Multiple Description Coding for Scalable Video Coding with Redundant Slice

Chunyu lin [1], Yao Zhao [1], Huihui Bai [1], Kwong Huang Goh [2]
Institute of Information Science, Beijing Jiaotong University[1]
Beijing Key Laboratory of Advanced Information Science and Network, China[1]
Signal Processing Department, Institute for Infocomm Research[2]
Agency for Science, Technology and Research (A*STAR), Singapore
Email: {cylin@bjtu.edu.cn, yzhao@bjtu.edu.cn, hhbai@bjtu.edu.cn, khgoh@i2r.a-star.edu.sg}

***Abstract*—Scalable video coding(SVC) is designed to provide adaptable capability for heterogeneous network structures with its scalabilities. Due to the introducing of inter-layer prediction, SVC coded bitstream is much vulnerable to the channel error. To solve this problem, a lot of schemes are proposed. However, they are either not compatible with the standard or just focus on one of scalabilities in SVC. In this paper, we propose a multiple description scheme that is compatible with SVC standard and supports all the scalabilities. The proposed scheme is based on the redundant slice in the SVC standard, in which each layer is composed of primary slice and redundant slice by interleaving. To deal with the mismatch error and propagated error, the quantization step of redundant slice and the mode of each macroblock is tuned by considering the prediction path in the same layer and between different layers, as well as the channel packet loss rate. The experimental results show that the proposed scheme is efficient for the packet loss channel.**

***Index Terms*—Scalable video coding, Multiple description coding, Redundant slice**

I. Introduction

Scalable video coding(SVC) is designed to meet different preferences and requirements for different kinds of end-users with one single bit-stream. In practice, appropriate sub-bit-streams can be extracted from a single bit-stream for the temporal, spatial and Signal-to-Noise Ratio (SNR) scalabilities. To provide the scalabilities, a video is coded into more than one layer, such as the base layer and enhancement layers. The enhancement layers can improve the performance of base layer with respect to frame rate, spatial resolution and/or video quality. With SVC, different receiving devices with their different decoding abilities and screen sizes can be supported, from the range of high definition(HD) TV to laptop and mobile phone. For the same kind of receiving devices, the connected ways may be different, such as cable, wireless local area network and 3G. Even for exactly the same device with the same connection, the bandwidth could be fluctuated. Hence, SVC is required to provide different salabilities for such heterogeneity.

It can also be noticed that error protection is very necessary for SVC bit-stream. On one hand, the current network is the best-effort and heterogeneous network that results in the packets delay or packet loss. On the other hand, one packet loss will not only affect the performance of its corresponding frame but also make the distortion propagate for a certain long way due to the motion prediction existing in one layer and between different layers. There are some error resilient tools existed in SVC, such as intra macroblock(MB) refresh, flexible macroblock order(FMO) and redundant slice [1]. However, these tools are only effective when the packet loss of channel is very low. A lot of schemes are proposed to protect the SVC bit-stream. The first kind of schemes are based on unequal error protection [2], [3], in which different layers are protected according to their importance. In [3], different numbers of Reed Solomon(RS) code are inserted for different layers. If the protection applies on the frame level, there is not enough packets for the spatial base layer, RS code cannot play its important role with less information packet. If RS code applies on more than one frames, there will be extra delay due to inter-layer prediction. In addition, the schemes based on UEP is less efficient with the burst packet loss case.

The second kind of schemes are based on multiple description coding(MDC) [4],[5],[6]. MDC is an effective scheme that can provide multimedia transmitted over non-prioritized networks. It can effectively combat packet loss without any retransmission thus satisfying the demand of real-time services and relieving the network congestion. In MDC, the same source signal will be encoded into more than one bit-streams that are called descriptions. When only one description is received, the signal can be constructed with acceptable quality. When more descriptions are received, the reconstructed quality will be much better. Generally, the balance case will be considered, that is, each description has the same bit-rate and distortion. Due to the prediction between different layers, MDC for SVC is not easy to implement or less efficient. In addition, most of them cannot be compatible with SVC standard or only supports fewer scalabilities. In [4] and [5], only enhancement layer is protected with MDC, while in [6], only the residual data in the base layer are formed into MDC. In this paper, we propose an efficient MDC scheme for all the SVC layer that can protect the bit-stream according to its importance and be compatible with the standard.

