**, **MSU averaging** and **ITU-T averaging**. You can choose any combination of these methods; results from any method that you choose are added to the final file "average_mark.csv". This file is placed in folder of *.tsk file in subfolder with the name of task.

Simple averaging

Results from experts are just averaged. No additional processing.

MSU averaging

In MSU method each expert gets mark. Mark is from 0 to 1, the higher the better. This mark is based on consistency of his results when one pair of films is shown several times. Results from expert with low mark have smaller weight in averaged results.

ITU-T averaging

In this case results are processed according to [2].

References

1. Tool page on the internet

http://www.compression.ru/video/quality_measure/perceptual_video_quality_tool_en.html

2. ITU-R BT.500-11 "Methodology for the subjective assessment of the quality of television pictures"

3. MSU Video Quality Measurement Tool, http://www.compression.ru/video/quality_measure/video_measurement_tool_en.html