SCENE CHANGE ADAPTATION FOR SCALABLE VIDEO CODING

Tea Anselmo, Daniele Alfonso

Advanced System Technology Labs., STMicroelectronics
Via Olivetti 2, 20041, Agrate Brianza, Italy
email: tea.anselmo@st.com, daniele.alfonso@st.com

ABSTRACT

This document presents a Scene Change Adaptation method for the Scalable Video Coding extension of H.264/AVC. Our method is based on a scene change detection algorithm that identifies transitions in video sequences by using the motion information provided by the pre-analysis phase of a Fast Motion Estimation algorithm based on spatial-temporal motion correlation. Intra coding picture is imposed at the end of the Group Of Pictures containing the scene change, thus dynamically adapting the sequence structure to the scene content and preventing the propagation of the prediction error due to scene change. The proposed algorithm has been also combined with a Buffer-Based Constant Bit-Rate control algorithm, which ensures HRD buffer compliance and target bit-rate requirements. Simulation results show that the SCD algorithm can improve the global coding efficiency and the local visual quality when a scene change occurs.

1. INTRODUCTION

Scalable Video Coding (SVC) is one of the latest video standards developed by the Joint Video Team (JVT) of the ITU-T Video Coding Expert Group (VCEG) and the ISO/IEC Moving Picture Expert Group (MPEG) as an extension of H.264/AVC [1][2]. SVC provides scalable video streams, which are composed of a base layer and one or more enhancement layers. Enhancement layers may enhance the temporal, spatial or SNR resolutions of the base layer representation, thus adapting the stream to a variety of end-users in terms of capabilities and applications. The JVT provides a reference software, the Joint Scalable Video Model (JSVM), which implements a fully scalable encoder [3].

The encoded bit-stream must respect transmission or storage constraints, such as channel bandwidth or limited memory availability. Moreover, to maximize the end user experience, the encoder is expected to optimize the subjectively perceived image quality and to keep it constant throughout the whole encoding process.

Given the encoding configuration set, the encoder can achieve the above mentioned requirements by optimally adapting the quantization parameter and the coding modes based on the actual scene content. But in real-world videos and movie sequences the scene content may vary quite often, thus compromising motion estimation process and consequently worsening the visual quality, as described in Section 2.

In order to improve visual quality at scene change points and provide a more constant image quality over time, Scene Change Detection (SCD) is fundamental. Moreover, SCD is also useful for supporting and improving other application algorithms, such as video indexing, to enable fast browsing and retrieval of sub-sequences of interest to the user, or as bit-rate control, since it allows a more efficient HRD (Hypothetical Reference Decoder) buffer management.

In this paper we propose a new effective SCD method with extremely low additional computation to dynamically adapt the sequence structure to the scene content, ensuring good reference pictures for motion estimation and providing regular random access points to the bit-stream.

The rest of the paper is organized as follows. Section 2 analyzes scene change influence and describes our proposed method. Simulation results are shown in Section 3 to verify the algorithm effectiveness and conclusions are reported in Section 4.

2. SCD ALGORITHM

We define a scene to be a sequence of pictures that appears to be continuously captured by the same camera. A scene change happens when the correlation between two subsequent frames is small or their relative motion is larger than the search range of the Motion Estimation (ME). If the scene has been changed, the motion estimation will fail and many macroblocks (MBs) in the picture will be coded as Intra, thus causing a considerable and unexpected increment of bit-rate.

In H.264/AVC there are basically three types of video pictures: Intra picture (I), prediction-coded picture (P) and bidirectionally-predicted picture (B). JSVM software provides temporal scalability with the concept of hierarchical B pictures [1]. The dyadic temporal enhancement layers are typically coded as B pictures, where forward and backward references are restricted to the nearest temporally preceding and succeeding pictures belonging to lower temporal layers. Typically, only the coarsest temporal layer pictures are I or P coded and are known as key pictures. The subsequence between two key pictures, including the next I or P picture, is referred to as a Group Of Pictures (GOP) while the distance between two I coded pictures is known as Intra Period (IP).

The I pictures are coded independently exploiting spatial correlation only and need more bits than P and B pictures. Since P and B pictures are coded by motion compensated prediction, their quality depends not only on the quantization step but also on motion estimation accuracy. Analyzing more

in detail the influence of a scene change, we observe that the effect depends on its position within the GOP. By way of example, we suppose to have a single layer with a hierarchical GOP of 4 pictures length, as depicted in Figure 1, and we state to have no more than one scene change per GOP.