II. Proposed scheme

In our scheme, the spatial and temporal salabilities will be considered. Because in MDC, when more descriptions are received, the better quality can be provided, this is also one

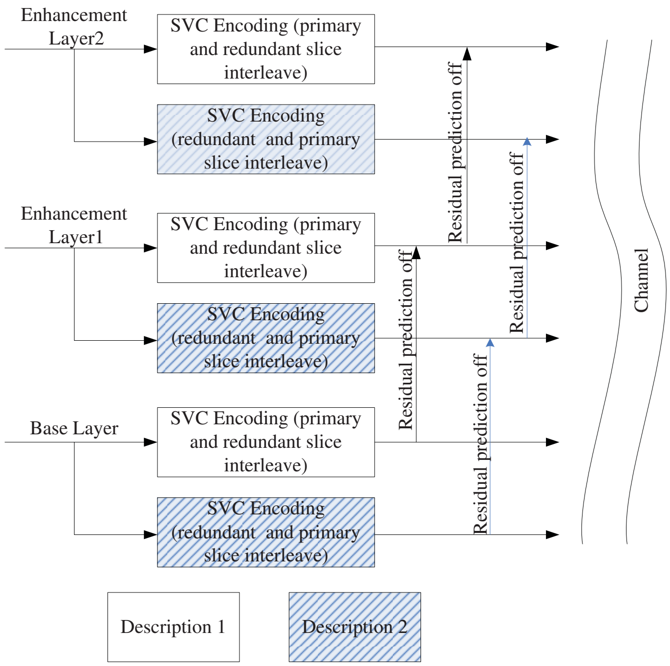

Fig. 1. The diagram of the proposed scheme.

way to support quality scalability. Our scheme is shown in Fig. 1, in which MDC scheme with three layers is taken for an example.

A. Multiple description coding for each layer based on redundant slice

Similar to the idea of redundant slice based MDC scheme(RS-MDC)[7] and the MDC scheme with redundant macroblock [8], we use the redundant slice to form the multiple description for SVC. Each description is composed of half primary slice and half redundant slice for each layer. Generally, the redundant slice has larger quantization step than its primary version, so there will be mismatch error if the primary slice is lost and the redundant version is used instead. Hence, an effective scheme for the quantization step of redundant slice should be designed to reduce the mismatch. In addition, except for intra and inter mode, there are inter-layer mode in SVC coding. The inter-layer prediction further includes inter-layer intra texture prediction mode, inter-layer motion prediction mode and inter-layer residual prediction mode. We will describe the proposed redundant quantization method considering these modes. The quantization step for the redundant slice is determined from a simplified rule in [7]. But, here the inter-layer prediction is taken into consideration.

$$QP_{r,f} = QP_p - 3log(p\phi_f) \tag{1}$$

Where QP_r and QP_p denotes the redundant and primary quantization step respectively. Parameter p represents the channel packet loss rate (PLR). Subscript f represents the current frame number. ϕ is a function related with error propagation shown as following

$$\phi_f = (1 - e^{-\alpha(N+W-f+1)})/(1 - e^{-\alpha}) \tag{2}$$

Where N and W represents the weight that are determined by the prediction times and ratios in the same layer and inter-layer respectively. The parameter α models the extent of error propagation, which will be fixed as 0.4 from empirical value. The QP_r for redundant slice will be kept in the range of [0 51]. In addition, QP_r will not be smaller than its corresponding QP_p. For a specified slice in the current frame f, formula (1) means that the more a slice is referenced by others and the larger PLR is, the smaller the QP should be assigned to provide more protection. Here, the reference considers both the same layer and inter-layer. The frames in the front of the same temporal layer or from the lower spatial layer will get more protection.

