



# Spatio-temporal error concealment technique for high order multiple description coding schemes including subjective assessment

Ahmed Aldahdooh, Marcus Barkowsky, Patrick Le Callet

## ► To cite this version:

Ahmed Aldahdooh, Marcus Barkowsky, Patrick Le Callet. Spatio-temporal error concealment technique for high order multiple description coding schemes including subjective assessment. 2016 Eighth International Conference on Quality of Multimedia Experience (QoMEX), Jun 2016, Lisbon, Portugal. 10.1109/QoMEX.2016.7498929 . hal-01346406

**HAL Id: hal-01346406**

**<https://hal.science/hal-01346406>**

Submitted on 18 Jul 2016

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Spatio-temporal Error Concealment Technique for High Order Multiple Description Coding Schemes Including Subjective Assessment

Ahmed Aldahdooh, Marcus Barkowsky, and Patrick Le Callet  
LUNAM Universite de Nantes, IRCCyN UMR CNRS 6597, Nantes, France  
Email: firstname.lastname@univ-nantes.fr

**Abstract**—Error resilience (ER) is an important tool in video coding to maximize the quality of Experience (QoE). The prediction process in video coding became complex which yields an unsatisfying video quality when NALunit packets are lost in error-prone channels. There are different ER techniques and multiple description coding (MDC) is one of the promising technique for this problem. MDC is categorized into different types and, in this paper, we focus on temporal MDC techniques. In this paper, a new temporal MDC scheme is proposed. In the encoding process, the encoded descriptions contain primary frames and secondary frames (redundant representations). The secondary frames represent the MVs that are predicted from previous primary frames such that the residual signal is set to zero and is not part of the rate distortion optimization. In the decoding process of the lost frames, a weighted average error concealment (EC) strategy is proposed to conceal these frames. The proposed scheme is subjectively evaluated along with other schemes and the results show that the proposed scheme is significantly different from most of other temporal MDC schemes.

**Keywords**—Error Resilience (ER), Source Coding (SC), Error Concealment (EC), Multiple description coding (MDC), Subjective experiment, HEVC, UHD.

## I. INTRODUCTION

Video applications became popular and the videos might be sent via error-prone channels. The decoded video quality might not be satisfying if one or more packets are lost. The main goal of video coding like high efficiency video coding (HEVC) [1] is to minimize the coding distortion for a target bitrate. This requires a complex prediction process to remove the redundant information in the video signal [2]. As a result, the error resilience in HEVC is decreased compared to H.264/AVC due to the increase of temporal dependency [3]. Several error resilience techniques are introduced in the literature [4], [5], [6]. Layer Coding (LC) and Multiple Description Coding (MDC) are both efficient in terms of error resilience. In LC, if the base layer is lost or corrupted and despite the presence of enhancement layers, the output video sequence will be degraded seriously. To mitigate this problem, different solutions might be applied here. One of them is to protect the base layer using forward error correction (FEC). This is useful in packet corruption with specific number of errors. Another solution is to retransmit the lost packet when feedback channels are available. The best solution is to use a hybrid scheme. Nevertheless, LC may not be convenient to real time applications so MDC is a promising solution to deal with these drawbacks of the LC. In MDC, the video sequence is encoded into two or more different bit streams called descriptions. One of the most important design principles of MDC is that each description has to deliver videos with acceptable quality even

if it is the only description received by the decoder and the highest quality will be achieved if all descriptions are received. A comprehensive review of multiple description coding can be found in [7], [8], [9]. It was shown that the multiple description coding is an effective and promising technique for error resilience for several reasons. First, it is suitable for real time applications since feedback is not required which simplifies the network design. Second, it performs better than other error resilience approaches in high error loss rates [10], [11]. In this paper, a temporal domain multiple description coding is studied.

Each MDC scheme defines two processes, the first is how to generate the descriptions at the encoder side and how to combine them at the decoder side. The second process is how to do the error concealment when one packet, or more, of a description is lost. Some schemes introduce side information to provide additional or redundant information to help the decoder conceal the lost frame. A review of different temporal MDC schemes is discussed in Section II. These schemes are not efficient for the following reasons. First, in schemes that do not include any side information, the error propagation will be annoying especially if the intra-period is large and if the sequence has high motion intensity. Second, in schemes that do include side information, the coding efficiency will be decreased and the error propagation will be noticeable. Third, these schemes are less efficient in  $n$ -MDC (*when  $n > 2$* ) since the side information is not fully utilized. For instance, in the case of 4-MDC with side information, each frame has one primary data and three redundant data and if the primary data is lost, one of redundant data will be utilized and the remaining two will not. In this paper, a new scheme is proposed in which the redundant data is represented in a different context and a new weighted average algorithm for error concealment is also introduced in which all the redundant data is utilized if the primary data is lost. The proposed scheme, as discussed in Section IV, is characterized by lightweight complexity, standard compatibility, redundant data tuning, and suitability for  $n$ -MDC ( $n \geq 2$ ). The proposed scheme along side with other schemes are tested in a subjective experiment.

