World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:5, No:11, 2011

Impact of Fixation Time on Subjective Video Quality Metric: a New Proposal for Lossy Compression Impairment Assessment

M. G. Albanesi, R. Amadeo

***Abstract*—**In this paper, a new approach for quality assessment tasks in lossy compressed digital video is proposed. The research activity is based on the visual fixation data recorded by an eye tracker. The method involved both a new paradigm for subjective quality evaluation and the subsequent statistical analysis to match subjective scores provided by the observer to the data obtained from the eye tracker experiments. The study brings improvements to the state of the art, as it solves some problems highlighted in literature. The experiments prove that data obtained from an eye tracker can be used to classify videos according to the level of impairment due to compression. The paper presents the methodology, the experimental results and their interpretation. Conclusions suggest that the eye tracker can be useful in quality assessment, if data are collected and analyzed in a proper way.

***Keywords*—**eye tracker, video compression, video quality assessment, visual attention

I. Introduction

THIS paper addresses the problem of evaluating visual quality of digital video clips. The concept of visual quality, both for still images and videos, has been investigated since the very beginning of digitalization of media [1]. In particular, a great interest in quality assessment metrics appeared when image compression standards appeared [2].

Nowadays, a great deal of research has been focused on suggesting new metrics and/or methodologies to evaluate quality impairments, in still images and video. In this work, discussion is limited to the case of digital video, where impairments can be caused by several issues: compression, frame loss, packet loss over a noisy channel, reduction of resolution (frame rate or color depth), and so on. One of the most important sources of quality degradation is lossy compression, as compression is fundamental for a great part of multimedia storing, retrieval and transmission. In this paper, H.264 compression standard is considered, but the method is quite general and can be applied to other type of compression engines. Moreover, it can be applied also to other kinds of quality impairments, even if more experiments should be necessary to confirm the same efficacy (see Section VI. Conclusions). To assess video quality, a subjective approach has been chosen, even if the final results are used to propose a quantitative metric based on data collected from the experimental tasks.

M. G. Albanesi is with the Computer Department of Faculty of Engineering, University of Pavia (phone: +39-+382-985919; fax: +39+382-985350; e-mail: mariagrazia.albanesi@unipv.it).

R. Amadeo is with the Computer Department of Faculty of Engineering, University of Pavia (phone: +39-+382-985523; fax: +39+382-985350; e-mail: riccardo.amadeo@gmail.com).

This challenge could be called *from a subjective to an objective quality metric*, which is the focus of a Master Thesis [3], which provided the experimental activity on which this paper is based.

The rationale which underlies the core of the research activity here reported is to investigate the efficacy of eye trackers to obtain hints on how the human observer judges differently a collection of videos, for different compression ratio (bit rates). Each bit rate corresponds to a certain level of degradation of video visual quality, due to lossy compression. However, it is important to state that quality impairments depend not only on the compression ratio, but also on some specific features of video visual content; in particular, the presence of movement (objects moving fast in the scene, such as in football or tennis games) and, in some extent, the *richness* of color. This can be motivated by the fact that video quality assessment is a complex task, which involves not only visual perception of the *sensor* (the human eye), but also visual attention and cognitive process, which are driven by a possible a-priori knowledge of the scene and by unconsciousness focused regions of interest. For example, attention is captured mainly by the presence of faces in the scene or the ball moving in a football field.

Previous works in literature [4], [5], are essentially based on two statements: (a) it seems reasonable to consider the position of the gaze (*gaze plots*) of a human observer as an indicator of which region of interest is more relevant. This underlies the hypothesis that an impairment appearing in one of these regions is more annoying than one outside. This hypothesis is substantially at the base of all the previous works in literature [6]. (b) If we want to simulate the behavior of a human observer in a specific metric, which is better than an eye tracker to record all the eye activity while observing the video?

However, even if the two ideas seem reasonable and interesting, results reported in literature [7] give somehow deluding conclusions. In fact, metrics where saliency maps derived from eye tracker data seem not to be a good solution to discriminate levels of visual quality in different impaired version of video.

A possible explanation of this fact, as confirmed by the experimental results here reported in Section III, is that two main problems can be founded in previous works: (a) using a distortion measure based on a saliency map does not improve the performance of a metric significantly; (b) saliency maps based on gaze plot do not differ significantly if we compare videos of increasing impairments. In few words, human observer still keeps on gazing at the *same* regions, regardless the compression ratio. This fact suggests that visual attention

is more a cognitive process, which takes into account not only the visual stimulus impacting the retina, but also its semantic interpretation. In this paper, the main effort is to overcome part of these two problems, finding a solution in which the eye tracker can still be useful to provide an evaluation of compressed video, as it is fully explained in section A. *Novelties of this proposal*.

