

ESTIMATION OF THE UTILITIES OF THE NAL UNITS IN H.264/AVC SCALABLE VIDEO BITSTREAMS

Bin Zhang, Mathias Wien and Jens-Rainer Ohm

Institute of Communications Engineering, RWTH Aachen University
RWTH Aachen, D-52056, Aachen, Germany

phone: + (49) 241 8027671, fax: + (49) 241 8022196, email: {bzhang, wien, ohm}@ient.rwth-aachen.de

ABSTRACT

In our previous work [1,2], we proposed a framework to provide optimized unequal error protection (UEP) with Reed-Solomon codes to the H.264/AVC scalable video bitstreams to get robust video streaming for uni-/multicast applications. In that framework, we need to measure the real utilities of the network abstraction layer (NAL) units in a separate process before the transmission of the bitstream. This process requires, however, multiple decoding processes of the whole bitstream, which makes the framework not directly applicable to a live streaming system, where the video content is captured and must be transmitted within a constrained delay. In this work, we present a simple but very effective algorithm to estimate the utilities of the NAL units on the enhancement layer. With the estimation algorithm, we don't need additional decoding processes, except those with the complete base layer and enhancement layer, which are integrated in the encoder. We performed intensive tests with the estimation algorithm for different sequences in the UEP framework. Our simulation results show that good UEP performance can be achieved with the estimated utilities. The loss of performance using the estimated utilities against the real utilities is very low, if the estimation period is appropriately selected.

1. INTRODUCTION

With the advent of the 3G and 4G mobile networks, real-time video streaming applications are becoming more and more important for the mobile users. The challenge to provide a reliable video streaming service through the wireless channel is to combat the loss of video packets over the channel, due to channel fading, connection outage, network congestions etc. Packet loss in the streaming video may cause serious visual annoyance to the users. In our previous work [1, 2], we proposed a framework using H.264/AVC scalable video coding (SVC) [3] and unequal error protection (UEP) to achieve an optimal graceful degradation of the video quality at the side of receivers for both unicast and multicast applications. One of the key components in the framework is the analyzer of the utilities and costs of the network abstraction layer (NAL) units. The utility is a quantitative indicator of the importance of one NAL unit. It indicates the contribution of the NAL unit to the overall quality of the sequence. The cost defined in the framework is the size of the NAL unit. Knowing the utilities and costs of all NAL units, we can derive the optimal Reed-Solomon (RS) code for each NAL unit adapted to the instantaneous channel conditions at the receiver side. In our previous work, we demonstrated that the optimized error protection can enhance the robustness for unicast and multicast video streaming.

The utility analyzer analyzes the utilities and costs of all NAL units in a bitstream by means of multiple decoding processes, which must be performed separately after the regular encoding process. This makes the framework not directly applicable to live streaming applications, where the video is captured and transmitted within a constrained delay. Our idea to solve this problem is to find an al-

gorithm to estimate the utilities of the enhancement layer (EL) NAL units based on the real utilities of the base layer (BL) NAL units. In this way, we can reduce the required decoding processes to those with the complete quality layers, which are performed in the regular encoding process. Furthermore, the estimation of the utilities can be integrated into the encoder, which effectively extends the applicability of our UEP framework to live streaming applications, such as video conferencing. The complexity of the estimation algorithm should be as low as possible to support real-time applications, since the delay induced by the estimation method is strictly constrained by the schedule of playing out the video.

Some algorithms which estimate the picture distortion at the decoder side have been proposed in the literature. In [4], a pixel-based algorithm called ROPE was proposed for the H.263+ codec. With the ROPE algorithm, the first and second moments of the pixel values, depending on the packet loss rate, are recursively calculated to determine the picture distortion. Similar to the ROPE algorithm, in [5], another pixel-based was developed for the H.264 SVC codec. Since both of the algorithms are pixel-based and recursive, the complexity constraint in our application scenario can be hardly fulfilled. Moreover, the picture distortion derived with both algorithms is dependent on the channel conditions, therefore, cannot be used as a direct indicator of the importance of the NAL unit.

The rest of the paper is organized as follows. In Section 2, we briefly introduce the UEP problem which we addressed in our previous work. In Section 3, we discuss how to exactly measure the utilities. In Section 4, we discuss the algorithm to estimate the utilities of the BL NAL units. We show some simulation results in Section 5. The paper is concluded with Section 6.

