Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG
(ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6)
16th Meeting: Poznań, PL, 24-29 July, 2005

Document: JVT-P084
Filename: JVT-P084.doc
NOTE – That first character after the dash up there in the document number is an "O", not a zero.

Title: **CE6 response of Samsung Electronics: in-depth comparison of closed-loop and open-loop MCTF structures**

Status: Input Document to JVT

Purpose: CE responses

Author(s) or Contact(s): Woo-Jin Han Email: wjhan.han@samsung.com

Yeongtong-Gu Suwon-City Gyeonggi-Do South Korea

Digital Media R&D Center

Samsung Electronics Co., Ltd.

Source: Samsung Electronics Co., Ltd.

Abstract

It is well-known that the open-loop MCTF with update-step structure provides a state-of-the art performance as well as the excellent scalability feature although most standardized video codecs utilize the closed-loop concept. In this contribution, we will investigate the various properties between closed-loop and open-loop framework including the single-layer and multi-layer tests. Furthermore, we propose a new scheme for utilizing both good properties of the closed-loop scheme and the update-step by combining the closed-loop prediction and the switched update-step technique. From the test results of combined scalability configuration, it can be concluded that the new scheme can combine the closed-loop prediction and the update-step successfully and improve the PSNR values up to 0.8 dB. The followings are the items to be investigated:

- Current implementation of temporal decomposition structure (open-loop hierarchical B and MCTF).
- **Tool 1**: closed-loop residual re-estimation technique (JVT-O062) to reduce the prediction-step mismatch
- Single-layer performance comparision between closed-loop and open-loop frameworks.
- Multi-layer performance comparision between closed-loop and open-loop frameworks.
- **Tool 2**: Switched update-step technique to combine the closed-loop prediction into MCTF framework.
- Experimental results of closed-loop prediction and switched update-step.
- Necessary syntax, semantics, and decoding process changes

1. Current implementation of temporal decomposition structure in JSVM2

1.1 Open-loop MCTF structure in enhancement layer

Current implementation of MCTF in JSVM2 is based on 5-3 MCTF with open-loop structure. MCTF process is performed from the highest level to the lowest level. In one level, every odd number frame is predicted by the suitable reference pictures to generate the corresponding high-pass frame, and it is re-used for the update-step to remove the high-frequency component to the reference pictures. In the encoder-side, the prediction-step and the update-step are performed sequentially for each temporal level. This process is performed until only one low-pass frame remains. This low-pass frame is encoded by closed-loop fashion taking the previous decoded low-pass frame as a reference frame. Finally, all other high-pass pictures are encoded. It should be noted that the motion vector is estimated by using the original low-pass references, which are not encoded. Therefore, it is called as the open-loop MCTF structure. Figure 1 shows a MCTF implementation in JSVM2.

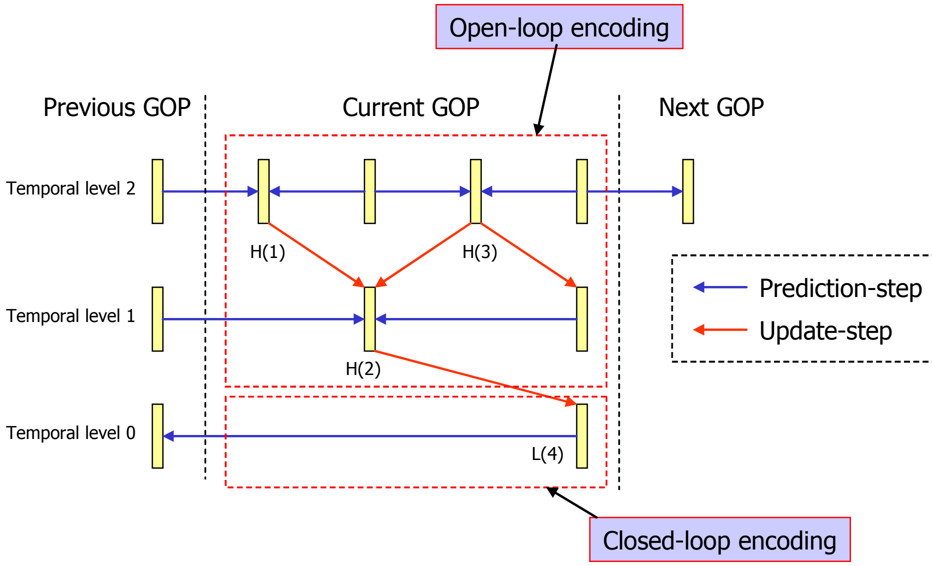


Figure 1 MCTF implementation in JSVM2.

In the decoder-side, the inverse MCTF process is invoked, which means, the inverse update-step is used for adding the high-frequency component to the low-pass pictures and the inverse prediction-step is used for reconstructing the decoded image. It should be noted that the processing order is reversed compared to the encoder-side. It guarantees the near perfect reconstruction in the open-loop structure if sufficient bit-rates are given. The following code briefly shows a processing order in the MCTF encoding process. Although the actual decoding process is somewhat complex mainly due to the optimal buffer management, the fundamental concept of the decoding process is exactly reverse form of the MCTF encoding process.

```
//===== MCTF decomposition =====
for( iLevel = 0; iLevel < m_uiDecompositionStages; iLevel++ )
{
  RNOK( xMotionEstimationStage   ( iLevel ) );      <- 1) Open-loop ME
  RNOK( xDecompositionStage      ( iLevel ) );      <- 2) Prediction-step and Update-step
  RNOK( xSetScalingFactorsMCTF   ( iLevel ) );
  //{{Quality level estimation and modified truncation- JVTO044 and m12007
  //France Telecom R&D-(nathalie.cammas@francetelecom.com)
  RNOK( xSetScalingFactorsMCTF_FT( iLevel ) );
  //}}Quality level estimation and modified truncation- JVTO044 and m12007
}

printf("\nCODING:\n");

//===== encode low-pass pictures =====
RNOK( xEncodeLowPassPictures    ( rcAccessUnitList ) );      <- 3) Low-pass frame encoding

//===== encode high-pass pictures =====
for( iLevel = m_uiDecompositionStages; iLevel-- > 0; )
{
  RNOK( xEncodeHighPassPictures    ( rcAccessUnitList, iLevel ) );      <- 4) High-pass frames encoding
}
```

Figure 2 Open-loop MCTF encoding process in JSVM2 code.

1.2 Open-loop hierarchical B structure in base-layer

In the base-layer, open-loop hierarchical B structure is currently used for the base-layer compatibility to the H.264 standards. In the hierarchical B structure, the temporal hierarchy is represented as the relationship between B pictures. It is very similar to the MCTF structure without update-steps. Although the JSVM2 software allows the use of MCTF structure even in the base-layer, it is not a part of WD. However, we will provide single-layer experimental results of both the MCTF structure and the hierarchical B structure in the experimental results section.

2. Tool 1: Closed-loop residual re-estimation [JVT-O062] [1]

2.1 Prediction-step mismatch in open-loop MCTF structure

Most significant differences between the MCTF and the closed-loop prediction result from the open-loop structure and update-step. Open-loop MCTF uses low-pass reference pictures for motion estimation and the prediction-step, instead of the decoded reference pictures. Although the update-step can decrease the drift error due to the mismatch of the reference pictures, but it still exists especially for pictures with high-energy residuals. We have proposed a new method to overcome the prediction-step mismatch in open-loop MCTF structure.

2.2 Closed-loop residual re-estimation technique

Figure 3 shows the residual re-estimation for reducing the prediction-step mismatch. After MCTF process we can obtain one low-pass frame and multiple high-pass pictures. The basic purpose of the closed-loop residual re-estimation is to simulate the decoder-side in terms of MCTF process. At the first-step, the low-pass frame is encoded and decoded as usual. Now, from the temporal level 1 to the highest temporal level, every high-pass frame is re-estimated by using the reconstructed reference pictures just same to the closed-loop prediction in the conventional video codec. After that, the re-estimated high-pass residual is encoded and recovered to reconstruct the decoded image for the prediction-step in the upper temporal level. In this way, the reconstructed reference pictures can be always obtained since the frame processing order goes from the lowest temporal layer to the highest temporal layer.

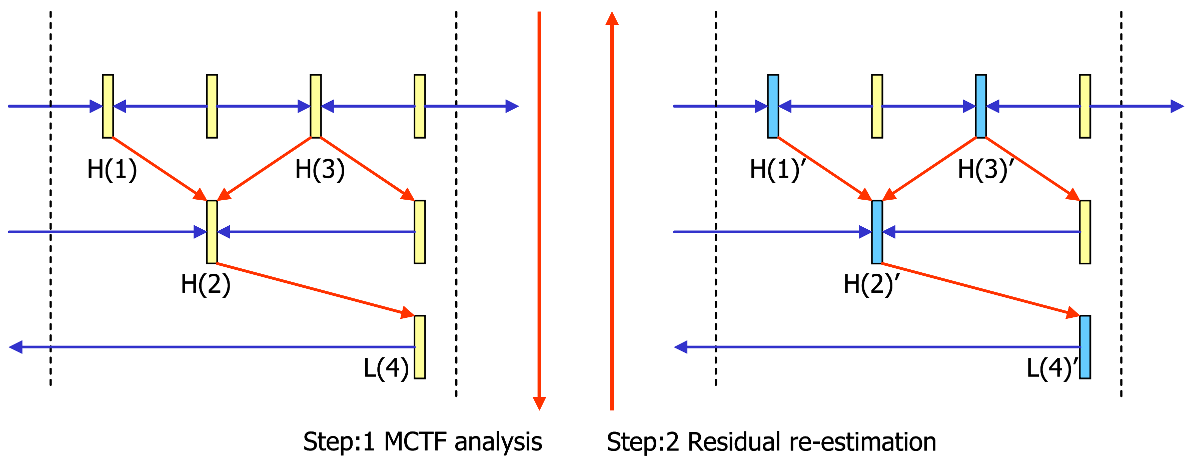


Figure 3 Residual re-estimation process after MCTF analysis.

If the update-step is not used, this process is equal to the conventional closed-loop prediction except it uses the open-loop motion estimation using the non-encoded reference pictures. We will provide some experimental results between open-loop motion estimation and closed-loop motion estimation using the decoded reference pictures in the motion estimation stage. With the update-step, the complete processing order is as follows:

Closed-loop residual re-estimation process [Encoder]

1. Perform normal MCTF process.
2. Encode the low-pass frame.
3. Re-estimate high-pass residual of temporal level N increasing from 1 to the decomposition level-1.
4. Encode the high-pass residual.
5. Perform inverse update-step using the encoded high-pass residual in Step 4.
6. Perform inverse prediction-step to reconstruct the decoded picture.
7. Repeat 3-6 until all high-pass residuals are re-estimated.

Note that the Step 5 and 6 will be modified in the section: switched MCTF decoding process.

Figure 4 Closed-loop residual re-estimation in the context of MCTF process.

Figure 4 shows a processing order of the closed-loop residual re-estimation. In Figure 4, the inverse update-step is included in the composition stage, which reconstructs the decoded picture. It should be noted that the quantized high-pass residual is used in the inverse update-step to simulate the decoder-side. Although this processing order is effective to reduce the prediction-step mismatch, it remains one problem: mismatch between two predictions of the closed-loop residual re-estimation step and the inverse prediction-step in the composition stage. In the closed-loop residual re-estimation step, the reference frames before the inverse update-step is used, whereas the modified reference frames after the inverse update-step is used in the inverse prediction-step. This mismatch is addressed in the other section: switched MCTF decoding process.

3. Single-layer experimental results

Before progressing further, we will provide the single-layer experimental results using the following systems to investigate the properties of the update-step and the open/closed-loop structures. Since the closed-loop residual re-estimation described in the previous section has a mismatch problem in the update-step, which will be solved later, we only investigate the performance when the update-step is not used for closed-loop residual re-estimation process.

- Open-loop Hierarchical B (OHB) – JSVM2 base-layer structure
- Open-loop MCTF without update-step (OMP)
- Open-loop MCTF with update-step (OMU) – JSVM2 enhancement-layer structure
- Closed-loop Hierarchical B with closed-loop motion estimation (CHB)
- Closed-loop MCTF without update-step with open-loop motion estimation (CMP)

The experimental condition was defined as:

QP values	22, 28, 34, 40
Sequences	bus_cif_30, football_cif_30, foreman_cif_30, mobile_cif_30, city_4cif_60, crew_4cif_60, harbour_4cif_60, soccer_4cif_60
GOP size	16 for cif 30 FPS sequences, 32 for 4cif 60 FPS sequences.

Due to the limited time, we have only performed the experiments of CIF sequences. However, the performance trend is rather clear to reveal the facts.

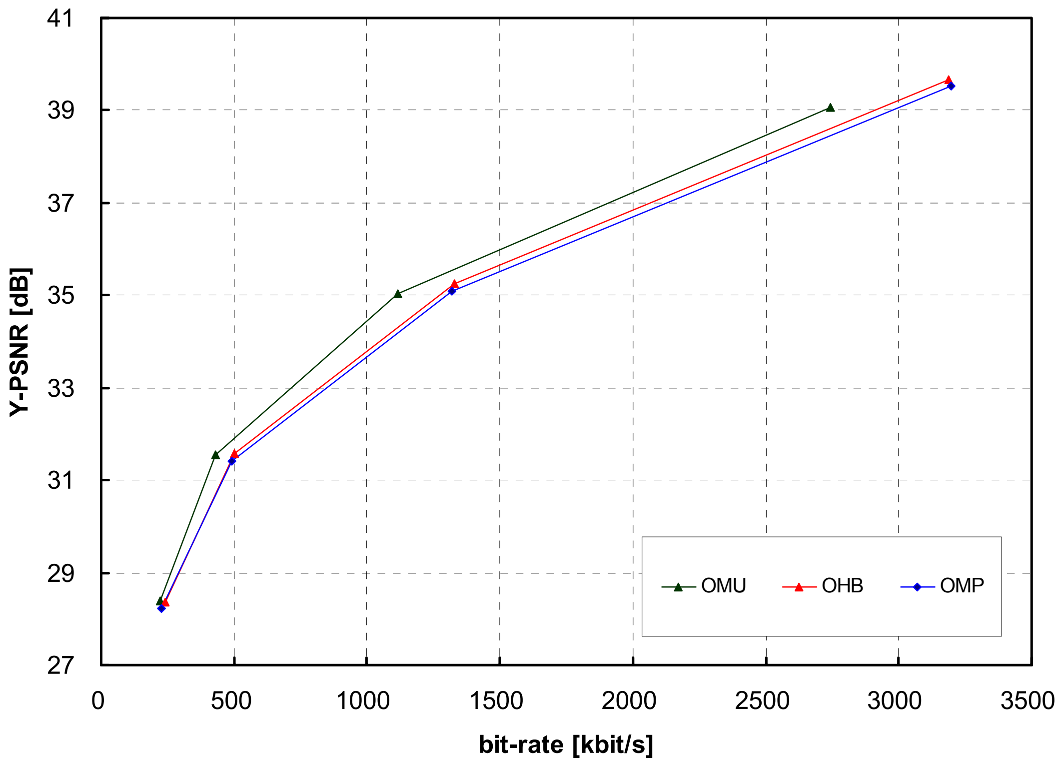


Figure 5 Single-layer performance of MOBILE sequence (OMU, OHB, and OMP).

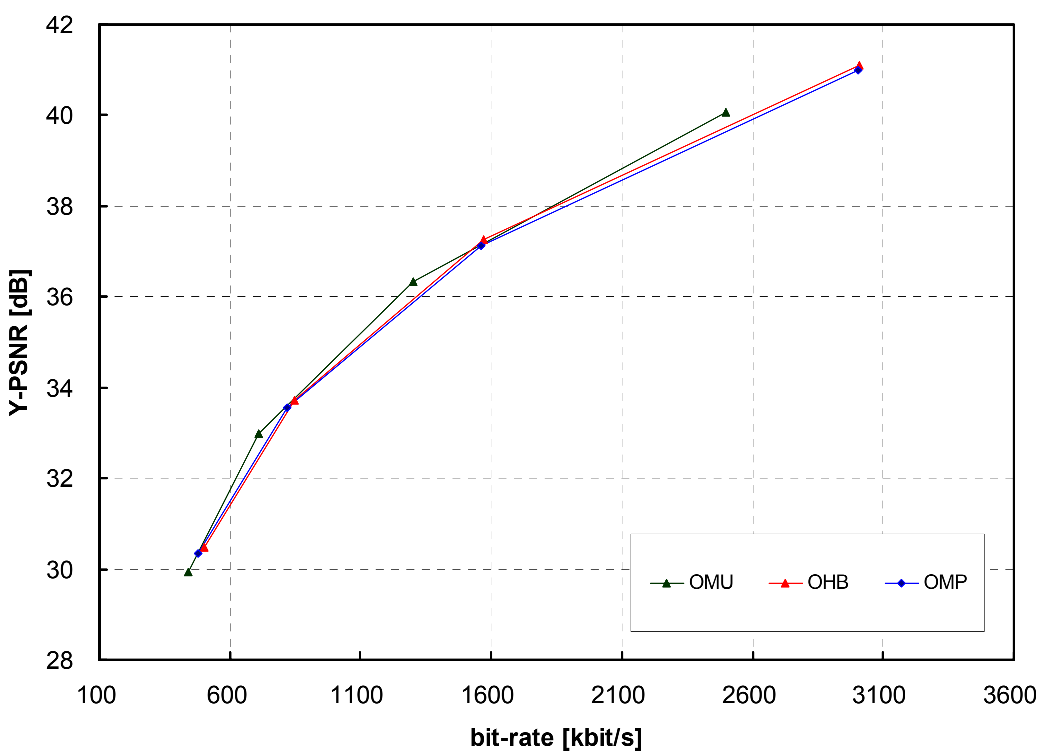


Figure 6 Single-layer performance of FOOTBALL sequence (OMU, OHB, and OMP).