The rest of this paper is organized as follow: a brief overview of related work is introduced in Section II. In Section IV, the proposed MDC scheme will be illustrated. Experimental setup and the results are presented in Section V. In Section VI, we sum up with the conclusion.

## II. RELATED WORK

The temporal MDC schemes with their error concealment techniques are categorized into three classes; the first class is referring to the schemes that do not have any side information, the second class is referring to the schemes that introduce some

additional data for each frame, while the third class is referring to the schemes that include a redundant frames for each primary frame. Table I shows the list of some MDC schemes and the corresponding hypothetical reference circuits (HRCs) as used later in this work. Apostolopoulos in [12] reviewed the first class of the schemes. All schemes in this class share the same encoding and decoding processes and differ in error concealment strategy. Suppose that an even frame is lost. Copying the previous even frame from the distorted description to replace the lost even frame in the buffer (HRC00, HRC01, HRC09), copying the previous odd frame from the undistorted description (HRC02, HRC10), averaging the previous and the next odd frames from the undistorted description (HRC03, HRC11), scaling the MVs of the next odd frame from the odd description by  $\frac{1}{2}$  and use them to do the motion compensation process using the previous odd frame of the undistorted description, namely inplaceMC (HRC04, HRC12), and generating the MVs using the available previous and next odd frames, namely MCinterp (HRC05, HRC13), are the error concealment strategies that are reviewed in [12]. In the second class of schemes, a side information is introduced. This side information can be a duplicate of MVs of each frame in the description or a duplicate of I-frames (HRC06, HRC14). In [13], a different scheme is proposed in which each description contains alternatively even/odd frames and odd frames in even description are containing the motion information only predicted from the previous even frame (HRC08, HRC16). While in the third class of the temporal MDC schemes, a complete frame is used as side information. Radulovic et al. [14], suggested that each description alternatively contains a fine quantization frame (even) followed by coarse quantization frame (odd) (HRC07, HRC15).

### III. PROBLEM STATEMENT

The 2-MDC scheme that is encoded in low-delay configuration (IPPPP) is used in order to provide a good illustration of different schemes. The descriptions in the 2-MDC are generated as follows; the sequence frame rate is down sampled by two to generate even/odd descriptions and each has its own encoding loop. In HRC01/09, the distorted description will continue to decode normally, therefore, the effect of error propagation due to the correlation reduction in one description will be highly noticeable. On the other hand, in HRC02/10, the effect of error propagation will be reduced relative to the HRC01/09 respectively but still is not efficient in sequences that have large motion intensity. HRC03/11 yield a blurred concealed frame which is also not an appropriate technique to use when there are spatial and temporal variations in the sequence. While HRC04/12 work well under the assumption that the motion is completely smooth which is not the case in most of the video sequences. HRC05/13 use the technique mentioned in [15] which employs the phase correlation motion estimation technique to conceal the lost frame. Though it adds extra complexity to calculate the MVs, it still suffers from blurriness and post-processing for the concealed frame is required. In HRC06/08/14/16, two important information are not included that have a vital impact in the concealed frame, the residual signal and the intra-block modes in inter-frames. The first class of MDC schemes perform well in term of coding efficiency since no side information is used but it does not provide a satisfactory video quality especially if there are errors in both descriptions and if the video has high motion intensity. In the second class, a trade-off between

TABLE I: List of hypothetical reference circuit (HRC). Check mark (✓) means that the HRC is subjectively evaluated while times mark (X) not. The dash mark (-) means that the HRC is not applicable.

EC technique	SD	2-MDC	4-MDC
Copy previous frame from the same description	✓(HRC00)	✓(HRC01)	✓(HRC09)
Copy previous frame from the another description	-	✓(HRC02)	X(HRC10)
The average between the previous and next frames from the another description	-	✓(HRC03)	X(HRC11)
Scale the MV of next frame from another description and use them to conceal the frame	-	X(HRC04)	✓(HRC12)
Average the two concealed frames the another description using predicted MVs using Phase correlation algorithm	-	X(HRC05)	✓(HRC13)
Use the duplicate I-frames and MVs	-	✓(HRC06)	X(HRC14)
Use the duplicate degraded-frames	-	✓(HRC07)	X(HRC14)
Use the MVs of redundant frames	-	X(HRC08)	✓(HRC16)
Weighted average of the concealed frames using the proposed strategy	-	-	✓(HRC17)

quality and coding efficiency is achieved by including the MVs and excluding the residual signal and the intra-block modes. While in the third part the trade-off is achieved by using the coarse frames. Unfortunately, the second and the third parts are not convenient in more than two-description schemes since not all redundant data is utilized. Therefore, these designs principles are taking into consideration in the proposed scheme. Like other schemes except HRC05/13, the complexity is lightweight since the weighted average is applied and the weights are stored in the decoding side. Standard compatibility and redundant data tuning are also preserved. Finally, the scaling to higher number of descriptions is also considered.