A. Novelties of this proposal

By referring to the two main problems described in the Introduction, a new methodology has been developed. Its main features are the following:

(a) A reference set of digital video widely used in literature and freely downloadable has been used. This choice is motivated by the necessity of assuring the possibility of comparing results to the ones described in literature, as the reference set is fundamental in any subjective quality assessment experiments.

Two levels of degradation covering a medium and coarse visual quality impairment has been considered. Thus, for each video, there are three versions: no impaired full reference video (the original), medium quality impaired video, and coarse quality impaired video.

(b) In opposition to the algorithms proposed in literature, this method does not consider the position of the gaze of the human observer in a scene (see Fig. 1). Even if the gaze plot is one of the data provided by the eye tracker, the proposed method does not use it to create a sort of saliency map. Instead, it considers *the duration of the fixation time* (also provided by the instrument) as an indicator of video quality impairment. The rationale under this choice is that, if a scene is full of defects (typically blur or block distortion), the attention of the eye is more attracted by these annoying details; if the impairment is particularly severe, the eye can lose the attention on the semantic interpretation of the scene. For example, for the video Football, in the version of coarse quality, the eye cannot follow in a proper way the oval ball as in Fig. 1, because it is immersed in a very annoying region of squared blocks (see Fig. 2). By comparing Fig. 1 to Fig. 2, it is clear that in high quality video the eye is closer to the position of the ball when it changes abruptly the trajectory. In fact the blue lines show an example of very different *scan paths* provided by the eye tracker, due to the different quality levels.

(c) A standard subjective evaluation [8] is performed for all the three version of the videos (see section *II.D The Methodology*), our evaluation is of type "No reference", as the observer watching the impaired versions of the video has not the reference version to make comparison to. Also the subsequent statistical analysis does not compares data form full reference video to the other version, therefore, the proposed method is fully "no reference". At the end of the statistical analysis, a classification of the video is performed, where obviously the reference video obtains always the best score.

(d) In order to eliminate the a-priori knowledge which is possible when an human observer watches several versions of the same video, the proposed method prepare a playlist of the video for each observer, in such a way that a single tester does not see the same video (even if under different versions) twice. This gets rid of a sort of memory or learning effect, already observed by other authors [7].

(e) To take into account the different visual content of video, especially if considered from a semantic point of view, the method classified the video according to *movement* and *color* (scene with rapid/slow moving objects and scene with rich/poor color) and statistical analysis has been performed for each class.

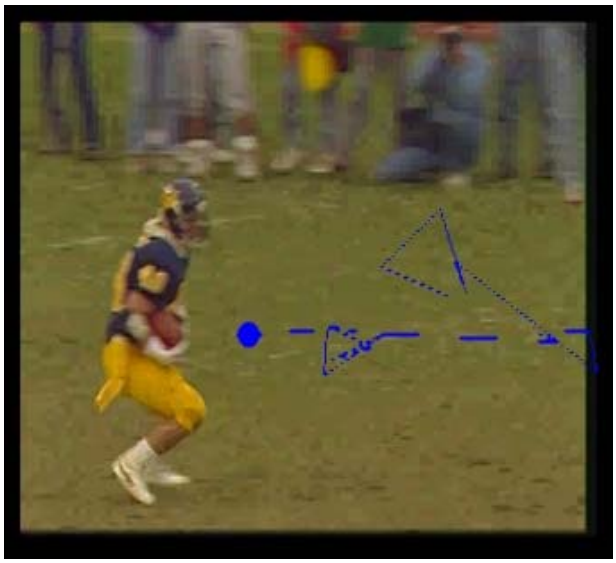

Fig. 1 An example of gaze plot on one frame of full reference high quality video.The blue line represents the position and the movement of the gaze of the human eye

Fig. 2 Same frame of Fig. 1, at coarse quality: the eye cannot keep on track on the ball

II. Eye Tracker and Quality Assessment Tasks

A. Video Sequences

To optimize the subjective scores, the number of video samples presented to each viewer was set to nineteen. This number, rather high, was intended to make the tester unconsciously adjust his/her voting range to best fit the quality evaluation needs. The files are the usual video sequences used in literature for video quality assessment, both objective and subjective. They have been downloaded from available online public libraries [9], [10], [11]. The original files are YUV sequences, 4:2:0, in CIF resolution (352x288) at 30 fps. The alphabetical order of the sequences keywords are hereunder reported along with the videos total duration:

1) Bridge_close (66 s)
2) Bus (5 s)
3) Coastguard (10 s)
4) Container (10 s)
5) Crew (10 s)
6) Flower (8 s)
7) Football (8 s)
8) Foreman (10 s)
9) Hall (66 s)
10) Highway (10 s)
11) Ice (8 s)
12) Mobile (10 s)
13) Mother_daughter (10 s)
14) News (10 s)
15)Paris (35 s)
16) Silent (10 s)
17) Stefan (3 s)
18) Tempete (8 s)
19) Waterfall (8s)

All the videos are identified by the usual name that can be found in specific literature. The difference in duration was needed to diversify also in this aspect the video sequences presented to the observer, in order to maximize his/her concentration while performing the test. Every sequence has been then analyzed, trying to give a preliminary classification of the importance of color and movement in the semantic of the each of them. The goal is to diversify the library just created, giving a rough and subjective categorization of the semantic of the sequences. The greatest problem is that, to obtain the best possible categorization, a set of subjective experiments it is mandatory to ask to testers to classify videos. This accuracy, thus, would have required a lot of time and resources (tester chosen for categorization would not be allowed to participate to the true testing phase, because after seeing a group of video a tester was not meant to see it again). For this reason, categorization of video has been realized without involving tester, but by the authors themselves: This preliminary categorization will be validated by the users during the test phase. The topic under discussion while preparing this preliminary categorization is: how much can the viewer be struck by the movement of the video? How much can he be struck by the colors? Every video was watched and rated using a three point scale (“slightly relevant”, “relevant”, “very relevant”) both for movement and color. The equivalent questions will be then inserted in the questionnaire in the experimental tasks (see Section D. *The Methodology*).

The tweaking of the sequences was to be made by compression. First of all, the YUV sequences were converted into .avi files. This process allowed a greater flexibility in choosing how to perform the outright compression. The software chose for compression was MeGUI, under GNU license. The chosen codec was H.264, also known as MPEG4-AVC (advanced video coding). The compression has been applied to create a library of impaired sequences in which each sequence is compressed twice. The compression engine is used by keeping all the parameters constant, except for the bit-rate. From each reference video two impaired sequences have been created. Of course, the lower this parameter is set, the more is the compression needed, so the transformation led to a more perceptible loss of quality. The two target bitrates were, 450 bps and 150 bps. To complete the library, the 19 reference video were also included in it, setting a total number of 57 video sequences.

B. Subjects for the Tests

The most important part of a subjective video quality assessment experiment is the human component. In choosing the participant to the activity it is very important to care about the tester vision (normal or corrected-to-normal), about their experience in both video quality assessment activities and in general utilization of video interfaces on a computer, and also about their gender, age and cultural background. All these factors are fundamental to have accurate and useful result [8].

The subjects, who voluntarily participated to the tests, were 8 females and 10 males, of age varying from 22 to 27 years old. Their vision was normal or corrected-to-normal and they had no experience in subjective video quality assessment. They, however, had normal or good experience in using IT interfaces to watch videos both online and offline. They were undergraduate or graduate college students, and they were not paid for their contribution.

C. Eye tracking protocol

The Eye Tracking apparatus is located at the Engineering Faculty of the University of Pavia (Computer Vision and Multimedia Laboratory). It is a TOBII device which runs on a PC with Windows XP and it is configured with a double monitor set-up. To calibrate it in the best possible way, it was necessary to adjust the screen resolution and the monitor colors as suggested in the User Manual [12]. To assure accuracy, the calibration phase is repeated at the beginning of experiment for each tester. The sampling frequency is 50 Hz and the accuracy of measurements is of the order of less than 1 degree of visual angle.

The most important choice is what type of experiments best fits the purposes of this work. The used methodology is the Absolute Category Rating with Hidden Reference [8]. There are many motivations for this choice: first of all, ACR methodology needs video sequences to be presented one at a time. This feature fits well with the interface offered by the

Eye Tracking software, TOBII Clearview. Another reason which oriented this decision towards the ACR-HR method was the fact that it could be used without reference. Recall that one of the novelty of this approach is that, in order to exclude the memory effect, each viewer must watch each sequence (regardless the distortion) just only once. Under this requirement, when asked to evaluate an impaired video, the observer has to be unaware about the reference video which generated the impaired version. Moreover, the ACR-HR experiment has statistical reliability, because it has smaller normalized mean of 95% confidence interval of MOS than other methods [13]. Following these considerations, ACR–HR may be considered the easiest and the more effective experimental way to assess subjectively video quality in a no-reference paradigm.