2. UNEQUAL ERROR PROTECTION FOR H.264/AVC SVC

In this section, we briefly discuss the proposed UEP framework. The SVC streams are encoded with one base layer and one medium grain scalable (MGS) [3] enhancement layer. Our investigation in previous work demonstrated that the two-layer setting is a good tradeoff between the reduced compression efficiency and the gained robustness for graceful degradation. The priority encoding transmission (PET) [6] scheme is applied to provide UEP. As in Figure 1, all NAL units belonging to the same group of pictures (GOP) are first collected in a PET matrix. Then parity symbols generated with RS codes are applied row-wise to the video source symbols. Finally, the transmission packets are generated from the columns of the PET matrix. The NAL units are protected with the optimized RS codes of different strength. Generally, the protection strength of the RS codes decreases from the BL to the EL NAL units. Within each quality layer, stronger RS codes are assigned to NAL units on lower temporal level. With the PET scheme, it is guaranteed that the NAL units belonging to one frame referenced by other NAL units will never get lost prior to those NAL units. Moreover, the hierarchical dependency in SVC GOP is well preserved among the received NAL units

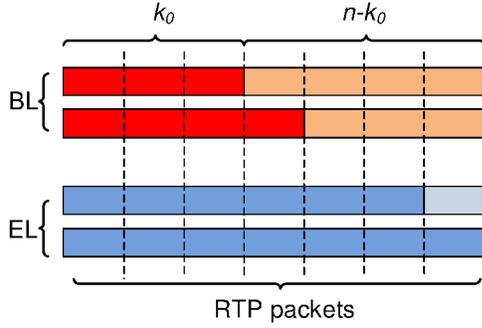


Fig. 1. The priority encoding transmission scheme.

at the receiver side. This prevents severe drift effect due to the loss of frames within a GOP.

The optimal RS codes for the NAL units in one GOP can be determined by solving the following optimization problem.

$$\bar{\mathbf{t}} = \arg \max_{\mathbf{t} \in \mathbf{T}} \sum_{q=1}^Q u_q p_{t_q}, \quad (1)$$

$$\text{subject to : } \sum_{q=1}^Q c_q r_{t_q} \leq C_{\max}, \quad (2)$$

$$1 \leq t_Q \leq t_{Q-1} \leq \dots \leq t_1 \leq n+1. \quad (3)$$

The NAL units in the GOP are indexed increasingly with increasing quality layer IDs and temporal level IDs. The GOP contains a total number of Q NAL units. u_q and c_q denote the utility and cost of the q^{th} NAL unit in the GOP. p_{t_q} and r_{t_q} are the probability of successfully transmitting the NAL unit and the redundancy factor with the specified RS code indexed by t_q .

We maximize the expected utility on the GOP basis by selecting a set of optimized RS codes denoted as $\mathbf{t} = [t_1, \dots, t_Q]$. The RS code index t_q ranges from 1 to $n+1$ in the order of increasing protection strength. $t_q = 1$ indicates the discarding of one NAL unit at the server side. $t_q \geq 2$ indicates the usage of the RS code $(n - t_q + 2, n)$, where n is the total number of the symbols including parity symbols and $n - t_q + 2$ is the number of source symbols.

C_{\max} is the total cost budget of the whole GOP for source coding and channel coding. Two constraints must be fulfilled: 1) The total cost of the GOP should not exceed the specified cost budget. 2) The strength of the selected RS codes should decrease from BL to EL NAL units and from the lowest temporal level to the highest one according to the dependency in the hierarchical GOP.

In the framework, we assume that the user sends feedback about the instantaneous channel conditions to the server periodically. If the channel condition changes, new set of $\{p_{t_q}\}$ will be calculated. Furthermore, optimized RS codes adapted to the changing channel conditions can be determined.

3. DETERMINATION OF THE REAL UTILITIES

The utility is a quantitative indicator of the importance of one NAL unit. It is defined as the reduction of the total sequence distortion contributed by this NAL unit. Since the distortion in one picture can be influenced by the distortion in its reference pictures, we cannot calculate the accurate utility of one NAL unit if the availability of the NAL units belonging to its reference frames is uncertain. In our framework, the loss patterns of the NAL units associated with different channel realizations are highly restricted to our favor due to

UEP and PET. Firstly, through the UEP, stronger protection is applied to the NAL units in the reference frames than in the predicted frames and this holds consistently true for different quality layers and temporal levels. Secondly, by means of the interleaving process applied in the PET scheme, we can preserve the dependency between the received frames in the hierarchical GOP irrespective of the channel realizations. With this restriction on the loss patterns, we can simplify the procedures to determine the utilities of the NAL units significantly.

