ERROR-RESILIENT DELIVERY OF REGION OF INTEREST VIDEO USING MULTIPLE REPRESENTATION CODING

Sourabh Khire† *Arturo Rodriguez*⋆ *Scott Robertson*† *Nikil Jayant*†

† School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA
⋆ Video Technology Group, Cisco Systems, Lawrenceville, GA

ABSTRACT

Video delivery over wireless and cellular networks can be severely impaired due to bandwidth limitations and also due to the presence of burst losses and signal-loss intervals. This paper proposes unequal protection of Region of interest (ROI) encoded videos for transmission over low-bitrate and error-prone channels. Here, a flexible and interactive ROI is introduced in a low-complexity, and standard-compliant fashion to encode the critical regions within each frame at a higher quality as compared to the background. Further, the reconstruction fidelity of the high-quality ROI is improved by unequally protecting the ROI using an error-resilient video delivery scheme known as Multiple Representation Coding (MRC). Simulation results indicate that the proposed strategy can facilitate a graceful recovery of the ROI in the presence of burst and signal losses.

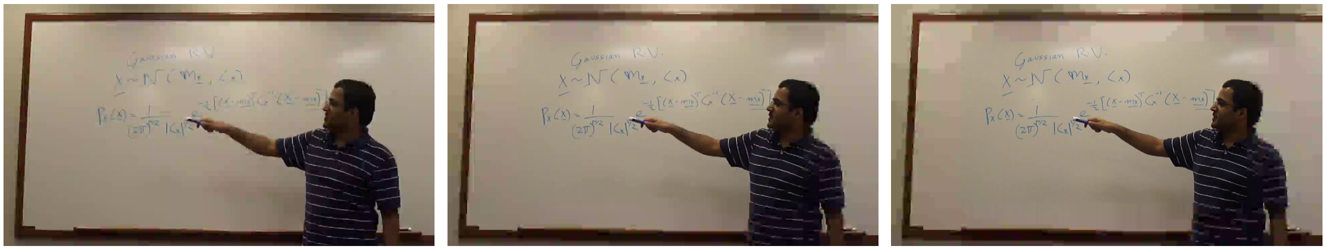
Index Terms— Region of Interest video coding, Multiple Representation Coding, Unequal protection.

1. INTRODUCTION

The relative ease in capturing high-quality videos afforded by the current generation of cheap and powerful mobile devices has led to a rapid increase in the amount of video content created and shared by consumers. Unfortunately, in the mobile multimedia communication paradigm, much of this video data is distributed over channels prone to jitter, delay and bandwidth fluctuations, and abundant packet losses. Enabling high-quality video communication for supporting video-based services in telemedicine, television and tele-education, over such unfriendly networks remains a huge technological challenge.

Region of interest (ROI) video coding is one possible solution to facilitate transmission of high-bitrate videos over bandwidth-constrained resources. By encoding the ROI with more number of bits as compared to the rest of the frame, or the background (BG), the ROI can be delivered with a higher visual quality to the end-user. ROI-based video coding has been successfully implemented for enabling applications in video surveillance [1], video telephony [2] and telemedicine [3, 4]. A recent study has also shown that low-bitrate videos encoded with a ROI support have a higher perceptual quality as compared to uniformly encoded videos [5]. Clearly, ROI-based video coding in low bitrate scenarios can artificially increase the bandwidth perceived by the video delivery application. Several approaches enabling ROI video coding have been previously proposed in literature. An iterative approach can be used to assign the highest possible compression level to the BG, and the lowest possible compression level to the ROI without exceeding the target bitrate [6]. ROI support can also be established implicitly, by blurring the BG, and then uniformly encoding the sequence with a standard encoder [3]. Several approaches employ the Flexible Macroblock Ordering (FMO) tool in the H.264 standard for enabling ROI support [7, 8]. However, FMO is known to reduce the coding efficiency, and increase the decoder complexity, and hence is seldom employed in practice. Instead of using complex (iterative), or inefficient (FMO-based) schemes, a flexible, interactive, and a low-complexity approach to establish ROI support was previously proposed by the authors [9]. This scheme will also be utilized in this paper.