D.Methodology

The testers have been divided into groups (A, B, and C, see Table I), each one of them watching to a different playlist from the others. Three playlists have been defined, each including video versions referred to three compression rates (and so three quality levels): high, medium and no compression meant low, medium and high quality sequences. The playlists for groups A, B and C have another important feature: every video is to be included just once in each playlist. Moreover, the three playlist were assembled with always the same video in the same position in all of them. For example, if *Bridge_close* was placed as the opening sequence for playlist A, no matter which compression rate in this case, the same video was placed as opening in B and C. The idea behind this procedure is that when watching an unknown video eye and brain are reasonably interested in semantic, which is to say that they observe and concentrate on main part and main meaning of the video, with little care for details. On the other hand, when watching a known or well known video sequence the eye may be more interested in details than in the main object of the video. These different interests can easy translate into different scan paths, which can give different and maybe useless data regarding the points of fixation on the screen when the video was played. In [6], for example, the memory effect is said to be "unlikely to influence the perceived quality" due to the small duration of videos, but it could not be completely excluded.

To collect data from the experiments, it has been prepared an ad-hoc questionnaire to be submitted to every tester after watching each stimulus of his/her playlist.

The questionnaire is composed by several questions, divided into sections. The first section consists in just one (fundamental) question: testers is asked to evaluate the perceived quality using the 5 point rating scale typical for ACR5-HR experiments. The possible answers are "bad", "poor", "fair", "good" and "excellent"; corresponding to a numerical value of MOS (Mean Opinion Score) from 1 (bad) to 5 (excellent).

Second section is related to the validation of movement and color classification made while building the playlists. The observer is asked to score the content of the video, by referring to movement and color, according to a three value scale: "slightly relevant" (numerical meaning 1), "relevant" (2) and "very relevant" (3). The same scale has been used in the preliminary categorization.

TABLE I
THE THREE PLAYLISTS. REF: FULL REFERENCE VIDEO (UNCOMPRESSED ORIGINAL VIDEO), 150: 150 BPS BIT RATE, 450: 450 BPS BIT RATE. THE NAME OF THE SEQUENCES ARE REPORTED IN SECTION II.A

N.	Playlist	Group A	Group B	Group C
1	Foreman	*ref*	*150*	*450*
2	Silent	*150*	*ref*	*450*
3	Flower	*450*	*150*	*ref*
4	Bus	*450*	*ref*	*150*
5	Tempete	*ref*	*450*	*150*
6	Bridge_close	*ref*	*150*	*450*
7	Ice	*150*	*ref*	*450*
8	Coastguard	*450*	*ref*	*150*
9	Mother_daughter	*150*	*450*	*ref*
10	Football	*ref*	*150*	*450*
11	Crew	*450*	*ref*	*150*
12	Paris	*450*	*150*	*ref*
13	Container	*ref*	*450*	*150*
14	Highway	*150*	*450*	*ref*
15	Waterfall	*150*	*ref*	*450*
16	Hall	*450*	*150*	*ref*
17	Stefan	*ref*	*450*	*150*
18	News	*ref*	*150*	*450*
19	Mobile	*150*	*450*	*ref*

Third section aims at classifying the kind of disturb perceived. Two questions are asked, under the condition that the perceived quality was not high, more specifically if the answer has been "bad", "poor" or "fair". The questions asks how much the bad quality of the stimulus is related to impairments on the movement or on the colors, in the viewer's opinion. Also in this section the rating scale used was the three point discrete one used in section two, with the meaning of "low" (numerical value 1), "medium" (2) and "high" (3) perceived impairments. If no quality degradation has been perceived consciously by the observer, the score is set to zero, meaning that the observer is able to perceive the poor quality, but he/she cannot judge if movement or color (or both) are seriously compromised. It was necessary to consider this last case when calculating the mean score related to the two kind of semantic aspect (movement and color richness).

Final section of the questionnaire leaves to the user the freedom to say everything that came to his/her mind while watching each sequence, of course if related to the subject under discussion.

