Performance assessment of the adaptive GoP-size extension of the wireless softCast video scheme

Anthony Trioux, François-Xavier Coudoux, Patrick Corlay, M Gharbi

Performance Assessment of the Adaptive GoP-size extension of the Wireless SoftCast Video Scheme

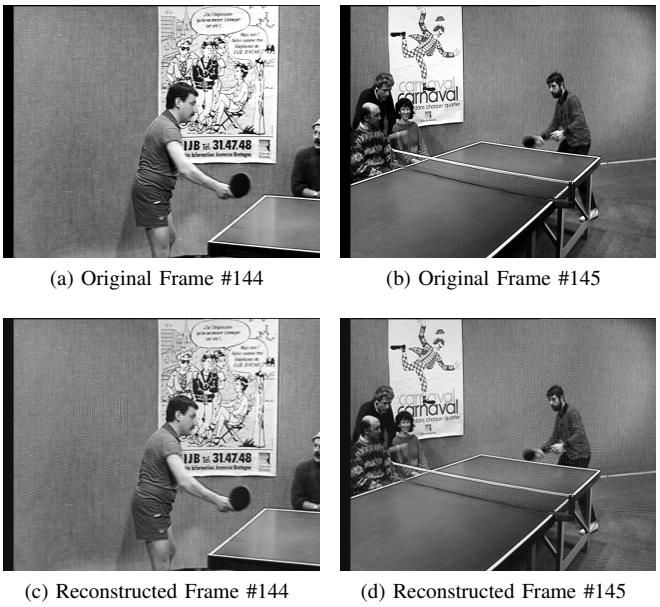
Anthony Trioux, François-Xavier Coudoux, Patrick Corlay, Mohamed Gharbi
UMR 8520 - IEMN, DOAE
Univ. Polytechnique Hauts-de-France, CNRS, Univ. Lille, YNCREA, Centrale Lille
F-59313 Valenciennes, France
{anthony.trioux, francois-xavier.coudoux, patrick.corlay, mohamed.gharbi}@uphf.fr

***Abstract*—The SoftCast video transmission scheme has been proposed as a promising alternative to traditional video broadcasting systems in wireless error-prone environments. However, in its original form, the scheme may introduce annoying artifacts such as temporal quality variations or the so-called ghost effect. To eliminate this artifact as well as improving the received quality, we recently proposed an updated version of SoftCast: Adaptive GoP-size mechanism based on Content and Cut detection for SoftCast (AGCC-SoftCast) [1]. In this paper, we provide additional results and verify the validity of the algorithm considering different video content and resolution. We also analyze the behaviors of the scheme considering different type of cuts: 1) Abrupt cuts i.e., a cut that appear between two different scenes, 2) Soft cuts that may appear in a same scene and 3) Intraframe cuts that represents a shot change inside a limited portion of the frame. The performance of AGCC-SoftCast are compared to the original SoftCast scheme under different Channel Signal-to-Noise Ratios (CSNR) and Compression Ratios (CR). Traditional objective metrics are considered such as: the Peak Signal-to-Noise Ratio (PSNR), the Structural SIMilarity (SSIM) and the Multi-Scale SSIM (MS-SSIM). In addition, the recent Video Multi-method Assessment Fusion (VMAF) metric proposed by Netflix is also used. Regardless of the CSNR and CR values, results clearly highlight the importance of the AGCC algorithm in a SoftCast wireless video transmission scheme. Depending on the transmitted video content, results show that improvements up to 11.6dB in terms of PSNR and 67.45 in terms of VMAF can be obtained, especially at the cut boundaries.**

***Index Terms*—Ghost effect, Cuts, Adaptive GoP-size, SoftCast, Temporal information index**

I. Introduction

Broadcast video content when considering heterogeneity of each user's channel represents a challenge due to the fact that each user is subject to unreliable and different wireless channels that vary over time. Current broadcast systems are based on video codec such as H.264/AVC [2], HEVC [3] or their scalable extensions H.264/SVC, SHVC [4], [5]. However, they do not provide sufficient scalability since they require a permanent adaptation of the source and channel coding parameters by the transmitter. Indeed, they are adjusted to match an available bitrate that is given under predicted or assumed channel state. Due to the heterogeneity of each user's channel, receivers whose channel conditions are degraded are subject to significant visual disturbances (e.g. freeze) while receivers experiencing a better channel than the estimated one cannot take full advantage of it.



(a) Original Frame #144 (b) Original Frame #145

(c) Reconstructed Frame #144 (d) Reconstructed Frame #145

Fig. 1: Illustration of the ghost effect. Simulation parameters: CR = 0.25, GoP-size=32. First row: Original frame #144-145 of the *Tennis_SD* sequence. Second row: Reconstructed frame after SoftCast compression process (no transmission).

In the last few years, the so-called SoftCast scheme [6] has emerged to tackle these problems. Indeed, the received video quality offered by SoftCast scales linearly with the Channel Signal-to-Noise Ratio (CSNR) [7] providing quality of service even in the presence of suddenly degraded channel quality. This property comes from the linear processing applied to the pixels, avoiding quantization or entropy coding, and the transmission carried out without channel coding.

