Performance Evaluation of HEVC With Intra-Refresh for Wireless Video Surveillance

Jason Jake E. Tan,[1] Edison A. Roxas,[1] Angelo R. dela Cruz,[1] Argel A. Bandala,[2] and Ryan Rhay P. Vicerra[2,*]

Abstract — **This paper presents the result of the performance evaluation for a wireless video surveillance using the High Efficiency Video Coding (HEVC) standard. The video test sequences are converted into HEVC bit streams by the HM6.0 encoder. The HEVC bit streams were transmitted using a wireless network simulator setup for video surveillance. The wireless network setup introduced error to the HEVC bit stream. Common wireless transmission losses include cross traffic and multipath fading. The intra-refresh error resilient technique is utilized in the HEVC encoder to mitigate the effects of wireless channel. Finally, the effects of adjusting different parameters of the HEVC encoder, namely, the intra-refresh period and quantization parameter were also evaluated in this paper.**

Keywords: **HEVC, wireless video surveillance, packet loss**

I. Introduction

Wireless video surveillance has many applications including security, identification, monitoring, etc., and is used in many establishments. However, wireless transmission comes with some typical problems for data transmission. One of these problems is that wireless networks are more prone to noise compared to wired networks. The noise can be caused by obstruction, signal attenuation and interference, etc. [1]. In order to obtain high-quality videos in a wireless environment, error robustness is required [2]. Another problem that comes with video surveillance is that videos take up large amount of data. These videos must be efficiently compressed in order to reduce the amount of data [3]. It is also a must for a video surveillance to have real-time access. Therefore, it is important for video surveillance encoders to have low computational complexity in order for the video to have minimal delay [1][2].

For efficient compression, High Efficiency Video Coding (HEVC) is the most preferable to use. HEVC standard was made by a partnership called Joint Collaborative Team on Video Coding [2]. The partnership includes ITU-T Video Coding Experts Group and the International Organization for Standardization and the International Electrotechnical Commission Moving Picture Experts Group [2]. HEVC standard can provide better coding efficiency and also provides up to 50% bit-rate savings for the same video quality relative to the previous standards including H.264 [2][3].

The software that we adopted to simulate HEVC is the Test Model under Consideration (TMuC, version HM6.0) [4][5][6]. We use HM6.0 to encode and decode test sequences. The encoded sequences are subjected to noise simulated under Network Simulator 2 (NS2). We modify the HEVC decoder to implement a non-motion compensated error concealment technique. Results show the effects of packet loss on HEVC videos during transmission. The performance of HEVC with intra-refresh over wireless video surveillance was evaluated through the simulations. The results of the simulations measures the objective quality of the video based on its peak signal to noise ratio (PSNR) [7].

The rest of the paper is organized as follows. Section II discusses the design considerations including video transmission architecture and configurations. Section III discusses wireless network model used and the configurations for its simulation. Section IV discusses the results of the experiments, and Section V concludes the paper and also includes recommendations for future works.

II. Design Considerations

The architecture for the simulated video transmission is shown in Figure 1. The first block consists of the input

[1]Jason Jake E. Tan, Edison A. Roxas, Angelo R. dela Cruz, University of Santo Tomas, Manila, Philippines

[2]Argel A. Bandala, and Ryan Rhay P. Vicerra, De La Salle University, Manila, Philippines (e-mail: ryan.vicerra@dlsu.edu.ph)

video sequence. The input video sequence is then encoded using an HEVC encoder. After encoding, the encoded video sequence is transmitted over a wireless network. On the other end, the transmitted video sequence will now be decoded by a video decoder and then be reconstructed [4].

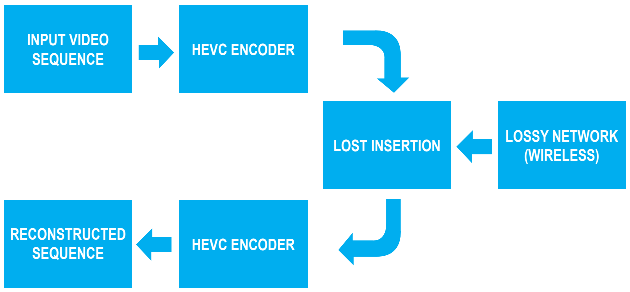


Fig. 1. Video transmission architecture.