III. EXPERIMENTAL RESULTS

Subjective evaluation has been performed for all the tester and the playlists, thus generating a huge set of data, varying from the subjective marks of each tester to the gaze point tracked by the apparatus, from the duration of each fixation to the video of each scanpath. The data have been collected and analyzed in tables on a spreadsheet (see Fig. 3). As it has been already pointed out, the interesting parameter is the duration

of the fixations, so the first step was to transfer all the textual data to the spreadsheet in order to have them ordered by the information they carry out. The fixations of each tester were divided into three groups, according to the playlist seen by the test subject, and on those tables for each viewer, the mean fixation time for each stimulus (MFTpS) is computed. Then, if n is the number of fixations per stimulus and ft_i each fixation duration (i varies in the range from 1 to n), we define:

$$MFTpS = \frac{\sum_{i=1}^{n} ft_i}{n} \quad (1)$$

It is necessary to compute the mean value of MFTpS over each group of viewer (Mean Fixation Time, MFT). As there are six viewer in each group the MFT value is:

$$MFT = \frac{\sum_{j=1}^{6} MFTpS_j}{6} \quad (2)$$

After doing the preliminary elaboration of the data and correcting the errors when necessary (for example by excluding data referring to black frames between two consecutive stimulus of the playlists), the successive step is the full understanding and deep elaboration of experimental results, in order to link subjective data to objective measurements. This is the object of next section, Data Interpretation. Table II summarize a conceptual flow chart of the whole methodology described so far.

Data properties:
Recording date: 06/04/2011
Recording time : 12:04:14:375 (corresponds to time 0)
Study: TLV-ts
Subject: XXXXXXXXXX
Recording: 12-4
Screen resolution: 1280 x 1024
Coordinate unit: Pixels

Filter settings:
Eye: Average
Validity level: Normal
Fixation radius: 30
Min duration: 100

Fix number	Timestamp	Duration	GazepointX	GazepointY
1	107	279	746	572
2	406	877	626	521
3	1303	339	545	521
4	1662	1037	516	541
5	2719	359	596	537
6	3337	199	617	559
7	3556	538	589	618
8	4114	279	523	609
9	4414	319	521	517
10	4792	279	632	539
MFTpS		450,5		

Fig. 3 A small chunk of the huge amount of data provided by the eye tracker for each video sequence of the playlists

TABLE II
CONCEPTUAL FLOW CHART OF THE METHODOLOGY FOR THE EXPERIMENTAL ACTIVITY

Choice of the ACR experiment	Simple Statistically reliable and effective
Selection of 19 video samples	High number to simplify the training phase of test subjects Taken from public databases
Tweak through compression	Creation of 133 different sequences, at 7 compression rates Compression: aim was to reduce Bit Rate from Ref. to 150 b/s
Preliminary categorization	Score to movement and colour importance in each video Categorization to be validated with experiments
A, B and C playlist creation	Each playlist had no repeated videos in it Videos were in A,B, C with different compression rates
Questionnaire creation	Asking for MOS, categorization parameters, disturbs Followed by collecting table, both in italian
Testers selection	College preparation, used to watching videos on computers Inexperienced in video quality assessment field
Experimental activity	Simple training phase for testers Answered orally to questions
Raw output elaboration	Tables with duration of fixations Number of fixations

IV. Data Interpretation

A. Statistical Analysis

The goal of this phase is to analyze data collected by the eye tracker and subsequently processed (data outcoming the step "raw output elaboration" of Table II). The statistical analysis regards the following variables:

MOS – Fig. 4 shows the MOS obtained by averaging the MOS over all the observers, for the entire set of videos. The fact that the blue line representing the MOS for reference (best quality) videos is almost always the highest of the three, while the green line (150 bit-rate compression, worst possible quality expected) is almost always the lowest of the three (with the red line between the blue and the green ones), validates the initial assumption and shows how the chosen compression rates usually fulfilled their aim of creating three different perceivable levels of impairment.

Subjective Color Score and Subjective Movement Score (SCS and SMS) – The video sequences have been to partitioned in two groups for each feature (movement and color). The videos in which color has been perceived as *relevant* have SCS > 2 and the videos in which movement is relevant have SMS > 2. The values have been obtained by the tester's answers. The videos which resulted with high color relevance are number 3, 5, 7, 10, 15, 19 while the ones with high movement (or temporal activity) relevance are number 2, 4, 7, 8, 10, 12, 14, 17, 19, according to the playlist index of TABLE I.