However, in its original form, the scheme may introduce annoying artifacts such as temporal quality variations or the so-called ghost effect illustrated in Fig. 1. The ghost effect is characterized by a superposition of the edges (high frequencies) between the frames before and after a cut inside a video. It is due to the use of a temporal DCT and the compression applied at the transmitter to match the available bandwidth for transmission. To represent the bandwidth limitation, the Compression Ratio [1] defined as: CR=M/N is introduced, where M represent the amount of data that can be transmitted

(a) Frame #169 (b) Frame #170 (c) Frame #655 (d) Frame #656

Fig. 2: Illustration of abrupt cuts inside the *Mixed_HD720p* video sequence from [1].

(a) Frame #144 (b) Frame #145 (c) Frame #161 (d) Frame #162

Fig. 3: Illustration of soft cuts inside the *Tennis_SD* video sequence from [10].

(a) Frame #90 (b) Frame #91 (c) Frame #150 (d) Frame #151

Fig. 4: Illustration of intra-frame cuts inside the *News_CIF* video sequence from [10].

and N the total amount of data. These parameters are further defined in Section II. When CR=1, no compression is applied, the ghost effect does not appear, however temporal quality variations are noticed. As the CR decrease, the ghost effect becomes more and more visible and is spread around the duration of the GoP. To eliminate this artifact as well as improving the received quality, we recently proposed an Adaptive GoP-size mechanism based on Content and Cut detection for SoftCast (AGCC-SoftCast) [1]. This extension takes into account the temporal complexity of the transmitted content, detects cuts that may appear inside a video and adjusts the GoP-size according to a classification process further detailed in Section II.

When considering abrupt cuts as displayed in Fig. 2) (a cut that appears between two different scenes putted together), we showed that our algorithm allows improvements up to 16dB in terms of Peak Signal-to-Noise Ratio (PSNR) and up to 0.55 in terms of Structural SIMilarity (SSIM) at the cut boundaries. In this paper, we further investigate the performance of the scheme. Specifically, we consider:

- Different video content and resolution (HD1080p sequences (class B) and WVGA sequences (class C) from [8]),
- Different types of cuts such as soft cuts, i.e., a cut that may appear in a same scene (Fig. 3) or intraframe cuts, i.e., a shot change inside the same scene (Fig. 4),
- Two additional objective metrics: the Multi Scale-SSIM (MS-SSIM) as well as the Video Multi-method Assessment Fusion (VMAF) metrics due to their high correlation with human judgements in a SoftCast context [9].

The rest of this paper is organized as follows: Section 2 gives an overview of the AGCC-SoftCast scheme. In Section 3, we assess the validity of the classification process proposed in [1] considering different video content and resolution. The algorithm is compared to the classical SoftCast scheme in Section 4 considering different types of cuts. Conclusions are presented in Section 5.

II. AGCC-SoftCast Review

The basic scheme of AGCC-SoftCast [1] is introduced in Fig. 5. AGCC-SoftCast uses Group of Pictures (GoP) of 8 frames at the input of a content analysis and cut detection process. The purpose of this step is to avoid to encode a cut inside a GoP as well as selecting the optimal GoP-size, i.e., the GoP-size that gives the best trade-off between received quality and coding complexity cost. In SoftCast, the complexity cost increases with K the number of frames in a GoP [1] according to $O(Klog(K))$. The analysis is performed using a modified version of the temporal information index proposed by the ITU-T to evaluate the temporal complexity of a video. It is defined as follows [1], [11]:

$$\text{TI} = mean_{time}\{\sigma_{\text{FD}}(k)\}, \tag{1}$$

where $\sigma_{\text{FD}}(k)$ represents the instantaneous TI index and equals $std_{space}[F_k(i,j) - F_{k-1}(i,j)]$. $F_k(i,j)$ represents the k^{th} frame and (i,j) the corresponding spatial coordinates.

Precisely, the cut detection process is based on a comparison between a fixed threshold value (defined by extensive simulations to 10) and the instantaneous TI index $\sigma_{\text{FD}}(k)$. However, to avoid false detection that may arise due to rapid changes in a single shot (e.g., sports content), we proposed to first remove the moving average $\text{TI}_{mov}(k)$) to the instantaneous TI index before comparing the resulting value to the threshold. If a cut is detected inside the frames of the current GoP, they are separated into two groups: the first one is added to the previous GoP and the last constitutes the new GoP that will be created. This version is actually defined by the Adaptive GoP-size based on Cut detection (AGCut-SoftCast) algorithm.

To fully take advantage of the decorrelation property of the temporal DCT and if the hardware capacities (such as buffer or processor) allow to use larger GoP-size ($\geq$16), the AGCC-SoftCast algorithm can be used. In this case and independently of the cut detection process, the size of the GoP is automatically adjusted according to Table I to provide the best trade off between complexity cost and received quality. This look-up table was generated in [1] based on extensive frame by frame empirical analysis on CIF and HD720p video sequences from [10].

TABLE I: Look-up table from [1] to perform the GoP-size adaptation based on threshold over the TI_{mean} indexes.