### IV. THE PROPOSED MDC SCHEME

In this section, the proposed MDC scheme is explained. Firstly, the encoding process and the corresponding decoding process are presented, then the error concealment algorithm is elaborated. The 4-MDC is used as an example to elaborate the two processes. In the 4-MDC, the video sequence frame rate is downsampled by 4 and each description contains one fourth of the original sequence.

#### A. Encoding and decoding processes

Figure 1 depicts the encoding process. Each description contains primary frames which represent the frames of the usual 4-MDC and secondary frames which represent frames of other descriptions and located between two primary frames. The primary frames are encoded using low-delay configuration and the secondary frames are predicted from the previous primary frame in the same description. As a result, each frame of the original sequence is represented with a primary frame and three secondary frames that can be sent in-stream or out as side information. The following rules are applied in the

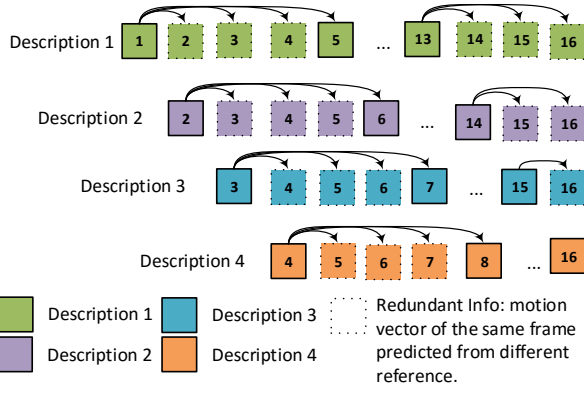


Fig. 1: 4-MDC with redundant data/side information. The solid-border square represents the primary frame. The dot-border square represents the redundant representations to be sent in the stream or as a side information. The arrows represent the prediction process; the redundant frames between two primary frames are predicted from the previous primary frame. Only motion vectors are transmitted. The primary frames represents the low-delay configuration.

encoding of secondary frames. First, intra blocks in inter-frames are not allowed which enforces reconstructing the lost frame using only MVs. Second, the residual is set to zero during the encoding process and rate-distortion optimization is used to decide the best splitting in terms of distortion. Such way requires more signaling in the stream which may increase the bitrate. Many solutions can be applied to tune the amount of redundant data (MVs) either by reducing the sub-pixel accuracy to  $\frac{1}{5}$ -pixel or integer-pixel accuracy or by using one of the algorithms that prioritize the MVs [16]. At the decoding side, four side-decoders are used to decode the descriptions and in case of no error the central decoder assembles the primary frames from the side-decoders and send them to the display buffer.

### B. Error concealment process

When one primary frame is lost the central decoder initializes the error concealment process. Figure 2 depicts the error concealment steps. In the first step, each secondary frame is decoded normally in the side decoders. Then, the lost primary frame is replaced with a weighted average of the three available secondary frames. The weights are applied on the pixel level and they are a function of the temporal distance ( $d$ ) and the number of pixels ( $n$ ) in the partition unit (PU) that the pixel belongs to. Number of pixels in the PU ranges from 4096 ( $64 \times 64$ ) down to 16 ( $4 \times 4$ ). That yields 13 different amounts of pixels in the PU. Since there are 3 redundant/secondary frames, 2197 ( $13^3$ ) combinations are counted. In addition, because secondary frames are predicted from previous, frames 3 distances are counted. In total, 6591 ( $3 \times 2197$ ) combinations of  $(d_i, n_i)$  are counted.