Besides bandwidth limitations, mobile and wireless video access can also be hampered by the presence of burst and signal losses. On a cellular network, users may experience a complete loss of signal, for example, when riding an elevator, or driving through a tunnel. Multiple techniques such as Forward Error Correction (FEC) [10], FMO in H.264/AVC [11] and Multiple Description Coding (MDC) [12, 13] have been proposed to facilitate error-resilient video delivery over unreliable channels. However, all these approaches face limitations in the presence of burst errors or signal-loss intervals. FEC is limited by its all or nothing performance: if the loss length exceeds the correction capability, almost nothing is available to the receiver. In MDC, several independently decodable "descriptions" are generated from the source video, and are transmitted over multiple paths to the receiver [14]. The strength of the so-called multipath transport (MPT) approach lies in the assumption that spatio-temporally co-located segments of multiple descriptions are less likely to be simultaneously impaired when routed over multiple disjoint paths [15]. However the MPT approach is unsuitable for systems where it is not possible, or is inconvenient to establish multiple paths between the source and the receiver. More-



(a) QP-offset=0 (No ROI) (b) QP-offset=5 (Moderate ROI) (c) QP-offset=10 (Strong ROI)

Fig. 1: Effect of changing QP-offset value on the ROI and BG quality of the *Equation* sequence.

over, the receiver implementation in a MDC system tends to be highly complex due to the non-deterministic arrival times of the descriptions transmitted over different paths. To address these limitations, authors of this paper have recently proposed a scheme for error-resilient video delivery, known as Multiple Representation Coding (MRC) [16]. Instead of relying on path-diversity, the MRC approach temporally disperses multiple representations on a single transmitted stream to facilitate a graceful recovery from impairments caused by a burst and signal losses.

All the approaches mentioned previously attempt error-resilient delivery of the entire frame. However, in applications such as distance-learning, where the transmitted video consists of a well-defined ROI, it might be more efficient to facilitate error-free delivery of just the ROI. In [17], the picture is split into foreground and background sub-pictures. The macroblocks (MBs) in the background sub-picture are more heavily quantized as compared to those in the foreground sub-picture. Further, unequal error protection (UEP) is employed at the packet level to facilitate error-recovery of the foreground sub-picture. In [18], the frame is partitioned into slices of "low", "medium", and "high" importance using FMO, and then unequally protected using Reed Solomon codes. MDC has been used on images containing multiple ROIs encoded using the set partitioning in hierarchical trees algorithm [19]. Each ROI is then placed on a separate description and transmitted. However, only the ROIs placed on descriptions that are received error-free can be fully recovered. Given the previously described limitations of error correction codes and MDC, this paper proposes employing the MRC scheme to facilitate error-resilient delivery of the ROI in presence of burst losses or signal loss intervals.

2. REGION OF INTEREST SUPPORT

In this paper the quantization parameter (QP) of each MB in the frame is modified to establish the ROI. Since QP is inversely proportional to the bitrate, more bits can be assigned to the ROI, by decreasing the QP of all the MBs occupied by the ROI. Conversely, fewer bits can be assigned to the BG by increasing the QP of all the MBs in the BG. To establish the ROI, a user-defined, positive or negative 'QP-offset' is assigned to each MB of the frame. This QP-offset is added on top of the QP decision made by the rate-control algorithm of the encoder. Fig. 1 shows the effect of using different QP-offsets. Further flexibility is introduced in the system by allowing the ROI to occupy five different locations in the frame: top-left, top-right, bottom-left, bottom-right and center. For example, the ROI is located in the center and the bottom-right locations for Figs. 1 and 2 respectively. The quality and the location of the ROI is signaled by the user (client). A detailed description of this system described can be found in [9].