We denote the decoded picture with the picture order number p as ψ_p and the NAL unit with the quality layer ID q and temporal level ID t belonging to that picture as $\phi_{t,q,p}$. p_{start} and p_{end} are the picture order numbers of the first and last picture in the current GOP, respectively. The picture order number p is related to the temporal level ID t as follows:

$$p = \text{mod}(2^{t_{\max}-t}, 2^{t_{\max}}) + i \cdot 2^{t_{\max}-t+1} + p_{start} \leq p_{end} \quad (4)$$

, where $i \in \mathbf{N}$.

Figure 2 illustrates the first GOP of an example bitstream with a GOP size of 8. To determine the real utilities of all NAL units, we need to perform multiple decoding processes with a number of selected subsets of all NAL units belonging to this GOP. We denote the subset, which includes the NAL units with a quality layer ID lower than q and those with the quality layer ID q but a temporal level ID lower than or equal to t , as $\Phi_{t,q} = \{\phi_{a,b,p} | b < q \text{ or } b = q \text{ and } a \leq t, p_{start} \leq p \leq p_{end}\}$. Then, the total subsets to be decoded comprise $\Phi_{t_{\max},0}, \Phi_{0,1}, \dots, \Phi_{t_{\max},1}, \dots, \Phi_{t_{\max},q_{\max}}$ with increasing size. For a bitstream with t_{\max} temporal levels and q_{\max} quality layers, a total number of $t_{\max} * (q_{\max} - 1) + 1$ subsets need to be decoded. The first subset consists of the complete BL NAL units. These subsets are decoded in the increasing order of their sizes in turn. For the example in Figure 2, $\Phi_{3,0}, \Phi_{0,1}, \Phi_{1,1}, \Phi_{2,1}$ and $\Phi_{3,1}$ need to be decoded in turn. Figure 3 illustrates the picture distortions (measured in logarithmic mean square error (LMSE)) associated with different subsets for two successive GOPs. The picture distortion in ψ_p associated with $\Phi_{t,q}$ is denoted as $D_{t,q,p}$. The additional reduction of the picture distortion in ψ_p contributed by $\Phi_{t,q}$ against the last decoding subset, i.e. $\Phi_{t-1,q}$ if $t > 0$ or $\Phi_{t_{\max},q-1}$ if $t = 0$, is denoted as $\Delta D_{t,q,p}$.

The dependency in the hierarchical GOP in Figure 2 can be modeled as a directed graph. If $\psi_{\bar{p}}$ uses ψ_p as the reference picture, then a directed edge is drawn from ψ_p to $\psi_{\bar{p}}$. If there exists a directed path consisting of multiple connected directed edges from ψ_p to $\psi_{\bar{p}}$, then the change of the distortion in ψ_p can affect the distortion in $\psi_{\bar{p}}$. The exact utility of the EL NAL unit $\phi_{t,q,p}$ can be calculated as $\Delta D_{t,q,p} + \omega \sum_{\psi_{\bar{p}} \in \Omega_p} \Delta D_{t,q,\bar{p}}$, where Ω_p is the set of pictures that are connected to ψ_p through a directed path and ω is a weight factor. If ψ_p associated with $\phi_{t,q,p}$ is not a key picture, then the pictures in Ω_p are all located in the current GOP and ω is set to 1. If ψ_p is a key picture, then Ω_p includes all B pictures in the current and the previous GOP if it exists. Assuming that the key pictures have an equal influence on the B pictures in between, we set the weight factor ω to $\frac{1}{2}$. In Figure 3, comparing the red solid and the green dashed distortion curves, both ψ_1 and ψ_9 (key pictures) have a direct quality improvement through $\phi_{0,1,1}$ and $\phi_{0,1,9}$, respectively and the B pictures in both GOPs all get a quality improvement as a result of the quality improvements in ψ_1 and ψ_9 . The utility of $\phi_{0,1,1}$ is $\Delta D_{0,1,1} + \frac{1}{2} \sum_{\psi_{\bar{p}} \in \Psi_{\{2,3,4,5,6,7,8\}}} \Delta D_{0,1,\bar{p}}$ and the utility of $\phi_{0,1,9}$ is $\Delta D_{0,1,9} + \frac{1}{2} \sum_{\psi_{\bar{p}} \in \Psi_{\{2,3,4,5,6,7,8,10,11,12,13,14,15,16\}}} \Delta D_{0,1,\bar{p}}$. We take $\phi_{2,1,3}$ as a non-key picture example. Comparing the dark red dashed and the purple dashed distortion curves, ψ_3 has a direct quality improvement. ψ_2 and ψ_4 get a quality improvement as a result of the improvement in ψ_3 . The utility of $\phi_{2,1,3}$ is $\sum_{\psi_{\bar{p}} \in \Psi_{\{2,3,4\}}} \Delta D_{2,1,\bar{p}}$.