Using temporal distance as one factor on the weight function has already been used in the literature in other contexts [14] and it is believed that the closest frame is not always the best match for the current frame, therefore other factors may have significant influence. In HEVC, the coding unit (CU) can be split using one of the eight supported PU modes. For more details and applied constrains, please refer to [2]. In the proposed EC, for each CU, at most 3 different splitting trees

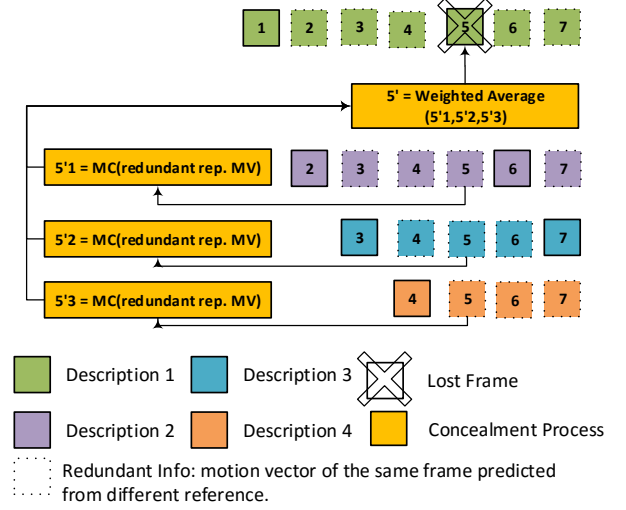


Fig. 2: Proposed error concealment procedures

are available that are optimized in term of distortion and the authors believe that it may have an impact reducing the overall distortion and error propagation in the concealment process. In order to train the weights, data samples are collected from video sequences. Each sample (each pixel in a primary frame) has tenth values  $(d_1, d_2, d_3, n_1, n_2, n_3, p_1, p_2, p_3, p_0)$ , where  $d_i$  represents the temporal distance of current pixel,  $n_i$  represents the number of pixels in the PU that the current pixel belongs to, and  $p_i$  represents the pixel value of secondary frames and primary frame respectively. Each combination  $(d_i, n_i)$  is considered as a unique condition and enumerated with the parameter  $k = [1, \dots, 6591]$ . Then the samples that share the same properties, i.e. the values of  $d_i$  and  $n_i$ , are grouped and then are split into train and validation sets to train the weights. The training can be expressed as:

$$W^k = \arg \min_{\{w_1^k, w_2^k, w_3^k\}} ((p_0 - \sum_{i=1}^3 w_i^k p_i)^2), \quad (1)$$

where,  $w_1^k + w_2^k + w_3^k = 1$

Where  $W^k = \{w_1^k, w_2^k, w_3^k\}$  are the weights that minimize the error in the validation samples. After that, the weights are stored in the decoder side.

## V. THE SUBJECTIVE EXPERIMENT

### A. Experimental setup

1) *Source video contents*: The source videos are selected from different content providers: 4 from Shanghai Jiao Tong University (SJTU) [17], 2 from Ultra Video Group [18], 2 from Sveriges Television AB (SVT) [19], 2 from Blender Foundation [20], and 2 from MediAventures [21]. The 12 source sequences are in ultra high definition (UHD) with a resolution of  $3840 \times 2160$  pixels. Figure 3 shows the thumbnails of the video sources. The frame rate of the video sequences varies from 25 frame per second (fps) to 120 fps. Each sequence is 10 seconds long. Video sequences cover different

video properties: motion intensity, camera motion type, spatial complexity, and colors.



Fig. 3: 12 UHD video sequences that are used in the subjective experiment

2) *Hypothetical reference circuit (HRC)*: In this experiment, the video sequences are encoded as single description (SD), 2 descriptions MDC, and 4 descriptions MDC. All are encoded with QP=32, intra period of 32, and motion search range of 64, 128, and 256 are used for single description, 2-MDC, and 4-MDC respectively. For HRC00/01/02/03/04/05/06/09/10/11/12/13/14, the restricted low-delay configuration is used (GOPSize=1), i.e. only the previous frame is used for prediction. For HRC07/08, the (IPP) GOP structure is used, while in HRC15/16/17, the (IPPPP) GOP structure is used, Figure 1. The same error pattern is inserted to all generated videos. The 34<sup>th</sup>, 50<sup>th</sup>, 162<sup>nd</sup>, and 178<sup>th</sup> NALUnits are dropped and concealed with different error concealment techniques as shown in Table I. In this experiment, each NALUnit represents one frame. For encoding, HM12.1 is used while for the decoding processes, the robust decoder [22] is used and has been adapted to call the appropriate error concealment strategy. A total of 12x18=216 processed video sequences (PVS) are generated. Since this number is large for a subjective experiment, not all of them are used. Each error concealment technique is applied to either 2-MDC or 4-MDC or both as shown in Table I. To sum up 12x11=132 PVS are used in the subjective experiment. Indeed, the redundancy overhead is increasing when MDC is used. In the above mentioned HRCs, the redundancy overhead is varied from a factor of 1.1 to 1.7 in term of bitrate in 2-MDC and from a factor of 1.5 to 3.5 in 4-MDC relative to SDC. One possible observation that might be obtained from the experiment is that when comparing two HRCs that are varied in the redundancy overhead, the HRC with lower overhead may have a better quality than the HRC of higher overhead.