TI_{mean} threshold	Optimal GoP-size
$\text{TI}_{mean} \leq$12	32
12$<\text{TI}_{mean}<$27	16
$\text{TI}_{mean} \geq$27	8

To select the optimal GoP-size we rely on the instantaneous TI index. Precisely, a local arithmetic mean TI_{mean} is defined and computed over the instantaneous TI index values. The

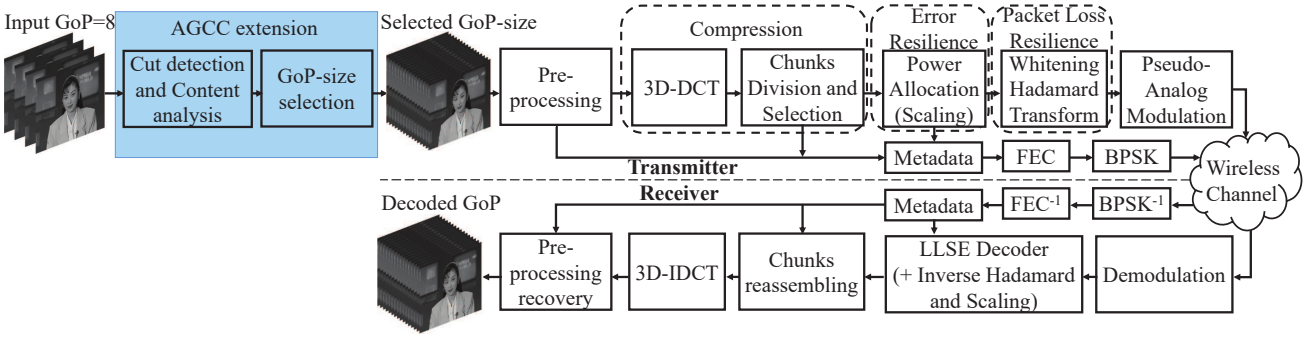


Fig. 5: Block diagram of the AGCC-SoftCast scheme.

latter is evaluated each 8 frames and is compared to the thresholds defined in Table I. According to the resulting TI_{mean} value, the 8 frames are either directly encoded or stored in a buffer whose size ranges between 8 and 32 frames.

Once the optimal GoP-size has been selected and the potential frames coming from the cut detection process added, the classical SoftCast algorithm is performed. First, a pre-processing is applied on the GoP. Specifically, the spatial average is subtracted from each image to reduce the energy of the transmitted data. By using this preprocessing method, the received quality is improved by up to 2.5 dB. We refer to [12] for further details.

Then, a 3D full-frame DCT is used as a decorrelation transform. The DCT frames are divided into N small rectangular blocks of transformed coefficients called *chunks*.

When the available channel bandwidth for the transmission is less than the signal bandwidth, *i.e.,* only $M < N$ chunks may be transmitted, SoftCast discards the $N - M$ chunks with less energy. This is generally the case especially for the transmission of High Definition (HD) content as mentioned in [6]. At the receiver side, these discarded chunks are replaced by null values [6].

The sixth block at the transmitter consists of a chunk scaling operation to match the transmission power constraints. The scaling coefficients are chosen so as to minimize the reconstruction Mean Square Error (MSE). A Hadamard transform is applied to the scaled chunks to provide packet loss resilience. This process transforms the chunks into slices. Each slice is a linear combination of all scaled-chunks. Finally, the slices are transmitted in a pseudo-analog manner using Raw-OFDM [6]. Classical channel coding is skipped. In parallel, the SoftCast transmitter sends an amount of data referred as metadata. These data consist of the mean and the variance of each transmitted chunk as well as a bitmap, indicating the positions of the discarded chunks into the GoP. Metadata are strongly protected and transmitted in a robust way (e.g., BPSK [6]) to ensure correct delivery and decoding.

At the receiver side, a Linear Least Square Error (LLSE) decoder is used to estimate the content of the chunks due to channel noise. Using the metadata, the decoded chunks are properly reassembled and undergo an inverse 3D-DCT, providing the corresponding GoP.

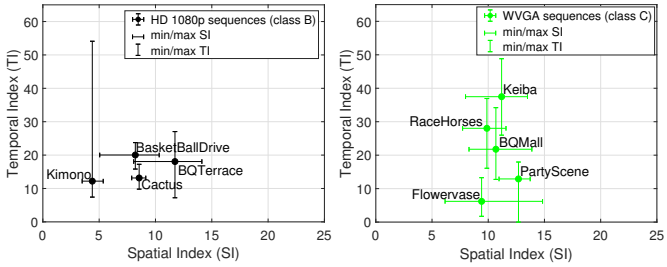


Fig. 6: Illustration of the SI,TI indexes for the selected video sequences. Dots correspond to the average values across all the video sequence. Vertical and horizontal bars represent respectively the min/max value of the TI index and the SI index of the video sequences. From left to right: HD1080p sequences (class B), WVGA sequences (class C) [8].

III. Preliminary Analysis

In this section, we first verify the validity of the look-up table (Table I) proposed in [1] on different video sequences and resolution: the class B (HD1080p: 1920×1080 pixels) and C (WVGA sequences: 832×480 pixels) that were used by the MPEG committee for the standardization of HEVC [8].

The complexity of the selected videos is represented in Fig. 6 according to the spatiotemporal information indexes (SI, TI). The SI index is defined as follows:

$$\text{SI} = mean_{time}\{std_{space}[Sobel(F_k(i,j))]\}, \qquad (2)$$

where $Sobel()$ denotes the Sobel filtering operation.