To determine the real utilities of the BL NAL units, we need to take the error concealment (EC) algorithm into account. In our UEP

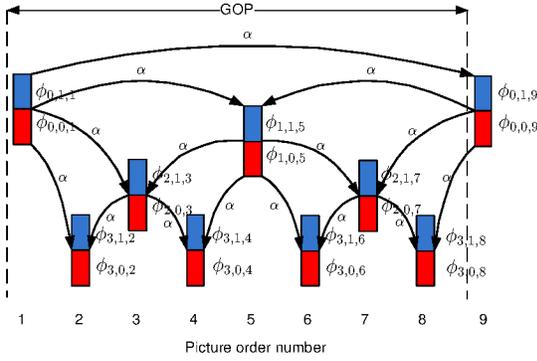


Fig. 2. One GOP of size 8 with one BL and one MGS EL.

framework, a lost frame is replaced by its temporally nearest picture that is stored in the decoded picture buffer (DPB). If more than one picture in the DPB are available for the lost picture, the one with the lowest temporal level ID is chosen. If multiple successive pictures are lost, the EC algorithm is first run for the lost pictures on the lowest temporal level. The already concealed pictures are stored in the DPB and can be further used by the EC algorithm. Since the loss patterns of the NAL units are constrained due to the PET and UEP scheme in our framework, the output pictures of the EC algorithm associated with different loss patterns have a deterministic nature. The EC output pictures corresponding to the subsets $\Phi_{3,0}$, $\Phi_{2,0}$, $\Phi_{1,0}$ and $\Phi_{0,0}$ for a bitstream with GOP size 8 are detailed in Table 1.

p	output pictures of the EC algorithm								
	1	2	3	4	5	6	7	8	9
$\Phi_{3,0}$	ψ_1	ψ_2	ψ_3	ψ_4	ψ_5	ψ_6	ψ_7	ψ_8	ψ_9
$\Phi_{2,0}$	ψ_1	ψ_1	ψ_3	ψ_5	ψ_5	ψ_5	ψ_7	ψ_9	ψ_9
$\Phi_{1,0}$	ψ_1	ψ_1	ψ_1	ψ_5	ψ_5	ψ_5	ψ_9	ψ_9	ψ_9
$\Phi_{0,0}$	ψ_1	ψ_1	ψ_1	ψ_1	ψ_1	ψ_9	ψ_9	ψ_9	ψ_9

Table 1. Output pictures for a GOP of size 8 corresponding to different subsets.

To determine the real utilities of the BL NAL units, we actually only need to decode the subset $\Phi_{3,0}$, because the output pictures of the EC algorithm associated with the other subsets consisting of the BL NAL units can be derived according to the Table 1. The reduction of the picture distortion $\Delta D_{t,0,p}$ can be calculated as $D_{t-1,0,p} - D_{t,0,p}$. And the utility of $\phi_{t,0,p}$ is $\sum_{\psi_p \in \Omega_p} \Delta D_{t,0,p}$.

4. ESTIMATION OF THE UTILITIES OF THE EL NAL UNITS

As discussed in the previous section, to determine the real utilities of the NAL units in the GOP, we need to perform multiple decoding processes on different subsets of these NAL units. Among all the decoding processes, the decoding processes with the subsets $\Phi_{t_{\max},0}$ and $\Phi_{t_{\max},1}$ are integrated into the encoding process. Therefore, the picture distortions $D_{t_{\max},0,p}$ associated with the complete BL and $D_{t_{\max},1,p}$ associated with the complete EL can be obtained directly after the encoding process. On the other hand, the other decoding processes must be performed with additional computational power and extra execution delay. These decoding processes make it difficult to apply our UEP strategy to live streaming applications. Therefore, we investigated the possibility to replace these additional decoding processes with an estimation process of the picture distortions for a

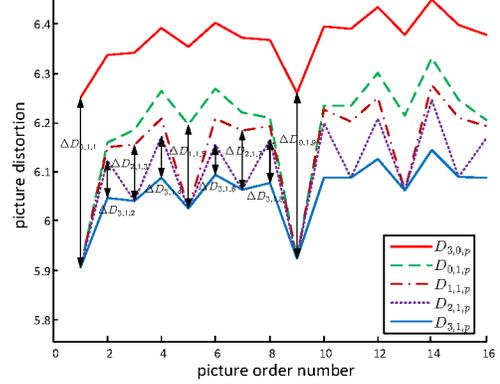


Fig. 3. Reduction of the picture distortion through the NAL units in the GOP.

large part of the video sequence. The estimation is based on the real picture distortions associated with the complete BL and EL NAL units.