3) *Testing conditions*: Since all processed videos were affected by error insertions, the pair comparison (PC) method from ITU-T Rec. P.910[23] was selected to obtain the subjective scale of the experiment. Not all HRCs are involved in the experiment to reduce the number of pairs, as mentioned in the previous subsection. The optimized square design (OSD) methodology was selected to reduce the number of pairs [24] in which the ranking of the stimuli in the test is known based on pre-test results or prior knowledge. In this experiment, the 3x4 rectangular matrix was selected for 11 HRCs and the ranking of the stimuli is defined by the authors' prior knowledge as shown in the R matrix: Where the matrix on the left represents the rank of the stimuli and the matrix on the right represents the corresponding HRC. The 12<sup>th</sup> cell of the matrix is filled

$$R = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 10 & 11 & 12 & 5 \\ 9 & 8 & 7 & 6 \end{bmatrix} \Rightarrow \begin{bmatrix} \text{HRC09} & \text{HRC12} & \text{HRC03} & \text{HRC01} \\ \text{HRC07} & \text{HRC17} & \text{HRC17} & \text{HRC13} \\ \text{HRC16} & \text{HRC06} & \text{HRC02} & \text{HRC00} \end{bmatrix}$$

with a repetition of the proposed error concealment strategy (HRC17). Due to OSD, the number of pairs is reduced from 11\*10/2=55 to 27 pairs for each content, thus 27x12=324 in total. Unfortunately, this number of pairs is still large. In order to reduce this number, the pairs that have very close quality, ((17,7),(16,6), and (2,0)) are viewed for each observer, and the other pairs are randomly and equally distributed between the observers. Note that the pair (HRC17,HRC17) is not considered in the experiment.

4) *Subjective assessment*: For each pair, the two stimuli are viewed one after another. The replay function was supported. The observer is asked for his preference for each pair in a forced choice manner. A playlist for each observer is prepared taking into consideration that the pairs that belong to the same content are not viewed consecutively, orders of the pairs are random, and the temporal order of the pairs is also switched between the observers. The viewing distance was 1.5 times the height of the screen. The experiment was explained to the observers using a training session prior to the test session. 4 pairs are selected from the PVSs for the training session without any explicit or implicit instruction on how to choose the preference. The test duration is about 75 minutes including training and breaks. All sequences are viewed at 60 fps therefore the video of 25 or 30 frame rate are up sampled by 2, i.e. the frames are duplicated. The PVS are displayed using 3840x2140 native screen resolution with 60 fps. The screen brand is Grundig FINEARTS 55 FLX 9490 SL with a 55-inch diagonal. The ITU Recommendations BT.709-5 [25] and BT.500-13 [26] are followed to adjust the screen color and brightness and to set up the testing room respectively. 46 non-expert observers participated in the experiment, 22 males and 24 females and the age average is 24 (18 to 38). The pairs ((17,7),(16,6), and (2,0)) are evaluated by the 46 observers and other pairs are evaluated with 11 or 12 observers. A vision check is performed before the experiment using far and color vision tests. Any observers with normal or corrected to normal visual acuity are allowed to do the experiment.

## B. Experimental results and discussion

In this subsection, subjective data is analyzed in terms of pair comparison raw data using Barnard's exact test [27] and in terms of pair comparison model using Bradley-Terry model [28].

1) *Barnard exact test*: Fisher's exact test and Barnard's exact test are both statistical exact/significance test of contingency tables. For 2x2 contingency tables, Barnard's exact test is claimed to be more powerful than Fisher's exact test. Its powerfulness came from its unconditional rule to calculate the p-value. Suppose that  $M = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ , is the Barnard's contingency table where, for each pair (A,B), a and d are equal to the number of observers that prefer A rather than B, and b and c are equal to the number of observers that prefer B rather than A. The M matrix is the input of the Barnard's test and the output is the p-value that is calculated on the 95% confidence interval.