3.1 Comparison of open-loop framework

Figure 5 and Figure 6 show the single-layer performance of MOBILE and FOOTBALL sequences using three open-loop structures, which are OMU: open-loop MCTF with update-step (JSVM2 enhancement layer), OHB: open-loop hierarchical B (JSVM2 base-layer), and OMP: open-loop MCTF without update-step. In MOBILE sequence, the update-step does some positive things in the open-loop framework, since the performance of OMU is clearly better compared to other systems. PSNR gap is about 0.5-0.8 dB. It is interesting that open-loop hierarchical B (OHB) provides slight improvement over open-loop MCTF without update-steps. In FOOTBALL sequence, the PSNR gap is much smaller than the case of BUS sequence. It indicates that the update-step does not have impact on sequences to have high motion mismatches. The results of other sequence are included in the excel file: PSNR-Single.xls. From the results of single-layer experiments for open-loop framework, a simple fact can be drawn:

- **In open-loop framework, update-step is useful for sequences of well-matched motion.**

3.2 Comparison of closed-loop framework

Figure 7 and Figure 8 show the single-layer performance of MOBILE and FOOTBALL sequences using one open-loop structure (OMU) and two closed-loop structure, which are CHB: closed-loop hierarchical B using closed-loop motion estimation, and CMP: closed-loop MCTF without update-step, which is described in the previous section. The main difference between CHB and CMP lies in the motion estimation procedure. In CHB, the motion vector is estimated by using the decoded reference pictures, whereas CMP uses the original reference pictures.

In both sequences, the use of closed-loop motion estimation (CHB) provides 0.1 – 0.2 dB PSNR gain over open-loop motion estimation (CMP). Furthermore, the closed-loop hierarchical B using closed-loop motion (CHB) provides comparable performance in MOBILE sequence compared to the open-loop MCTF with update-step (OMU) and clearly better performance in FOOTBALL sequence. It provides the effectiveness of the mismatch reduction by the closed-loop prediction technique. From the results of single-layer experiments for closed-loop framework, simple facts can be drawn:

- **Closed-loop prediction provides better performance for less matched motion to the best open-loop framework.**
- **Even for well-matched motion, it has comparable performance to best open-loop framework.**
- **Closed-loop motion estimation provides slight performance gain over the open-loop motion estimation in single-layer condition. (CHB > CMP)**

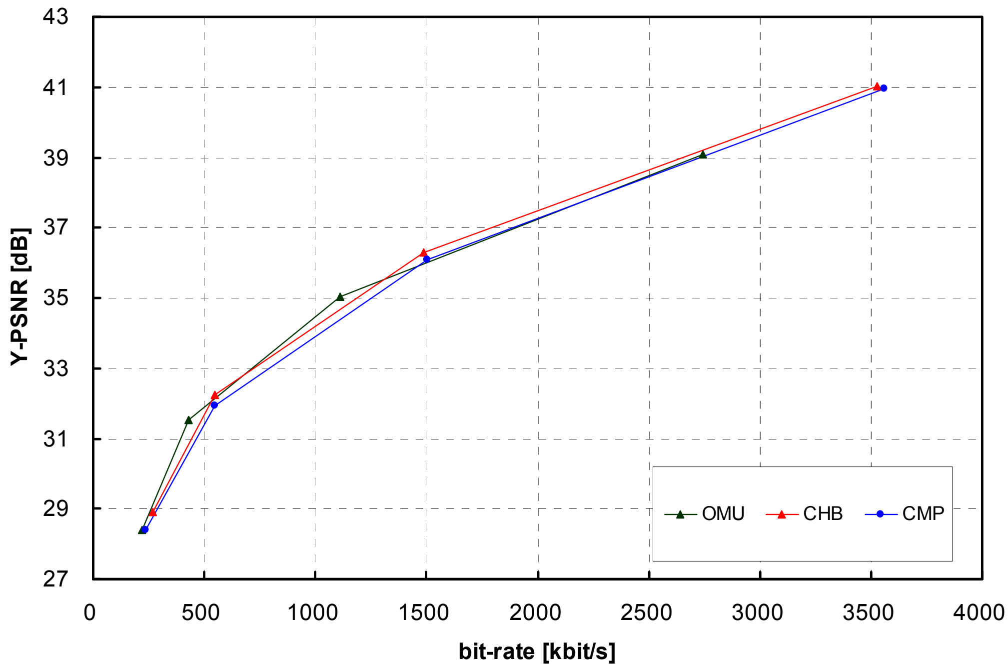


Figure 7 Single-layer performance of MOBILE sequence (OMU, CHB, and CMP).

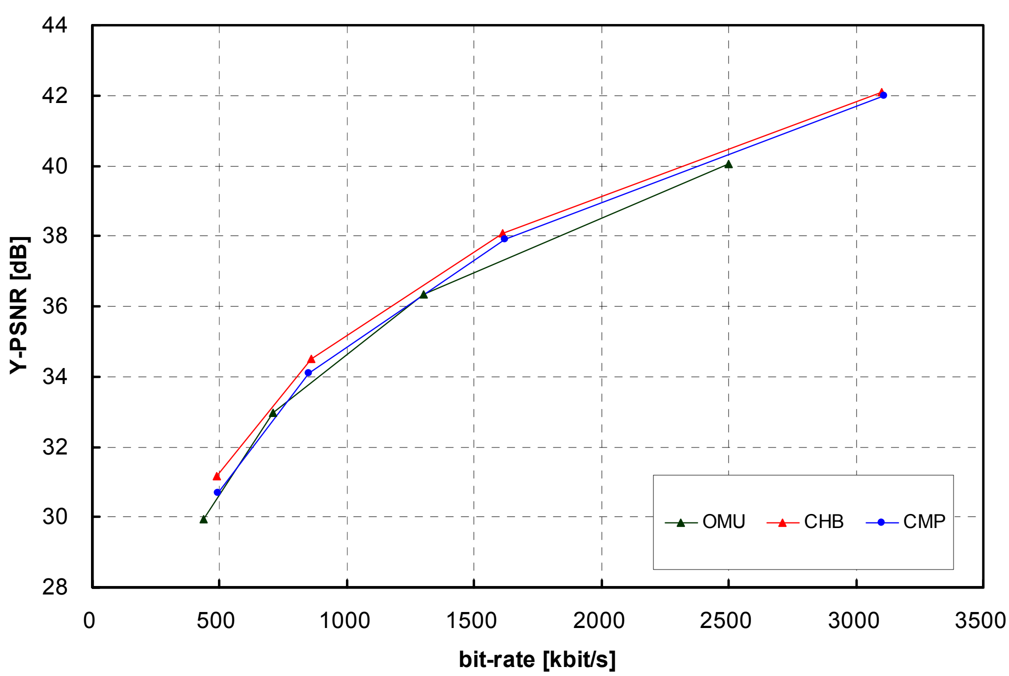


Figure 8 Single-layer performance of FOOTBALL sequence (OMU, CHB, and CMP).

4. Multi-layer experimental results

For the multi-layer experiments, we have used the "combined" configuration included in the newly released test configuration by testing conditions AHG. It uses one layer per each resolution and each layer has possibly multiple FGS layers. The higher layer has finer temporal resolution compared to the lower layer. It enables us to test three scalabilities: spatial, temporal, and SNR scalabilities.

From the results of single-layer experiments, we decided to consider three systems: ANC (open-loop MCTF with update-step; JSVM2 = OMU), CHB (closed-loop hierarchical B without update-step), and CMP (closed-loop MCTF prediction without update-step). For the base-layer, OMU uses the open-loop hierarchical B just like to the JSVM2, while other systems use CHB and CMP even in the base-layer since it has natural support of the H.264 compatibility.

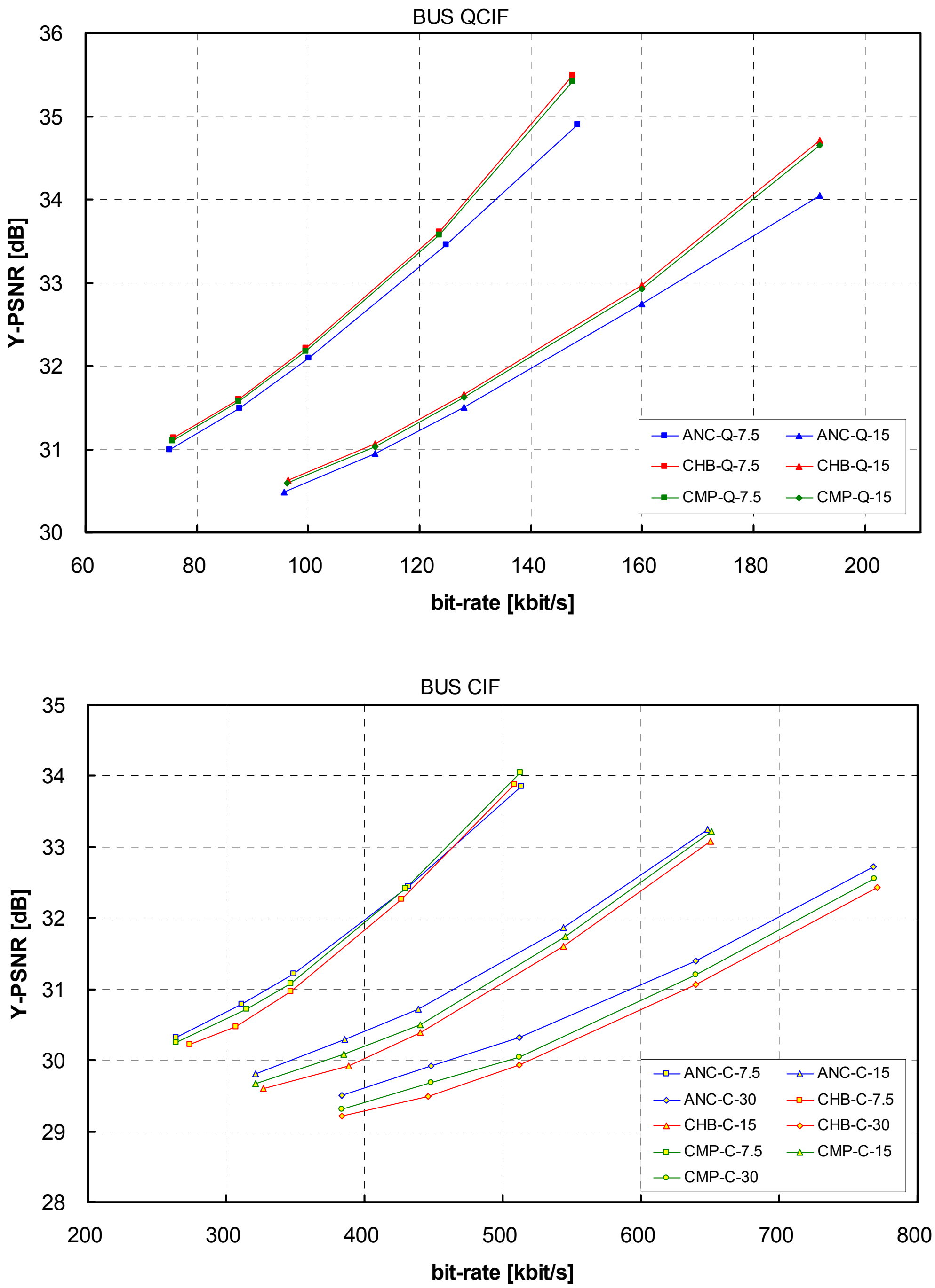


Figure 9 Combined performance results of BUS sequence (ANC, CHB, and CMP).

Figure 9 shows a very clear tendency of update-step's efficiency: for QCIF layer, which is the base-layer, closed-loop prediction is better than the update-step with open-loop prediction up to 0.6 dB. For CIF layer, update-step is better than two closed-loop prediction systems. Especially, the performance of closed-loop motion estimation (CHB) is only slightly better than that of open-loop motion estimation (CMP), while the open-loop motion estimation provides better performance in CIF layers up to 0.3 dB. It is mainly due to the motion consistency that is related to the motion vector prediction performance as already noted in JVT-O062 [1]. The closed-loop motion estimation is possibly the best way to estimate the motion in the single-layer, however, in the multi-layer structure, the motion vector itself is dependent on the texture Qp value in the closed-loop motion estimation, which makes the motion prediction in the upper layer from the base-layer becomes less effective.

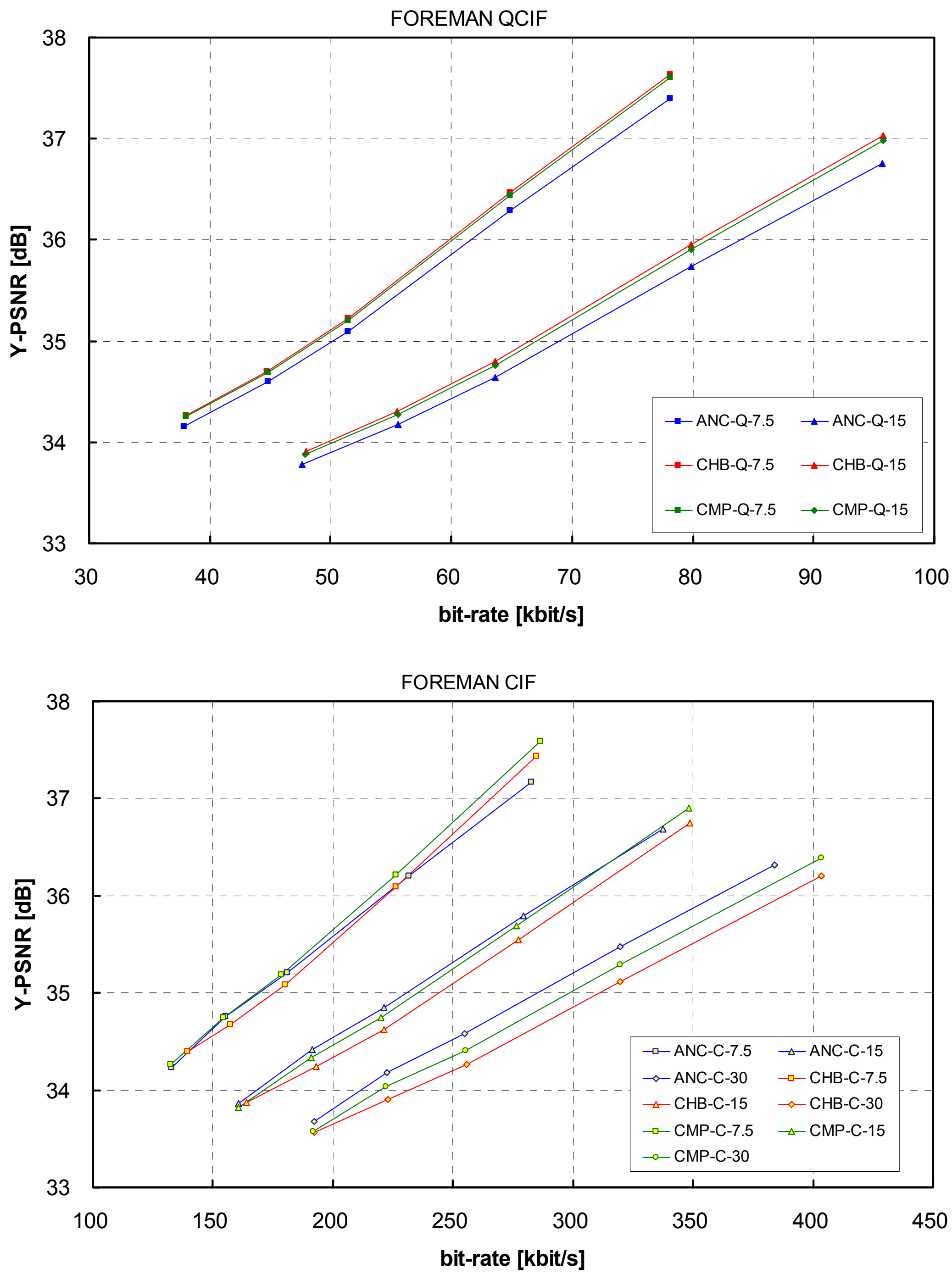
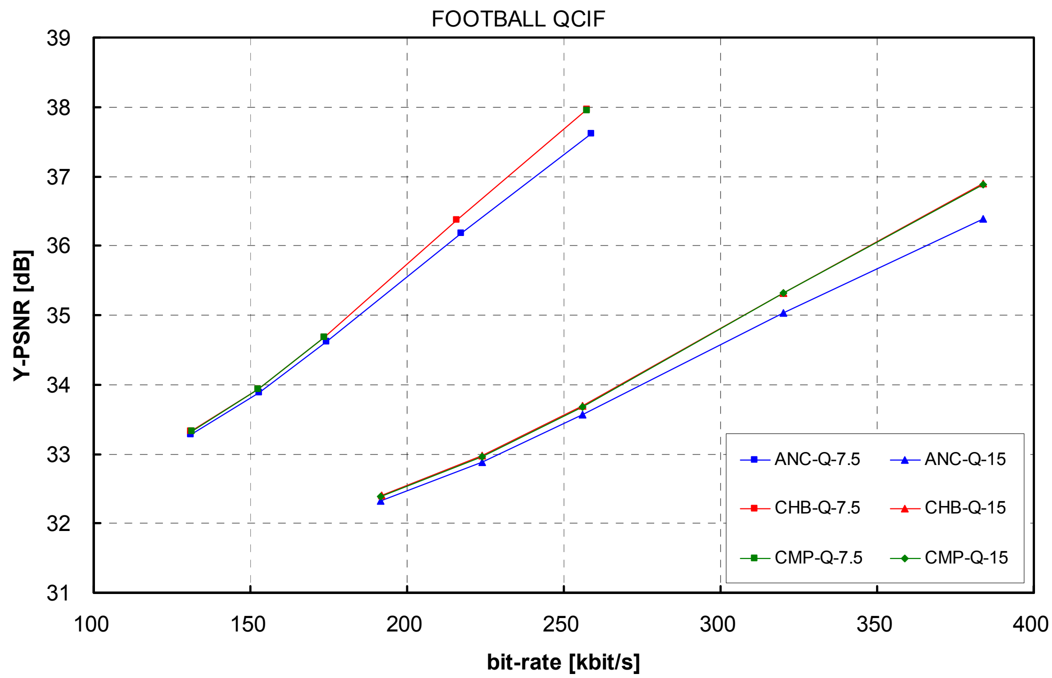