If a scene change happens just before an I picture, all the B pictures of the current GOP will be directly or indirectly influenced because the I picture is one of the available references. But the B pictures can also refer to the previous key picture of the last GOP, thus still providing a good prediction result. So there is little influence on the prediction efficacy when scene changes occur before an I picture.

Since hierarchical B pictures are predicted from neighbouring and temporally adjacent I, P or B pictures, the current B picture will have at least one reference belonging to the same scene. When scene change happens at a B picture, again we can get a satisfactory prediction.

P pictures are predicted from previous I or P coded anchor frames. If scene changes happen before a P, motion estimation will fail and many MBs will be Intra coded. In this case, a fixed compression ratio would be sustained only at the cost of a visible image quality loss. The prediction accuracy of this P would be low and consequently the quality of the pictures predicted from this P would be worsened. On the contrary, quality loss would be avoided at the cost of a decreased compression ratio, but this solution could not be applicable in case of strict bit-rate constraints. Resuming the above observations, encoder action is necessary only in case of scene change before a P picture.

An SCD algorithm can precisely identify the scene changed picture and try to limit quality loss and bit-rate increment by avoiding ineffective inter picture coding. Performance evaluations and characterization of a number of scene change algorithms are proposed in [4][5]. Colour histograms, edge change ratio, block motion information or video content variation are used to compute picture differences. An SCD method for Scalable Video Coding has been proposed in [6], but it is specifically designed for Motion Compensated Temporal Filtering (MCTF) structure, a video coding technique that was considered during the early development stages of SVC, but eventually abandoned. Generally these SCD methods require some extra computation to obtain scene characteristics.

A first version of the proposed algorithm has been previously implemented for H.264/AVC encoding [7]. It has been improved and adapted to SVC by maintaining hierarchical temporal scalability and allowing multiple layers encoding. The algorithm has been integrated in a JSVM 9.8 compliant encoder, which further includes proprietary Fast Motion Estimation [8] and Constant Bit-Rate (CBR) control procedures [11].

Our SCD algorithm does not need any ad hoc frame by frame computation because it uses the motion information provided by Coarse Search, the pre-analysis phase of a Fast ME algorithm, to identify changes in the scene content of a video sequence. The Fast ME algorithm is able to achieve performance very close to the one of Full Search Block Matching by using only a small fraction of its computational complexity. Actually the complexity of our ME algorithm is independent of search window size, so that the search region can be set equal to the entire picture size and wrong scene change detections due to limited search area are avoided. A more detailed description of the Fast ME algorithm can be found in [8].

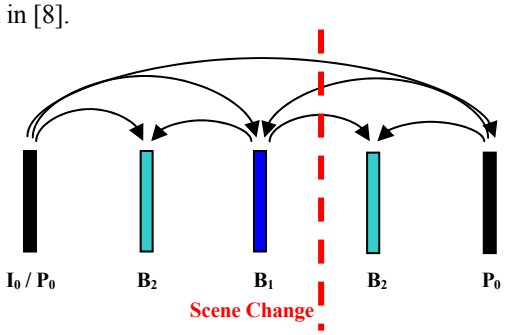

Figure 1 – Effect of scene change on hierarchical prediction for a 4 length GOP.

The Coarse Search analyses temporal correlation by testing a predefined set of motion vectors and assigns to each MB a first rough prediction, estimated with respect to the previous picture, regardless of the picture coding type. The SCD technique evaluates the coarse results to identify poor temporal prediction at MB level: if the scene change had happened, the ME will fail, leading to a large prediction error. So, for each MB, the algorithm compares the temporal correlation with the spatial one to identify the most costly convenient to be coded, as described in the following pseudo-code:

$$\begin{aligned} &if\left(\left(MAE > VAR\right)\&\left(SAD > T_1\right)\right) \\ &\quad potential\_INTRA\_MBs++; \end{aligned} \tag{1}$$

where MAE (Mean Absolute Error) and SAD (Sum of Absolute Differences) represent the temporal correlation and coding cost, while the variance (VAR) gives a measure of the spatial homogeneity. T_1 is an empirical threshold value obtained by several test cases. If the condition in (1) is true, meaning the temporal prediction is too costly, the current MB is marked as a potential Intra MB and the MB counter is incremented.