To achieve an accurate estimation, we perform the complete decoding processes for a GOP and derive a distortion propagation model from the real picture distortions associated with all decoding processes. Then, we apply the distortion propagation model to the next m GOPs to get the estimated picture distortions associated with decoding processes which are not integrated in the encoding process until a new model is derived from the following GOP. The parameter m depends on the picture contents in adjacent GOPs and can be adjusted dynamically in the encoding process at the server. With larger m , we can reduce the computational complexity and transmission delay in the server, but the distortion propagation model may become increasingly inaccurate for the subsequent GOPs.

Now let's discuss how to derive the distortion propagation model from a selected GOP. We assume that the reduction of distortion contributed by one EL NAL unit in the reference picture propagates to the directly predicted picture with an attenuation factor α . As in Figure 2, each directed edge leads to an attenuation of the reduction of distortion by α . If there exists a directed path which consists of i directed edges between two pictures, then the reduction of the picture distortion at the starting node of the path will be attenuated by α^i at the end node of the path. With this approach, we can formulate the estimated picture distortions with the contribution of additional EL NAL units as a function of α and the picture distortions before these additional EL NAL units are included. To create the distortion propagation model, we need to compare the estimated picture distortions with the real picture distortions associated with different subsets of the EL NAL units. Through minimizing the sum of the square error between the estimated and real picture distortion values, we can obtain the α as a model parameter, which will be used to estimate the picture distortions associated with the decoding subset with the same temporal ID in the next m GOPs. We determine the α parameter separately for the decoding subsets on the EL with different temporal IDs.

In the following, we would like to elaborate on the establishment of the model more concretely with a GOP of size 8 as illustrated in Figure 2. In the first step, we estimate the picture distortions associated with the decoding subset $\Phi_{0,1}$. The reduction of picture distortions in ψ_1 and ψ_9 can be calculated as follows:

$$\begin{aligned} \Delta D_{0,1,1} &= D_{3,0,1} - D_{3,1,1} \\ \Delta D_{0,1,9} &= D_{3,0,9} - D_{3,1,9} \end{aligned} \quad (5)$$

The estimated reduction of picture distortions for $2 \leq p \leq 8$ according to the distortion propagation model can be formulated as:

$$\begin{aligned}
\Delta D_{0,1,2}^{\text{est}} &= (\alpha^3 + \alpha^2 + \alpha)\Delta D_{0,1,1} + \alpha^3\Delta D_{0,1,9} \\
\Delta D_{0,1,3}^{\text{est}} &= (\alpha^2 + \alpha)\Delta D_{0,1,1} + \alpha^2\Delta D_{0,1,9} \\
\Delta D_{0,1,4}^{\text{est}} &= (\alpha^3 + 2\alpha^2)\Delta D_{0,1,1} + (\alpha^3 + \alpha^2)\Delta D_{0,1,9} \\
\Delta D_{0,1,5}^{\text{est}} &= \alpha\Delta D_{0,1,1} + \alpha\Delta D_{0,1,9} \\
\Delta D_{0,1,6}^{\text{est}} &= (\alpha^3 + \alpha^2)\Delta D_{0,1,1} + (\alpha^3 + 2\alpha^2)\Delta D_{0,1,9} \\
\Delta D_{0,1,7}^{\text{est}} &= \alpha^2\Delta D_{0,1,1} + (\alpha^2 + \alpha)\Delta D_{0,1,9} \\
\Delta D_{0,1,8}^{\text{est}} &= \alpha^3\Delta D_{0,1,1} + (\alpha^3 + \alpha^2 + \alpha)\Delta D_{0,1,9} \quad (6)
\end{aligned}$$