Figure 10 Combined performance results of FOREMAN sequence (ANC, CHB, and CMP).

For FOREMAN sequence, the same tendency can be observed. In QCIF layers, the closed-loop prediction is effective, while the update-step is effective in CIF layers. Furthermore, the open-loop motion estimation may be better approach in the multi-layer structure, since it guarantees the motion consistency across layers.



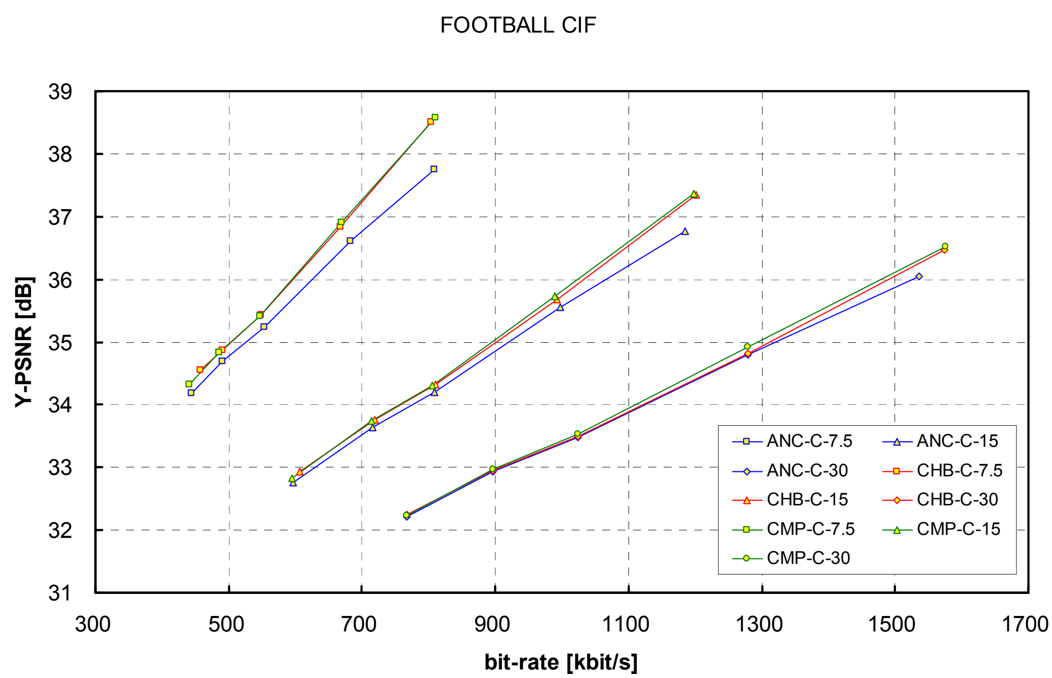
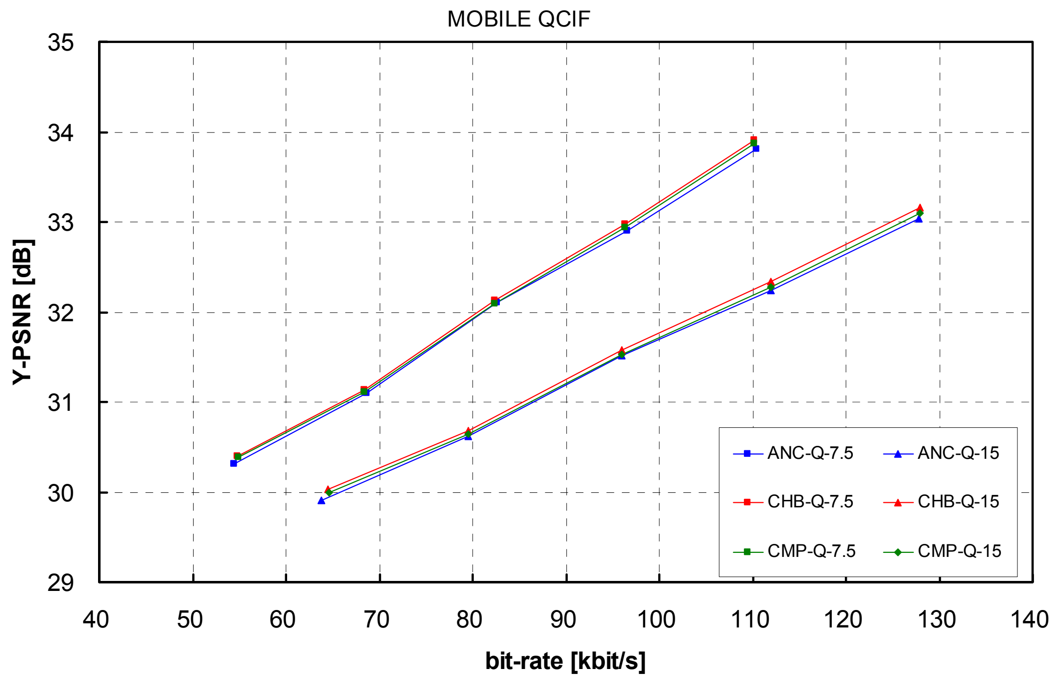


Figure 11 Combined performance results of FOOTBALL sequence (ANC, CHB, and CMP).

For FOOTBALL sequence, two closed-loop prediction systems are better than the JSVM2 anchor even in the CIF layers, since the update-step is less effective and the motion consistency across layers is not necessary in this sequence. However, even in this case, the open-loop motion estimation provides slight gain over the closed-loop motion estimation approach in CIF layers.



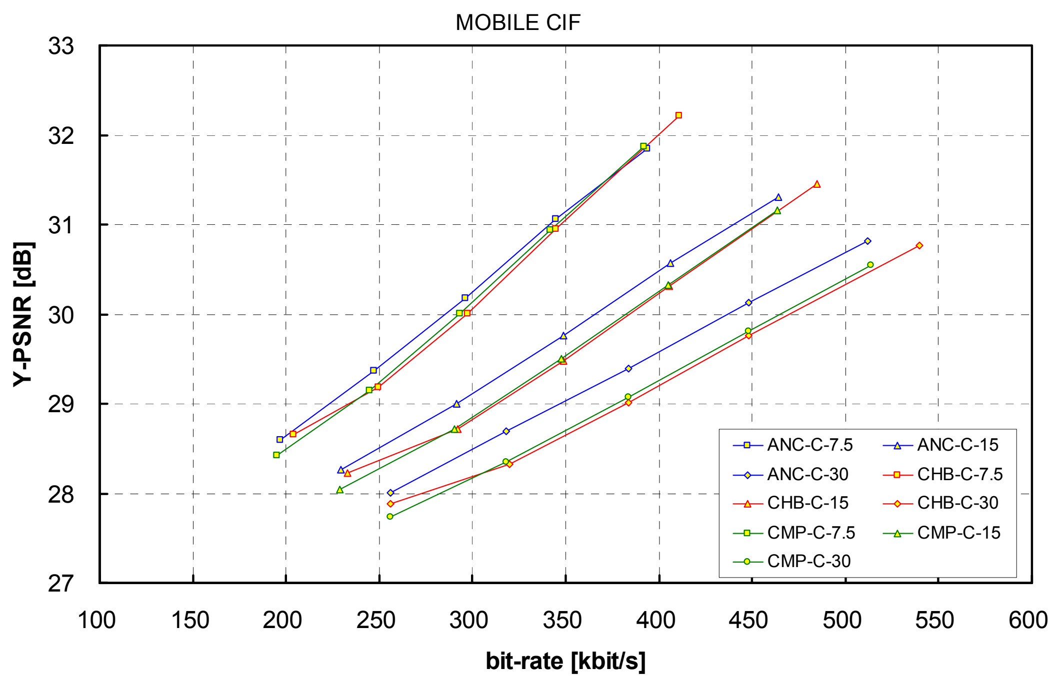


Figure 12 Combined performance results of MOBILE sequence (ANC, CHB, and CMP).

MOBILE sequence has very strong temporal correlations and high-detail textures, which maximizes the performance of the update-step. Due to this effect, the performance of JSVM2 (open-loop with update-step, ANC) is comparable to the closed-loop prediction even in the QCIF layers, and clearly better in the CIF layers. This effect is clearer in the highest frame-rate condition due to the stronger temporal correlation. PSNR difference is up to 0.5 dB and the matched bit-save is about 8-12%, which is clearly meaningful value. In addition, the open-loop motion estimation gives only slight improvements over the closed-loop motion estimation in this case, mainly due to the very small residual energy, which negates any performance differences between them. It indicates the exploitation of the update-step is needed to fill the performance gap. Therefore, it would be the best combination if we can exploit both the closed-loop prediction and the update-steps. Multi-layer experimental results show the following facts:

- **Closed-loop prediction techniques provide better performance in the base-layer.**
- **Open-loop motion estimation seems to be better choice compared to the closed-loop motion estimation in the multi-layer structure. (CMP > CHB)**
- **Update-step is useful for the enhancement layer especially for the temporally well-correlated sequences.**

5. Tool 2: Switched update-step in MCTF structure [JVT-O062] [1]

5.1 Mismatch in closed-loop residual re-estimation and update-step

Now, let's review the Tool 1: closed-loop residual re-estimation. Although the previous experimental results include the performance of Tool 1 without update-step, it is possible to use Tool 1 with the update-step. We name it as CMU. CMU re-estimates the high-pass residuals according to the processing order in the decoder-side. This residual re-estimation process uses the reconstructed pictures in the lower temporal level as the reference pictures. After the quantization process, this re-estimated residual is re-used for the inverse update-step, which updates the low-pass reference pictures. The problem is that the updated low-pass reference pictures are used in the inverse prediction-step, whereas the re-estimated residual is optimized by the non-updated low-pass reference pictures. Figure 13 shows PSNR graph of two layer experiment of MOBILE CIF@30Hz sequence using two systems: JSVM2 and closed-loop residual re-estimation with original update-step with various Qp values from 36 to 8. As shown in the figure, in the high to middle Qp range, the closed-loop residual re-estimation with update-step provides better performance compared to the open-loop MCTF with update-step (JSVM2), while the situation is reversed in the low Qp range, mainly due to the inherent mismatch between closed-loop residual re-estimation and the update-step.

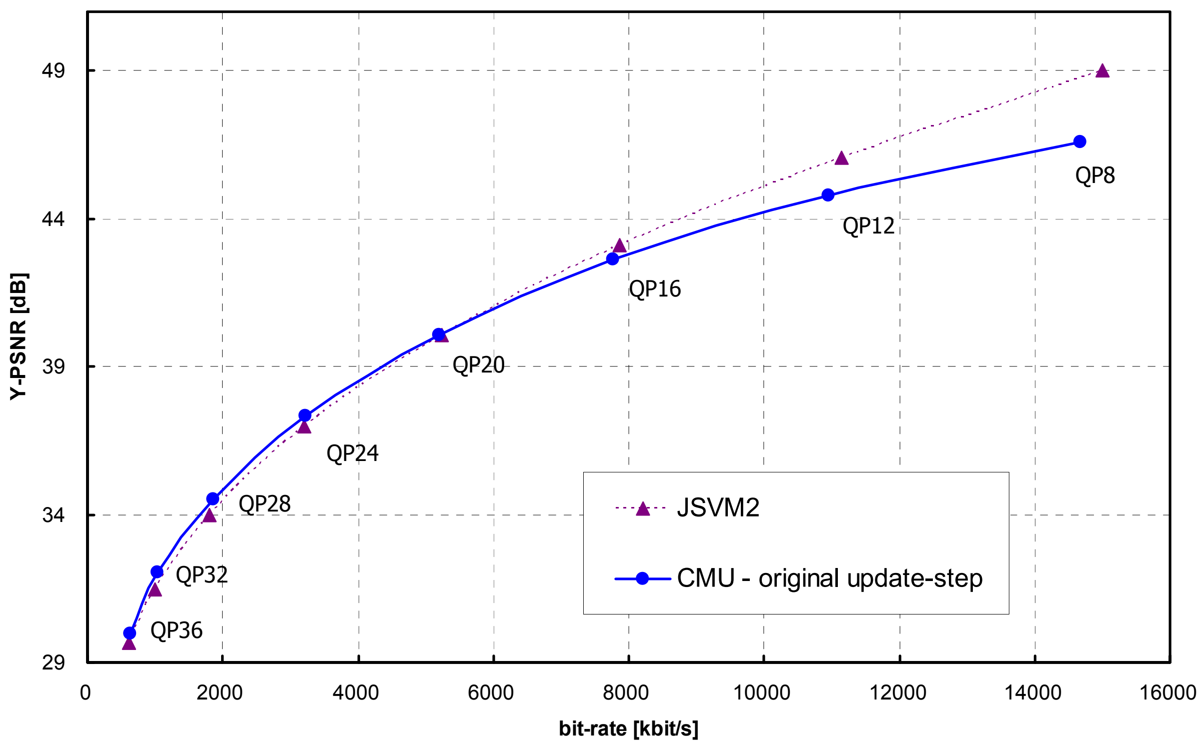


Figure 13 PSNR results of MOBILE sequence with various Qp values (JSVM2 and CMU)