Table II shows the results of applying the Barnard's test for

TABLE II: Barnard's exact test between two pairs per content

SRC/HRCs		1	2	3	4	5	6	7	8	9	10	11	12	#←	#→	#No	Sig.	Total sig.
12	9	←	-	-	→	←	←	←	←	-	-	←	←	7	1	4	×	←
3	9	←	-	←	-	←	-	-	←	←	←	-	←	7	0	5	-	←
3	12	-	→	←	←	→	-	→	←	-	←	-	-	4	3	5	×	←
1	9	←	→	←	-	-	←	-	←	-	-	←	-	5	1	6	×	←
1	12	-	→	←	-	-	→	→	→	-	→	→	→	1	6	5	×	→
1	3	→	→	→	→	→	→	-	→	→	→	-	→	0	10	2	→	→
13	1	←	←	→	←	←	-	←	-	←	-	←	-	7	1	4	×	←
0	1	←	←	-	←	←	-	←	←	←	←	←	-	9	0	3	←	←
0	13	-	-	←	-	←	←	-	←	-	←	←	-	6	0	6	-	←
2	3	→	-	-	→	-	-	-	→	-	-	-	-	0	3	9	-	→
2	0	→	→	→	→	→	→	→	→	→	→	→	→	0	12	0	→	→
6	12	←	←	-	-	-	-	-	←	←	-	-	←	5	0	7	-	←
6	0	-	-	→	→	-	→	-	-	-	→	-	-	0	4	8	-	→
6	2	←	←	→	→	-	-	-	←	←	-	-	←	5	2	5	×	←
16	9	←	←	-	←	←	←	←	←	←	←	←	←	11	0	1	←	←
16	0	←	←	-	→	←	-	←	-	←	-	←	-	6	1	5	×	←
16	2	←	←	→	-	←	←	←	-	←	-	←	←	8	1	3	×	←
16	6	←	←	←	←	←	←	←	←	←	←	←	←	12	0	0	←	←
7	9	←	←	←	←	←	←	←	-	←	→	←	←	10	1	1	×	←
7	13	←	←	-	-	←	←	-	→	←	→	←	←	7	2	3	×	←
7	16	→	-	←	-	←	-	-	→	-	-	→	-	2	3	7	×	-
17	12	←	←	←	-	←	←	←	←	←	←	←	←	11	0	1	←	←
17	3	←	←	-	←	←	←	←	-	←	←	←	←	10	0	2	←	←
17	13	-	←	←	←	-	←	←	←	←	←	←	←	10	0	2	←	←
17	2	←	←	-	←	-	←	←	←	←	←	-	←	9	0	3	←	←
17	6	←	←	←	-	←	←	←	-	←	←	←	←	10	0	2	←	←
17	7	←	←	←	←	-	-	←	←	←	←	←	-	9	0	3	←	←

each pair per content. It shows two types of analysis. First, the significant difference for each pair per content is calculated. The first and the second columns represent the pairs while columns labeled 1 to 12 represent the content. Second, the significant difference on the pair level is calculated using two methods. The first method counts the number of sequences that have significant difference for two HRCs and is summed up either in the “#←” or in the “#→”. It also counts the number of sequences that do not have significant difference for two HRCs which is recorded in the “#No” column. The significant difference between the new pair ((#← or #→), #No) is calculated and represented in “Sig.” column. The second method is to sum the votes for each pair across the video sequences and to calculate the significance of the difference between any two pairs. The result is represented in “Total Sig.” column. For instance, the first row which represents the pairs that belong to HRC12 and HRC09. HRC12 significantly differs from HRC09 in 7 sequences and HRC09 significantly differs from HRC12 in one sequence while there is no significant difference in 4 sequences. In this pair we cannot apply the Barnard's test on the pair level using the first method because HRC09 is significantly preferred in one sequence, while the second method shows that there is a significance preference for HRC12. The error concealment strategy preference for one video content is different for another video content and this is clear in different pairs. An important question here is raised “What is the impact of involving video properties to select the appropriate EC strategy to better enhance the QoE?”. One of the Barnard's test intuitive assumption is that if HRC17 is significantly different from HRC07 and HRC07 is significantly different from HRC09, we can say that the HRC17 is significantly superior compared to HRC09. Using this property, we can conclude that the proposed algorithms is significantly different from other HRCs except HRC 16 since there is no evidence of preference.

2) *Bradley-Terry model*: Bradley-Terry model [28] is a linear model that analyzes pair comparison preference in order to map their probabilities to scales. Given K stimuli, suppose that the pair  $(A_i, A_j)$  are two stimuli, and  $X_i$ ,  $X_j$  are the number of  $A_i$  beats  $A_j$  and the number of  $A_j$  beats  $A_i$  respectively. The probability that the observers choose  $A_i$  over  $A_j$  is  $P(X_i > X_j)$  and it is defined as:

$$P(X_i > X_j) \equiv \pi_{ij} = \frac{\pi_i}{\pi_i + \pi_j}, i \neq j \quad (2)$$

Where  $\pi_i > 0$  and  $\sum_{i=1}^K \pi_i = 1$ . The value that describes a stimulus ( $A_i$ ) on the scale is calculated as  $V_i = \log(\pi_i)$ . Since the  $\pi_i$  value is less than one, the Bradley-Terry score  $V_i$  is a negative value.