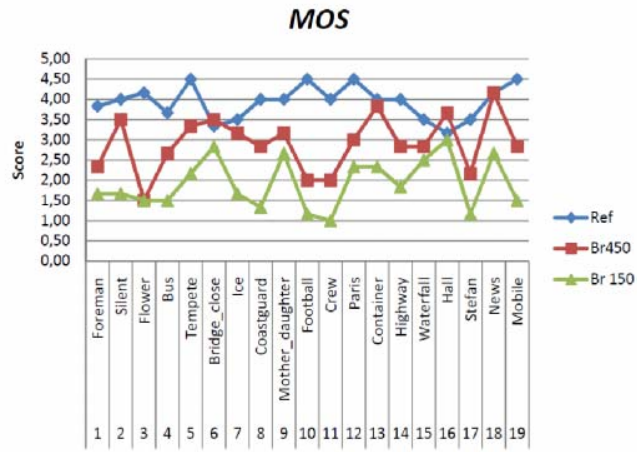

Fig. 4 Mean Opinion Score, for the whole set of videos. Blue: reference (uncompressed) video, Red: compressed video at bit rate 450 bit/s, Green: compressed video at bit rate 150 bit/s

MFT – By referring to eq. (2) the MFT has been computed for each video and for each quality level, by averaging the values over the observer groups in the playlists. All the values are expressed in ms. Fig. 5 shows the obtained values, plotted over the entire set of video sequences.

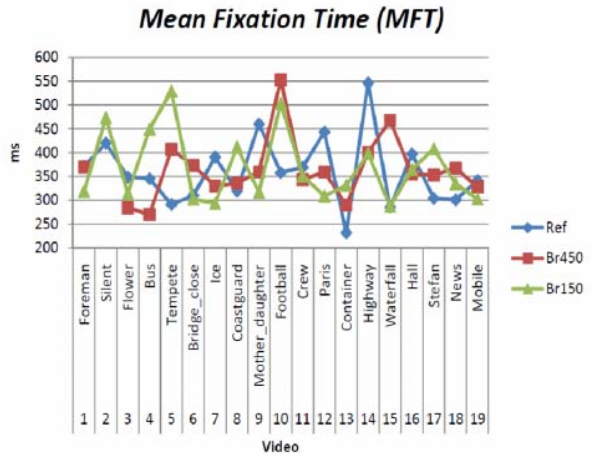

Fig. 5 Mean fixation time (MFT, see eq. 2) for the whole set of videos, for the three classes of quality

Standard deviation computed on MFT parameter (SDoFT) - In this work, a first interpretation of the data suggests that the standard deviation of fixation time could be the best parameter to evaluate the quality degradation introduced by compression. The idea behind this choice is to determine if, when watching an high quality video, the expected duration of fixation is more or less variable if compared to the same measures referring to low quality videos. Unfortunately, this first hypothesis is not confirmed by experimental results. In fact, Fig. 5 does not show any direct correlation between the MOS and the parameter here under analysis. In fact, by comparing Fig. 4 to Fig.5, it is clear that, while for MOS values there is a clear behavior related to quality, in Fig. 5 values are intermingled without any relation.

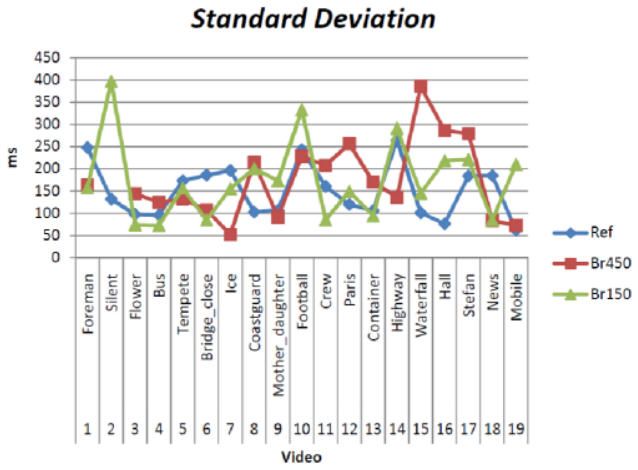

Fig. 6 Standard Deviation of MFT (SDoFT), for the whole set of videos. Blue: reference (uncompressed) video, Red: compressed video at bit rate 450 bit/s, Green: compressed video at bit rate 150 bit/s

The same behavior has been observed also by splitting the general curves of Fig. 5 for each class of quality degradation and for each categorization according to movements and color predominance (the results do not add any further consideration, so they are not reported here, the whole set of graphs are in [3]). The standard deviation does not look directly to be connected to the perceived quality level, even if it looks like there can be some connection between the perceived quality and the *expected value* of the standard deviation. This observation suggest to perform another analysis, with third level statistics, which revealed itself to be the best indicator of quality degradation. The standard deviation of MFT for reference quality video looks much more regular than the one calculated starting from data related to lower video quality. This idea has been confirmed by experimental results because, as it can be seen in Table III, the Standard Deviation's expected value for highest quality video is considerably lower than the other two calculated values, which are related to the lower quality stimuli.