Video compression efficiency of HEVC is achieved by exploiting the redundancy of information in both the spatial (within a frame) and temporal dimensions (among neighboring frames) of the video. Due to the high correlation within and among neighboring frames, predictive coding schemes are employed to exploit this redundancy. A predictive coding scheme includes motion estimation and compensation, which are both key parts of video compression.

Almost all video coding standards use these coding schemes such as the MPEG series and the most recent HEVC. While the predictive coding schemes are able to reach high compression ratios, they are highly susceptible to the propagation of errors. The illustration of error propagation over time is shown in Figure 2.

Fig. 2. Illustration error propagation over time. Image concept taken from [8].

Because of the low error robustness of HEVC, the use of an error resilient technique is needed [9]. Among the various error control schemes of video encoders, the intra-refresh method is preferred because of its low complexity implementation and its efficiency in reducing transmission distortion [10][11].

This method can effectively mitigate the error propagation due to packet loss and improve the video quality significantly [7].

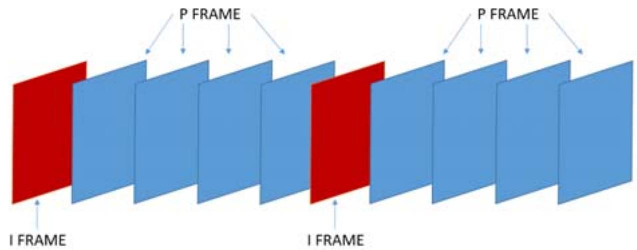


Fig. 3. Illustration of insertion of I-frames to minimize error propagation over time. Image concept taken from [8].

HEVC divides a picture, or a video frame, into coding tree blocks, or CTBs, and it is arranged in a raster order. Much like the previous standards, a sequence of CTBs can be put together to form a slice. Therefore, a picture can be divided into a different number of slices. It is very much possible to make the whole picture a slice, in which it has its specific network abstraction layer (NAL) unit [4][12].

One reason that makes video coding standards a lossy compression is the quantization parameter (QP). The QP determines the quantity of spatial details retained. The quality of the decoded video is proportional to the amount of spatial detail saved, with or without errors. A low QP ensures that most spatial details are saved albeit with a high storage requirement, while a high QP decreases the spatial details but now requires less storage space.

A. Video Sequences

Three video test sequences are used to represent a wide range of motion intensity. The Akiyo test sequence represents a video sequence with a minimal amount of motion. The female reporter is reading news only by moving her lips and eyes. The Foreman test sequence represents a video sequence with high motion content. The Foreman video contains a monologue of a man, and at the end, there is a continuous scene change. Meanwhile, the Container sequence has medium motion but high spatial texture and parallel to many object boundary edges. For our experiments, these three video sequences are used for the entire simulation. Figure 4 shows a sample frame of each video sequence.

(a) (b) (c)

Fig. 4. Sample frame of each video sequence (a) Akiyo, (b) Container, and (c) Foreman.

These video sequences are in CIF (352 × 288) resolution [3] in YUV format [9], which contains 300 frames. In a wireless video surveillance, Akiyo can represent a video with minimal movements from a static camera. For Container, it characterizes a video with a constantly moving environment from a static camera. And for Foreman, it can represent a video with a dynamic camera.

B. Configuration

The HEVC encoder is configured in such a way that a single frame is treated as a single slice. This makes the size of the slice matched to that of the maximum transmission unit (MTU) of the network in which the video is streamed [7]. Finally, the encoder sets the NAL unit to the MTU, where a NAL unit represents a packet ergo represents a video frame.

The reference software for encoding and decoding video test sequences using HEVC is called HM (HEVC Test Model) [11]. The purpose of an HM encoder is mainly to provide a common reference implementation of an HEVC encoder [13].

The configurations used for the HM encoder are as follows. The HM encoder that we used was HM version 6.0. HM provides various configuration profiles with different specifications. From the profiles provided, intra main configuration was used in encoding. All 300 frames of the video sequences were encoded with a frame rate equal to 30. A sequence of I-frame and P-frames were encoded. An I-frame is an encoded video frame in which it is independently decoded. It basically contains all the information to decode itself albeit it requires more storage requirement. On the other hand, a P-frame is an encoded video frame where a predictive coding scheme was used. It is highly dependent on information of its neighboring frame to be decoded. It contains less information making it require less storage.