A. Simulation Setup

Three GoP-sizes of 8, 16 and 32 frames and two CRs of 0.25 (75% of discarded chunks) and 1 (no compression applied) are considered. Transmissions through AWGN channels are simulated and represented by the CSNR value varying from 0 to 30dB by 5dB step as in [1]. Each frame is split into 64 chunks as in [1], [7], [12]. As classically done in the literature [7], [9], [12], only the luminance is considered hereafter.

B. Simulation Results

Simulation results are displayed in Table II and Table III. In the original paper [1], we used an informal threshold of 0.4 dB, below which the MPEG committee considers that the

TABLE II: Resulting average PSNR and SSIM scores over the whole sequence for different GoP-sizes and CSNR values with CR = 1 and CR=0.25. Class C video sequences from the JCT-VC [8].

Simulation Setup			CSNR(dB) 0		10		20	
			PSNR(dB)	SSIM	PSNR(dB)	SSIM	PSNR(dB)	SSIM
GoP = 8	CR=1	*BQ Mall*	**28,38**	**0,756**	**36,94**	**0,938**	**46,40**	**0,992**
		Flower Vase	30,86	0,764	39,45	0,941	48,76	0,992
		Keiba	**27,49**	**0,710**	**36,32**	**0,925**	**45,84**	**0,990**
		PartyScene	26,63	0,741	34,89	0,934	44,42	0,991
		Racehorse	**27,68**	**0,713**	**36,00**	**0,919**	**45,50**	**0,989**
	CR=0.25	*BQ Mall*	**24,86**	**0,636**	**31,63**	**0,854**	**35,35**	**0,938**
		Flower Vase	27,27	0,654	34,10	0,870	37,99	0,962
		Keiba	**23,46**	**0,574**	**31,11**	**0,825**	**36,55**	**0,945**
		PartyScene	23,76	0,618	28,98	0,841	31,13	0,918
		Racehorse	**24,67**	**0,608**	**29,97**	**0,817**	**32,09**	**0,902**
GoP = 16	CR=1	*BQ Mall*	28,69	0,765	37,22	0,941	46,67	0,992
		Flower Vase	31,57	0,783	40,12	0,948	49,38	0,993
		Keiba	27,57	0,711	36,40	0,925	45,92	0,990
		PartyScene	**27,04**	**0,754**	**35,30**	**0,939**	**44,81**	**0,992**
		Racehorse	27,74	0,714	36,05	0,920	45,55	0,989
	CR=0.25	*BQ Mall*	25,180	0,648	31,912	0,860	35,579	0,939
		Flower Vase	28,05	0,679	34,77	0,883	38,56	0,965
		Keiba	23,58	0,577	31,19	0,827	36,57	0,945
		PartyScene	**24,175**	**0,635**	**29,418**	**0,851**	**31,621**	**0,924**
		Racehorse	24,77	0,611	29,99	0,817	32,04	0,900
GoP = 32	CR=1	*BQ Mall*	28,83	0,769	37,35	0,942	46,79	0,992
		Flower Vase	**32,02**	**0,794**	**40,55**	**0,951**	**49,77**	**0,993**
		Keiba	27,63	0,710	36,43	0,924	45,95	0,990
		PartyScene	27,29	0,760	35,53	0,941	45,04	0,992
		Racehorse	27,77	0,714	36,06	0,919	45,56	0,989
	CR=0.25	*BQ Mall*	25,35	0,654	32,02	0,863	35,59	0,939
		Flower Vase	**28,56**	**0,694**	**35,19**	**0,890**	**38,86**	**0,967**
		Keiba	23,64	0,576	31,23	0,826	36,58	0,945
		PartyScene	24,43	0,644	29,67	0,856	31,88	0,926
		Racehorse	24,81	0,611	29,99	0,817	32,00	0,899

TABLE III: Resulting average PSNR and SSIM scores over the whole sequence for different GoP-sizes and CSNR values with CR = 1 and CR=0.25. Class B video sequences from the JCT-VC [8].

Simulation Setup			CSNR(dB) 0		10		20	
			PSNR(dB)	SSIM	PSNR(dB)	SSIM	PSNR(dB)	SSIM
GoP = 8	CR=1	*BasketBallDrive*	**30,40**	**0,781**	**38,91**	**0,949**	**48,24**	**0,993**
		BQ Terrace	27,88	0,723	36,40	0,931	45,88	0,991
		Cactus	29,56	0,776	38,09	0,947	47,49	0,993
		Kimono	33,98	0,894	42,68	0,979	51,61	0,997
	CR=0,25	*BasketBallDrive*	**26,93**	**0,672**	**33,77**	**0,861**	**37,64**	**0,930**
		BQ Terrace	24,40	0,590	31,21	0,823	34,92	0,913
		Cactus	25,98	0,654	32,87	0,857	36,59	0,927
		Kimono	29,95	0,807	37,69	0,937	42,52	0,969
GoP = 16	CR=1	*BasketBallDrive*	30,53	0,783	39,01	0,949	48,34	0,993
		BQ Terrace	**28,48**	**0,741**	**36,95**	**0,937**	**46,40**	**0,992**
		Cactus	30,25	0,794	38,73	0,953	48,07	0,994
		Kimono	**34,41**	**0,900**	**43,06**	**0,980**	**51,94**	**0,997**
	CR=0,25	*BasketBallDrive*	27,09	0,676	33,86	0,862	37,62	0,930
		BQ Terrace	**25,08**	**0,614**	**31,73**	**0,834**	**35,22**	**0,916**
		Cactus	26,81	0,683	33,44	0,868	36,86	0,930
		Kimono	**30,47**	**0,819**	**38,06**	**0,940**	**42,66**	**0,969**
GoP = 32	CR=1	*BasketBallDrive*	30,53	0,781	39,00	0,949	48,33	0,993
		BQ Terrace	28,80	0,749	37,23	0,939	46,67	0,992
		Cactus	**30,78**	**0,808**	**39,21**	**0,956**	**48,51**	**0,994**
		Kimono	34,60	0,902	43,22	0,980	52,07	0,997
	CR=0,25	*BasketBallDrive*	27,14	0,675	33,83	0,860	37,52	0,929
		BQ Terrace	25,47	0,628	31,96	0,840	35,26	0,917
		Cactus	**27,45**	**0,707**	**33,87**	**0,876**	**37,03**	**0,932**
		Kimono	30,72	0,825	38,22	0,941	42,70	0,969