TABLE III
EXPECTED VALUE OF STANDARD DEVIATION OF MEAN FIXATION TIME FOR THE THREE CLASSES OF QUALITY

	Ref	Br450	Br150
Standard Deviation Expected Value	149,3888	174,2288	173,8378

TABLE IV
STANDARD DEVIATION OF SDOFT PARAMETER FOR THE THREE CLASSES OF QUALITY

	Ref	Br450	Br150
Standard Deviation of SDoFT	60,6719	88,2987	90,7023

If we consider the statistical moments of the third order, this behavior is more accentuated, as shown in Table IV, and this regularity is confirmed even for high relevant movement video sequences (see table V), while is less evident considering color component (see Table VI).

TABLE V
STANDARD DEVIATION OF SDOFT PARAMETER FOR THE THREE CLASSES OF QUALITY, FOR VIDEO OF HIGH MOVEMENT ACTIVITY (SMS >2)

Third Level Parameters, SMS>2	Ref	Br450	Br150
Standard Deviation Expected Value	155,4752	170,6233	225,5708
Standard Deviation of SDoFT	69,7361	85,7656	100,3062

TABLE VI
STANDARD DEVIATION OF SDOFT PARAMETER FOR THE THREE CLASSES OF QUALITY, FOR VIDEO WITH RELEVANT COLOR (SCS >2)

Third Level Parameters, SCS>2	Ref	Br450	Br150
Standard Deviation Expected Value	145,4335	169,0899	178,9579
Standard Deviation of SDoFT	69,4837	122,8213	87,1634

This suggests the most relevant conclusion of the statistical analysis; the third order statistics of fixation time is a good indicator of quality degradation for a set of video sequences where no assumption about semantic content characterization is made, in terms of movement or color. However, if we consider this kind of characterization, the third order statistics is more suitable to movement relevance, rather than color richness.

V. APPLICATION OF THE METHOD

As pointed out in section I.A, this method is suitable to a "no-reference" quality assessment task, even if in the set of video sequences reference-high quality video can be included (as in our experiments). However, this a-priori knowledge is not relevant. In fact, the method requires only the categorization of a generic set of video sequences in classes (for example, video compressed at different bit rates, video displayed at a particular resolution, and so on). Of course, as our interest is in studying quality assessment, the categorization has to respect some hypothesis of quality degradation. For each class, a subjective test is performed according to the methodology summarized in Table II, and a third order statistics is computed on the values of fixation times provided by the eye tracker. Then, the classes are ranked according to the standard deviation of SDoFT values, and the rank corresponds to the progressive loss of quality. In the present work, the method has been applied to classes of differently compressed video, thus confirming the progressive loss of perceived quality. However, in other cases, the rank is not so obvious. For example, which quality is better perceived between a Full HD video on a 42" TV or a reduced resolution video on a 32" TV set? The answer is not so clear in advance. In this case, the categorization of the video set is in two classes, one of 1920 X 1080 videos watched on a 42" display, the second a reduced version 1280 X 720 on 32" display. The statistical analysis following the experiments will give a meaningful rank related to video quality, as perceived by human observers. In fact, experiments here reported confirm that the third order statistical analysis of the parameter here proposed (fixation time) reflects the behavior of the MOS of the human observer, thus it is a valid method to rank set of video sequences in relative order of perceived quality.

VI. CONCLUSIONS

The main goal of this work is to find a new use for the eye tracker in the research field of subjective quality assessment of video sequences. The duration of the fixation time, rather than their position on the screen, is the parameter here proposed to evaluate the observer reaction to stimuli of different quality levels. Experimental results reported in Sections III and IV show that there is a correlation between some of the analyzed parameters and the MOS obtained for each video, hinting that the apparatus may be used best to study the video quality using its ability to study a "time only" dependent parameter, and not a "time and space" dependent one.

Its usefulness, moreover, is probably enhanced by the choices made while setting up the work. The elimination of the memory effect that was not taken into account in previous works in literature, may be the most important trait that identifies the experimental activity.