The picture distortions $D_{0,1,p}$ associated with the decoding subset $\Phi_{0,1}$ are available in the GOP that is used to create the distortion propagation model. The target reduction of the picture distortions in this step is $\Delta D_{0,1,p} = D_{3,0,p} - D_{0,1,p}$. By minimizing the sum of the difference between the target and the estimated reduction of distortions, the α parameter can be determined for $\Phi_{0,1}$ as follows:

$$\alpha_0^* = \arg \min_{\alpha > 0} \sum_{p \in \{2,3,4,5,6,7,8\}} (\Delta D_{0,1,p} - \Delta D_{0,1,p}^{\text{est}})^2 \quad (7)$$

We denote the solution of the least square problem 7 as α_0^* , which is the positive root of the Equation 8. The Equations 6 with α replaced with α_0^* are utilized in the following GOPs to estimate the reduction of picture distortions associated with $\Phi_{0,1}$. The estimated picture distortion $D_{0,1,p}^{\text{est}}$ associated with $\Phi_{0,1}$ is $D_{3,0,p} - \Delta D_{0,1,p}^{\text{est}}$.

$$\sum_{p \in \{2,3,4,5,6,7,8\}} (\Delta D_{0,1,p}^{\text{est}} - \Delta D_{0,1,p}) \frac{d\Delta D_{0,1,p}^{\text{est}}}{d\alpha} = 0 \quad (8)$$

In the second step, we estimate the picture distortions associated with the decoding subset $\Phi_{1,1}$. The reduction of picture distortions in ψ_5 can be obtained as $\Delta D_{1,1,5} = D_{0,1,5}^{\text{est}} - D_{3,1,5}$. The estimated reduction of picture distortions for $p = 2, 3, 4, 6, 7, 8$ can be formulated as:

$$\begin{aligned}
\Delta D_{1,1,2}^{\text{est}} &= \alpha^2\Delta D_{1,1,5} \\
\Delta D_{1,1,3}^{\text{est}} &= \alpha\Delta D_{1,1,5} \\
\Delta D_{1,1,4}^{\text{est}} &= (\alpha^2 + \alpha)\Delta D_{1,1,5} \\
\Delta D_{1,1,6}^{\text{est}} &= (\alpha^2 + \alpha)\Delta D_{1,1,5} \\
\Delta D_{1,1,7}^{\text{est}} &= \alpha\Delta D_{1,1,5} \\
\Delta D_{1,1,8}^{\text{est}} &= \alpha^2\Delta D_{1,1,5} \quad (9)
\end{aligned}$$

The target reduction of the picture distortion in this step is $\Delta D_{1,1,p} = D_{0,1,p}^{\text{est}} - D_{1,1,p}$ and the least square problem can be formulated as:

$$\alpha_1^* = \arg \min_{\alpha > 0} \sum_{p \in \{2,3,4,6,7,8\}} (\Delta D_{1,1,p} - \Delta D_{1,1,p}^{\text{est}})^2 \quad (10)$$

The solution of the least square problem 10 is denoted as α_1^* . The Equations 9 with α_1^* deliver the estimates of the reduction of picture distortions associated with $\Phi_{1,1}$. The estimated picture distortion is $D_{1,1,p}^{\text{est}} = D_{0,1,p}^{\text{est}} - \Delta D_{1,1,p}^{\text{est}}$.

In the last step, we estimate the picture distortions associated with the decoding subset $\Phi_{2,1}$. The reduction of the picture distortions in ψ_3 and ψ_7 are $\Delta D_{2,1,\{3,7\}} = D_{1,1,\{3,7\}}^{\text{est}} - D_{2,1,\{3,7\}}$.

The estimated reduction of picture distortion for $p = 2, 4, 6, 8$ can be formulated as:

$$\begin{aligned}
\Delta D_{2,1,2}^{\text{est}} &= \alpha\Delta D_{2,1,3} \\
\Delta D_{2,1,4}^{\text{est}} &= \alpha\Delta D_{2,1,3} \\
\Delta D_{2,1,6}^{\text{est}} &= \alpha\Delta D_{2,1,7} \\
\Delta D_{2,1,8}^{\text{est}} &= \alpha\Delta D_{2,1,7} \quad (11)
\end{aligned}$$

The target reduction of the picture distortions in this step is $\Delta D_{2,1,p} = D_{1,1,p}^{\text{est}} - D_{2,1,p}$ and the least square problem is:

$$\alpha_2^* = \arg \min_{\alpha > 0} \sum_{p \in \{2,4,6,8\}} (\Delta D_{2,1,p}^{\text{est}} - \Delta D_{2,1,p})^2 \quad (12)$$

The Equations 11 with α_2^* deliver the estimates of the reduction of picture distortion associated with $\Phi_{2,1}$. The estimated picture distortion for $\Phi_{2,1,p}$ is $D_{2,1,p}^{\text{est}} = D_{1,1,p}^{\text{est}} - \Delta D_{2,1,p}^{\text{est}}$.