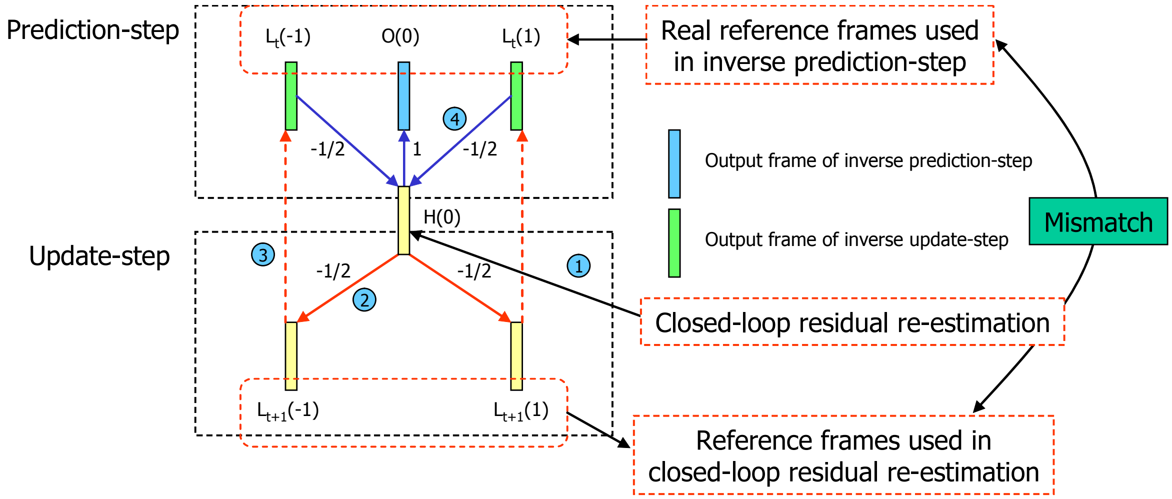


Figure 14 Closed-loop residual re-estimation and the *original* decoding process

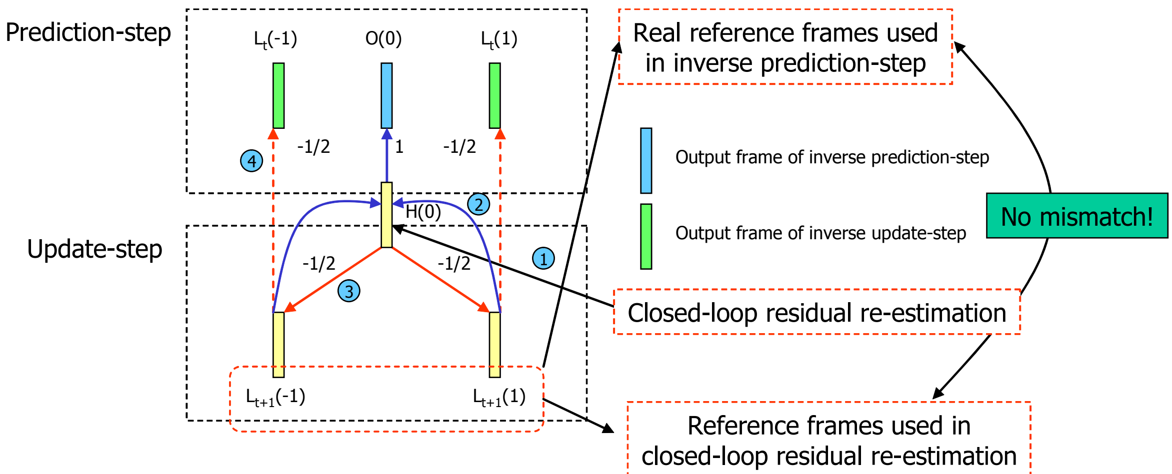


Figure 15 Closed-loop residual re-estimation and the *switched* decoding process

5.2 Simple solution to the problem: switched MCTF decoding process

Figure 14 shows a diagram of closed-loop residual re-estimation with the original update-step process. As already noted in the previous section, there is a mismatch as shown in the figure. To solve the problem, we propose a new processing order of the MCTF decoding process, called as ***switched MCTF decoding***, which is used in the decoder and the decoder-part in the encoder, which is used in the closed-loop residual re-estimation. It should be noted that the MCTF analysis remains here without any change. Just decoding process and the decoding part in the encoder is changed.

Figure 15 shows a diagram of closed-loop residual re-estimation with the switched update-step process. In the switched update-step process, the prediction-step is performed eariler than the update-step in the decoder-side while the update-step is usually performed eariler in the original MCTF decoding process. In Figure 15, the closed-loop residual re-estimation is performed first, which gives the re-estimated and decoded high-pass residual. In the next stage, the inverse prediction-step is performed by using the reference frames used in the closed-loop residual re-estimation process. It eliminates the mismatch occurred in the inverse prediction-step. After that, the high-pass residual is used for updating the low-pass reference frames for the higher temporal layer. Summarized procedure of the closed-loop residual re-estimation and switched update-step is the followings:

<u>Closed-loop residual re-estimation with switched update-step [Encoder]</u>

1. Performs normal MCTF process.
2. Encodes the low-pass frame.

3. Re-estimates high-pass residual of temporal level N increased from 1 to the decomposition level-1.
4. Encodes the high-pass residual.
5. Performs inverse prediction-step to reconstruct the picture using the reference frames in Step 3
6. Performs inverse update-step using the encoded high-pass residual in Step 4.
7. Repeat 3-6 until all high-pass residuals are re-estimated.

The only difference between the switched MCTF decoding and the original method is the order exchange of prediction and update-steps. However, since Tool 2 is always used with Tool 1, the re-estimated and decoded high-pass residual is used in the inverse update-step, which means that the new update-step is performed as the *closed-loop fashion* with no mismatch between encoder and decoder. We define the new system using closed-loop residual re-estimation and the switched update-step as CMU+. Figure 16 shows PSNR results. The results are very clear. The switched update-step can effectively eliminate the mismatch problem in the closed-loop residual re-estimation and the update-step process. CMU+ shows a steady performance gain compared to the JSVM2 over all Qp ranges.

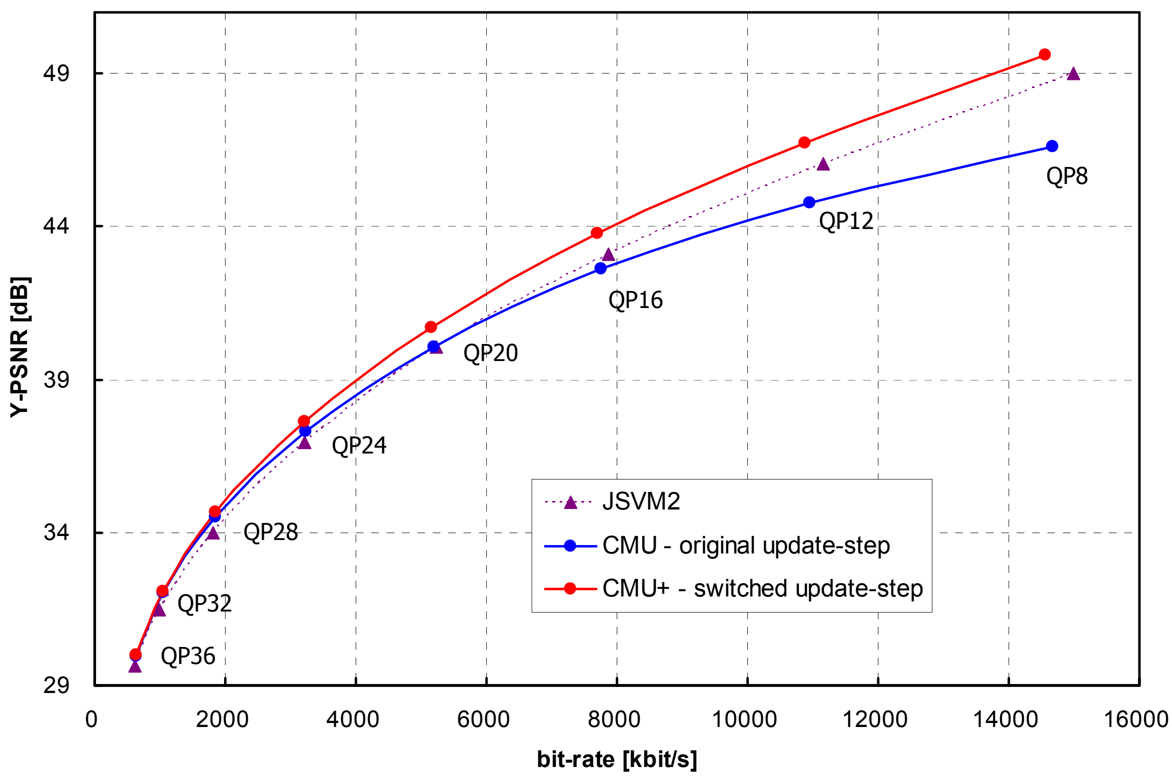


Figure 16 PSNR results of MOBILE sequence with various Qp values (JSVM2, CMU and CMU+)

5.3 Complexity issues on MCTF update-step process: update-step skipping in decoder

One of the major issues on MCTF update-step process is the complexity problem. Since the update-step requires the quarter-pel interpolation and motion compensation, MCTF decoding process has nearly double complexity compared to the system without update-step. To reduce the complexity, we proposed a ***decoder-side update-step skipping*** concept, which means the update-step may be omitted in the decoder-side. In other words, in the encoder-side, full update-steps are used and some update-steps are disabled in the decoder-side. It surely affects the performance. Now, let's compare the performance drop with update-step skipping. It should be noted that the partial update-steps are used in the closed-loop residual re-estimation procedure for the synchronization to the decoder.

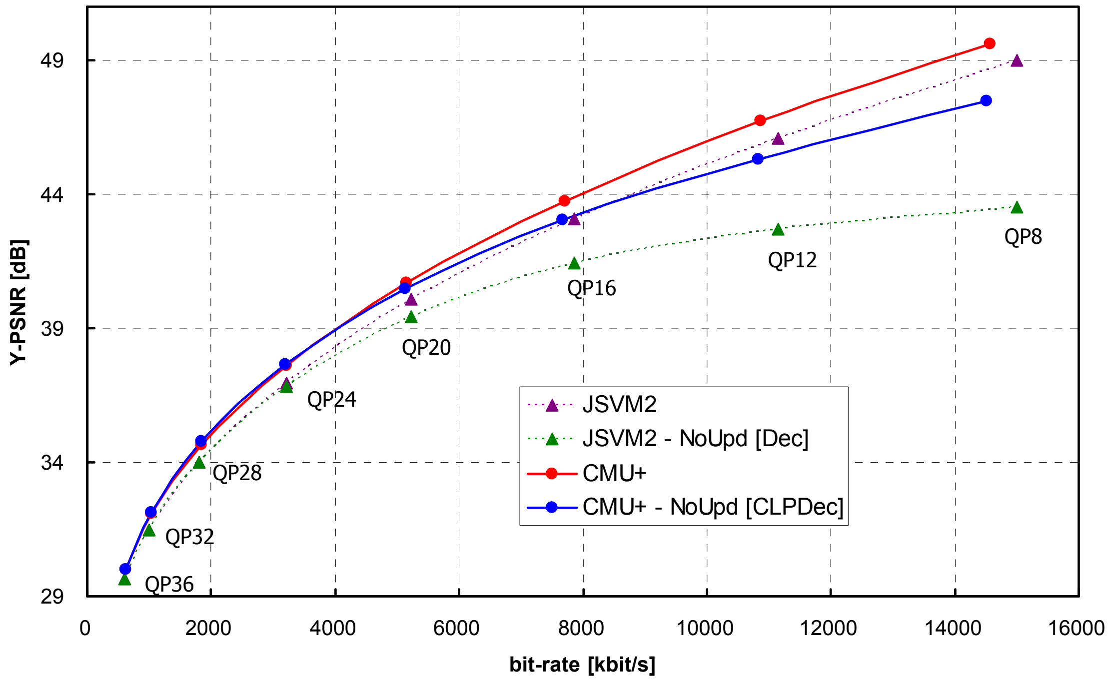


Figure 17 PSNR results of MOBILE sequence with various Qp values without any update-step in decoder-side (JSVM2 and CMU+)

As shown in Figure 17, JSVM2 shows a lot of performance drop up to 5 dB in the lower Qp ranges from 20. However, CMU+ shows much smaller performance drop and its performance is still better than JSVM2 when Qp values are larger than 16. It is useful results, since for most applications requiring the low complexity the relatively large Qp values are used due to the bandwidth or the storage limitation. The reason of smaller performance drop in the CMU+ is the closed-loop residual re-estimation procedure. In this case, in the first phase of the MCTF process defined in Step 1 of previous section, full update-steps are used; however, partial update-steps are used in Step 3 – 6 to synchronize the decoder-side. The closed-loop residual re-estimation procedure re-estimates the high-pass residual to reflect the update-step skipping, thus the decoder-side update-step skipping has smaller performance drop. Figure 18 shows the comparison diagram of update-step skipping.

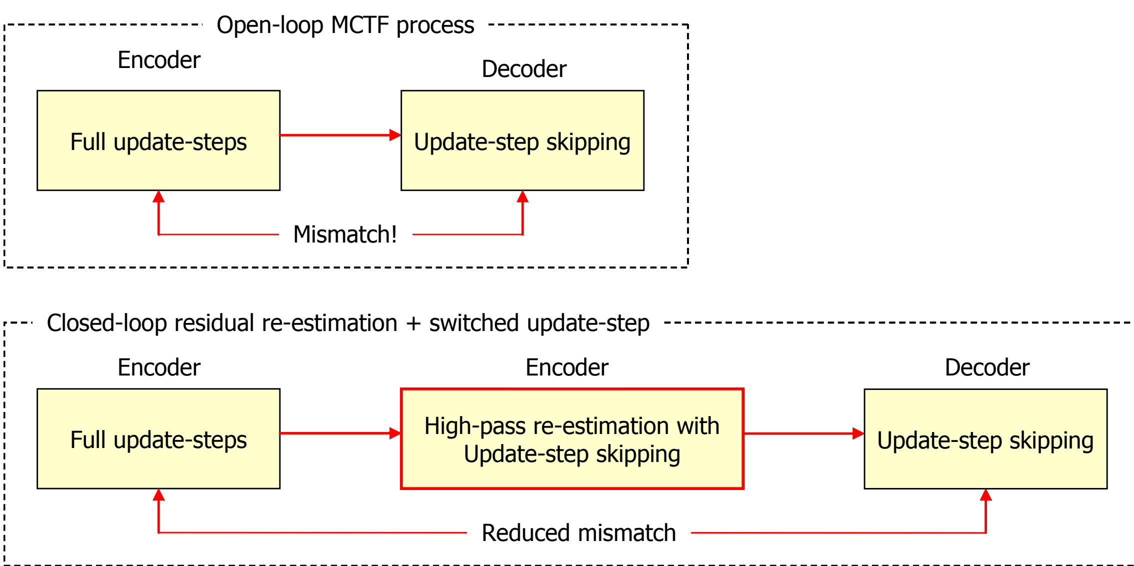


Figure 18 Concept of update-step skipping: comparison between open-loop MCTF and closed-loop residual re-estimation

It is clear that the degree of update-step skipping and the performance have trade-off relationship. If we skip all update-steps in the decoder-side, the decoding complexity of one layer is just the same to the well-known closed-loop video decoder. Thus if all update-steps are skipped in the decoder-side, the total complexity of SVC standards can be reduced at the level of the H.264 decoder with the single-loop decoding feature [2]. However, it is more flexible choice if we can choose the trade-off parameters between the complexity and the performance. For example, we can allow the update-steps for only several temporal levels. If only anchor frame in the temporal 0 is allowed to be updated, total number of update-steps can be reduced to one per each temporal layer with performance improvements compared to the all update-step skipping.

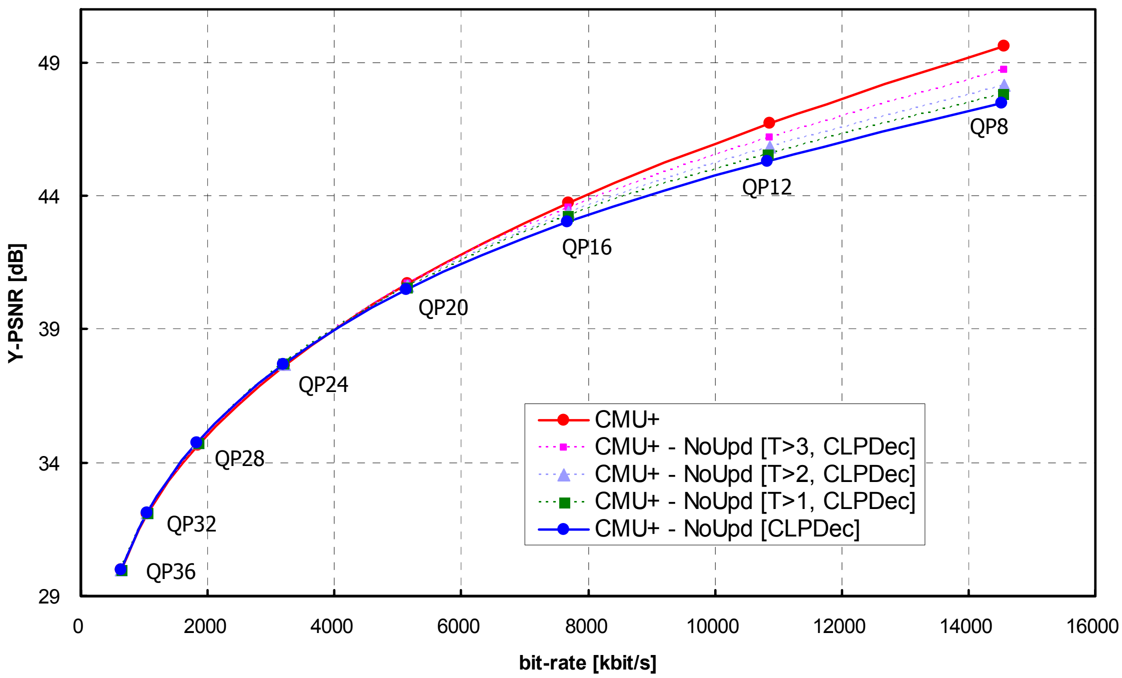


Figure 19 Performance of various kinds of update-skipping according to temporal level.