difference is visually unnoticeable [13], to decide the optimal GoP-size for each sequence. The same criterion is here applied and the optimal GoP-size for each sequence is indicated in bold in the tables. The higher the objective metrics are, the better the quality received.

As observed in Table II and Table III, it can be seen that in average, increasing the GoP-size for video sequences with low spatio-temporal complexity increases the reconstructed quality, whereas for videos with strong spatio-temporal complexity, a smaller GoP-size is sufficient since the improvement is limited.

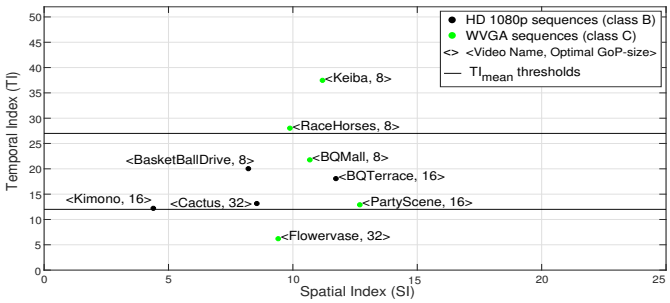


Fig. 7: Illustration of the optimal GoP-size over spatio-temporal indexes (SI, TI) for the selected video sequences. Black and green dots correspond to the average values of the SI, TI indexes for the class B and C video sequences, respectively. The label of each dot refers to the following couple data: <Video name, Optimal GoP-size>.

These results are in accordance with the ones provided in [1]. From Table II and Table III, we can establish a visual synthesis in Fig. 7. In average, the thresholds proposed in [1] are still relevant for the considered videos even if the latter were not used in the original paper to determine and fix the thresholds. The *BasketBallDrive, Cactus and BQMall* sequences may seems to be not well classified however, the displayed optimal GoP-size corresponds to an average over the whole sequence. In our solution, a local GoP-size adaptation is instead performed to reduce improper decisions. Still, we discuss in the following such wrong/improper decisions.

Two situations exist:

- First: A sequence is classified with a GoP-size of 16 whereas a GoP-size of 8 represents the optimal GoP-size. In such case, there is no quality loss as shown in the results (e.g., the *BQ Mall* sequence in Table II). Nevertheless, the improvement brought by using such GoP-size is considered meaningless (e.g. 28.38dB to 28.69≈0.3dB gain for a CSNR=0dB and CR=1) and a GoP-size of 8 could have been chosen with other thresholds. In such case, improper decision leads to a useless increase of the complexity cost but does not impact the received quality.
- Second: A sequence is classified with a GoP-size of 16 whereas a GoP-size of 32 represents the optimal GoP-size (e.g. *Cactus* in Table III). In such case, the maximal quality improvement is not reached but is limited (30.25dB instead of 30.78dB≈0.5dB loss for a CSNR=0dB and CR=1).

As demonstrated, such improper decisions considering a whole sequence only have a very limited impact on the global performance. This impact is further reduced by performing a local GoP-size adaptation.

IV. Performance Analysis of the AGCC-SoftCast scheme

We now evaluate the performance of the AGCC-SoftCast scheme considering three different types of cuts as displayed in Fig. 2, Fig. 3 and Fig. 4. In addition to the previous used metrics, we add the MS-SSIM [14] and VMAF [15] ones

TABLE IV: Evaluation of the average and maximum gain brought by the AGCC extension in comparison to the classical SoftCast scheme considering different GoP-size (8, 16 and 32).