For the inter-layer intra prediction mode, the current MB will use the reconstructed lower layer MB as the prediction in the current layer. This mode only exists when lower layer uses intra mode. However, the intra mode is mostly selected in the intra frame, in which QP_r of redundant intra slice is almost similar to its primary version according to (1) because intra frame are in the front of a GOP. Hence, the mismatch error for this inter-layer mode is mitigated a lot. For the inter-layer motion prediction mode, all the redundant MBs will be forced to have the same motion information with their corresponding primary MB. There will be no mismatch error even if the redundant slice is employed. The inter-layer residual prediction mode is a big problem due to the different quantization steps of primary and redundant slice. To solve this problem, this mode will be disabled when the redundant QP and primary QP have a big gap, that is, $QP_r - QP_p > 6$.

In conclusion, the mismatch error will be largely mitigated with the above rules. It should be noted that our scheme will be inclined to protect low layer and the slices in the front of a GOP more. The whole scheme can provide a certain good quality for the lower layers on most cases.

B. The coding mode selection

When encoding a MB in each layer, it will decide to use inter-layer prediction or the other prediction modes in the same layer. Instinctively, if a slice in the base layer is protected heavily, the corresponding MB in the enhancement layer should use the base layer as inter-layer prediction to stop the error propagation error. In practice, the mode selection for each MB in SVC is based on its rate-distortion(RD) cost. For the packet loss case, the PLR is considered in our MDC scheme. The RD cost function is still as

$$L = D + \lambda \times R \tag{3}$$

However, here D includes both the distortion from coding and the possible distortion due to packet loss. The mode of the MB in the base layer will not be affected by other layer, therefore the expected distortion can be estimated as

$$\begin{aligned}\widetilde{D}(f,n,o) &= (1-p)(D^c(f,n,o)+D^{ep}(f,n,o)) \\ &+ p(1-p)(D^r(f,n,o)+D^{ep}(f,n,o)) \\ &+ p^2 D^{ec}(f,n)\end{aligned} \quad (4)$$

where f, n and o denotes fth frame, nth macroblock with coding mode o. The superscript c, ep and ec represent the coding distortion, error propagated distortion and error concealment distortion respectively. D^c and D^r are the primary coding distortion and the redundant coding distortion. Finally, we assume that the PLR is p that applies to all the layers. Hence, at probability $1-p$, the source is reconstructed with primary encoded distortion and error propagated distortion. At probability $p(1-p)$, it is reconstructed with redundant encoded distortion and error propagated distortion. Obviously, the coding distortion D^c, D^r and D^{ec} are easy to be calculated. In addition, D^{ec} is independent of coding mode. For the error propagated distortion D^{ep}, it is generated by using the error reconstructed pixels as reference. Since H.264/SVC supports motion compensation with block sizes ranging from 4×4 to 16×16, the minimum size 4×4 is selected as the basic unit, which means each MB is composed of 16 basic units. If a MB is composed of block with other size (different from 4×4), then the block can be divided into basic units with the same motion information. In conclusion, the error propagated distortion for each MB is composed of 16 basic units and each unit will use 4 blocks as reference at most, which can be represented by the following formula.

$$\begin{aligned}D^{ep}(f,n,o) &= \sum_{i=1}^{16} D_i^{ep}(f,n,i,o) \\ &= \sum_{i=1}^{16}\sum_{k=1}^{4} w_k D_i^{ep}(f_k,n_k,i_k,o_k)\end{aligned} \quad (5)$$

Where w_i denotes the weight of each reference block that is proportional to the area being referenced. D^{ep} can be obtained during the encoding of each picture and is calculated by iteration. For convenience, assume the current block will only use its previous frame and corresponding position as prediction, then D^{ep} can be calculated during its encoding process as

$$\begin{aligned}D^{ep}(f,n,i) &= (1-p)D^{ep}(f-1,n,i)+p(1-p)(D_{pr}^m(f,n,i) \\ &+ D^{ep}(f-1,n,i))+p^2(D_{p\_ec}^m(f,n,i)+D_{ec\_ep}^m(f,n,i))\end{aligned} \quad (6)$$

Where D_{pr}^m represents the mismatch error between primary and redundant coding. $D_{p\_ec}^m$ denotes the mismatch between the primary decoding and error concealment reconstruction. $D_{ec\_ep}^m$ is caused by the error concealment and the possible error propagation distortion in the reference frame that used for error concealment.