At the end of each picture, SCD computes a smoothness parameter, which is a novelty with respect to previous implementation in [7]. The smoothness parameter represents the correlation lying between the motion vectors of adjacent MBs of the same picture: it is used to keep track of motion complexity and thus to avoid false detections due to high motion shots. For the pictures within the same shot, the ME produces coherent motion vectors to describe realistic movement of objects between contiguous pictures. On the contrary, in case of scene change, the motion fields are incoherent because they don't refer to a consistent movement. The smoothness parameter for the *i*-th picture is obtained by averaging for every MB the vectors components of four neighbouring MBs, as depicted in Figure 2 and shown in the equation below:

$$\begin{aligned} &Smooth_i = \frac{1}{4 \cdot N}\sum_{m=1}^{N}\sum_{n=1}^{4}\left|\Delta x_{m,n}\right| + \left|\Delta y_{m,n}\right| \\ &Smooth_i - \frac{1}{k}\sum_{j=i-1}^{i-k} Smooth_j > T_2 \end{aligned} \tag{2}$$

where N is the total number of MBs per picture, $\Delta x_{m,n}$ and $\Delta y_{m,n}$ are the horizontal and vertical components of the motion vectors of neighbouring MBs. To identify local motion discontinuity, the $Smooth_i$ value is compared with the average smoothness of a local window of k pictures. If the difference between current smoothness and the average of the past k smoothness values is greater than an empirical threshold T_2, then a motion complexity discontinuity is pointed out. T_2 threshold selection is critical, a trade-off must be found so that it is high enough to avoid false hits caused by high motion scene and low enough to avoid miss rate. $T_2 = 1.5$ works well for the video test set.

When both the number of hypothetical Intra MBs in (1) exceeds 40% of the entire picture amount and a complexity discontinuity is verified in (2), a scene change is said to happen.

After the identification of a scene change, the next key picture is forced to be Intra coded, thus inducing a Variable Intra Period (VIP). By dynamically forcing an Intra picture coding, the temporal correlation is properly exploited within each Intra Period and motion estimation is prevented across the scene change, at a very low complexity cost. Moreover the temporal scalability is preserved, thus avoiding misalignment with enhancement layers at higher temporal scalability.

Since I pictures need more bits to be coded, repetitive scene changes can significantly worsen the compression ratio and irreparably compromise rate constraints. As proposed in [9], when the IP of length M containing a scene change is stopped, the new started IP combines the remaining pictures of the previous one and the new M pictures, changing the I picture in the next IP to P picture. Moreover, a minimum distance of 4 pictures between two successive scene change is imposed to allow the sliding window refresh for smoothness calculation and to avoid oversize bit-rate increment. On the contrary, an upper IP bound is necessary to limit the prediction error propagation. The last remark is even more necessary in case of low bit-rate applications, so that IP is bounded to a maximum length of 64 pictures. In case of unconstrained IP, meaning only the first picture is Intra coded, if a scene change occurs, the key picture is forced to I, without any other concern for IP calculation.

3. SIMULATION RESULTS

In order to verify the SCD algorithm effectiveness, we used seven test sequences, obtained by mixing a set of video segments, characterized by various amount of movement complexity and different resolutions, as reported in Table 1.

The first simulation test has been performed at fixed quantization parameter (QP), disabling any rate control algorithm. By varying only the QP, other configuration parameters being equal, we used $QP_I=QP_P=[10, 25, 40]$, $QP_{B1} = QP_I + 3$, $QP_{B2} = QP_I + 4$, $QP_{B3} = QP_I + 5$.

We analysed two different sequence structures, with IP and GOP couples equal to [16,4] and [32,8]. For both configurations and regardless of the QP, we obtained the same performance, which are evaluated by three basic numbers:

- **hit rate** is the ratio of correctly detected scene changes to its actual number,
- **miss rate** is the ratio of missed scene changes to the actual number of scene changes,
- **false rate** is the ratio of incorrectly detected scene changes to the actual number of scene changes.

Table 1 – Test sequences.

Resolution	Pictures Number	Scene Changes
CIF (325x228)	2213	10
PAL (720x576)	420	21
720p_A (1280x720)	750	14
720p_B (1280x720)	420	19
1080p_A (1920x1080)	600	5
1080p_B (1920x1080)	280	11
Tot.	**4683**	**80**

Table 2 – Efficiency of the proposed SCD algorithm.