Figure 19 shows the performance when various kinds of decoder-side update-skipping are used. It is clear that the performance drop increases with the smaller number of update-steps. It should be noted that the performance gap is only meaningful for the low range of Qp values, especially larger than 20. For most low-complexity applications, Qp values larger than 20 is hardly used. It indicates that the update-step skipping with the closed-loop residual re-estimation can be a good choice for the complexity and performance trade-off.

5.4 Application aspects: decoder-side complexity scalability

Currently, most contributions of the scalable video coding focus on the coding performance improvements. However, if we adjust the complexity in the decoder-side with the same bit-stream, it is very useful for the real-world applications. For example, we can obtain the full quality decoded images from the same bit-stream when the complexity constraints are not severe and get the less quality decoded images when the complexity constraints are very strict. The former case covers the high-powered PC and dedicated decoding hardware, while the latter case covers the mobile devices or less-powered software DSP codec. Although the current SVC standards can reduce the complexity by reducing the spatial resolution, temporal resolution, or the bit-rates, it has not been studied to reduce the decoder's complexity without bit rates and spatial, temporal changes.

Skipping update-step can be used for this purpose. If we use full update-steps in the closed-loop residual re-estimation and the switched update-step in the encoder side, and only partial update-steps in the decoder-side, there is a mismatch between the encoder and the decoder; however, considerable amounts of complexity can be saved. For example, even if an encoder generates bit-stream using full updates-steps, the decoder can skip some update-steps to reduce the complexity. In this case, if the decoder utilizes full update-steps, full quality decoded images can be reconstructed, but the quality degrades in a graceful manner as some update-steps are skipped simultaneously.

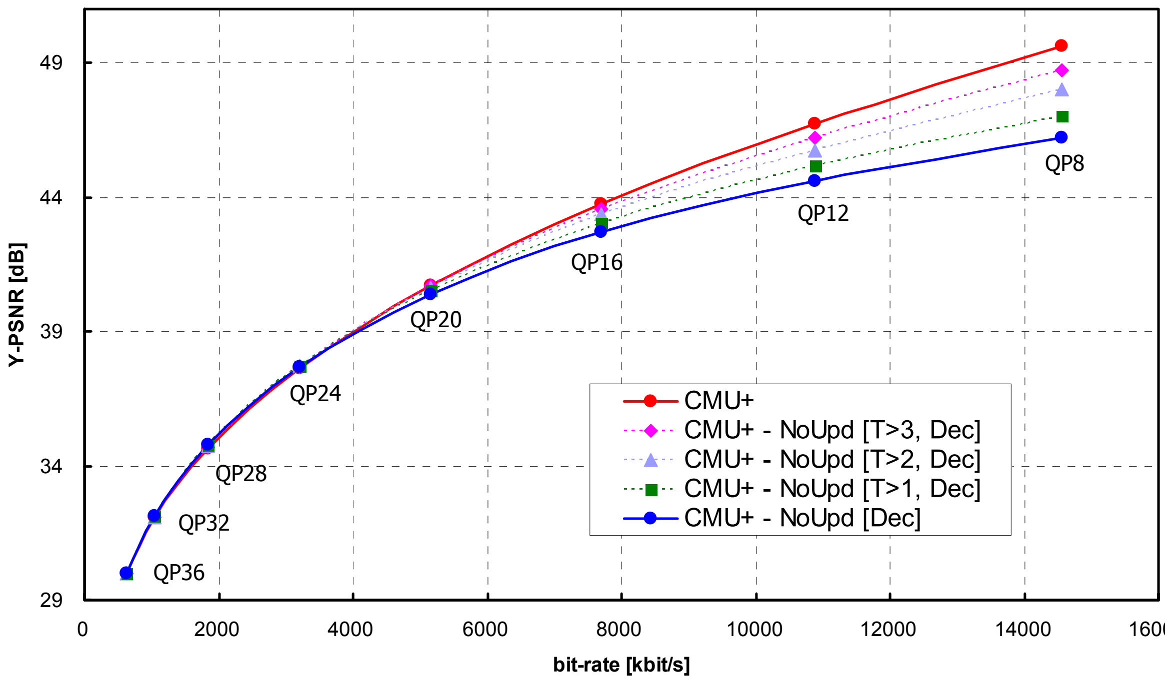
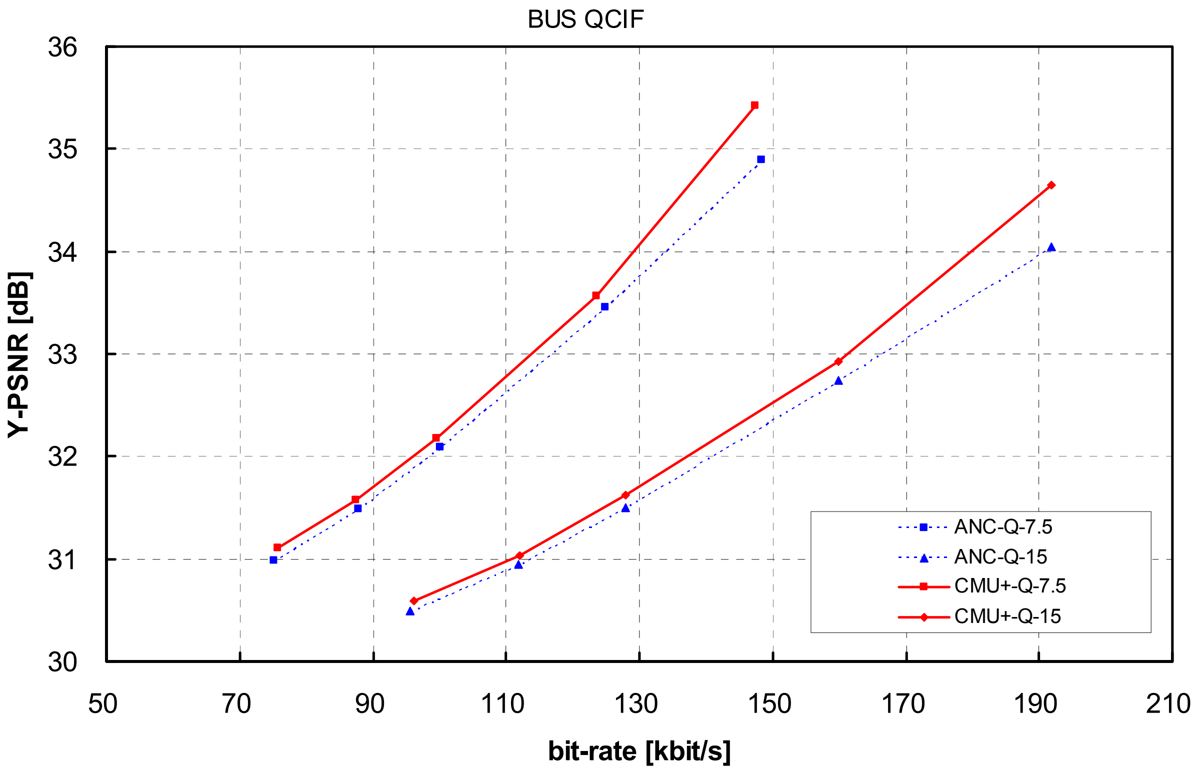


Figure 20 PSNR graph of various update-step skipping condition using CMU+.

Figure 20 shows the results of PSNR trends with various skipping conditions of update-step in the decoder-side when CMU+ is used. It should be noted that the encoder uses full update-steps in both MCTF analysis, closed-loop residual re-estimation, and the switched update-steps and only the decoder skips the update-step of the specific temporal levels. The performance drop at the low Qp value is relatively large compared to Figure 19, which skips update-step in the decoder part in the encoder and the actual decoder together. When we skip all update-steps, the PSNR difference is up to 3.4 dB at the Qp value '8' although the complexity is only half. However, it should be noted that the PSNR difference is negligible when the Qp value is larger than 20, which is sufficiently small Qp value for most applications. Therefore, if the contents are encoded with the Qp value larger than 20, it does not matter to skip the update-steps in the decoder-side to reduce the complexity. On the other hand, if the contents are encoded with the sufficiently small Qp value for the professional work, the bit-stream can be decoded with a full quality at the high-performance machine as well as the low-performance mobile device, which decodes the contents with a degraded quality. This is very useful property for many applications. SVC standards should consider this issue in the future. As shown in Figure 17, the proposed closed-loop residual re-estimation and the switched update-step show desirable properties of smaller PSNR drop when the update-step is skipped mainly due to the smaller encoder-decoder mismatch compared to the JSVM2. Thus the proposed method is more preferable on the aspects of the decoder-side complexity scalability.

6. Experimental results of Tool 1 and 2: closed-loop re-estimation with switched update-step

We used the "combined" configuration to evaluate the performance of new tools. Because the complexity problem of the update-step is very important issue, we only allow one update-step to the anchor frame per each level in the closed-loop residual re-estimation and the decoder-side. In this case, only 4 update-steps are additionally performed for GOP size of 32. The decoder complexity increases by only 13% (35 motion compensation instead of 31) compared to the case of no update-steps. Three difference systems are compared: ANC (JSVM2), CMP (closed-loop residual re-estimation without update-step), and CMU+ (closed-loop residual re-estimation with switched update-step). For the base-layer, CMU+ and CMP use only closed-loop residual re-estimation and do not use update-step to support the base-layer compatibility although CMU+ and all update-step skipping can be used for the base-layer. Thus CMP is not included in the results of QCIF layer since CMP and CMU+ has the same PSNR values. Finally, for 4CIF sequences, we will present only CIF and 4CIF layers while QCIF layers are included in the attached excel file, "PSNR-Combined-4CIF-CU.xls."



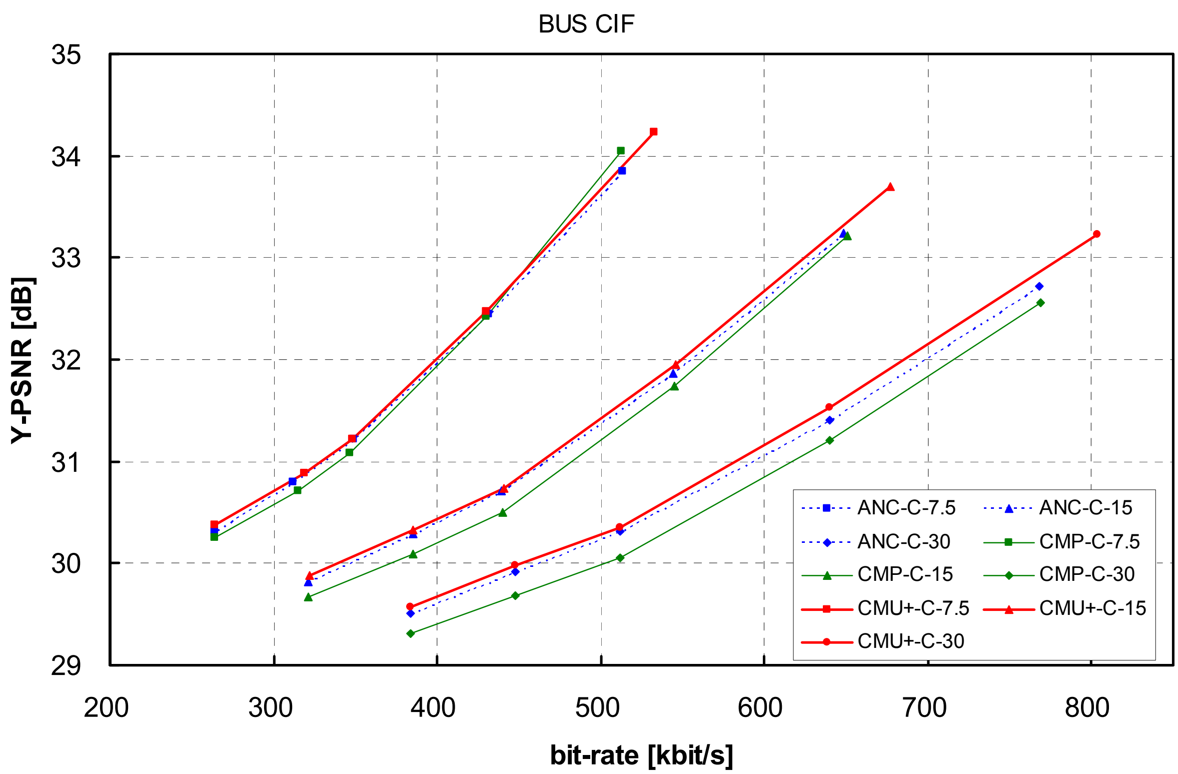
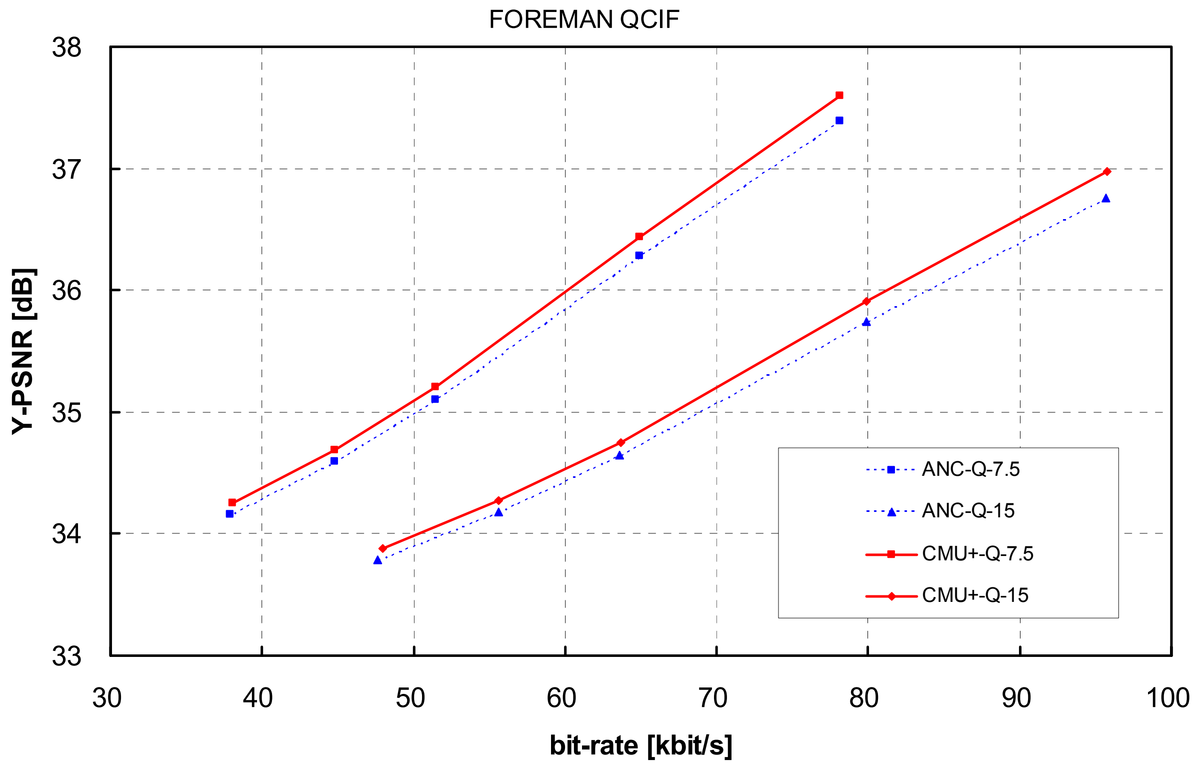


Figure 21 PSNR graph of BUS sequences for JSVM2, CMP, and CMU+.