For the intra coded block, the previous propagated error will be reset once. Formula (6) becomes

$$\begin{aligned}D^{ep}(f,n,i) &= p(1-p)D_{pr}^m(f,n,i)) \\ &+ p^2(D_{p\_ec}^m(f,n,i)+D_{ec\_ep}^m(f,n,i))\end{aligned} \quad (7)$$

For the inter-layer prediction mode, the estimated cost should consider the propagated distortion from the lower layer. If inter-layer intra prediction is selected, the calculation of D^{ep} will use its lower layer instead of its previous frame. If inter-layer motion prediction mode is used, there will be no mismatch error for the predicted motion because the primary and redundant decoding share the same motion information in the reference layer. However, the compensated blocks is still in the same layer. Hence, unless the reference blocks in the lower layer is received, the concealment will be employed. Formula (6) becomes

$$\begin{aligned}D^{ep}(f,n,i) &= (1-p^2)(1-p)D^{ep}(f-1,n,i) \\ &+ (1-p^2)p(1-p)(D_{pr}^m(f,n,i)+D^{ep}(f-1,n,i)) \\ &+ (2p^2-p^4)(D_{p\_ec}^m(f,n,i)+D_{ec\_ep}^m(f,n,i))\end{aligned} \quad (8)$$

Notice only motion information comes from lower layer, the propagated distortion is still from the previous frame $f-1$. For the inter-layer residual prediction, the first part of error propagated distortion still comes from formula (6). The second part results from the loss of the residual in the lower layer. On one hand, we protect the lower layer with smaller QP gap between primary and redundant slice. On the other hand, the residual prediction mode is disabled conditionally. Hence, the second part will not be considered due to its complex expression and less affection.

III. EXPERIMENTAL RESULTS

In this section, the proposed scheme is tested on different channel case with different PLR. In the simulations, two 4CIF sequences *Crew* and *City* are selected. Three spatial layers are encoded, which are QCIF, CIF and 4CIF resolutions. The frame rate for each layer is 15 fps, 30 fps and 60 fps. The GOP size is 16 for all the layers. Only I and P frames are used in our simulation. The quantization parameters(QPs) range from 22 to 38 to cover certain RD interval and the *ExplicitQPCascading* is used as {-2 1 2 3 4 5} for different temporal layers. Each frame is coded with 3 primary slices and 3 redundant slices. The QPs for the redundant slices are determined from (1) and the coding mode is selected according to the RD cost.

Fig. 2, Fig. 3 and Fig. 4 show the results of side and central performance of the proposed scheme. For comparison, the results of single description case are also included. Here, the central and side results are obtained with PLR=10%. It should be noticed that the side description is formed by interleaving the primary and redundant slice. Hence, the results of side description means only 50% packets are received. The 4CIF case is reconstructed with fps=60, i.e. full temporal layer. Because the highest temporal layer will not be referenced by any frame, the redundant QPs obtained from (1) will be much larger than that of its primary version, which results in large mismatch for the results of 4CIF. However, if the 4CIF is reconstructed with lower temporal layer, the results can be much better. This can be observed from the CIF and QCIF case that are reconstructed with fps=30 and fps=15 respectively. Furthermore, the inserted redundancy is also affected by the