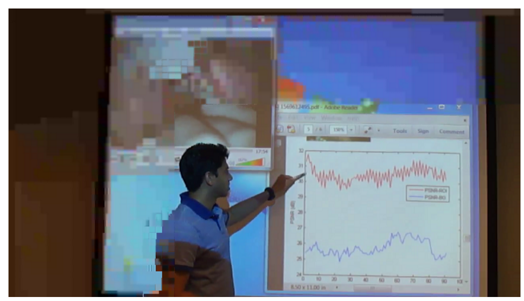
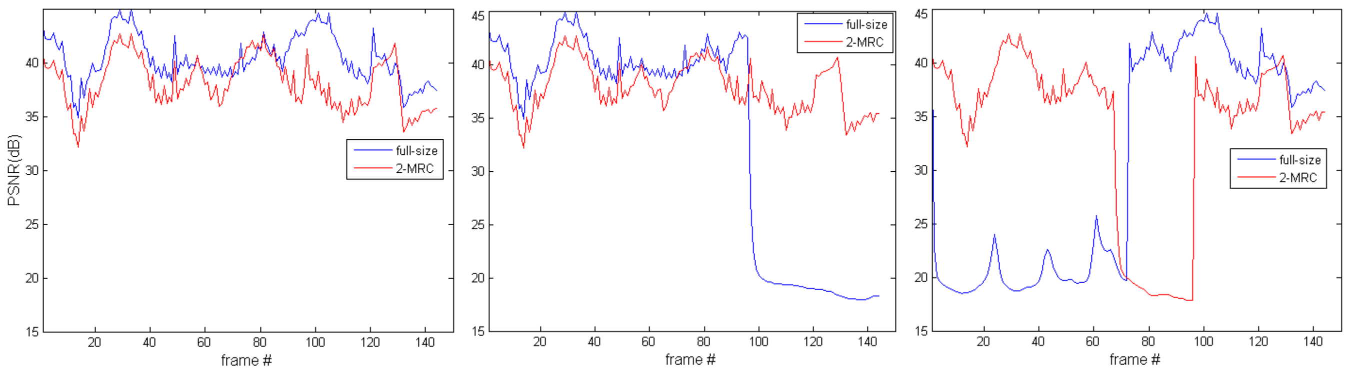


Fig. 2: Bottom-right ROI in the *Presentation* sequence.

3. MULTIPLE REPRESENTATION CODING

Multiple Representation Coding (MRC) [16] is an effective approach for error-resilient video streaming over channels prone to burst losses and signal-loss intervals (see results in Fig. 3). The MRC scheme involves creating multiple downsampled representations from the source video as seen in Fig. 4. These multiple representations are then encoded and interleaved on a single transmitted stream using the "GOP interleaver" (GOP: Group of Pictures).

The key of the MRC scheme is to temporally disperse data via the GOP interleaver. If the full-size video is impaired by a burst loss spanning multiple GOPs, then multiple seconds of the sequence are lost, and the decoder has to request data re-transmission, or rely on some naive error-concealment algorithms. However, in the MRC scheme, there are at least two downsampled representations corresponding to every second of the video. By ensuring that the same burst loss does not



(a) BLL=0, 40.38 dB, 38.12 dB

(b) BLL≈ $\frac{1}{2}$ GOP, 33.43 dB, 37.85 dB

(c) BLL≈1 GOP, 30.46 dB, 33.80 dB

Fig. 3: The *Sunflower* sequence encoded with average bitrate of *2100 kbps* for both the full-size and the 2-MRC schemes. The caption of each sub-figure indicates the burst loss length in units of GOP-length and the average PSNR of the full-size and the 2-MRC sequences.

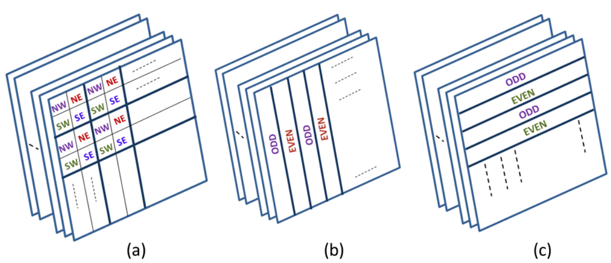


Fig. 4: Different downsampling configurations for MRC.

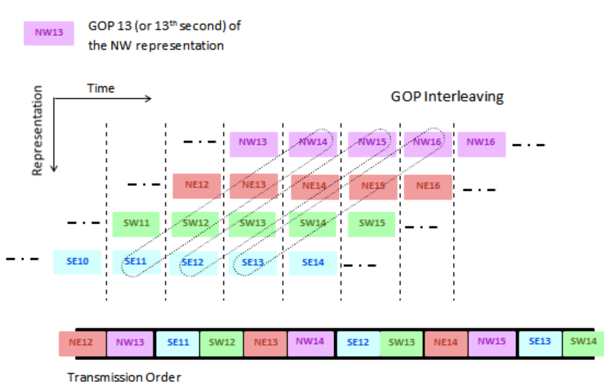


Fig. 5: GOP Interleaving with 4-MRC.

impair co-located frames of all the representations, the error-concealment at the receiver can yield higher picture fidelity. The GOP interleaver disperses the multiple representations in the 4-MRC configuration as shown in Fig. 5. On careful examination of the transmission order shown in Fig. 5, it can be seen that temporally co-located GOPs of different representations are never adjacent to each other, and hence, are less likely to be impaired by the same burst loss. The readers are encouraged to review the work by Khire *et al.* [16] for a detailed description of the MRC scheme.