For BUS sequences, CMU+ shows the best performance. In QCIF layer, the PSNR gain is up to 0.6 dB and 0.15 dB in CIF layer compared to the JSVM2. Especially, in CIF layer, the CMU+ has 0.31 dB higher than CMP, which means that the switched update-step works properly.



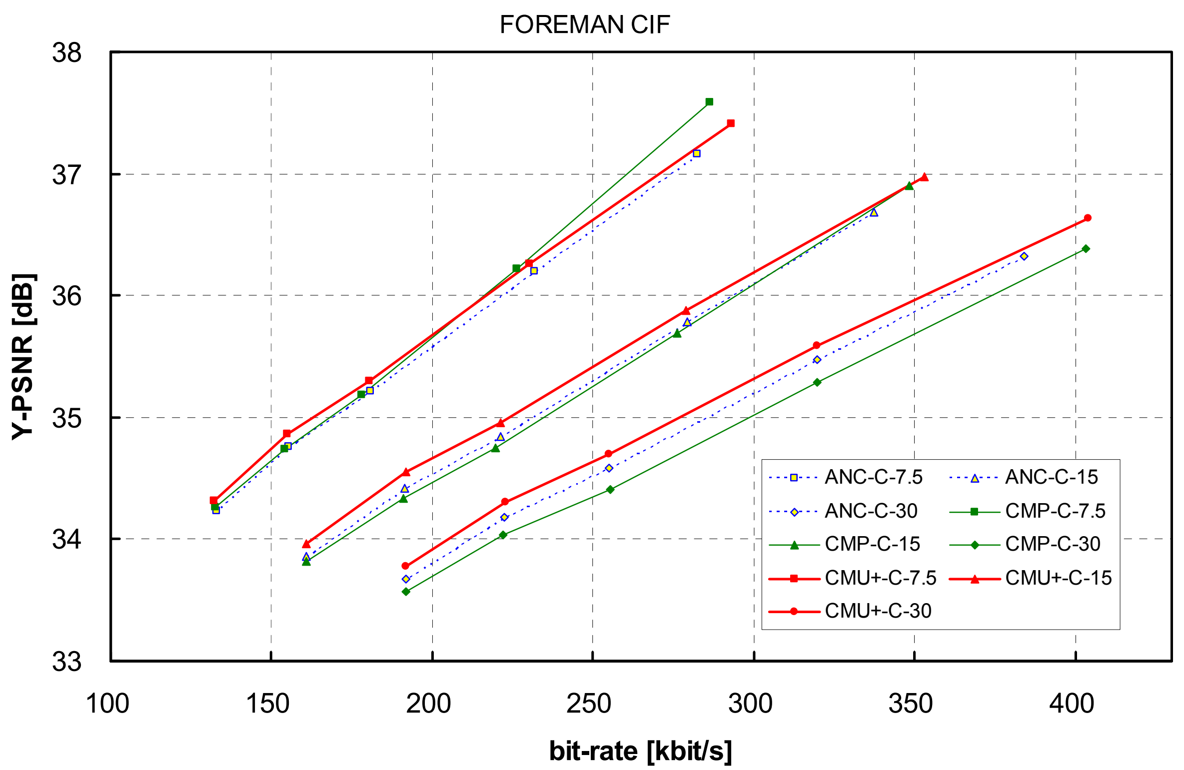
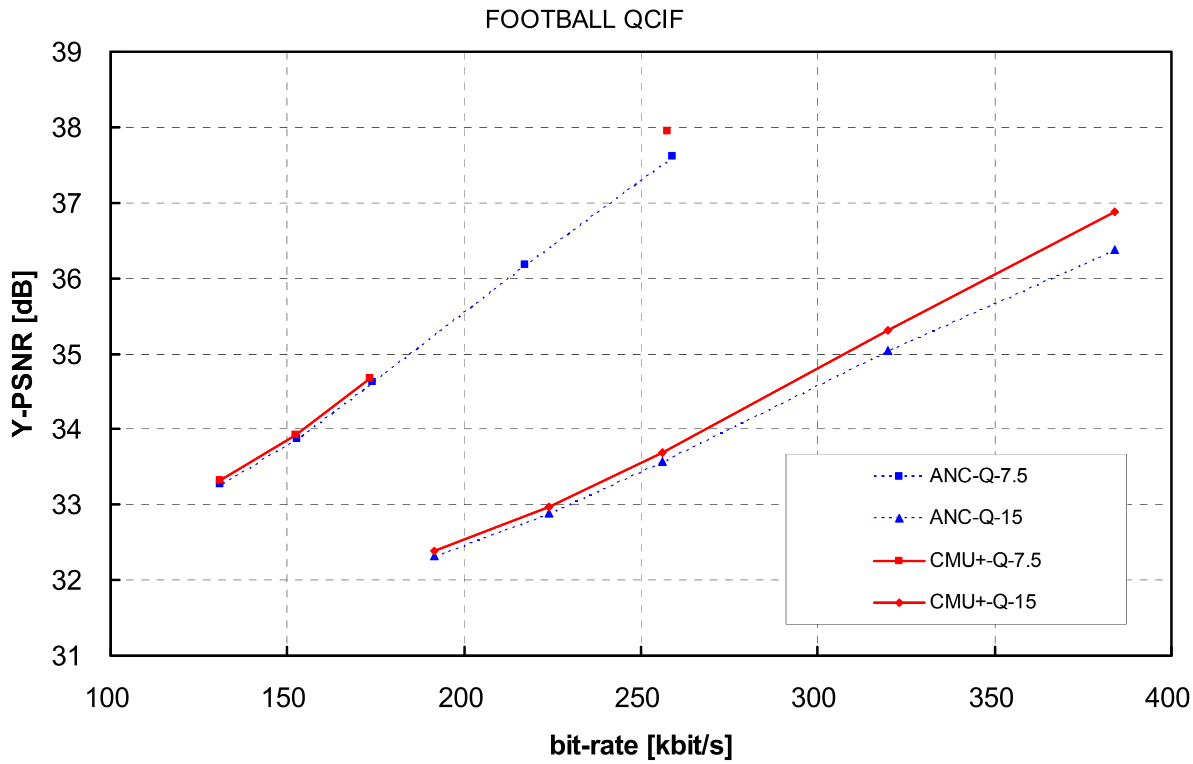


Figure 22 PSNR graph of FOREMAN sequences for JSVM2, CMP, and CMU+.

FOREMAN sequence shows the similar pattern except that the CMP shows the best performance in CIF@7.5Hz layer at the highest FGS layer. It can be explained that the effect of the update-step is sometimes negative in the very low frame-rate condition such as 7.5Hz. At the highest frame-rate condition in CIF layer, the PSNR gain of CMU+ is up to 0.31 dB compared to CMP and 0.13 dB compared to JSVM2. The performance gap compared to JSVM2 is not large, however, it should be noted that the complexity of CMU+ is much less than that of JSVM2.



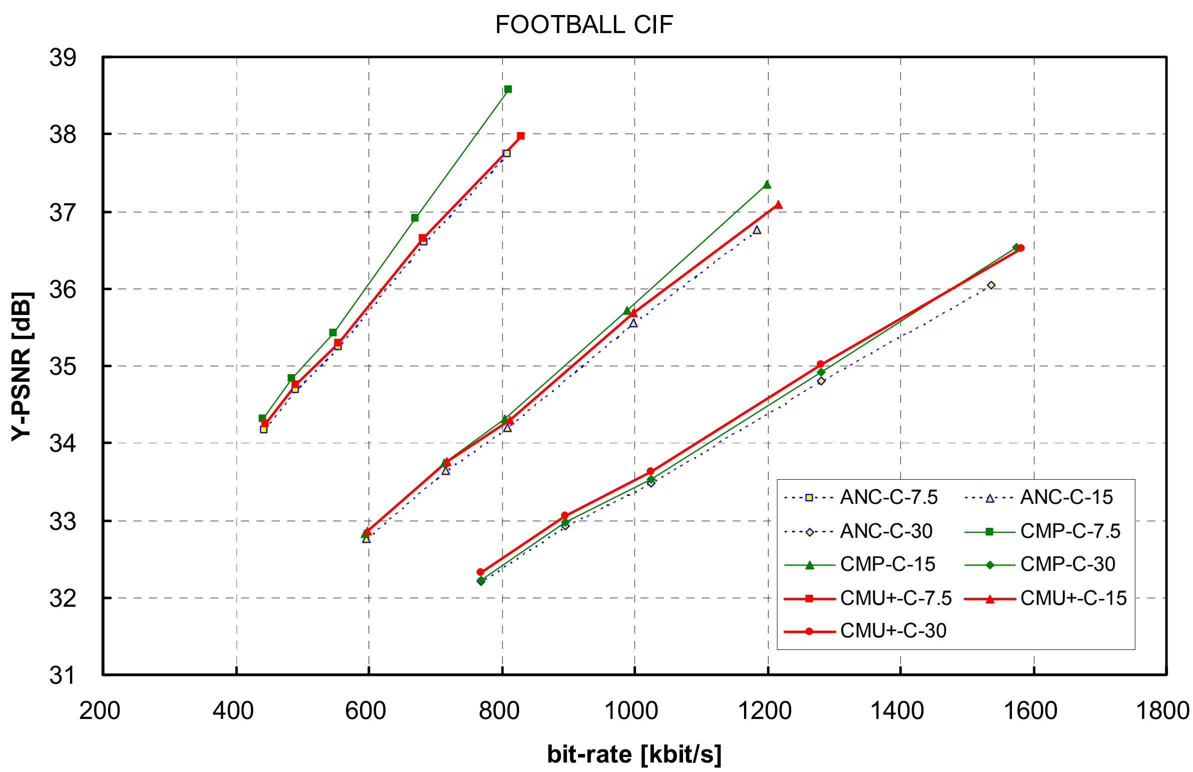
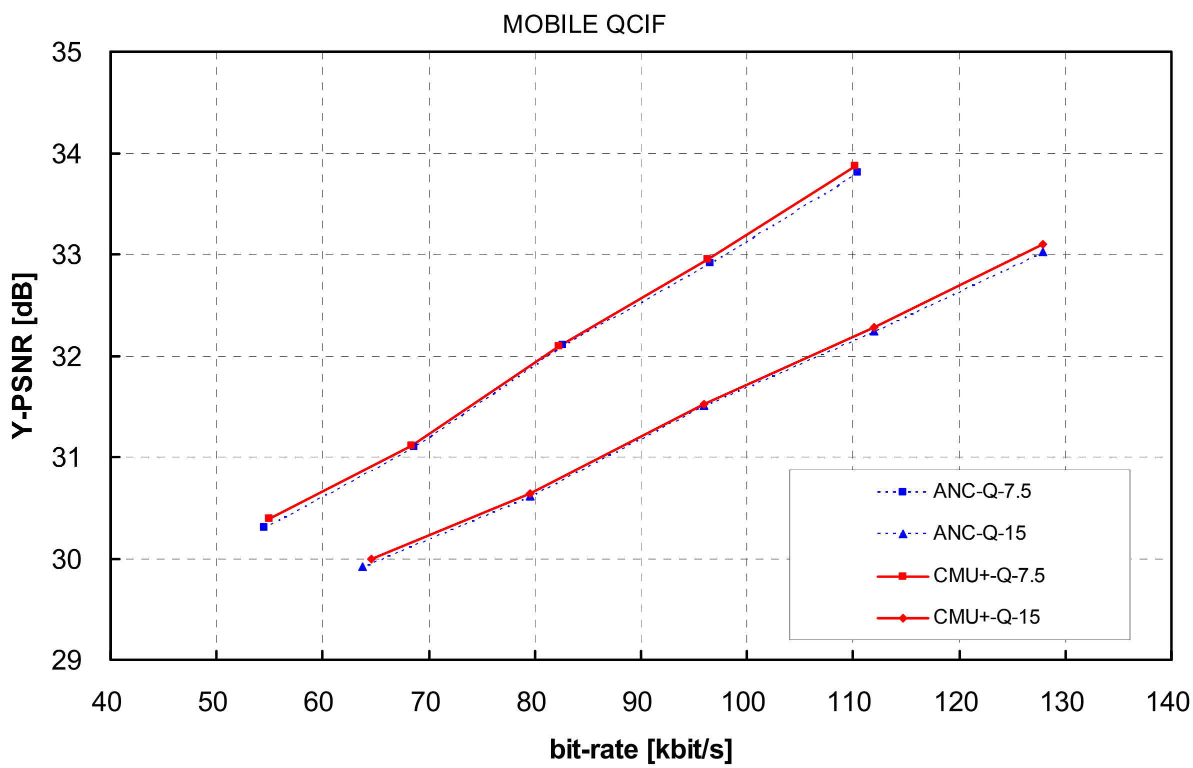


Figure 23 PSNR graph of FOOTBALL sequences for JSVM2, CMP, and CMU+.

Due to unknown reason, we cannot obtain one test-point in FOOTBALL QCIF@7.5Hz for both CMP and CMU+. All other test-points can be decoded normally. For FOOTBALL sequence, CMP shows the best performance in CIF@7.5Hz and CMP@15Hz, whereas CMU+ takes the best position in CIF@30Hz. It shows that the combination of CMP and CMU+ can further improve the average PSNR, thus it should be supported in a flexible way.



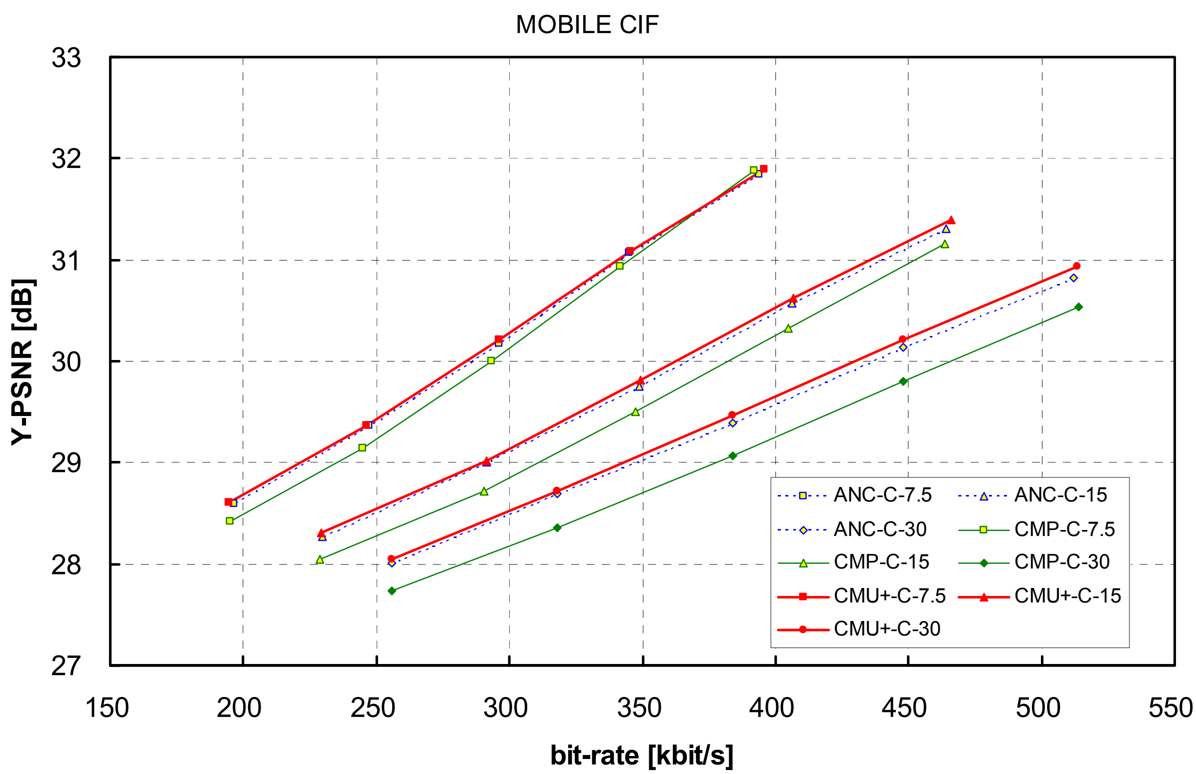
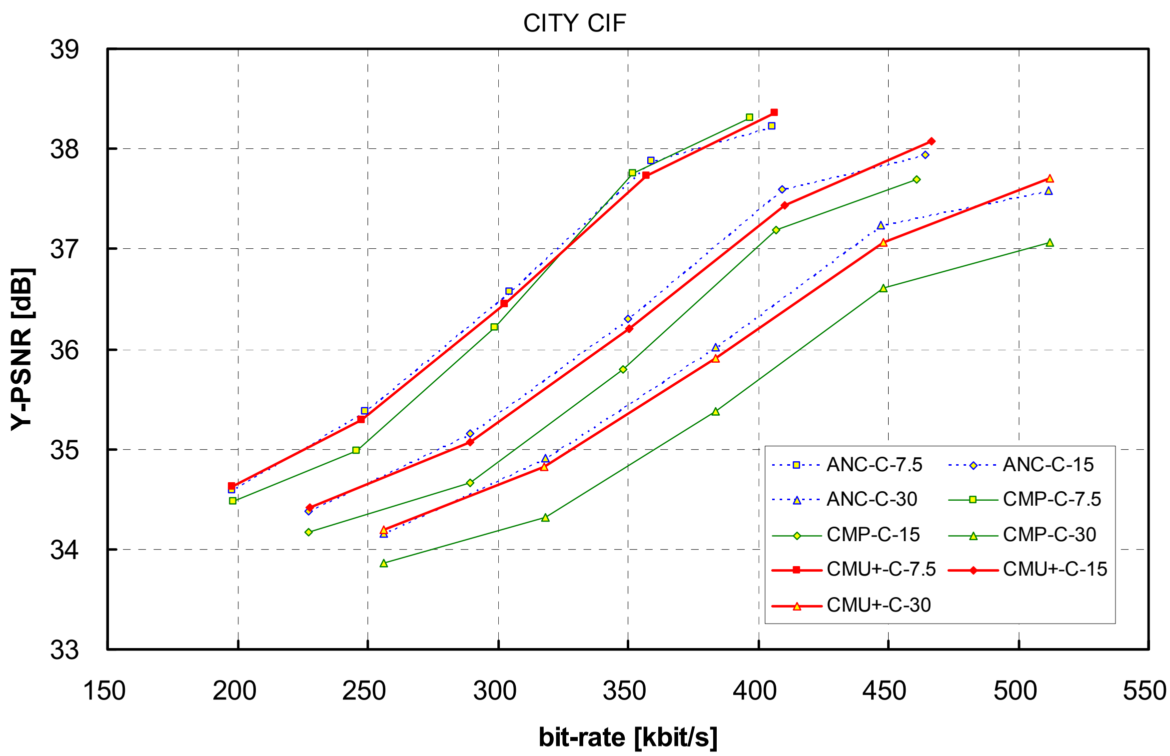