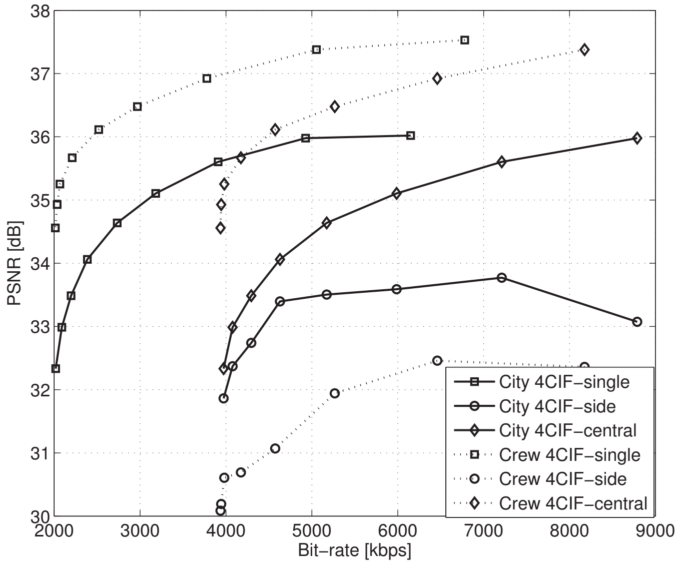

Fig. 2. Side and central performance comparison(4CIF).

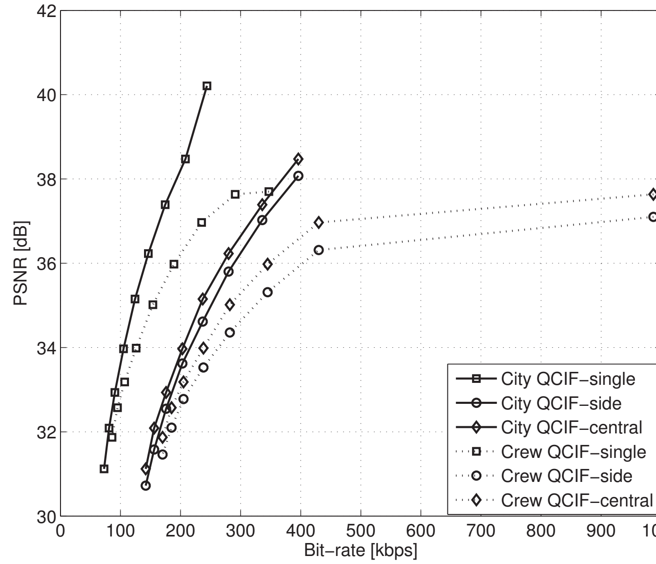

Fig. 4. Side and central performance comparison(QCIF).

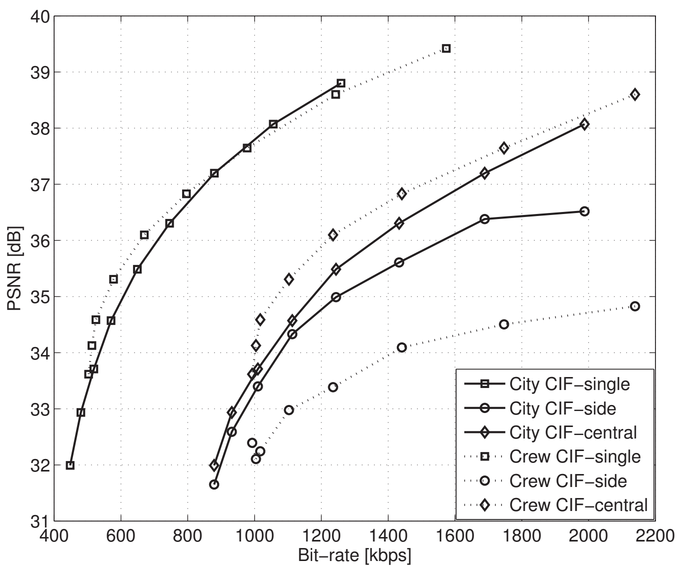

Fig. 3. Side and central performance comparison(CIF).