For the application of the intra-refresh method, an I-frame will be encoded at specified intervals. The intra periods were set to 10, 20, 30, 40, and 50. The QP is set to typical values of 22, 27, 32, and 37 in video encoding [14].

III. WIRELESS NETWORK

The typical network topology used in wireless video surveillance is a mesh topology [1]. A mesh topology could achieve high system performance. Its benefits include a good data throughput, power efficiency, and conservation of energy.

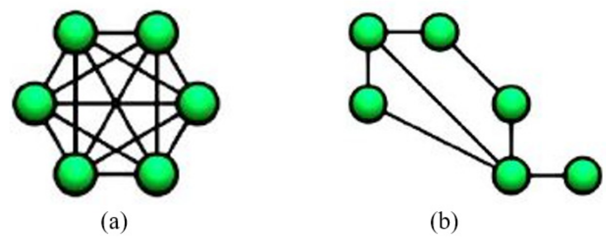


Fig. 5. Sample structure for mesh topology: (a) full mesh and (b) partial mesh. Image taken from [15].

The simulation of a wireless network was observed using NS2 as shown in Figure 6. In a mesh network, a 2-Mbps constant bit-rate traffic flows from node 0 to node 19. The distance of nodes for outdoors is around 80 to 90 m, and the distance of nodes for inside the building is around 15 to 18 m [16].

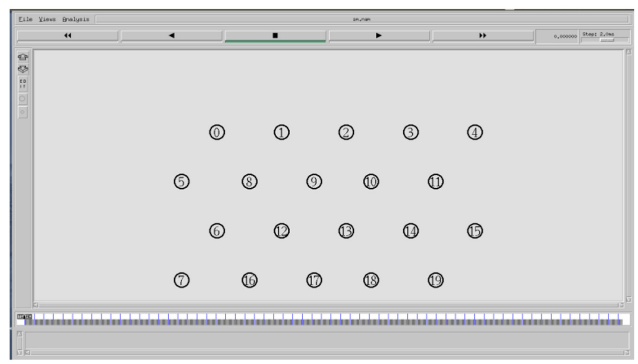


Fig. 6. Screenshot of mesh topology in NS2.

From the typical values of path loss and shadowing deviation in a wireless network environment, we used the following configurations shown in Table I [16].

TABLE I
TYPICAL VALUES OF PATH LOSS BETA (Β) AND SHADOWING DEVIATION (Σ_{DB}) FOR WIRELESS NETWORK ENVIRONMENT

ENVIRONMENT	β	σ_{db}(dB)
In building, obstructed – office, soft partition	4.0	9.6
In building obstructed – office, hard partition	4.0	7.0
Outdoor – outdoor	2.7	4.0

An NS2 trace file will be generated from the simulation wherein the error pattern of the given network can be acquired. The error pattern from the trace file was extracted via MATLAB. The error pattern obtained from the various network environment which comprises 1s (no error) and 0s (with error), as shown in Figure 7. NAL unit loss software will be used for the insertion of loss in the HEVC bit stream file produced by the encoder [17].

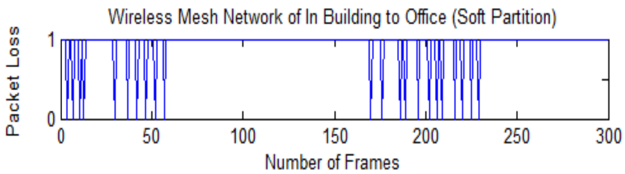


Fig. 7. Packet loss example of in building to office (soft partition) in wireless mesh network.

The HM decoder will be used in reconstructing the video with the HEVC bit stream file produced by the NAL unit loss software. In terms of error concealment, the decoder will copy the previous frame if the current frame is under error.

(a) (b)

Fig. 8. Decoded frame of Foreman (a) without error and (b) with intra period of 10 with transmission from in building to office (soft partition) in wireless mesh network.

In Figure 8 is the comparison of the sample frame from the decoded video without error and a video with error concealment introduced.

$$\mathrm{PSNR}(i,j) = 10\log_{10}\frac{\left(2^P-1\right)^2}{\mathrm{MSE}(i,j)} \tag{1}$$

$$\mathrm{MSE}(i,j) = \frac{1}{MN}\sum_{x=1}^{M}\sum_{y=1}^{N}\left[f_i(x,y)-g_j(x,y)\right]^2 \tag{2}$$

Using equations (1) and (2), the PSNRs of the decoded video sequences are measured and analyzed [18].