In this experiment, the Bradley-Terry (BT) test is used to obtain the HRCs scale for each content [29]. Figure 4 shows the results. The scale is offset such that HRC00, which represents the single description coding, is set to zero to easily read the figures. The confidence intervals in the subplots belong to the fitting model and does not represents the observer's confidence. In [30], the author shows a method to calculate the significance between the BT scales. Since the BT scale depends on the goodness of the fitting, the significance results between the HRCs is not necessarily coherent with Barnard's test which is an unconditional test.

## VI. CONCLUSION

In this paper, two main contributions are introduced: First, a new temporal MDC scheme which is characterized by standard compatibility, redundancy tuning, lightweight complexity, and suitability for  $n$ -MDC schemes. Second, the subjective experiment that shows the preference of the proposed scheme against other MDC schemes is introduced. The proposed scheme is significantly preferred to the other temporal MDC schemes, but the number of sequences is too



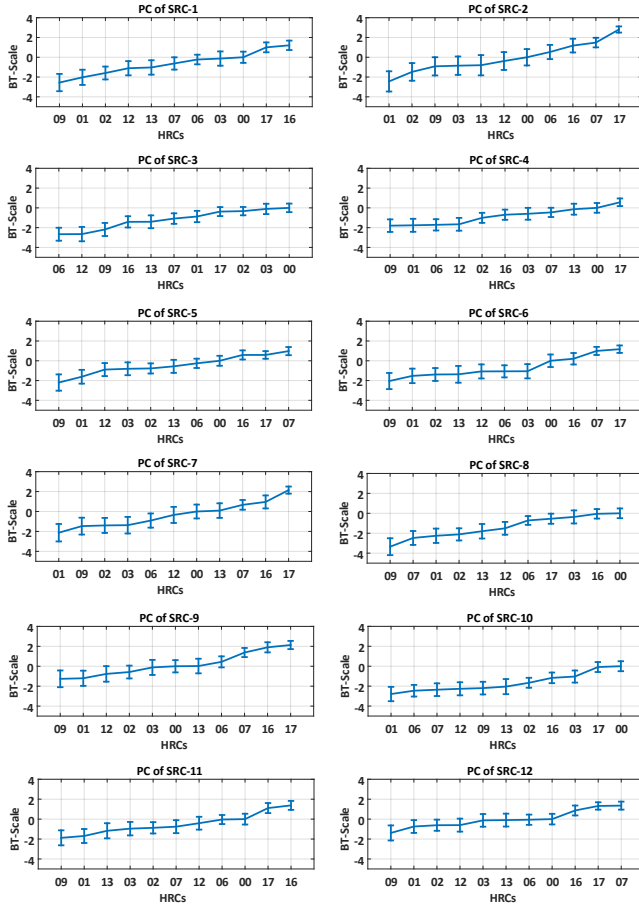


Fig. 4: Bradley-Terry Scale for each video content

small for generalization. In addition, this paper also highlights the fact that the preference of the MDC scheme depends on the video content itself. Hence, more investigations are required to identify these content features.

#### ACKNOWLEDGMENT

This work is supported by the Marie Skłodowska-Curie under the PROVISION (PeRceptually Optimised Video Compression) project bearing Grant Number 608231 and Call Identifier: FP7-PEOPLE-2013-ITN and by UltraHD-4U project.

#### REFERENCES

- [1] ITU, "ITU-R h.265 : High efficiency video coding," 2013.
- [2] G. J. Sullivan, J. Ohm, W.-J. Han, and T. Wiegand, "Overview of the High Efficiency Video Coding (HEVC) standard," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 22, no. 12, pp. 1649–1668, 2012.
- [3] B. Oztas, M. T. Pourazad, P. Nasiopoulos, and V. C. Leung, "A study on the HEVC performance over lossy networks," in *Electronics, Circuits and Systems (ICECS), 2012 19th IEEE International Conference on*. IEEE, 2012, pp. 785–788.
- [4] Y. Wang, J. Ostermann, and Y.-Q. Zhang, *Video processing and communications*. Prentice Hall Upper Saddle River, 2002, vol. 5.
- [5] S. Wenger, "H. 264/avc over ip," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 13, no. 7, pp. 645–656, 2003.