Simulation Setup			Gain							
			PSNR(dB)		SSIM		MS-SSIM		VMAF	
			Avg	Max	Avg	Max	Avg	Max	Avg	Max
Mixed_HD CSNR=0dB (# of cuts=8)	CR=1	AGCC-GoP32	0.88	10.31	0.019	0.301	0.019	0.281	3.86	46.05
		AGCC-GoP16	0.97	11.60	0.016	0.338	0.016	0.322	4.47	47.5
		AGCC-GoP8	1.48	10.44	0.024	0.313	0.023	0.297	7.49	43.97
	CR=0,25	AGCC-GoP32	0.77	8.69	0.025	0.355	0.024	0.326	4.12	49.33
		AGCC-GoP16	0.96	9.51	0.026	0.380	0.025	0.353	5.56	52.44
		AGCC-GoP8	1.58	8.74	0.045	0.360	0.041	0.336	9.82	51.26
News_CIF CSNR=10dB (# of cuts=3)	CR=1	AGCC-GoP32	0.06	2.63	0.001	0.017	0.001	0.013	0.01	5.03
		AGCC-GoP16	0.85	3.19	0.005	0.027	0.004	0.021	1.98	7.86
		AGCC-GoP8	2.03	3.67	0.014	0.034	0.011	0.027	5.16	10.08
	CR=0,25	AGCC-GoP32	0.06	2.33	0.001	0.037	0.001	0.031	0.04	6.39
		AGCC-GoP16	0.89	2.81	0.014	0.060	0.011	0.049	3.57	10.01
		AGCC-GoP8	2.07	3.33	0.036	0.073	0.029	0.06	9.14	13.85
Tennis_SD CSNR=20dB (# of cuts=4)	CR=1	AGCC-GoP32	0.54	6.17	0.001	0.013	0.001	0.013	0.6	7.47
		AGCC-GoP16	0.24	6.16	0.001	0.013	0.001	0.013	0.3	7.47
		AGCC-GoP8	0.45	6.38	0.001	0.014	0.001	0.014	0.43	7.87
	CR=0,25	AGCC-GoP32	0.97	10.96	0.008	0.178	0.008	0.174	3.95	67.45
		AGCC-GoP16	0.28	10.49	0.002	0.15	0.002	0.154	1.05	53.57
		AGCC-GoP8	0.39	10.14	0.002	0.13	0.003	0.139	1.64	50.67

as suggested in [9] due to their strong correlation with the subjective scores. The VMAF metric ranges between [0-100] whereas the (MS)-SSIM metrics range between [0-1]. The higher the objective metrics are, the better the quality received.

The first considered video sequence in this section is denoted by *Mixed_HD720p* and contains eight different 720p (1280×720) subsequences from [10], resulting in eight abrupt cuts. It has been generated to cover a large portion of the SI, TI map and is further detailed in [1]. The second and third video sequences come from [10]. They represent respectively a CIF (352×288) video sequence with three intraframe cuts and a SD (720×576) video sequence with four soft cuts.

For each metric, we show in Table IV, the average and maximum gain between the AGCC extension and the original SoftCast scheme considering different simulation setups (video sequence, CSNR value, CR value). We first note that all the cuts inside the different video have been perfectly detected by the AGCC algorithm.

Regardless of the simulation setup and the considered metric, results show that the AGCC extension provides in average a better quality than the classical SoftCast scheme. When CR=1 and CSNR=20dB, the average and maximum gains for the *Tennis_HD* sequence may seems limited, however, it is due to the fact that at such high CSNR value, the noise during transmission is limited, therefore the metrics already reach their maximum values. The maximum gains obtained with the *News_CIF* sequence are not as high as the other videos, although still satisfactory. This is probably due to the fact that the cut only happens in a limited portion of the frame. Depending on the transmitted video content, results show that maximum improvements up to 11.6dB in terms of PSNR, 0.35 in terms of (MS)-SSIM and 67.4 in terms of VMAF can be obtained, especially at the cut boundaries.

Examples of visual comparison are given in Fig. 8 and Fig. 9. Due to limited space, we only show the visual representation of the GoP-size of 32 frames as it usually gives the best reconstructed quality [7]. Furthermore, the MS-SSIM index is not indicated since it gives similar trends to the SSIM one. The ghost effect is perfectly cancelled for both video with the AGCC algorithm. Huge quality improvement can be noticed for *Tennis_SD* whereas it is limited for the *News_CIF* sequence. As mentioned above, it is probably due to the fact that the cut only happens in a limited portion of the frame.

V. Conclusion

In this paper, we review the AGCC-SoftCast scheme and provide additional results considering different video content and resolution as well as different types of cuts: abrupt, soft and intraframe. The validity of the AGCC-SoftCast scheme is assessed considering four objective metrics including: PSNR, SSIM, MS-SSIM and VMAF. The performance of AGCC-SoftCast are compared to the original SoftCast scheme considering different CSNR, CR and GoP-size values. Results are in accordance with the original paper and highlight the importance of the AGCC extension in a SoftCast context. Regardless of the considered types of cuts, we show that they are perfectly detected by the proposed algorithm. Depending on the transmitted video content, results show that improvements up to 11.6dB in terms of PSNR and 67.45 in terms of VMAF can be obtained, especially at the cut boundaries.

References

[1] A. Trioux, F.-X. Coudoux, P. Corlay, and M. Gharbi, "Temporal information based gop adaptation for linear video delivery schemes," *Signal Proc.: Image Commun.*, vol. 82, p. 115734, 2020.

[2] I. E. G. Richardson, *The H.264 advanced video compression standard*, 2nd ed. Chichester: Wiley, 2010.

[3] G. J. Sullivan, J. R. Ohm, W. J. Han, and T. Wiegand, "Overview of the High Efficiency Video Coding (HEVC) Standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1649–1668, Dec. 2012.

[4] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the scalable video coding extension of the h. 264/avc standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 9, pp. 1103–1120, 2007.