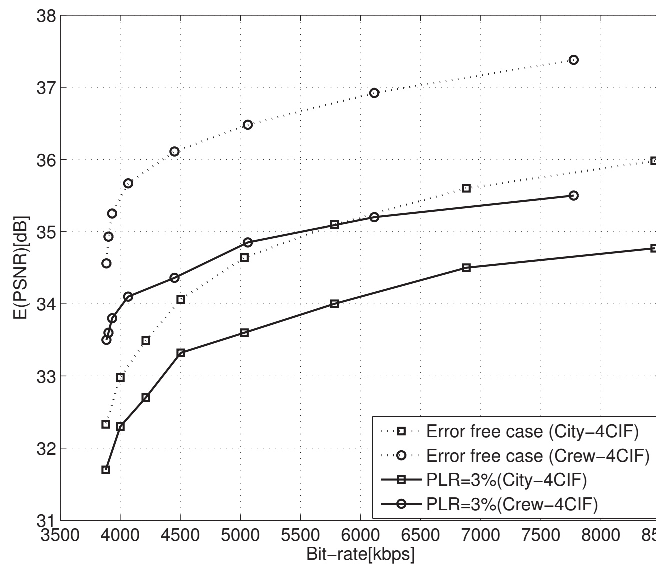

Fig. 5. The performance with channel packet loss(4CIF)

QPs due to the cascading structure. For example, if the primary QP=38, the temporal layer 4 will be 42. It is obvious that the mismatch will be larger with QP group (QP_r=25, QP_p=4) than the QP group (QP_r=43, QP_p=42). Finally, it can be seen that the ratio of inserted redundancy is larger at lower temporal layer due to (1). This is reasonable because lower temporal layer will be referenced more.

Fig. 5, Fig. 6 and Fig. 7 provide the results of the proposed scheme at PLR=3%. For comparison, the performance of the whole scheme without any packet loss is also shown in the figures. It can be seen that the mismatch between the error case and no error case is much smaller at lower temporal layer, which is due to the higher protection for the lower layer. The mismatch is a little larger at higher layer, however, it still provides acceptable quality. In fact, user can select lower frame rate to get a better quality. The whole scheme provides a robust bitstream for different requirement.

IV. Conclusion

A multiple description coding for SVC scheme based on redundant slice is proposed in this paper. In the scheme, the quantization step of redundant slice is tuned according to the motion prediction path and channel packet loss rate. To reduce the mismatch error further, the motion prediction information of the redundant macroblock is forced to be the same as that of its primary version. The mode of each macroblock is decided by considering the error propagation path and channel packet loss rate in the MDC framework. Simulation results show the

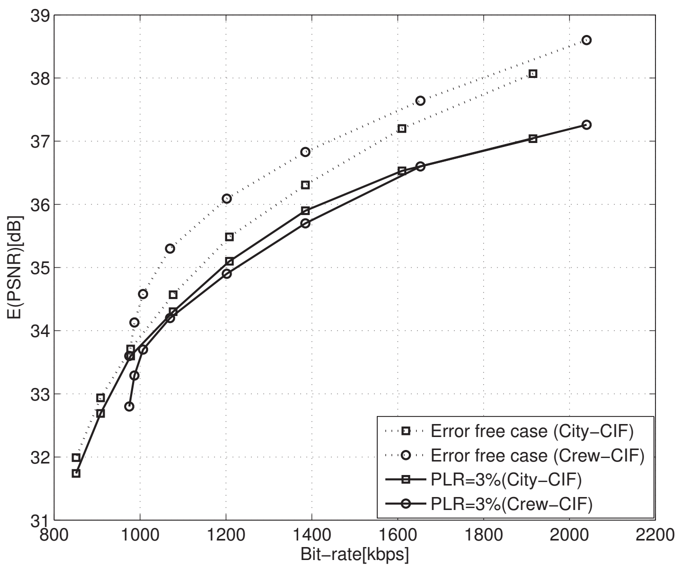

Fig. 6. The performance with channel packet loss(CIF)

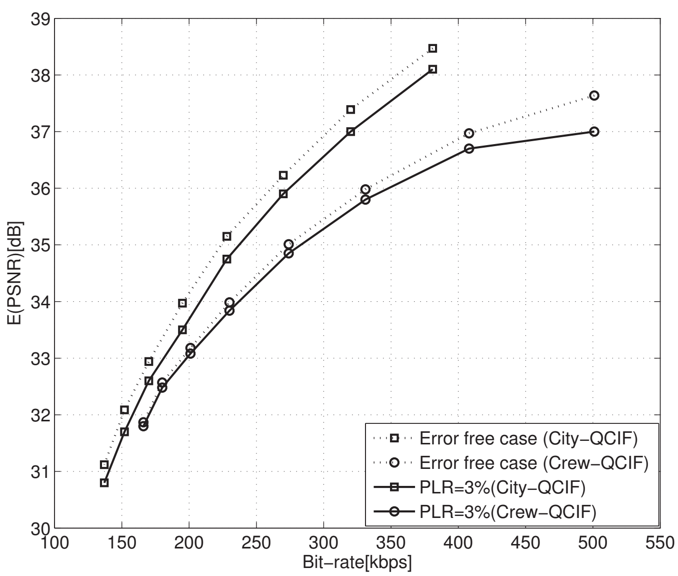

Fig. 7. The performance with channel packet loss(QCIF)

efficiency of the proposed scheme.

Acknowledgment

This work was supported National Basic Research Program of China (Grant No. 2012CB316400), National Science Foundation of China (Grant No. 61025013), Sino-Singapore JRP (Grant No. 2010DFA11010), National Natural Science Foundation of China (Grant No. 60972085).

References

[1] Y. Guo, Y. Chen, Y.-K. Wang, H. Li, M. Hannuksela, and M. Gabbouj, "Error Resilient Coding and Error Concealment in Scalable Video Coding," *IEEE Transactions on Circuits and Systems for Video Technology*, 2009.

[2] A. Naghdinezhad, A. Farmahini-Farahani, M. Hashemi, and O. Fatemi, "An Adaptive Unequal Error Protection Method for Error Resilient Scalable Video Coding Using Particle Swarm," in *IEEE International Conference on Signal Processing and Communications*, Nov. 2007, pp. 396 –399.

[3] A. Naghdinezhad, M. R. Hashemi, and O. Fatemi, "Unequal Error Protection for the Scalable Extension of H.264/AVC Using Genetic Algorithm," in *Advances in Computer Science and Engineering*. Springer Berlin Heidelberg, 2009, vol. 6, pp. 194–202.

[4] D. Alfonso, R. Bernardini, L. Celetto, R. Rinaldo, and P. Zontone, "Multiple Description for Robust Scalable Video Coding," in *16th IEEE International Conference on Image Processing (ICIP)*, Nov. 2009, pp. 905 –908.

[5] H. Mansour, P. Nasiopoulos, and V. Leung, "An Efficient Multiple Description Coding Scheme for the Scalable Extension of H.264/AVC (SVC)," in *IEEE International Symposium on Signal Processing and Information Technology,*, Aug. 2006.

[6] Z. Zhao, J. Ostermann, and H. Chen, "Low Complexity Multiple Description Coding for the Scalable Extension of H.264/AVC," in *Picture Coding Symposium, 2009. PCS 2009*, may 2009, pp. 1–4.

[7] T. Tillo, M. Grangetto, and G. Olmo, "Redundant Slice Optimal Allocation for H.264 Multiple Description Coding," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 18, no. 1, pp. 59 –70, Jan. 2008.

[8] C. Lin, T. Tillo, Y. Zhao, and B. Jeon, "Multiple Description Coding for H.264/AVC With Redundancy Allocation at Macro Block Level," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 21, no. 5, pp. 589 –600, May 2011.