- [6] T. Schierl, M. M. Hannuksela, Y.-K. Wang, and S. Wenger, "System layer integration of high efficiency video coding," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 22, no. 12, pp. 1871–1884, 2012.
- [7] V. K. Goyal, "Multiple description coding: Compression meets the network," *Signal Processing Magazine, IEEE*, vol. 18, no. 5, pp. 74–93, 2001.
- [8] Y. Wang, A. R. Reibman, and S. Lin, "Multiple description coding for video delivery," *Proceedings of the IEEE*, vol. 93, no. 1, pp. 57–70, 2005.
- [9] M. Kazemi, S. Shirmohammadi, and K. H. Sadeghi, "A review of multiple description coding techniques for error-resilient video delivery," *Multimedia Systems*, vol. 20, no. 3, pp. 283–309, 2014.
- [10] Y.-C. Lee, J. Kim, Y. Altunbasak, and R. M. Mersereau, "Layered coded vs. multiple description coded video over error-prone networks," *Signal Processing: Image Communication*, vol. 18, no. 5, pp. 337–356, 2003.
- [11] J. Chakareski, S. Han, and B. Girod, "Layered coding vs. multiple descriptions for video streaming over multiple paths," *Multimedia Systems*, vol. 10, no. 4, pp. 275–285, 2005.
- [12] J. G. Apostolopoulos, "Reliable video communication over lossy packet networks using multiple state encoding and path diversity," in *Photonics West 2001-Electronic Imaging*. International Society for Optics and Photonics, 2000, pp. 392–409.
- [13] R. Kibria and J. Kim, "H. 264/avc-based multiple description coding for wireless video transmission," in *International Conference on Communications*, 2008.
- [14] I. Radulovic, P. Frossard, Y.-K. Wang, M. M. Hannuksela, and A. Halapuro, "Multiple description video coding with H. 264/AVC redundant pictures," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 20, no. 1, pp. 144–148, 2010.
- [15] G. Thomas, "Television motion measurement for datv and other applications," *NASA STI/Recon Technical Report N*, vol. 88, p. 13496, 1987.
- [16] J. Carreira, E. Ekmekcioglu, A. Kondo, P. Assuncao, S. Faria, and V. De Silva, "Selective motion vector redundancies for improved error resilience in hevcc," in *Image Processing (ICIP), 2014 IEEE International Conference on*. IEEE, 2014, pp. 2457–2461.
- [17] L. Song, X. Tang, W. Zhang, X. Yang, and P. Xia, "The SJTU 4K video sequence dataset," in *QoMEX*, 2013, pp. 34–35.
- [18] Ultra Video Group, *Ultra Video Group 4K sequences*, accessed September 15, 2015, <http://ultravideo.cs.tut.fi/#testsequences>.
- [19] *SVT sequences hosted in VQEG*, accessed September 15, 2015, [ftp://vqeg.its.bldrdoc.gov/HDTV/SVT\\_MultiFormat/](ftp://vqeg.its.bldrdoc.gov/HDTV/SVT_MultiFormat/).
- [20] *Blender Foundation 4K sequences*, accessed September 15, 2015, <http://bbb3d.renderfarming.net/download.html>.
- [21] *MediAVentures*, 2013, <http://www.mediaventures.be/>.
- [22] M. Barkowsky, E. Masala, G. Van Wallendaal, K. Brunnström, N. Staels, and P. Le Callet, "Objective video quality assessment towards large scale video database enhanced model development," *IEICE Transactions on Communications*, vol. 98, no. 1, pp. 2–11, 2015.
- [23] ITU, "ITU-T recommendation p.910: Subjective video quality assessment methods for multimedia applications," 2008.
- [24] J. Li, M. Barkowsky, and P. Le Callet, "Subjective assessment methodology for preference of experience in 3d tv," in *IVMSP Workshop, 2013 IEEE 11th*. IEEE, 2013, pp. 1–4.
- [25] ITU, "ITU-R recommendation bt.709-5: Parameter values for the HDTV standards for production and international programme exchange," 2002.
- [26] ITU, "ITU-R recommendation bt.500-13: Methodology for the assessment of the quality of television pictures," 2012.
- [27] G. Barnard, "A new test for  $2 \times 2$  tables," *Nature*, vol. 156, p. 177, 1945.
- [28] R. A. Bradley and M. E. Terry, "Rank analysis of incomplete block designs: I. the method of paired comparisons," *Biometrika*, vol. 39, no. 3/4, pp. 324–345, 1952.
- [29] F. Wickelmaier and C. Schmid, "A matlab function to estimate choice model parameters from paired-comparison data," *Behavior Research Methods, Instruments, & Computers*, vol. 36, no. 1, pp. 29–40, 2004.
- [30] J. C. Handley, "Comparative analysis of bradley-terry and thurstone-mosteller paired comparison models for image quality assessment," in *PICS*, vol. 1. Citeseer, 2001, pp. 108–112.