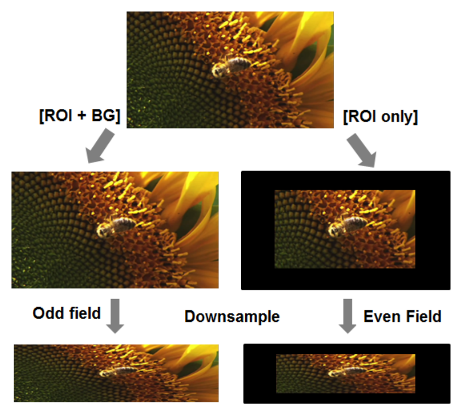


Fig. 6: Process of generating two source representations from the source video.

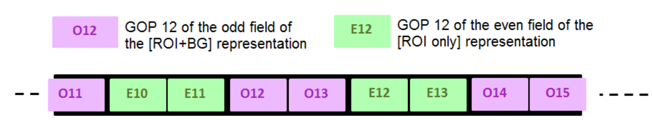
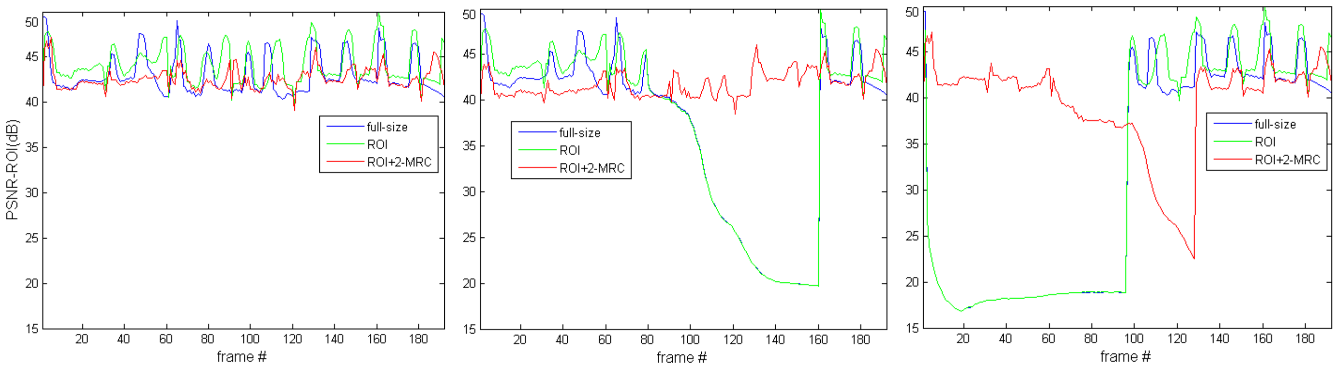


Fig. 7: GOP Interleaving for unequal protection of ROI with the 2-MRC configuration (Fig. 4c).

Fig. 3 shows the per-frame PSNR for a sequence encoded with the full-size and the 2-MRC schemes. As seen from the Fig. 3a, the MRC scheme performs poorly as compared to the full-size scheme in the absence of burst losses. This is due to the redundancy introduced by the MRC scheme, and the PSNR-loss introduced by the anti-aliasing filter applied prior to down-sampling the source frame [16]. However, as seen from Figs. 3b and 3c, as the length of the burst loss increases, the MRC scheme has a considerable advantage over the full-size scheme. The "valleys" in the PSNR-curves indicate complete frame-loss. The "valleys" for the MRC scheme are either non-existent (Fig. 3b), or occur for smaller dura-



(a) BLL=0, 42.95 dB, 44.14 dB, 42.61 dB (b) BLL≈ $\frac{3}{2}$ GOP, 36.81 dB, 37.42 dB, 41.75 dB (c) BLL≈ $\frac{5}{2}$ GOP, 31.10 dB, 31.62 dB, 39.33 dB

Fig. 8: The *Presentation* sequence encoded with average bitrate of *2700 kbps* for both the full-size, ROI, and the ROI+2MRC schemes (Fig. 7) schemes. The caption of each sub-figure indicates the burst loss length in units of GOP-length followed by average ROI-PSNR of the full-size, ROI, and the ROI+2MRC videos.

tions (Fig. 3c) because of the introduced GOP-interleaving.