[5] J. M. Boyce, Y. Ye, J. Chen, and A. K. Ramasubramonian, "Overview of shvc: Scalable extensions of the high efficiency video coding standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 1, pp. 20–34, 2015.

[6] S. Jakubczak and D. Katabi, "A cross-layer design for scalable mobile video," in *Proc. of the 17th annual international conf. on Mobile computing and networking (MobiCom)*, Sep. 2011, pp. 289–300.

[7] R. Xiong, F. Wu, J. Xu, X. Fan, and al., "Analysis of decorrelation transform gain for uncoded wireless image and video communication," *IEEE Trans. Image Process.*, vol. 25, no. 4, pp. 1820–1833, Apr. 2016.

[8] M. Wien, *High Efficiency Video Coding: Coding Tools and Specification*, ser. Signals and Communication Technology. Berlin Heidelberg: Springer-Verlag, 2015.

[9] A. Trioux, G. Valensize, M. Cagnazzo, M. Kieffer, F.-X. Coudoux, P. Corlay, and M. Gharbi, "Subjective and objective quality assessment of the softcast video transmission scheme," in *IEEE Visual Commun. Image Process. (VCIP)*, Dec. 2020.

[10] "Xiph.org :: Derf's Test Media Collection." [Online]. Available: https://media.xiph.org/video/derf/

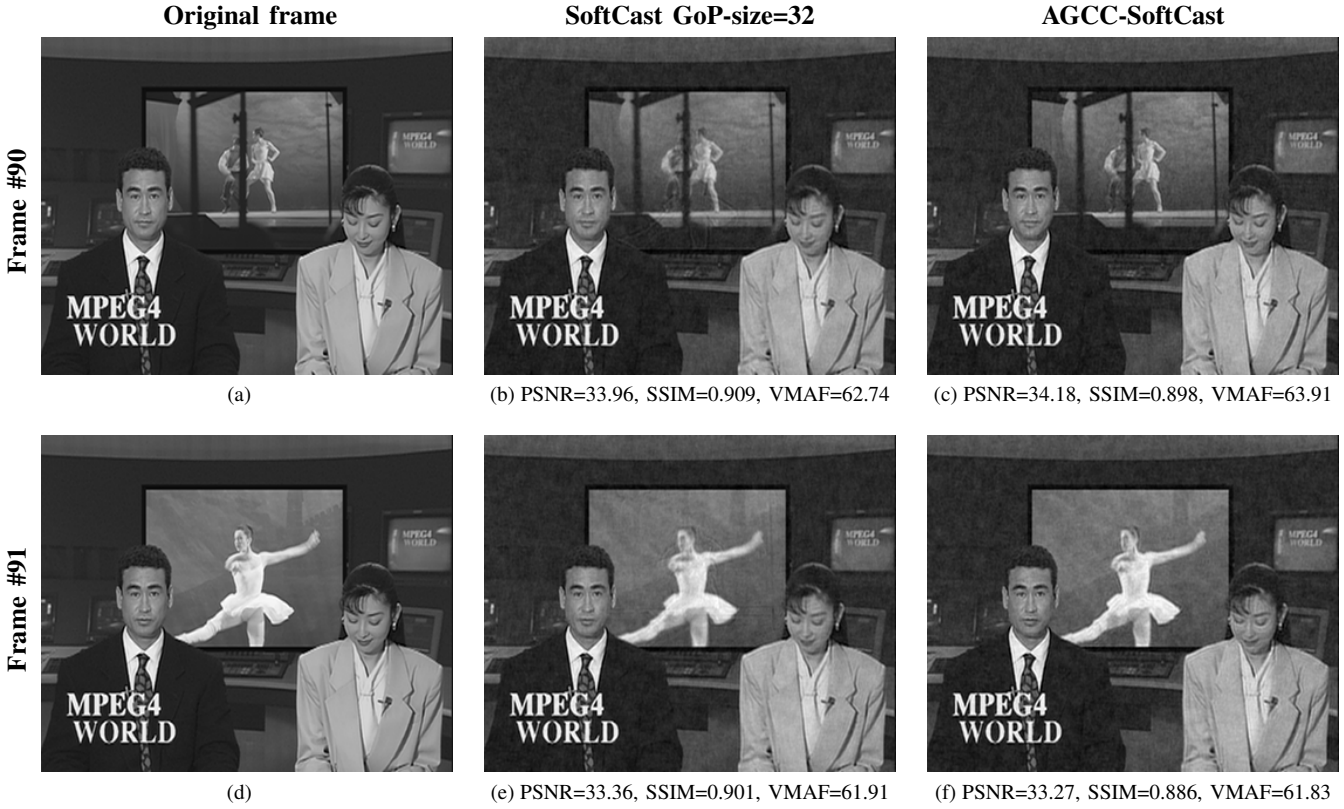
[11] P. ITU-T RECOMMENDATION, "Subjective video quality assessment methods for multimedia applications," Sep. 1999.

[12] A. Trioux, F.-X. Coudoux, P. Corlay, and M. Gharbi, "A comparative preprocessing study for softcast video transmission," in *Proc. IEEE Int. Symp. on Signal Image and Video Commun. (ISIVC)*, Nov. 2018.

[13] D. Salomon and G. Motta, *Handbook of Data Compression*, 5th ed. London: Springer-Verlag, 2010.