5. EXPERIMENTAL RESULTS

We tested the utility estimation algorithm in our UEP framework with the CARPHONE, CITY, CREW, FOOTBALL, FOREMAN, HARBOUR, MOBILE, SHUTTLE, SOCCER, TEMPETE, VCAR sequences. All sequences were encoded with the JSVM (version 9.14) software into one BL and one MGS EL with the quantization parameter (QP) setting: $QP_{\text{BL}} = 36$, $QP_{\text{EL}} = 30$. The GOP size was set to 8 and the key pictures were intra-coded. The channel transmission bitrate was adapted to the bitrate of the video bitstreams in the simulation, such that $R_{\text{src}}/R_{\text{ch}} \approx 70\%$ holds. The packet loss rate was varied from 2% to 40% with a step-size of 2%. The packet loss correlation factor was set to 0.00 and 0.20 to emulate the transmission channel with random packet loss and burst packet loss, respectively. We used the set of RS codes $RS(60, k)$ ($1 \leq k \leq 60$) for UEP protection. For each sequence, we utilized the real utilities of the NAL units and the estimated utilities with m equal to 3, 10, 37. For $m = 37$, the real utilities were measured only for the NAL units in the first GOP, the utilities of the NAL units in the rest GOPs were estimated with the model created from the first GOP. Optimized RS codes were derived for the real and estimated utilities separately. We ran 200 simulations for each channel setting.

In Figure 4, the averaged PSNR of the FOREMAN sequence that is protected with the optimized RS codes is plotted against the changing packet loss rate on the channel. Different curves represent the UEP results with the RS codes which are optimized by means of the real and estimated utilities, respectively. As we can see, the loss of the UEP performance associated with the estimated utilities against the real utilities is very small for the FOREMAN sequence. The UEP performances of the estimated utilities with m equal to 3, 10, 37 are quite close and the best performance among them is achieved with $m = 3$. This can be explained by the fact that strong foreground motion exists through the whole FOREMAN sequence and the model created from different GOPs can be well applied to the rest of the sequence. Similar results were also observed with the other test sequences except the VCAR sequence. The VCAR sequence contains the scene of a man driving an old timer in a forest area. At the beginning of the sequence, the car is almost static and begins to move. In Figure 5, we show the UEP performance for the VCAR sequence with the real and estimated utilities. As we can see, the UEP performances with $m = 3, 10$ are quite close to the UEP performance with the real utilities, while the UEP performance with $m = 37$ is significantly worse than with the real utilities. The reason is that the scene in the first GOP is almost static, which is quite different from the scene in the rest of the sequence. Therefore, it is not appropriate to apply the model built from the first GOP to all the

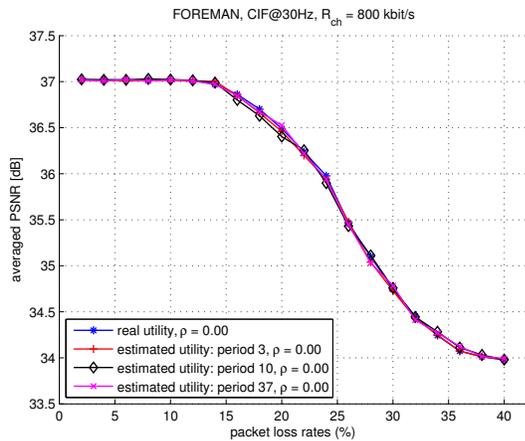


Fig. 4. Comparison of the UEP performance with the real and estimated utilities with $m = 3, 10, 37$ for FOREMAN.

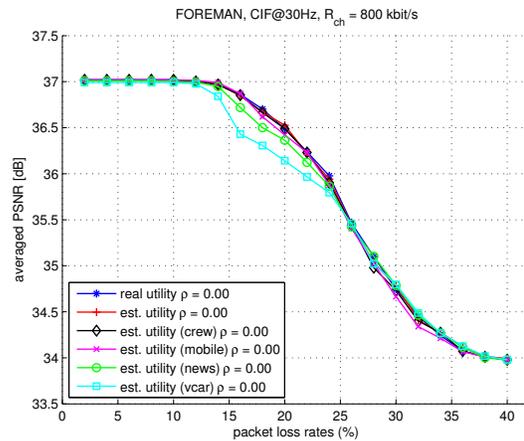


Fig. 6. Comparison of the UEP performance with the real utility and estimated utilities from a different sequence for FOREMAN.