4. UNEQUAL PROTECTION OF THE ROI USING MULTIPLE REPRESENTATION CODING

Video transmission over lossy channels can be improved by prioritizing error-free delivery of the ROI over the rest of the frame. This paper proposes using the MRC scheme for achieving that goal. For a ROI-encoded video using the 2-MRC configuration (ROI+2MRC), two independently decodable, downsampled representations are generated as shown in Fig. 6. These representations are then encoded using the ROI-based encoder described in Section 2, and the interleaved by the GOP-interleaver shown in Fig. 7. As seen from the Figs. 6 and 7, the downsampled ROI is encoded two times (once by each representation), and hence can be recovered with a higher fidelity as compared to the BG. Since the BG is downsampled and encoded by only one representation, the complete BG of the source frame is *not* available to the receiver even if all representations are received error free. Thus, the BG always needs to be upsampled to the full-size at the receiver. If just one representation-[ROI+BG] is received, then the full-size ROI, and the BG can be reconstructed by simple interpolation. However, if just one representation-[ROI only] is received, then only the ROI can be reconstructed using interpolation. Since no BG information is available for reconstruction, the lost BG is concealed using BG information from previously received frames. This case clearly demonstrates that the ROI is unequally protected as compared to the BG using the MRC scheme.

5. RESULTS AND DISCUSSIONS

The video database for all experiments consisted of the *Carphone*, *Miss America*, *Sunflower* and the *ParkJoy* sequences [20]. A database consisting of sequences such as *Equation* (Fig. 1) and *Presentation* (Fig. 2), representative of a distance learning application were also captured for conducting experiments. For studying the performance of the proposed scheme, encoded sequences were subject to bursts of packet losses. To introduce packet losses, bitstreams encoded using the full-size, ROI and ROI+2MRC schemes, were packed into fixed-size "packets" of 512 bytes. These packets were then dropped to simulate losses. Several packet-loss traces, each consisting of a single burst, with loss length approximately equaling zero to four GOPs of the full-size sequence, and occurring at random locations on the bitstream were generated. The "previous frame copy" method was used to conceal frames that were entirely lost. Fig. 8 shows the per-frame PSNR in the ROI (PSNR-ROI) for the *Presentation* sequence encoded with average bitrate of around *2700 kbps* (for all the three schemes). In the absence of burst losses (Fig. 8a) the ROI-encoding scheme outperforms the other two schemes as expected. However, as the loss length increases, the ROI+2MRC scheme significantly improves the fidelity of the reconstructed ROI, resulting in much graceful recovery of the ROI. Similar results were obtained using all other sequences in our database.

6. CONCLUSION

This paper presented a flexible, interactive, and standard-compliant method to introduce Region of Interest support for enabling video streaming over low-bandwidth channels. An error-resilient video delivery scheme known as Multiple Representation Coding, which employed a novel "GOP interleaver" to temporally disperse multiple downsampled representations of the source video over a single transmitted stream, was also presented in this paper. Finally, it was demonstrated that unequal protection of the ROI-encoded sequences using the MRC scheme can facilitate a graceful recovery of the ROI from burst losses and signal loss intervals.

7. REFERENCES

[1] P. Baccichet, X. Zhu, and B. Girod, "Network-aware H.264/AVC region-of-interest coding for a multi-camera wireless surveillance network," in *Proc. of Picture Coding Symposium*, April 2006.

[2] Chia-Wen Lin, Yao-Jen Chang, and Yung-Chang Chen, "Low-complexity face-assisted video coding," in *Proc. International Conference on Image Processing*, September 2000, vol. 2, pp. 207–210.

[3] N. Tsapatsoulis, C. Loizou, and C. Pattichis, "Region of interest video coding for low bit-rate transmission of carotid ultrasound videos over 3g wireless networks," in *Proc. 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, August 2007, pp. 3717 –3720.