IV. EXPERIMENTAL RESULTS

A. Encoding Complexity

To analyze the complexity introduced in encoding the video with intra-refresh method, varying the intra period values can determine the complexity of the encoder by observing its encoding time.

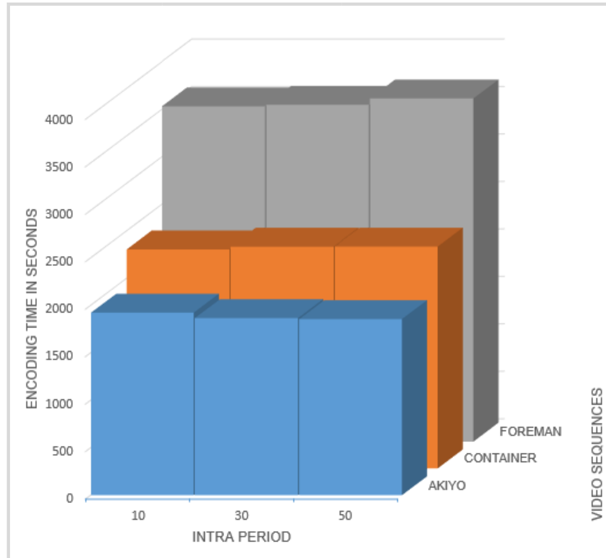


Fig. 9. Encoding time of all video sequences.

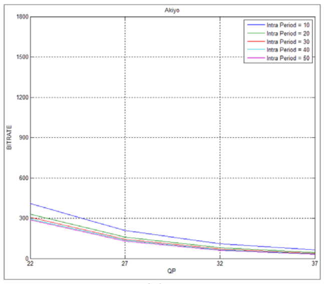
Increasing intra period means decreasing the number of intra frames encoded. In Figure 9, as the number of intra period increases (I-frames decrease), it can be observed that the encoding time increases because P-frames consume more time due to the predictive coding complexity.

In terms of motion intensity, high motion test sequences are longer to encode due to the search algorithm of the predictive coding. For the medium and high motion content video sequence, increasing intra period increases the complexity of the encoder. But for low motion content, I-frame used for prediction takes more time than generating the P-frames. The reason for this is that low motion content sequences have high temporal redundancy where the P-frames result to a relatively faster search algorithm than using an intra coded frame.

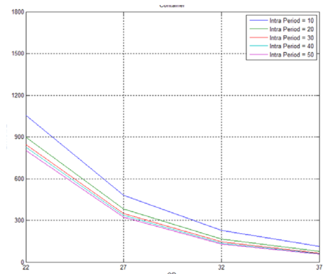
B. Bit Rate Versus QP

In wireless video surveillance, bitrate represents the required channel capacity to transmit the compressed sequences and the storage space required when storing the encoded video sequence.

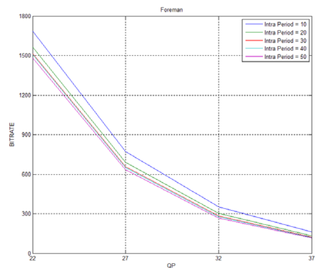
QP adjusts how much spatial detail will be preserved. Lower QP means more spatial detail is saved while the intra period defines the frequency of I-frames.



(a)



(b)



(c)

Fig. 10. Bitrate of video sequences with different QP values: (a) Akiyo, (b) Container, and (c) Foreman.

Figure 10 shows that lower QP means more spatial detail saved and thus more storage requirements. Foreman consumes more storage followed by Container then Akiyo. Videos with high motion content require more storage while low motion content requires less. Note also that higher frequency of I-frames consumes more storage.

C. QP Versus PSNR for Different Intra Periods

With no error, PSNR decreases as QP increases regardless of intra period due to the high level of quantization. In a lossy condition, the PSNR decreases but is somehow maintained to a certain value as QP increases. Figure 11 shows the QP versusPSNR for different intra periods for Foreman.