The greatest importance of the Eye Tracking activity while performing a subjective video quality assessment experiment is probably in studying the duration, as said. The gaze maps which are another kind of output of the apparatus, even if more understandable and intuitive, are probably too complex both to be statistically compared among them and/or to be studied by themselves to get useful results, while the only measure of time of durations, even if it does not include the information about space, could lead to the useful conclusions presented in sections IV and V.

In order to take into account of a semantic characterization of the presented sequences, then, the proposed method does not study the position of the gaze ("ok, testers looked a lot here so this must be something important") but to ask them an opinion about color and movement importance in them. This methodology gives a characterization of the sequences with interesting results.

The conclusion is that data referring to high quality video fixations are expected to be more predictable than the low quality video ones. For high quality videos it is also possible to predict the fixation duration with a greater precision, given the fact that data showed an inferior Standard Deviation. This HVS behavior is more accentuated when considering the observation of a video in which movement assumes a leading role in catching the tester's attention. The same conclusion cannot be given to the color-dominant videos, even if the expected behavior of fixation time is also useful to predict quality degradation.Future works will start from these conclusions, in order to test the methodology also for other types of quality degradation, not only due to compression. Besides, new research are in progress in order to study how to combine semantic interpretation of the scene with the statistical approach here proposed.

REFERENCES

[1] F. Lukas Z. Budrikis, Picture Quality Prediction Based on a Visual Model, *IEEE Transaction on Communications*, vol. 30, issue 7, 1982, pp. 1679-1691.
[2] N. Nill, A Visual Model Weighted Cosine Transform for Image Compression and Quality Assessment, *IEEE Transaction on Communications*, vol 33 Issue 6, 1985, pp. 551-557.
[3] R. Amadeo. Compressed Video Quality Assessment: from Subjective to Objective Metrics, Master thesis, University of Pavia, Engineering Faculty, 2011.
[4] U. Reiter, J. Korhonen, Comparing apples and oranges: subjective quality assessment of streamed video with different types of distortion, *Proc. Of International Workshop on Quality of Multimedia Experience*, 2009, pp. 127-132.
[5] Y. Ou, Y. Zhou, Y.Wang, Perceptual quality of video with frame rate variation: a subjective study, 2010, *Proc. Of Acoustics Speech and Signal Processing (ICASSP)*, pp- 2446-2449.
[6] A. K. Moorthy, K. Seshadrinathan, R. Soundararajan and A. C. Bovik, "Wireless video quality assessment: A study of subjective scores and objective algorithms". *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 20, no.4, April 2010, pp. 513-516.
[7] O. Le Meur, A. Ninassi, P. Le Callet, D. Barba, Overt visual attention for free-viewing and quality assessment tasks: Impact of the regions of interest on a video quality metric, Signal Processing Image Communication, 2010, vo. 25, pp- 547-548.
[8] H. R. Wu, K. R- Rao: Digital Video Image Quality and Perceptual Coding, Taylor and Francis ed. 2006, pp. 125-151.
[9] http://trace.eas.asu.edu/yuv/ (Video trace library of Arizona State University)
[10] ftp://ftp.tnt.uni-hannover.de/pub/svc/testsequences/ (Hannover Liebnitz University video library)
[11] http://media.xiph.org/video/derf/
[12] User Manual - Tobii Eye Tracker, Clearview analysis software - February 2006
[13] T. Tominaga, T. Hayashi, J. Okamoto, A. Takahashi, Performance comparisons of subjective quality assessment methods for mobile video, *Proc. Of Quality of Multimedia Experience (QoMEX)*, 2010pp. 82-87.

M. G. Albanesi was born in Pavia in 1962 and was graduated cum laude in Electronic Engineering (University of Pavia, 1986) with a Master Thesis on image compression and visual perception. Later she obtained the Ph.D. in Electronic Engineering and Computer Science (Pavia, 1992) with a thesis on VLSI architecture for image compression.She worked at ST Microelectronics on silicon compiler to design dedicated devices for image processing. After this work experience, she joined the Computer Department of Faculty of Engineering of University of Pavia, first as senior researcher (since 1993), then as Associate Professor (since 1998). Her main actual field of interest are quality evaluation of visual media and user-experience driven application for media description and retrieval.Prof. Albanesi is associate Editor of Pattern Recognition and gives lectures in master courses for Ph.D students and corporate specialists for technological innovation in the field of visual multimedia data and applications.

R. Amadeo was born in Pavia in 1985 and was graduated at the second level degree (University of Pavia, 2011) with a thesis on subjective and objective video quality assessment.
He is currently applying for a position of Ph. D. at the University of Pavia, computer science department, to continue his previous research on quality assessment of visual multimedia information.