Figure 24 PSNR graph of MOBILE sequences for JSVM2, CMP, and CMU+.

For MOBILE sequence, QCIF performance of CMU+ is nearly same to the JSVM2 and 0.1 dB better than JSVM2 in CIF layers. Compared to CMP, CMU+ obtains PSNR gain up to 0.40 dB. This result shows that the CMU+ can do well even in the sequences having strong temporal correlations and good motion matches, and the closed-loop prediction concept is properly combined with the update-step by using the switched update-step.



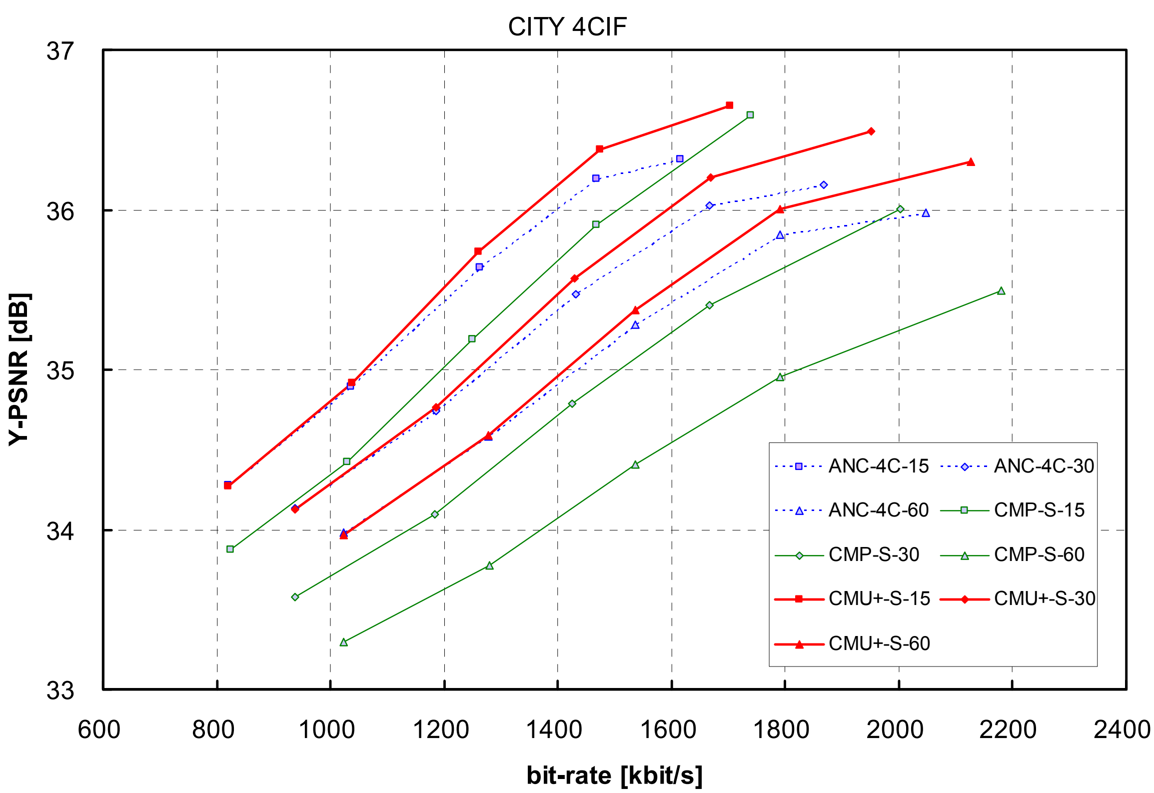
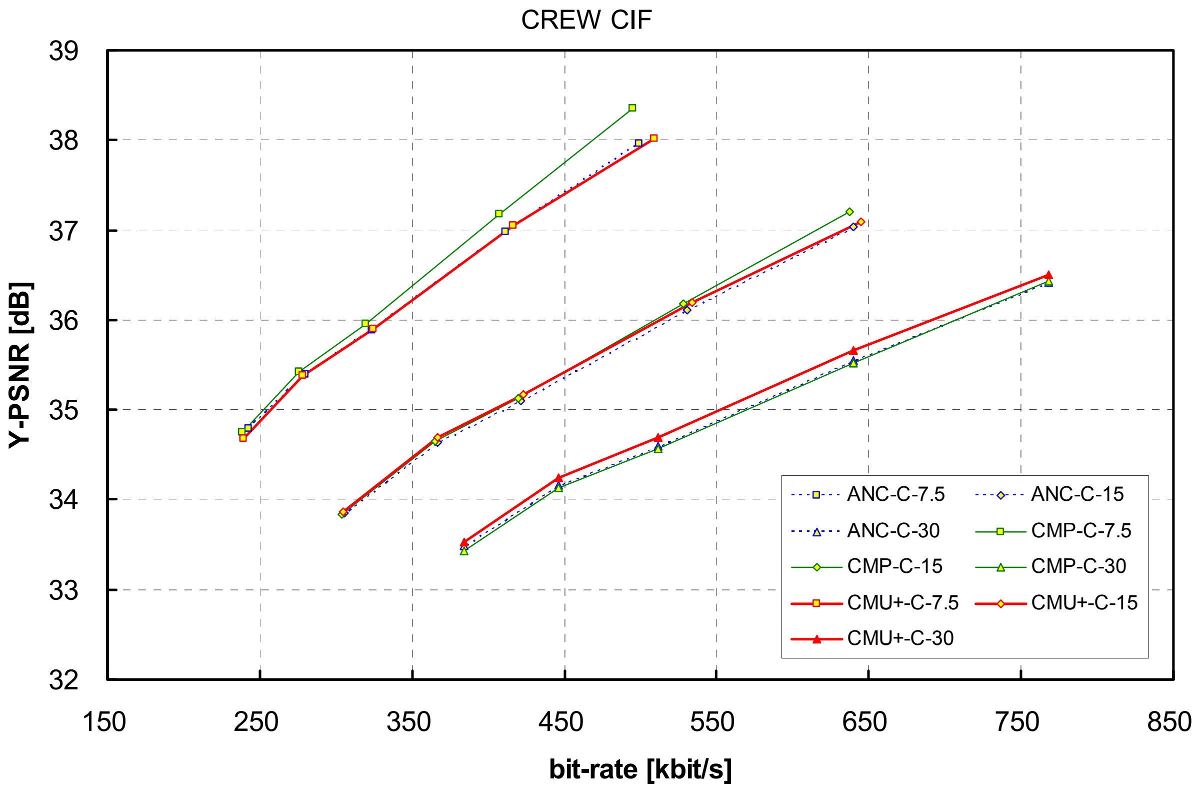


Figure 25 PSNR graph of CITY sequences for JSVM2, CMP, and CMU+.

CITY sequence has very steady motion fields and the texture detail is very high, which means the update-step may be useful. Therefore, CMP has clearly lower PSNR value up to 1.0 dB compared to CMU+ in 4CIF layer. It is interesting that CMU+ even surpasses the JSVM2 by a clear margin of 0.25 dB in 4CIF layer It shows that the closed-loop residual re-estimation concept is still valid when the update-step performs well.



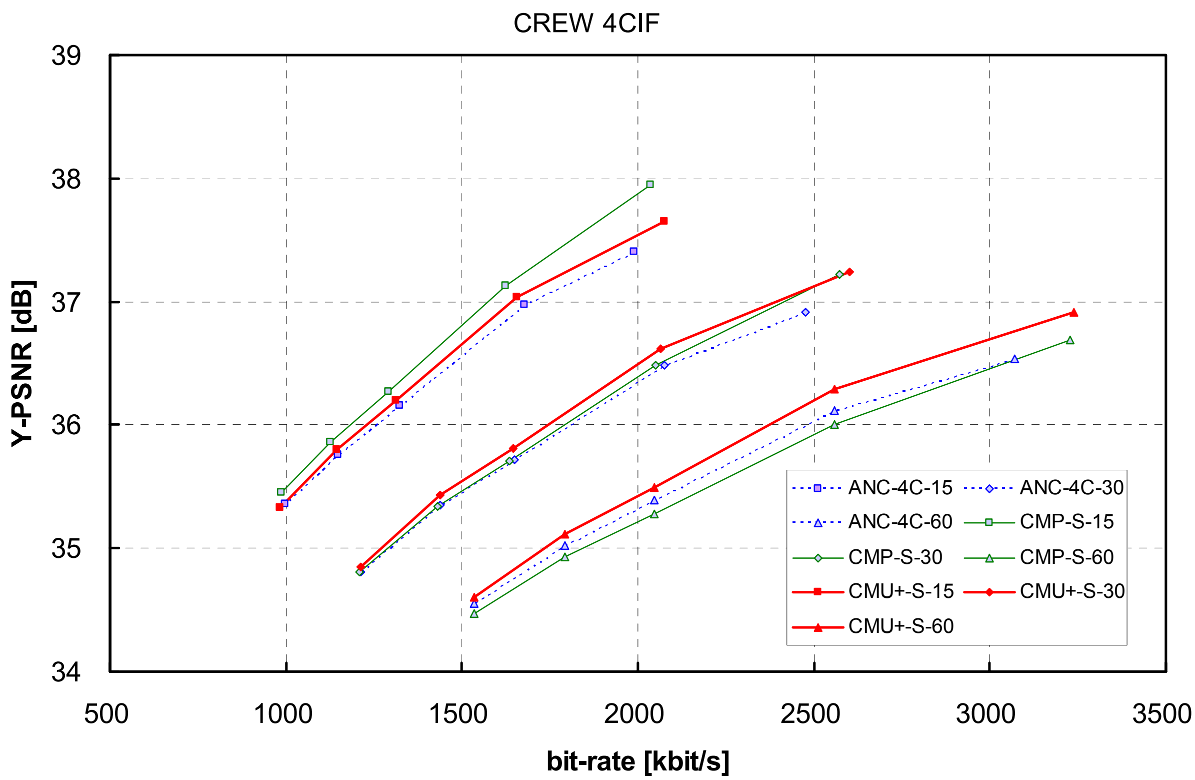
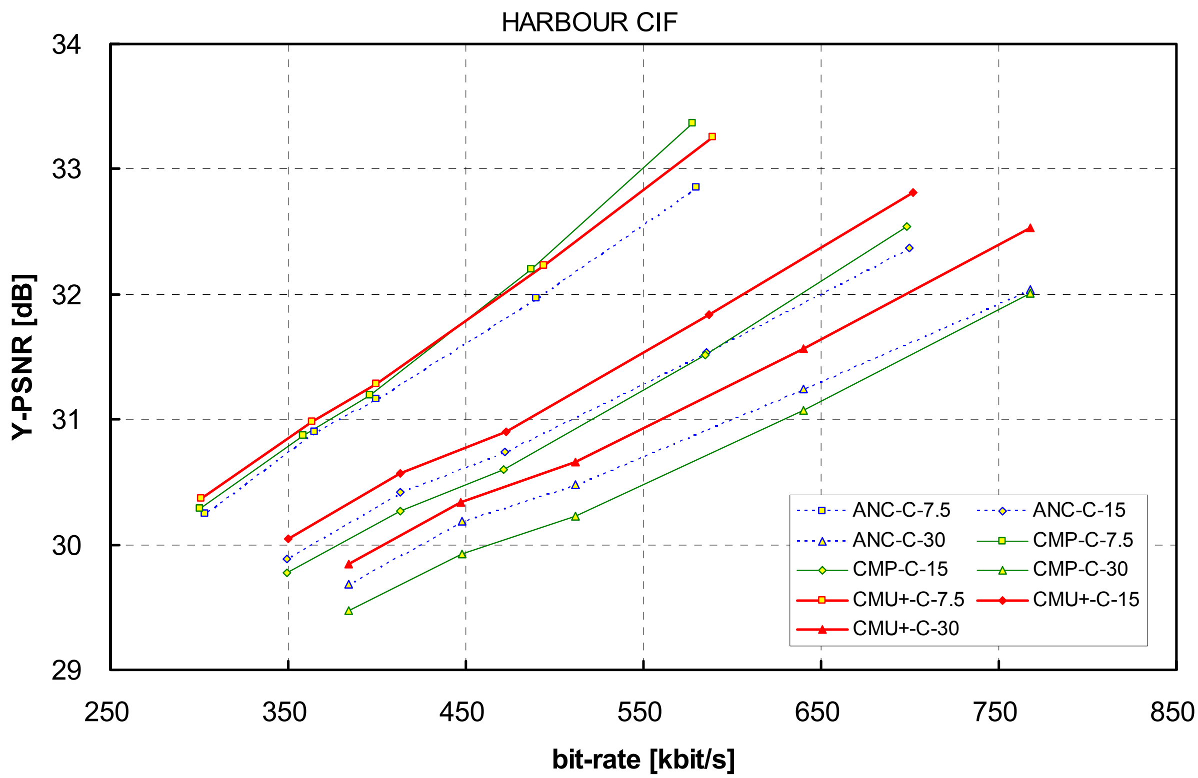


Figure 26 PSNR graph of CREW sequences for JSVM2, CMP, and CMU+.

It is interesting that for CREW sequence, which has many lightning effects, thus update-step is not effective, CMU+ outperforms CMP at 30Hz and 60Hz test points.