Total Scene Changes	Hit Rate	Miss Rate	False Rate
80	80 100%	0 0%	2 2.5%

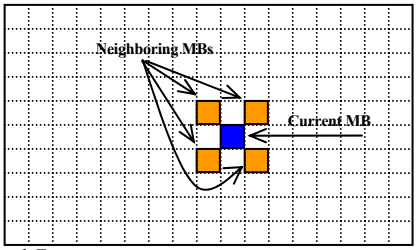

Figure 2 – Neighbouring MBs scheme to obtain Smoothness component of the current MB in the i-th picture.

The results reported in Table 2 show that the SCD algorithm is able to completely detect all the scene changes, which means a reliability of 100% and no missed hits for the analysed test sequences, considerably improved compared to 96% efficiency and 2 missed events reported in [7]. The two false hits are associated with the "Crew" segment of the 720P sequence, which contains photoflash noise. Actually camera flashlight is a critical event for SCD and future improvements of the algorithm will have to solve this issue.

Tables 3-4 compare the performance of our SCD algorithm aided by the VIP solution, with respect to fixed Intra period coding, for three different QP values (15, 25 and 35). The first column reports the results obtained with fixed Intra period ($IP_{REF}=16$), while the last three columns refer to our VIP method (IP equal to 16, 32 and unrestricted). As noticeable from simulation results, our approach causes a negligible Y-PSNR loss, limited to 0.21dB for the highest QP with unrestricted IP, and in general produces a reduced variance, assuring more constant quality throughout the coded sequence. The quality drop is due to the less frequent Intra refresh, which leads to higher cumulative motion compensated prediction error within the Intra Period. Of course the increased error propagation can be traded-off with the im-

proved compression by imposing an upper bound on the maximum IP length as needed by the application.

From bit-rate point of view, the encoder process combining SCD and VIP approaches can yield better compressing rate than conventional approach. For equal Intra period value, IP = 16, the compression gain of the proposed method lays between 0.55% and 7.5%, considering all the tested QP and resolutions. This means the VIP approach always provides an improvement of coding efficiency, depending on bit-rate. The compression gain is even more evident for IP=32 and unrestricted IP: the algorithm can give a benefit up to 24% bit-rate reduction.

The evident advantage of SCD combined with VIP technique can be exploited to improve Constant Bit-Rate (CBR) applications. The preserved bit-rate can lead to higher quality or help a more efficient HRD buffer management. CBR in JSVM model has been partly investigated in [10]: this JVT contribution has been implemented into JSVM software only for the base layer and it extends to SVC the same RC scheme already adopted within the H.264/AVC Joint Model (JM) reference software. However, in the current JSVM software, there are no scene change detection mechanisms, so when scene changes happen, the rate control fails in bit allocation, compromising motion estimation process, and visual quality is consequently worsened.

Hence we tested the proposed SCD and VIP technique with our proprietary CBR algorithm [11]. Our CBR algorithm is a single-pass algorithm, since the encoding process is done once per picture and there is no need of a pre-analysis phase to determine target bits allocation. It is based on the principle of buffer management and is suitable for multiple layer coding: the algorithm tries to achieve, at the end of each Intra period, the same buffer fullness that was before encoding the last Intra picture. As a consequence, it performs a constant bit-rate encoding, since every Intra period consists of about the same number of bits:

$$TargetBitsIntraPeriod = AvgBitPict \cdot IP$$

where *AvgBitPict* is the average amount of bits per picture obtained as the ratio of the target bit-rate and the sequence frame rate. After a scene change, the CBR algorithm has to update some internal parameters relative to the new IP, for example the *TargetBitsIntraPeriod* value and the buffer thresholds. By taking these solutions, the encoder is able to avoid quality loss and maintain, at the same time, good bit-rate control performance.

In order to keep comparable working conditions for both JSVM and the proposed rate control algorithms, in the JSVM configuration file the *MaxQPchange* parameter is set equal to 6, the *BasicUnit* is picture sized and the QP can vary from 1 to 51. The target bit-rate for both algorithms is 2 Mb/s.

The test sequence is in 720P format, 600 pictures long, and it includes 5 scene changes, one every 100 pictures. The results in Table 5 show that the final Y-PSNR qualities are almost the same, but variance is quite smaller for our algorithm, meaning more constant quality throughout the whole coded sequence. By analyzing more specifically the frame by frame PSNR values depicted in Figure 3, we notice some huge quality loss of the JSVM rate control for the pictures immediately after scene changes, causing an annoying visual quality degradation. On the contrary, our algorithm better controls the encoding process by dynamically adapting to the sequence scene content, thus assuring always a satisfactory visual quality.