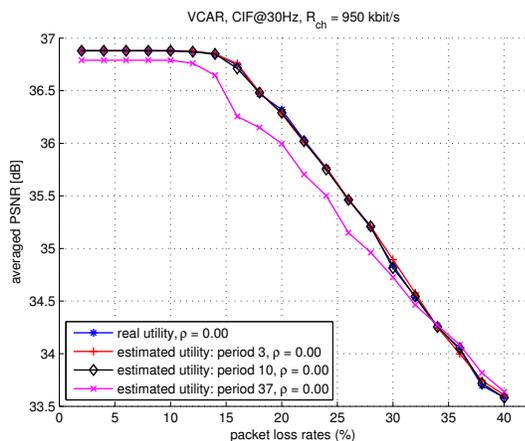


Fig. 5. Comparison of the UEP performance with the real and estimated utilities with $m = 3, 10, 37$ for VCAR.

other GOPs. From the results in these figures, we can conclude that a small m in the estimation process can prevent a mismatch between the UEP performance with the real and estimated utilities. According to our experiments, the model built from one GOP with a certain level of motion can be applied to GOPs with approximately the same level of motion. The level of motion in a GOP can be estimated, for example, from the averaged sum of square of the motion vectors in all frames belonging to this GOP.

We also investigated the possibility of estimating the utilities using a model created from different sequences. In Figure 6, we estimated the utilities of all NAL units in the FOREMAN sequence using models created from the first GOP of several different sequences. The UEP performances with the models created from NEWS and VCAR (low level of motion) are significantly worse than that with the real utilities. The UEP performances with the models created from CREW and MOBILE (comparable level of motion as FOREMAN) are quite close to that with the real utilities. These results demonstrate that the consistency of the level of motion in the GOP used to create the model and in the GOP to which the model is applied is a key factor for the applicability of the estimated utilities in the UEP framework.

6. CONCLUSIONS

We proposed an algorithm to estimate the utilities of the EL NAL units. With this algorithm, no additional decoding processes are needed in the encoder. This makes it possible to apply the adaptive UEP algorithm to real-time live streaming applications. Our experiments show that the performance of the estimation algorithm is related to the consistency of the motion level in the GOP used to create the distortion model and the GOPs to which the model is applied. The motion vectors in the frames of one GOP are good hints on the motion level of the whole GOP. With appropriately selected estimation period, very good UEP performance can be achieved with the estimated utilities.

7. REFERENCES

- [1] Bin Zhang, Mathias Wien, and Jens-Rainer Ohm, "A novel framework for robust video streaming based on H.264/AVC MGS coding and unequal error protection," in *Proc. Int. Sympo. on Intell. Sig. Process. and Comm. Syst. (ISPACS)*. Kanazawa, Japan, Dec. 2009.
- [2] Bin Zhang, Mathias Wien, and Jens-Rainer Ohm, "Optimized Channel Rate Allocation For H.264/AVC Scalable Video Multicast Streaming Over Heterogeneous Networks," in *Proc. IEEE Int. Conf. on Img. Process. (ICIP)*. Hong Kong, China, Sept. 2010.
- [3] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the Scalable Video Coding Extension of the H.264/AVC Standard," *IEEE Trans. Circ. and Syst. for Video Technology*, vol. 17, no. 9, pp. 1103–1120, Sept. 2007.
- [4] R. Zhang, S.L. Regunathan, and K. Rose, "Video coding with optimal inter/intra-mode switching for packet loss resilience," *IEEE Journal on Sel. Areas in Comm.*, vol. 18, no. 6, pp. 966–976, June 2000.
- [5] M.K. Jubran, M. Bansal, L.P. Kondi, and R. Grover, "Accurate distortion estimation and optimal bandwidth allocation for scalable h.264 video transmission over mimo systems," *Image Processing, IEEE Transactions on*, vol. 18, no. 1, pp. 106–116, Jan. 2009.
- [6] A. Albanese, J. Blomer, J. Edmonds, and M. Luby, "Priority encoding transmission," *IEEE Trans. Inf. Theory*, vol. 42, no. 6, pp. 1737–1744, Nov. 1996.