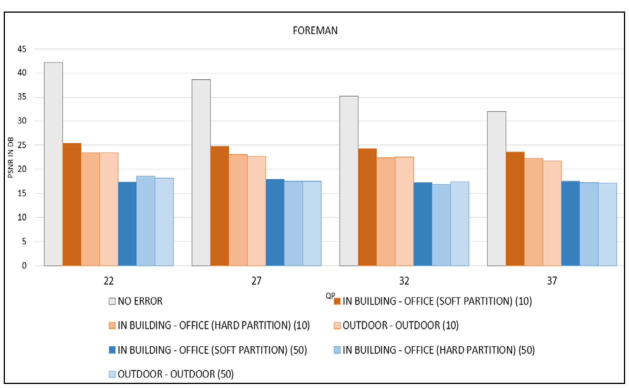


Fig. 11. Foreman with various configurations and different QP values with intra period set to 10 and 50.

Note that an intra period of 10 has less decrease in PSNR compared to an intra period of 50. Loss from quantization and network transmission causes minor variations from the results gathered for some cases in which a higher QP value produces a greater PSNR compared to lower QP values.

As quantization increases, the spatial detail decreases. So as the spatial detail decreases, the difference from the initial frame compared to the next frame is also reduced. Thus, after it underwent a lossy environment, some frames lost does not do any significant decrease in PSNR.

All tests show similar behaviors. But, in observing the PSNR of the decoded video without error, it can be perceived that the PSNR decreases as the QP decreases due to quantization.

D. Intra Period Versus PSNR for Different Wireless Configurations

Figure 12 shows the effect of quantization for the video sequences where it is observed to be greater for Foreman and less for Akiyo.

Using Foreman, there were averages of 42.11%, 69.11%, and 86.05% PSNR drop for intra periods of 10, 30, and 50, respectively. For intra periods from 10 to 30, there were averages of 18.50%, 16.21%, and 11.85% PSNR drop for Foreman, Container, and Akiyo, respectively.

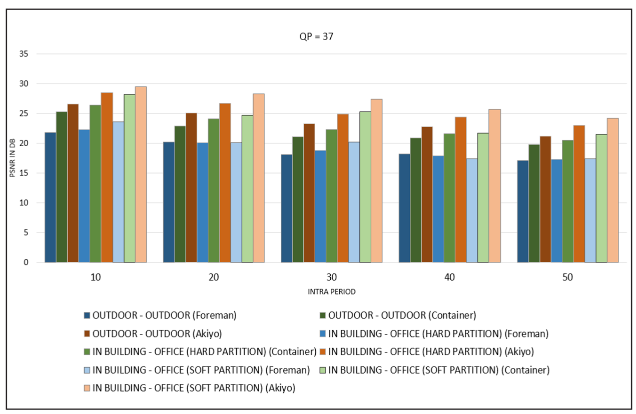


Fig. 12. Various configurations and different intra period values with QP set to 37 for all video sequences.

The decrease in PSNR when decreasing frequency of I-frame is more drastic for videos with higher motion content compared to those with less motion content.

E. Subjective Quality

At our discretion, the subjective quality shows that the effect of error is greater for high motion videos compared to low motion videos.

Fig. 13. Foreman with outdoor to outdoor configurations and intra period set to (a) 10, (b) 50, and (c) none.

In Figure 13, the tests show that a 6-dB decrease in PSNR between intra periods from 10 to 50 displays a significant effect on the quality of the video. Errors propagate greatly for an intra period of 50 compared to 10. This subjective quality supports the objective quality that there is a significant drop in PSNR for an intra period of 50 compared to 10.

V. Conclusion and Future Work

Storage requirement for encoded video sequence is affected by two major things. The first is by varying the QP. Increasing the value of QP reduces the storage space requirement at the expense of lesser spatial detail and lower video quality.

However, under lossy network, the difference in PSNR between different values of QP is at best tolerable. High motion video benefits greatly as it causes a significant decrease in storage requirement when increasing the QP. On the other hand, while low motion video also enjoys a decrease in storage requirement as QP increases, it is not as significant compared to high motion video.

The second is by varying the intra period or simply varying the frequency of I-frames in the encoded video sequence. Higher frequency of I-frames increases the storage requirement of the encoded video sequence. While it increases the storage requirements, its benefits lie in the error robustness it provides. Increasing the frequency of I-frames minimizes the effect of errors on the video quality when under a lossy network such as wireless channel.