4. CONCLUSIONS

In this paper we proposed an effective and low complexity scene change detection method, based on the motion information provided by a fast motion estimation algorithm. The method shows a 100% efficiency for hard cuts detection on the tested sequences. By adapting Intra period to scene changes, it allows a bit-rate reduction from 0.55% minimum to 24% maximum relative to the approach without SCD. The quality loss is limited to -0.21 dB in the worst case and to -0.06 dB on average. Further, the proposed method is useful for supporting new applications, as video indexing, and for improving other ones, as bit-rate control, by providing regular random access points to the bit-stream and ensuring good reference pictures and limiting prediction error propagation.

REFERENCES

[1] H. Schwarz, D. Marpe, and T. Wiegand,, "Overview of the Scalable Video Coding Extension of the H.264/AVC Standard," *IEEE Trans. On Circuits and Systems for Video Technology*, vol. 17, no.9, pp. 1103–1020, Sept. 2007.

[2] T. Wiegand, G. Sullivan, J. Reichel, H. Schwarz, and M. Wien, "Joint Draft ITU-T Rec. H.264 | ISO/IEC 14496-10 / Amd.3 Scalable Video Coding", *Joint Video Team*, JVT-X201, Geneva, CH, July 2007.

[3] JSVM 9.8 Software Package, CVS server for JSVM software.

[4] U. Gargi, R. Kasturi, and S.H. Strayer, "Performance Characterization of Video-Shot-Change Detection Methods", *IEEE Trans. On Circuits and Systems for Video Technology*, vol. 10, no. 1, Feb. 2000.

[5] R. Lienhart, "Comparison of Automatic Shot Boundary Detection Algorithms", in *Proc. of SPIE*, *VII Conf. on Storage and Retrieval for Still Image and Video Database*, vol. 3656, pp. 290–301, San Jose, CA, Jan. 1999.

[6] J.-R. Ding and J.-F. Yang, "Joint Adaptive GOP and SCD Coding for Improving H.264 Scalable Video Coding", *Multimedia Workshop, IX IEEE International Symposium on*, TW, Dec. 2007.

[7] D. Alfonso, B. Biffi, and L.Pezzoni, "Adaptive GOP size control in H.264/AVC encoding based on scene change detection", *7th NORSIG*, *Nordic Signal Processing Symposium*, Reykjavik, IS, June 2006.

[8] L. Lima, D. Alfonso, L. Pezzoni, R. Leonardi, "Low Complexity Motion Estimation for Scalable Video Coding extension of H.264/AVC", in *Proc. of Visual Communication and Image Processing (VCIP) 2009,* San Jose, CA, Jan. 2009.

[9] Y. Yu, J. Zhou, and Y. Wang, "A Fast Effective Scene Change Detection and Adaptive Rate Control Algo-

rithm", in *Proc. of ICIP, International Conf. on Image Processing*, vol. 2, pp. 379–382, Chicago, IL, Oct. 1998.

[10] Leontaris, A.M. Tourapis, "Rate Control for the Scalable Video Model", *Joint Video Team*, JVT-W043, San Jose, CA, April 2007.

[11] T. Anselmo and D. Alfonso, "Buffer-Based Constant Bit-Rate Control for Scalable Video Coding", in *Proc. of PCS 2007*, *Picture Coding Symposium*, Lisboa, PT, Nov. 2007.

Table 3 – CIF sequence: (left)Y-PSNR, quality loss and variance, (right) bit-rate values and percentage reduction of VIP approach (IP=16, 32, unconstrained) relative to fixed IP configuration (IP_{REF}=16).