[14] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *The 27th Asilomar Conf. on Signals, Syst. & Computers, 2003*, vol. 2, 2003, pp. 1398–1402.

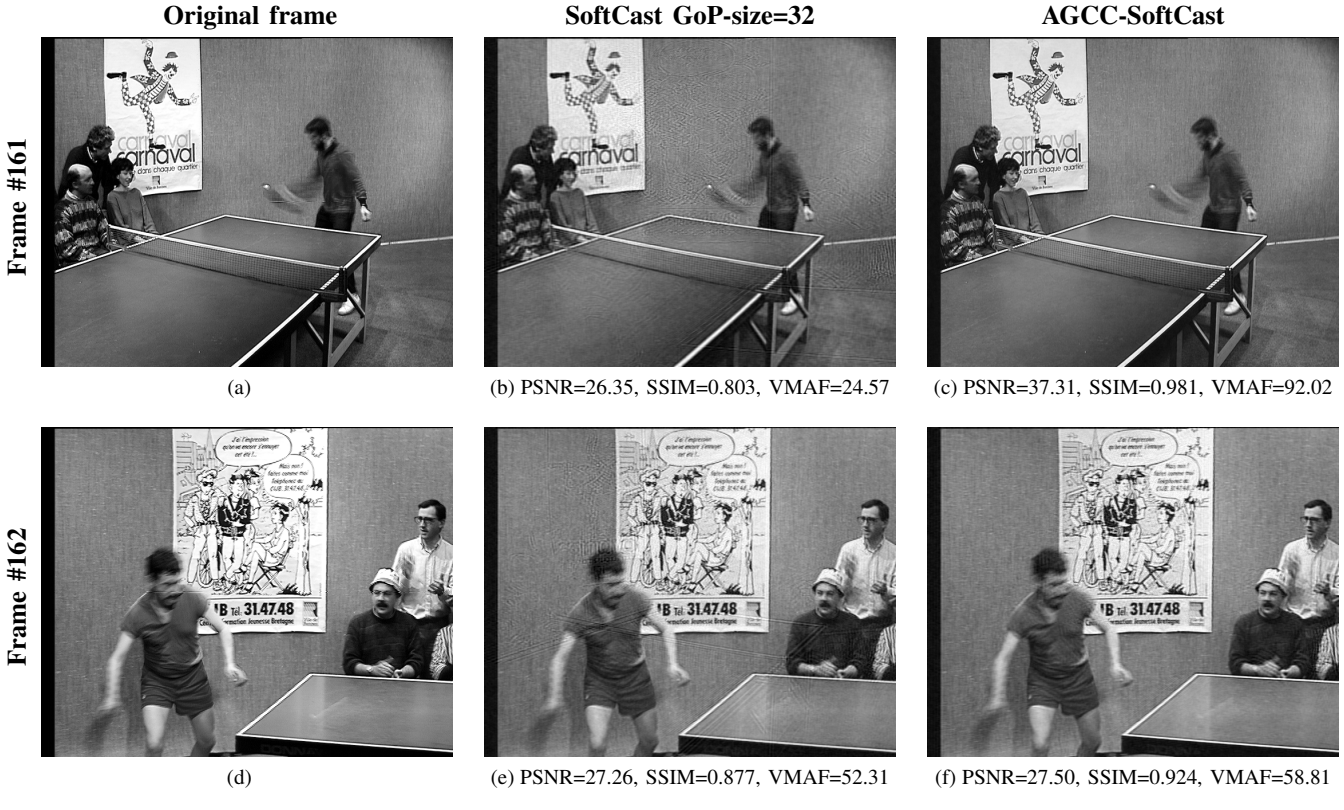
[15] C. G. Bampis *et al.*, "Spatiotemporal Feature Integration and Model Fusion for Full Reference Video Quality Assessment," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 29, no. 8, pp. 2256–2270, Aug. 2019.



(a) (b) PSNR=33.96, SSIM=0.909, VMAF=62.74 (c) PSNR=34.18, SSIM=0.898, VMAF=63.91

(d) (e) PSNR=33.36, SSIM=0.901, VMAF=61.91 (f) PSNR=33.27, SSIM=0.886, VMAF=61.83

Fig. 8: Visual quality comparison at a CSNR = 10dB, CR = 0.25 for the *News_CIF* sequence (Frames #90, 91). (a),(b),(c): Frame #90. (d),(e),(f): Frame #91. (a),(d): Original frames. (b),(e): SoftCast with fixed GoP-size of 32 frames. (c),(f): AGCC-SoftCast. The PSNR scores are expressed in decibels.



(a) (b) PSNR=26.35, SSIM=0.803, VMAF=24.57 (c) PSNR=37.31, SSIM=0.981, VMAF=92.02

(d) (e) PSNR=27.26, SSIM=0.877, VMAF=52.31 (f) PSNR=27.50, SSIM=0.924, VMAF=58.81

Fig. 9: Visual quality comparison at a CSNR = 20dB, CR = 0.25 for the *Tennis_SD* sequence (Frames #161, 162). (a),(b),(c): Frame #161. (d),(e),(f): Frame #162. (a),(d): Original frames. (b),(e): SoftCast with fixed GoP-size of 32 frames. (c),(f): AGCC-SoftCast. The PSNR scores are expressed in decibels.