It is safe to say that decreasing the frequency of I-frames equates to a decrease in video quality; however, there will come a point where the decrease is insignificant, if not tolerable. It is especially true for a video with low motion content compared to that with high motion content. The complexity introduced by changing the intra period is insignificant, and the complexity relies entirely on the performance of HEVC itself.

In summary, it is possible to vary the parameters of the encoder in which one of the requirements for a wireless video surveillance benefits without having to compromise the others significantly. The performance of HEVC with error resiliency techniques such as intra-refresh can be further evaluated by using the latest HM reference software with a different approach.

References

[1] Y. Ye, S. Ci, A. Katsaggelos, Y. Liu, and Y. Qian, “Wireless video surveillance: A survey,” *Access, IEEE*, vol. 1, pp. 646–660, Sept. 18, 2013.

[2] Cisco, “Why go wireless.” [Online]. Available: http://www.cisco.com/cisco/web/solutions/ small_business/resource_center/articles/work_from_anywhere/why_go_wireless/index.html. [Accessed: Apr.23,2015].

[3] G. Gualdi, A. Prati, and R. Cucchiara, “Video streaming for mobile video surveillance,” *IEEE Transactions on Multimedia*, vol. 10, no. 6, Oct. 2008.

[4] J. Tan, C. Bello, J. Fajardo, and A. dela Cruz, “Performance evaluation of HEVC over delay-prone wired networks,” in *IEEE, 2015 International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNICEM)*, pp. 1–5, Dec.9–12, 2015.

[5] S. Wenger, “Loss robustness report (AHG14),” JCT-VC Document, JCTVC-H0014, Feb. 2012.

[6] S. Oudin et al, “Block merging for quadtree-based video coding,” in *2011 IEEE International Conference on Multimedia and Expo (ICME)*, pp. 1–6, July 2011.

[7] J. Nightingale, Q. Wang, and C. Grecos, “HEVStream: A framework for streaming and evaluation of High Efficiency Video Coding (HEVC) content in loss-prone networks,”*IEEE Transactions on Consumer Electronics*, vol. 58, no. 2, May 2012.

[8] A. Vetro, J. Xin, and H. Sun, “Error resilience video transcoding for wireless communications,” *IEEE Wireless Communications*, vol. 12, no. 4, pp. 14–21, Aug. 2005.

[9] P. Seeling, M. Reisslein, and B. Kulapala, “Network performance evaluation using frame size and quality traces of single-layer and two-layer video: A tutorial,” *Communications Surveys & Tutorials, IEEE*, vol. 6, no. 3, pp. 58–78, Dec. 1, 2009.

[10] G. Correa, P. Assuncao, L. Agostini, and L. da Silva Cruz, “Performance and computational complexity assessment of high-efficiency video encoders,” *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 22, no. 12, pp. 1899–1909, 2012.

[11] Fraunhofer, Heinrich Hertz Institute, “High Encoding Video Coding (HEVC).” [Online]. Available: http://hevc.hhi.fraunhofer.de. [Accessed: Jan. 17, 2016].

[12] “HEVC test model under consideration (TMuC) revision 6.0,” JCTVC Contribution HM 6.0, March 2012.

[13] F. Bossen, B. Bross, K. Sühring, and D. Flynn, “HEVC complexity and implementation analysis” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 12, Dec.2012.

[14] F.Bossen, “Common test conditions and software reference configurations,” JCT-VC Document, JCTVC-G1200.

[15] ItrainOnline, “Wireless Basic Infrastructure Topology.” [Online]. Available: http://www.itrainonline.org/itrainonline/mmtk/wireless_en/04_Infrastructure_Topology/04_en_mmtk_wireless_basic-infrastructure-topology_slides.pdf. [Accessed: Jan. 17, 2016].

[16] K. Fall and K. Varadhan, “The NS manual,” May 2010.

[17] S. Wenger, "NAL Unit Loss Software," JCT-VC Document, JCTVCH0072, Feb. 2012.

[18] V. Pullano, A. Vanelli-Coralli, and G. Corazza, "PSNR evaluation and alignment recovery for mobile satellite video broadcasting," Advanced Satellite Multimedia Systems Conference (ASMS) and 12th Signal Processing for Space Communications Workshop (SPSC), 2012.