[4] S.P. Rao, N.S. Jayant, M.E. Stachura, E. Astapova, and A. Pearson-Shaver, "Delivering diagnostic quality video over mobile wireless networks for telemedicine," *International Journal of Telemedicine and Applications*, 2009, Article ID 406753, 9 pages, 2009. doi:10.1155/2009/406753.

[5] I. Himawan, Wei Song, and D. Tjondronegoro, "Impact of region-of-interest video coding on perceived quality in mobile video," in *Proc. IEEE International Conference on Multimedia and Expo*, July 2012, pp. 79–84.

[6] D. Chai, K.N. Ngan, and A. Bouzerdoum, "Foreground/background bit allocation for region-of-interest coding," in *Proc. of International Conference on Image Processing*, September 2000, vol. 2, pp. 923–926.

[7] Y. Dhondt, P. Lambert, S. Notebaert, and R. Van de Walle, "Flexible macroblock ordering as a content adaptation tool in H.264/AVC," *Proc. of the SPIE*, vol. 6015, pp. 44–52, October 2005.

[8] S. Van Leuven, K. Van Schevensteen, T. Dams, and P. Schelkens, "An implementation of multiple region-of-interest models in H.264/AVC," in *Signal Processing for Image Enhancement and Multimedia Processing*, vol. 31 of *Multimedia Systems and Applications Series*, pp. 215–225. Springer US, 2008.

[9] S. Khire, S. Robertson, N. Jayant, E.A. Wood, M.E. Stachura, and Col. T. Goksel, "Region-of-interest video coding for enabling surgical telementoring in low-bandwidth scenarios," in *Proc. 2012 Military Communications Conference*, October 2012, pp. 552–557.

[10] M. Johanson, "Adaptive forward error correction for real-time internet video," in *Proc. the 13th Packet Video Workshop*, 2003.

[11] S. Kumar, L. Xu, M. Mandal, and S. Panchanathan, "Error resiliency schemes in H.264/AVC standard," *Journal of Visual Communication and Image Representation*, vol. 17, no. 2, pp. 425–450, April 2006.

[12] V. K. Goyal, "Multiple description coding: Compression meets the network," *IEEE Signal Processing Magazine*, vol. 18, no. 5, pp. 74–93, September 2001.

[13] Y. Wang, A.R. Reibman, and S. Lin, "Multiple description coding for video delivery," in *Proc. of the IEEE*, January 2005, vol. 93, pp. 57–70.

[14] J. G. Apostolopoulos, "Reliable video communication over lossy packet networks using multiple state encoding and path diversity," in *Proc. Visual Communication and Image Processing*, January 2001, pp. 392–409.

[15] Shiwen Mao, Shunan Lin, S.S. Panwar, Yao Wang, and E. Celebi, "Video transport over ad hoc networks: multistream coding with multipath transport," *Selected Areas in Communications, IEEE Journal on*, vol. 21, no. 10, pp. 1721–1737, December 2003.

[16] S. Khire, N. Jayant, and A.A. Rodriguez, "Multiple representation coding for error-resilient video delivery," in *Proc. IEEE International Conference on Consumer Electronics - Berlin*, September 2012, pp. 6–10.

[17] M.M. Hannuksela, Y.-K. Wang, and M. Gabbouj, "Sub-picture: Roi coding and unequal error protection," in *Proc. International Conference on Image Processing*, June 2002, vol. 3, pp. 537–540.

[18] N. Thomos, S. Argyropoulos, N.V. Boulgouris, and M.G. Strintzis, "Robust transmission of h.264/avc video using adaptive slice grouping and unequal error protection," in *Proc. IEEE International Conference on Multimedia and Expo*, July 2006, pp. 593–596.

[19] M. Nystrom, J.D. Gibson, and J.B. Anderson, "Multiple description image coding using regions of interest," in *Conference Record of the Forty-First Asilomar Conference on Signals, Systems and Computers*, November 2007, pp. 925–929.

[20] T. K. Tan, G. Sullivan, and T. Wedi, "Recommended simulation conditions for coding efficiency experiments," ITU-T SC16/Q6, 36th VCEG Meeting, San Diego, USA, 8 - 10 July 2008, Document VCEG-AJ10.