QP \ Intra Period / Y-PSNR (*var*) loss [dB]	16 REF.	16	32	unc.
15	44,56 (*4,76*)	44,55 (*4,74*) -0,01	44,53 (*4,64*) -0,03	44,51 (*4,55*) -0,05
25	37,56 (*7,04*)	37,55 (*7,03*) -0,01	37,52 (*6,96*) -0,04	37,50 (*6,83*) -0,06
35	31,02 (*8,83*)	31,01 (*8,85*) -0,01	30,92 (*8,71*) -0,10	30,81 (*8,33*) -0,21

QP \ Intra Period / BR [kb/s] BR Red. [%]	16 REF.	16	32	unc.
15	3720,48	3700,20 -0,55%	3596,27 -3,34%	3506,06 -5,76%
25	1252,82	1240,09 -1,02%	1172,75 -6,39%	1115,86 -10,93%
35	353,34	347,34 -1,70%	315,26 -10,78%	288,27 -18,41%

Table 4 – 720P_A sequence: (left)Y-PSNR, quality loss and variance, (right) bit-rate values and percentage reduction of VIP approach (IP=16, 32, unconstrained) relative to fixed IP configuration (IP_{REF}=16).

QP \ Intra Period / Y-PSNR (*var*) loss [dB]	16 REF.	16	32	unc.
15	44,18 (*5,03*)	44,17 (*4,99*) -0,01	44,15 (*4,90*) -0,02	44,15 (*4,87*) -0,03
25	38,02 (*5,46*)	38,00 (*5,41*) -0,02	37,95 (*5,30*) -0,07	37,94 (*5,27*) -0,08
35	32,98 (*6,35*)	32,94 (*6,28*) -0,05	32,82 (*6,19*) -0,16	32,77 (*6,20*) -0,21

QP \ Intra Period / BR [kb/s] BR Red. [%]	16 REF.	16	32	unc.
15	25885,70	25727,56 -0,61%	25452,28 -1,67%	25365,62 -2,01%
25	5184,82	5010,01 -3,37%	4709,15 -9,17%	4622,91 -10,84%
35	1147,48	1061,43 -7,5%	915,94 -20,18%	868,26 -24,33%

Table 5 – Comparison of JSVM CBR algorithm to our CBR method including SCD and VIP techniques. Table reports Y-PSNR, variance, final bit-rate and bit-rate errors relative to target bit-rate (2 Mb/s).

	Y-PSNR [dB]	var	BR [kb/s]	BR Error [%]
JSVM	37.36	11.59	2143.67	+7.18%
CBR+SCD+VIP	37.39	7.84	2012.98	+0.65%

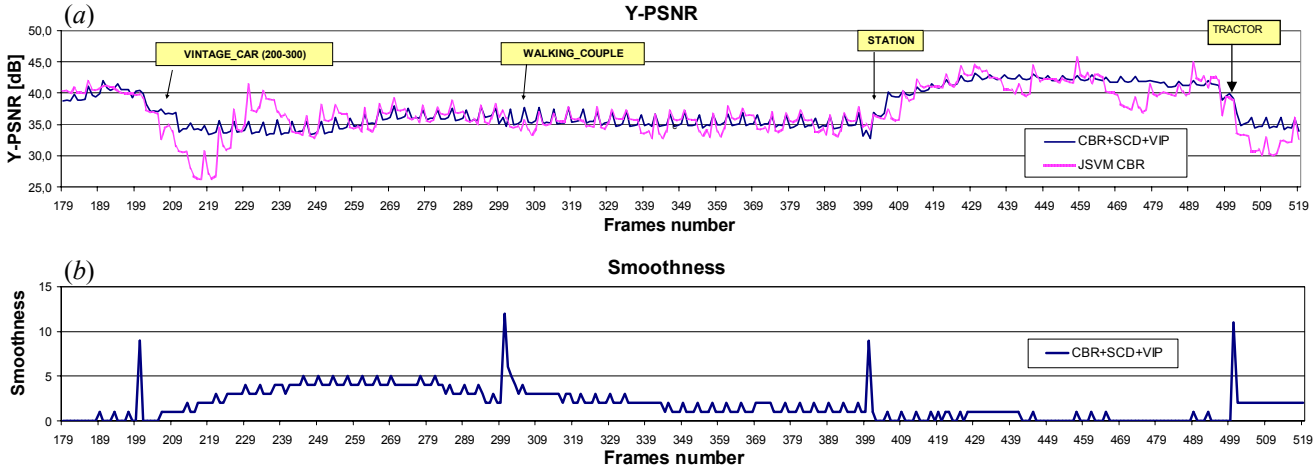

Figure 3 – (*a*) Picture by picture Y-PSNR for 720P sequence: JSVM CBR results (magenta) vs. proposed CBR+SCD+VIP method (blue). (*b*) Picture by picture Smoothness parameter for SCD method: smoothness peaks pinpoint scene change occurrences.