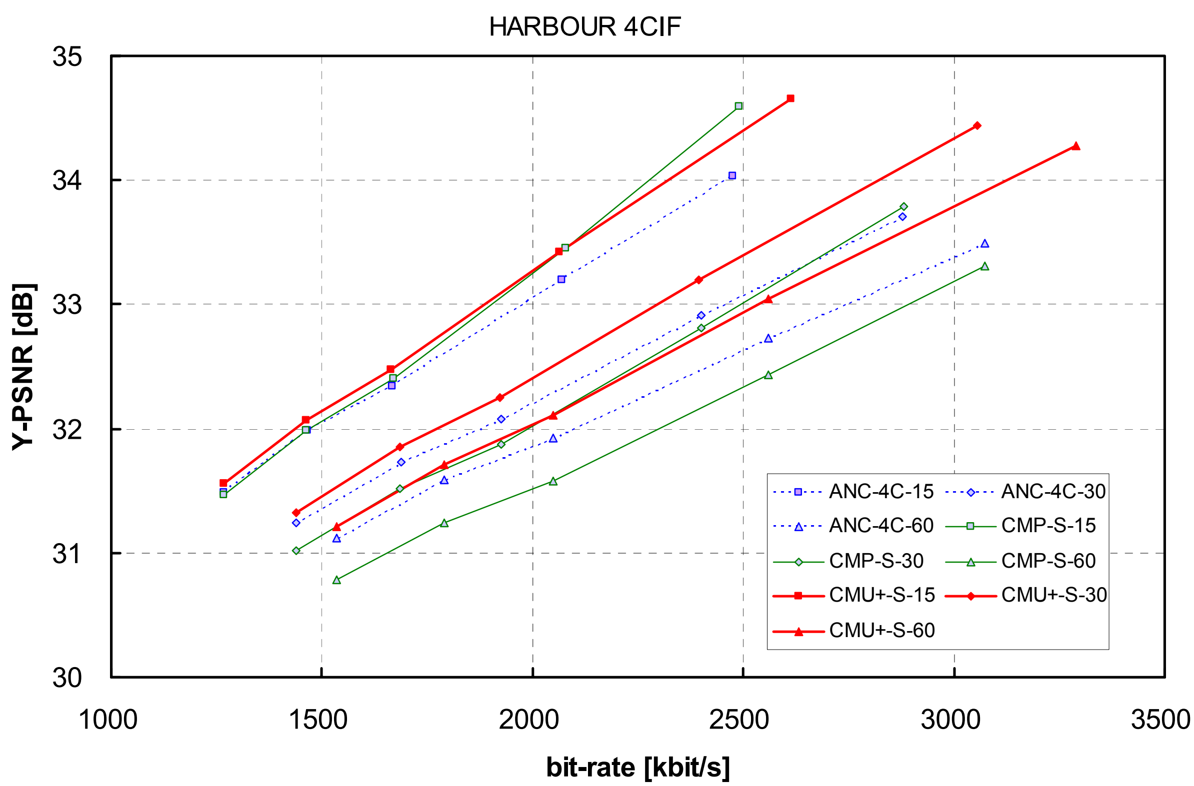
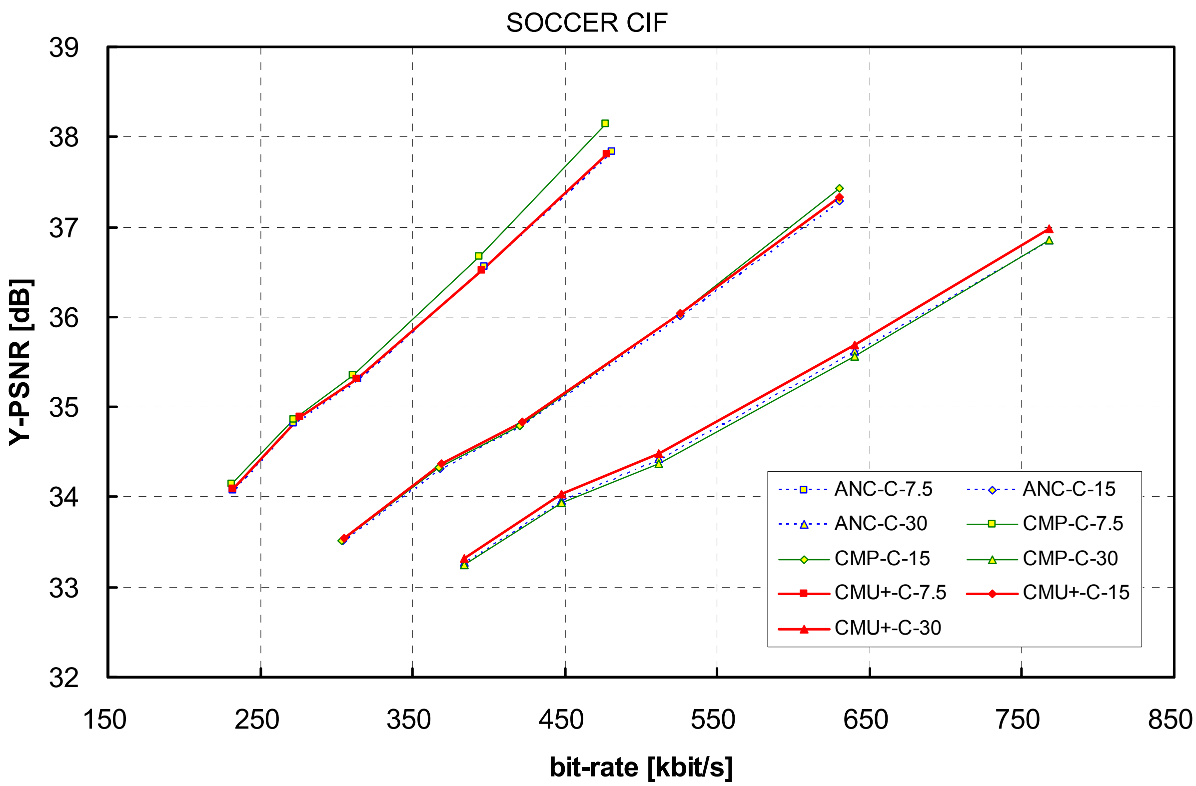


Figure 27 PSNR graph of HARBOUR sequences for JSVM2, CMP, and CMU+.

For HARBOUR sequence, the performance improvement of CMU+ is very clear. At the highest frame-rate in 4CIF layer, PSNR gain is up to 0.8 dB compared to JSVM2 and more than 1.0 dB compared to CMP.



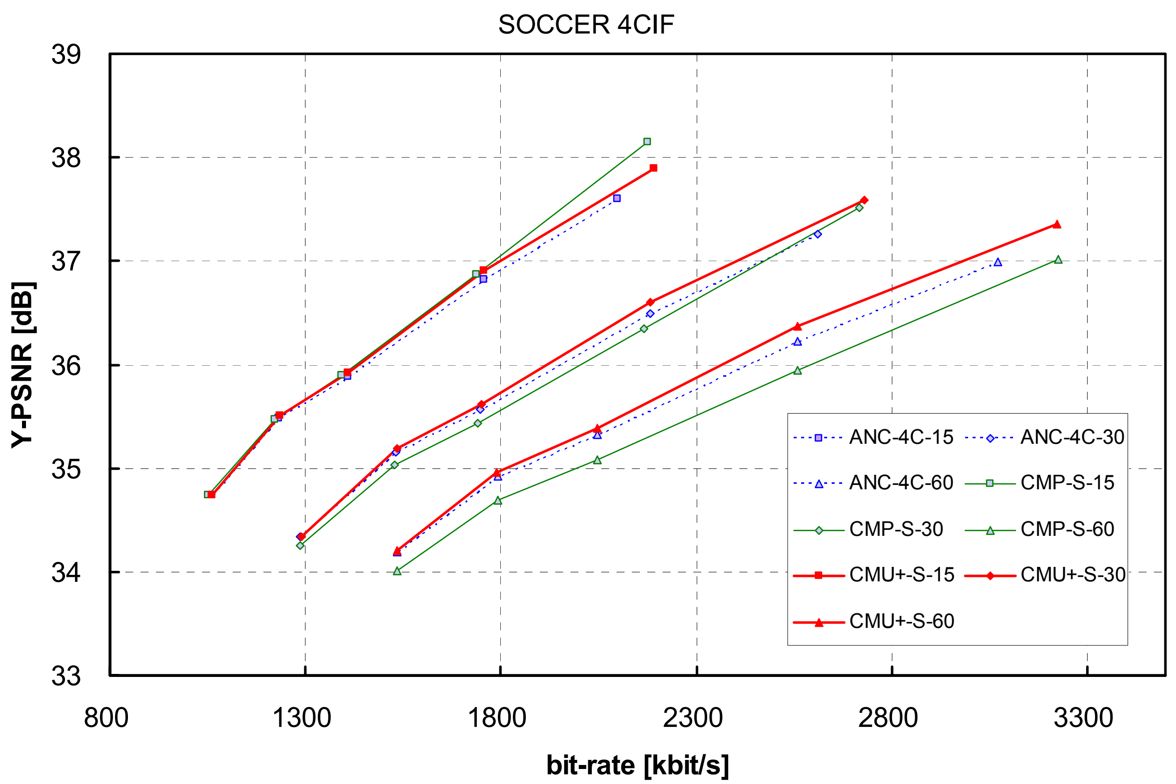


Figure 28 PSNR graph of SOCCER sequences for JSVM2, CMP, and CMU+.

Finally, in SOCCER sequence, CMU+ clearly outperforms CMP by 0.4 dB and slightly outperforms JSVM2 by 0.1 dB. From these results, it can be concluded that CMU+ really improves the performance compared to both JSVM2 and CMP.

7. Syntax, semantics, and decoding process changes

Syntax and semantics changes

There are no syntax and semantics changes even for the update-step skipping feature, since the update-step can be individually controlled in the current JSVM2 syntax for each slice.

Decoding process changes

As shown in the experiments, both Tool 1 and Tool2 should be combined although Tool 1 itself does not change the decoding process. Therefore, the decoding process should be changed to reflect the changes from Tool 2: switched update-step, which means that the decoding processing order should be changed. More concisely, in the decoder-side, the following sentences can be put in the draft:

- Inverse update-step is applied after the inverse prediction step using the decoded high-pass residuals.

Or, in other words, the following changes can be possible:

- Inverse prediction-step is performed by using the reference frames which are processed by the inverse update-step at the one-level lower temporal level. If the current temporal level is 1, non-updated reference frames are used.

It can be summarized as the following processes:

Closed-loop residual re-estimation with switched MCTF update-step [Decoder]

1. Decode the low-pass frame.
2. Decode the high-pass residual.
3. Perform inverse prediction-step to reconstruct the picture
4. Check whether the inverse update-step should be skipped. (update-step skipping feature)
5. Perform inverse update-step
6. Repeat 2-6 until all frames are reconstructed.

Non-normative: encoding process changes

Encoding process should be changed to reflect the closed-loop residual re-estimation and the switched update-step scheme. The motion compensated temporal filter section (Section 1.1.1 in JSVM2 document [3]) should be divided into two subsections, which are MCTF analysis and MCTF synthesis. In the MCTF analysis, no changes are required. MCTF synthesis section should cover the closed-loop residual re-estimation and the switched update-step. It can be summarized as the following processes:

<u>Closed-loop residual re-estimation with switched MCTF update-step [Encoder]</u>

1. Performs normal MCTF process.
2. Encodes the low-pass frame.
3. Re-estimates high-pass residual of temporal level N increased from 1 to the decomposition level-1.
4. Encodes the high-pass residual.
5. Performs inverse prediction-step to reconstruct the picture.using the reference frames in Step 3
6. Performs inverse update-step using the encoded high-pass residual in Step 4.
7. Repeat 3-6 until all high-pass residuals are re-estimated.

8. Conclusion

By applying the closed-loop concept to the MCTF prediction-step and modifying the update-step to remove the mismatch from the closed-loop prediction framework, we can improve the performance consistently up to clear margin of 0.6 dB (in HARBOUR 4CIF layers) with greatly reduced computational complexity. It results from the facts that the mismatch between an encoder and the decoder are removed by both the prediction and update steps. Especially, coding efficiency of the video sequences with fast motion and high residual energy can be improved by the closed-loop residual re-estimation procedure whereas the coding efficiency of the video sequences with slow motion and high texture can be improved by the proposed update-step.

References

[1] Woo-Jin Han and Bae-Keun Lee, "Closed-loop update-step in MCTF for scalable video coding," 15th JVT meeting, JVT-O062, Busan, Korea.

[2] Heiko Schwarz, Detlev Marpe, and Thomas Wiegand, "Further results on constrained inter-layer prediction," 15th JVT meeting, JVT-O073, Busan, Korea.

[3] Julien Reichel, Heiko Schwarz, and Mathias Wien, "Joint Scalable Video Model JSVM-2," 15th JVT meeting, JVT-O202, Busan, Korea.

[4] (Append for Proposal Documents)

JVT Patent Disclosure Form

International Telecommunication Union
Telecommunication Standardization Sector

International Organization for Standardization

International Electrotechnical Commission

Joint Video Coding Experts Group - *Patent Disclosure Form*

(Typically one per contribution and one per Standard | Recommendation)

Please send to:
JVT Rapporteur Gary Sullivan, Microsoft Corp., One Microsoft Way, Bldg. 9, Redmond WA 98052-6399, USA
Email (preferred): Gary.Sullivan@itu.int Fax: +1 425 706 7329 (+1 425 70MSFAX)

This form provides the ITU-T | ISO/IEC Joint Video Coding Experts Group (JVT) with information about the patent status of techniques used in or proposed for incorporation in a Recommendation | Standard. JVT requires that all technical contributions be accompanied with this form. *Anyone* with knowledge of any patent affecting the use of JVT work, of their own or of any other entity ("third parties"), is strongly encouraged to submit this form as well.

This information will be maintained in a "living list" by JVT during the progress of their work, on a best effort basis. If a given technical proposal is not incorporated in a Recommendation | Standard, the relevant patent information will be removed from the "living list". The intent is that the JVT experts should know in advance of any patent issues with particular proposals or techniques, so that these may be addressed well before final approval.

This is not a binding legal document; it is provided to JVT for information only, on a best effort, good faith basis. Please submit corrected or updated forms if your knowledge or situation changes.

This form is *not* a substitute for the *ITU ISO IEC Patent Statement and Licensing Declaration*, which should be submitted by Patent Holders to the ITU TSB Director and ISO Secretary General before final approval.

Submitting Organization or Person:

Organization name	**Samsung Electronics**
Mailing address	**416, Maetan 3-dong, Paldal-gu, Suwon-si, Gyeonggi-do, 442-742**
Country	**Korea**
Contact person	**Woo-Jin Han**
Telephone	+82-31-200-9319
Fax	+82-31-200-3427
Email	wjhan.han@samsung.com
Place and date of submission	

Relevant Recommendation | Standard and, if applicable, Contribution:

Name (ex: "JVT")	**JVT**
Title	**CE6 response of Samsung Electronics: in-depth comparison of closed-loop and open-loop MCTF structures**
Contribution number	

(Form continues on next page)

Disclosure information – Submitting Organization/Person (choose one box)

☐ 2.0 The submitter is not aware of having any granted, pending, or planned patents associated with the technical content of the Recommendation | Standard or Contribution.

or,

The submitter (Patent Holder) has granted, pending, or planned patents associated with the technical content of the Recommendation | Standard or Contribution. In which case,

☐ 2.1 The Patent Holder is prepared to grant – on the basis of reciprocity for the above Recommendation | Standard – a free license to an unrestricted number of applicants on a worldwide, non-discriminatory basis to manufacture, use and/or sell implementations of the above Recommendation | Standard.

■ 2.2 The Patent Holder is prepared to grant – on the basis of reciprocity for the above Recommendation | Standard – a license to an unrestricted number of applicants on a worldwide, non-discriminatory basis and on reasonable terms and conditions to manufacture, use and/ or sell implementations of the above Recommendation | Standard.

Such negotiations are left to the parties concerned and are performed outside the ITU | ISO/IEC.

☐ 2.2.1 The same as box 2.2 above, but in addition the Patent Holder is prepared to grant a "royalty-free" license to anyone on condition that all other patent holders do the same.

☐ 2.3 The Patent Holder is unwilling to grant licenses according to the provisions of either 2.1, 2.2, or 2.2.1 above. In this case, the following information must be provided as part of this declaration:

- patent registration/application number;
- an indication of which portions of the Recommendation | Standard are affected.
- a description of the patent claims covering the Recommendation | Standard;

In the case of any box ***other than 2.0*** *above, please provide the following:*

Patent number(s)/status ____

Inventor(s)/Assignee(s) ____

Relevance to JVT ____

Any other remarks: ____

(please provide attachments if more space is needed)

(form continues on next page)

Third party patent information – fill in based on your best knowledge of relevant patents granted, pending, or planned by other people or by organizations other than your own.

Disclosure information – Third Party Patents (choose one box)

☐ 3.1 The submitter is not aware of any granted, pending, or planned patents *held by third parties* associated with the technical content of the Recommendation | Standard or Contribution.

☐ 3.2 The submitter believes third parties may have granted, pending, or planned patents associated with the technical content of the Recommendation | Standard or Contribution.

For box 3.2, please provide as much information as is known (provide attachments if more space needed) - JVT will attempt to contact third parties to obtain more information:

3rd party name(s) ______________________________

Mailing address ______________________________
Country ______________________________
Contact person ______________________________
Telephone ______________________________
Fax ______________________________
Email ______________________________
Patent number/status ______________________________
Inventor/Assignee ______________________________
Relevance to JVT ______________________________

Any other comments or remarks: