

This paper has also been submitted to the 2nd ACM Workshop on Information Hiding and Multimedia Security

# Slice Groups for Post-Compression Region of Interest Encryption in SVC

Andreas Unterweger  
University of Salzburg  
Department of Computer Sciences  
Jakob-Haringer-Straße 2  
Salzburg, Austria  
aunterweg@cosy.sbg.ac.at

Andreas Uhl  
University of Salzburg  
Department of Computer Sciences  
Jakob-Haringer-Straße 2  
Salzburg, Austria  
uhl@cosy.sbg.ac.at

## ABSTRACT

In this paper, we assess the adequacy of slice groups for the reduction of drift which occurs in bit-stream-based region of interest encryption in SVC. For practical surveillance camera applications, we introduce the concept of all-grey base layers which simplify the encryption of regions of interest while obeying all standard-imposed base layer restrictions. Furthermore, we show that the use of slice groups is possible with relatively low overhead for most practical configurations with two or three spatial layers. In addition, we analyze the effect of spatial resolution on overhead, showing that an increase in resolution decreases the relative overhead.

## Categories and Subject Descriptors

I.4.2 [Image Processing and Computer Vision]: Compression (Coding)—SVC, Slice groups; E.3 [Data]: Data Encryption—Selective encryption

## General Terms

Experimentation, Measurement, Verification

## Keywords

SVC, Slice groups, Selective encryption, Region of Interest, Overhead

## 1. INTRODUCTION

In video surveillance and other applications, there is often the need to disguise people’s identities in order to protect their privacy. A common approach to achieve this is the selective encryption of people’s faces (also called region-based selective [15] or Region of Interest (RoI) encryption), i.e. encrypting all picture areas which contain a face, while leaving all other picture areas untouched.

This allows for reversible de-identification, i.e., the disguise of identities with the possibility to restore them by undoing the encryption. Restoring is typically only possible with a



**Figure 1: Examples for spatial and temporal drift:** The first, second and tenth frame (from left to right) of the *foreman* sequence (top: original, bottom: encrypted) where one macroblock block row around the eyes in the first frame (left) has been encrypted.

correct key which is possessed, e.g., by law enforcement authorities in case suspects of a crime need to be identified. Although several techniques for reversible de-identification exist, RoI encryption is one of the most common ones in video surveillance.

While RoI encryption can be applied before (e.g., [2, 3, 7]), during (e.g., [23, 20, 11]) or after compression (e.g., [24, 4, 6]), each with its own advantages and disadvantages [14], most approaches proposed so far focus either on encryption before or during compression. Although this makes drift, i.e., the propagation of parts of encrypted picture areas into non-encrypted ones through spatial and temporal prediction as depicted in figure 1, easier to manage, it does not allow using existing surveillance infrastructure whose input images and/or encoder cannot be modified.

Typically, surveillance cameras have compression hardware built in (as of 2013, Motion JPEG and H.264 are very common) which is used to reduce the bandwidth of the captured and transmitted video footage. Although this saves time and computational resources by not requiring additional encoding hardware, it makes modifications (like an additional encryption step) to the built-in compression hardware quasi impossible due to the often hard-wired encoder.

In order to be able to reuse this infrastructure notwithstanding, applying RoI encryption after compression has to be considered, reviving the drift issue. Therefore, in this paper, we try to assess the fitness of the slice group coding

tool of Scalable Video Coding (SVC) [17] for allowing to selectively encrypt picture areas and containing drift. For the sake of applicability, we consider a state-of-the-art video surveillance system which delivers SVC bit streams. We assume that the surveillance system detects faces (or other regions of interest) and places them in slice groups which are to be encrypted after compression. The main reason for using slice groups is their ability to contain drift to a certain extent, thereby simplifying RoI encryption. Note that slice groups have other uses as well, thereby extending the results of our investigations to scenarios which are not encryption-specific.

By evaluating the limitations and possibilities of slice group coding, we aim at determining whether or not the aforementioned setup simplifies the encryption process in terms of drift. Furthermore, we evaluate the overhead induced by this approach in order to determine whether or not it is of practical use, i.e., for example to be included into existing and/or future surveillance systems to simplify RoI encryption after compression.

Related work on RoI encryption in SVC is sparse. Two approaches are proposed in [21] and [10], albeit without considering or compensating for the effects of drift, which is an important matter. [11] deals with drift by imposing restrictions on the encoding process in terms of a limited motion estimation range as well as interpolation and upsampling constraints. Besides the reported significant increase in bit rate, this method cannot be applied on a bit stream level without recompression. Similarly, [19] proposes separate RoI coding by restricting motion estimation and inter-layer prediction, albeit without the explicit intention to do so for the sake of encryption. However, all of these approaches are in-compression encryption methods and cannot be applied at bit stream level.

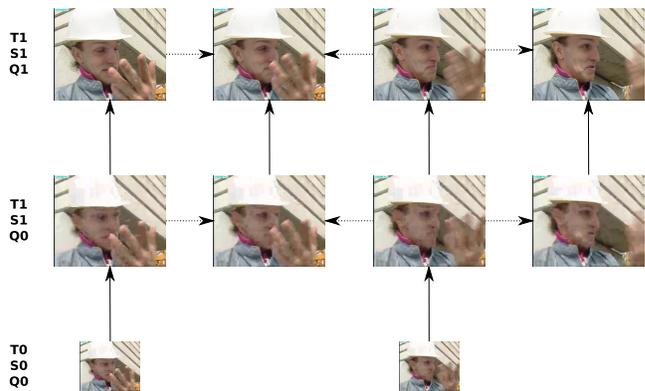
Apart from RoI-related experiments and analyses of SVC, the encryption of certain Network Abstraction Layer (NAL) units has been proposed in [13]. However, their proposed encryption approach yields bit streams which are no longer format compliant and can hence not be decoded anymore by a regular decoder. This is not desirable in surveillance applications as the background without the encrypted RoI should be visible and therefore decodable. Furthermore, the optimization of RoI across multiple layers to lower the total bit rate has been analyzed in [8].

Although slice groups have been used to deal with drift in a number of encryption approaches (e.g., [23, 5]), a detailed examination of its actual usefulness to contain different causes of drift has not been done so far. The overhead induced by some of the aforementioned encryption approaches has been analyzed, but this is not true for the general overhead introduced by slices groups which change from frame to frame to cover RoI. This is especially true for SVC.

This paper is structured as follows: In section 2, the key concepts of video coding with slice groups in SVC are described, followed by an analysis of their limitations in section 3. After evaluating several scenarios in terms of feasibility for video surveillance with encrypted RoI in section 4, we conclude our paper.

## 2. Scalable Video Coding

SVC is specified as the scalable extension of H.264, specified in its Annex G [9]. It allows for multiple so-called layers within one bit stream, which can be accessed or ex-



**Figure 2: SVC with multiple layers: The base layer with half the frame rate and a quarter of the picture size can be used to predict the first spatial enhancement layer, which itself can be used to predict a second temporal and subsequently a third SNR enhancement layer. Adopted from [17]**

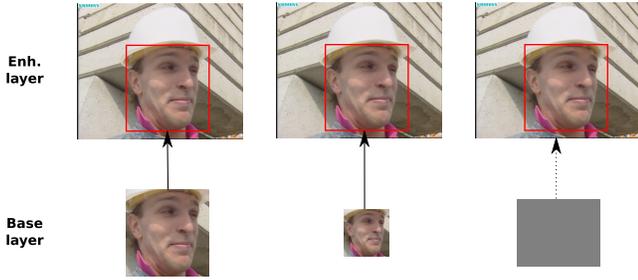
tracted depending on the capabilities of the device decoding the stream. Each layer differs from the others either by frame rate (temporal scalability), resolution (spatial scalability) or quality (Signal-to-Noise Ratio (SNR) scalability). The bottom-most layer is referred to as base layer and coded in a way that is compatible with (non-scalable) H.264.

All layers but the base layer can exploit inter-layer redundancies by using coded information of lower layers for prediction. The basis of this prediction for spatial and SNR scalability can either be filtered intra-coded samples (inter-layer intra prediction), motion vectors (inter-layer motion prediction) or inter-coded difference signal samples (inter-layer residual prediction), with details for each prediction type to be found in [18]. In contrast, temporal scalability is achieved through hierarchical inter prediction as explained in detail in [17].

Figure 2 shows an example of a scalable bit stream with multiple layers. The base layer (temporal layer 0 (T0), spatial layer 0 (S0) and SNR layer 0 (Q0)) has the lowest possible frame rate, resolution and quality and is used to predict the first spatial enhancement layer (T0, S1, Q0; not labeled) which doubles both, picture width and height. This enhancement layer is further used to predict an enhancement layer of the same resolution, but a doubled frame rate (T1, S1, Q0) as well as an enhancement layer with higher quality (T0, S1, Q1; not labeled) and subsequently a doubled frame rate (T1, S1, Q1).

In each layer, a coded picture is split into slices which can be summarized to slice groups of specific forms, depending on the so-called slice group map type. As RoI encryption requires a background left-over, i.e., a region of the picture which does not belong to any encrypted region of interest, only slice group map types 2 (foreground slice groups with left-over background) and 6 (explicit slice group specification) will be considered, as only they allow this. Since slice group map type 6 is practically identical to slice group map type 2 in this use case, we will only consider slice group map type 2 henceforth.

To exploit spatial and temporal redundancy, samples can be predicted from neighboring samples around the block to be



**Figure 3: Alternatives to slice groups in the base layer: Left and middle: Extended spatial scalability; right: all-grey base layer**

predicted (in the same picture) as well as from samples of arbitrary blocks in previously coded pictures. In the former case, predictions over slice borders are forbidden, thereby allowing all slices to be decoded independently and preventing spatial drift.

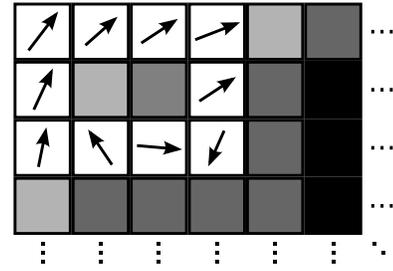
### 3. STANDARD-IMPOSED LIMITATIONS

The H.264 standard imposes restrictions on coding tools and parameter values by specifying profiles. As this paper discusses slice groups, we only consider profiles which allow the use of multiple slice groups in the first place. In this section, we investigate other relevant limitations imposed by those profiles.

For scalable bit streams, only the Scalable Baseline profile supports slice groups. Although it allows using up to eight slice groups in total, one slice group is considered to be the background, i.e., the remainder of what the other seven slice groups encode. In addition, entropy coding is limited to CAVLC, the number of slice groups cannot exceed seven (plus background) and B slices are not allowed. Furthermore, the base layer may not contain more than one slice group.

This is a severe limitation in an encryption scenario because this means that the regions of interest cannot be in separate slice groups in the base layer. Thus, either a different drift compensation approach for the base layer is required or an alternative to slice groups in the base layer has to be found. As the former is hard to achieve, we consider three additional alternatives to slice groups in the base layer as depicted in Figure 3.

One possibility is to use extended spatial scalability, depicted on the left and in the middle of Figure 3, where the base layer only contains the region of interest and the enhancement layer adds the rest of the video frame. Due to the limitations of the Scalable Baseline profile, the width and height ratios between the base layer and the corresponding region of interest in the enhancement layer have to be either 1 (Figure 3, left), 1.5 (not depicted) or 2 (Figure 3, middle). However, this setup is only useful if there is exactly one region of interest. Since this would impose a severe practical limitation, it is not considered in the remainder of this paper. Alternatively, we propose adding a base layer which is all-grey ( $Y = C_b = C_r = 128$ ) as shown in Figure 3, right. Since intra DC prediction and skip modes allow encoding such an artificial layer very compactly, its overhead is relatively small when using the maximum possible width and height ratios of 2, i.e., a base layer with half the width and



**Figure 4: Constrained intra prediction: In a P slice, intra blocks may not use inter blocks for prediction. The grey level of the depicted intra blocks denotes the number of allowed intra modes**

height of the enhancement layer.

However, it effectively reduces the number of usable spatial layers, which is limited to three in the Scalable Baseline profile, by one. This allows for a maximum of two non-grey spatial layers for actual video content. Depending on the use case, these two remaining layers may be sufficient to provide spatial scalability.

Despite the loss of one usable spatial layer, the grey base layer simplifies encryption by containing drift. Although the unavailability of slice groups in the base layer (see above) would normally make encryption harder (without the possibility of using slice groups to contain drift), the fact that the base layer is all grey does not require any encryption and does therefore not induce any drift.

Although there have been multiple proposals for region-of-interest support through slice groups in all layers [1, 22], the final version of the standard does not allow this. Similarly, the technique proposed in [12] to alternatively support regions of interests as enhancement layers is not supported. This paper limits the available options to the ones supported by the standard, i.e., the all-grey base layer introduced above as well as a regular (i.e., full-content) base layer for comparison.

Regrading further limitations imposed by the standard, we will focus on the combination of constrained intra prediction and constrained inter-layer prediction, which ensure single-loop decoding [16]. Since these two limitations severely limit the number of possibilities for prediction and therefore drift, they are crucial for the RoI encryption use case.

Constrained intra prediction limits the blocks which can be used for intra prediction. Figure 4 illustrates this in a P slice which contains inter (depicted by motion vectors) and intra (depicted by grey levels) macroblocks. Although the black intra blocks may use all possible intra prediction modes, the dark- and light-grey ones may not. For example, the light-grey macroblock at the top left may only use DC prediction since all other prediction directions would require predicting from one of the surrounding inter macroblocks.

SVC enforces constrained intra prediction in all layers which are used for inter-layer prediction so that inter-layer predicted samples do not require additional motion compensation in the base layer. Additionally, constrained inter-layer prediction ensures that inter-layer-predicted intra samples are not used for intra prediction themselves, as illustrated in Figure 5.

Inter-layer prediction allows using information from the base layer in the enhancement layer. If blocks are upsampled

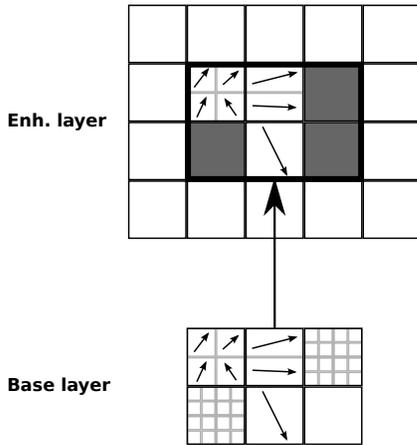


Figure 5: Constrained inter-layer prediction: Upsampled intra blocks (grey) must be reconstructed from base layer intra samples



Figure 6: Moving slice groups: Frames 1, 11 and 21 of the *foreman* sequence with one moving foreground slice group around the face (green) and one background slice group (remainder, turquoise)

through inter-layer intra prediction (grey blocks in Figure 5), the corresponding reference block in the base layer has to be an intra block as well. Constrained inter prediction in the base layer ensures that no additional motion compensation loop is required. Furthermore, if the enhancement layer is used for further inter layer prediction, the upsampled blocks may not be used for further intra prediction due to the constrained intra prediction requirement to avoid multi-loop decoding.

## 4. EXPERIMENTAL EVALUATION

In this section, we describe our experimental setup and results. We refer to the term of "moving slice groups" for RoI herein since the position of RoI may change from frame to frame, thereby changing the slice group positions accordingly, as illustrated in Figure 6. Recall that our use case is encryption, i.e., we assume that the moving slice groups will be encrypted at some point, as illustrated by example in figure 7.

### 4.1 Setup

In order to evaluate the effect of slice-group-based RoI for encryption, we added support for moving slice groups to the SVC reference software (*JSVM*) since it does not support this by itself.

In the *JSVM*, slice group coding is implemented partially, but not used. Therefore, it is enabled separately for all spatial layers but the base layer which does not support slice group coding (see section 3). This is done by setting the



Figure 7: Encrypted RoI: Frames 1, 11 and 21 of the *foreman* sequenc. The RoI in this example is the actor's face

slice group map type to 2 using the current layer's Picture Parameter Set (PPS) in *LayerEncoder::process*. In each layer, the RoI coordinates are calculated depending on the picture size and the corresponding slice group settings (number of slice groups, top-left and bottom-right coordinates) in the PPS are adapted accordingly. In order to determine the absolute frame number in the layer being processed, a helper variable is introduced which counts the number of processed Group Of Pictures (GOP). Together with the frame index of the current GOP (which is provided by the encoder), an absolute frame number can be calculated so that the corresponding RoI coordinates can determined. In order to signal the slice groups, one additional PPS per frame and enhancement layer is needed. Although the PPS changes described above take effect in the encoder immediately, the decoder needs to take notice of them by a PPS update. This requires inserting one PPS per frame per enhancement layer into the bit stream using the corresponding functions provided by the *NalUnitEncoder* class. Note that it is essential to use the *LayerEncoder::xAppendNewExtBinDataAccessor* and *LayerEncoder::addParameterSetBits* functions to assign the PPS NAL unit and its corresponding overhead to the correct layer.

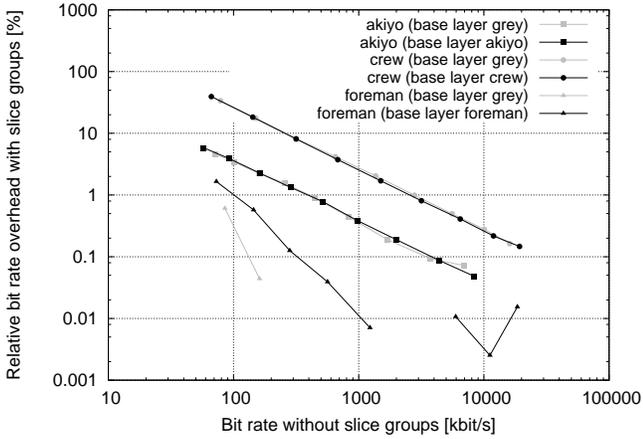
We use three test sequences with 300 frames each to simulate typical surveillance scenarios. *akiyo* has one RoI and very little motion, while *foreman* has a significant amount of motion and also one RoI. Conversely, the *crew* sequence has a changing number of RoI. Since a maximum of seven slice groups (RoI) is supported in SVC (see section 3), only the first top-left-most faces are considered, i.e., placed in a separate slice group. All faces were segmented manually by enclosing them in rectangles. The corresponding coordinates were rounded to the nearest macroblock border.

We use both, Common Intermediate Format (CIF) and 4CIF resolution, in order to determine the impact of spatial resolution on the measurements. While the following section gives a detailed description of the results for CIF resolution, section 4.3 describes the differences when using 4CIF resolution.

### 4.2 Overhead (CIF)

We use the GOP size of the default JSVM configuration, i.e., four. Since GOP structures with B frames are not allowed in combination with slice groups (see section 3), we use P frames instead. Thus, an  $(IPPP)^*$  GOP structure, i.e., a repeated sequence of one I frame and followed by three P frames, is used.

We encode the test sequences with a constant Quantization Parameter (QP) for both frame types and default settings with two and three dyadic spatial layers. The base layer is all grey (see section 3), although we test "classical" base layers (with the actual down-sized input video) as well for



**Figure 8: Overhead with slice group coding for different CIF sequences when using two dyadic spatial layers**

comparison. Inter-layer prediction is set to adaptive to allow for optimal coding efficiency.

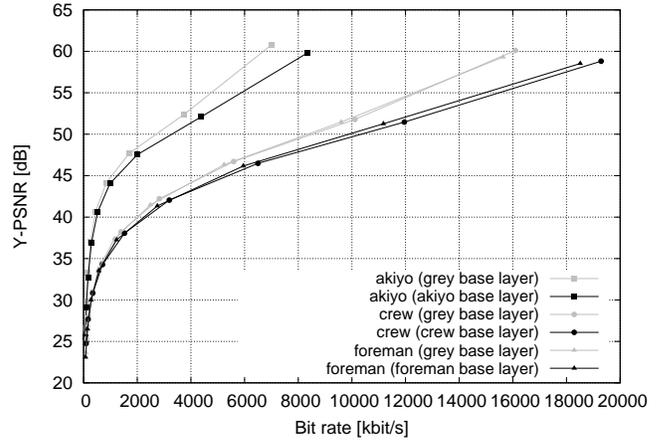
We encode the test sequences with a constant QP for all frame types and default settings. Using QPs between 3 and 51 with a step size of 6 to double the quantizer step size with each run allows covering the whole QP range. Each QP-sequence combination is encoded with and without slice groups. Since the difference in terms of distortion between the encoded sequences with and without slice groups is very small ( $< 0.15$  dB), we approximate the overhead introduced by slice group coding by comparing the corresponding bit rates directly.

As depicted in Figure 8, it is obvious that the *crew* sequence (depicted by circles in figure) exhibits the highest overhead in quasi all scenarios, since it requires the highest number of slice groups. Conversely, the *foreman* sequence exhibits the lowest overhead, since it requires only one additional slice group (apart from the background) for the first half of the sequence. It profits from scalability more than the other sequences, resulting in some very small negative overhead values ( $< 0.1\%$  absolute). Note that these values cannot be depicted properly due to the logarithmic Y axis.

In general, the overhead decreases with the bit rate, i.e., it increases with the QP. For low bit rates, slice group coding adds an unacceptable overhead of up to nearly one hundred per cent. Conversely, for bit rates which are higher than 500 kbit/s, all sequences but *crew* exhibit a small overhead of approximately 1% or less.

Using an all-grey base layer does not affect the overhead significantly due to the use of slice groups, except for very low bit rates, which are impractical. Compared to the classical base layer configuration, however, an all-grey base layer allows using slice-group-based encryption for SVC in the first place, since slice groups cannot be used in the base layer (see section 3).

Figure 9 shows a rate-distortion plot for the two-layer case with slice groups, where the Y-PSNR values are those of the enhancement layer. The plot allows comparing the all-grey base layer with a classical base layer. It is obvious that the all-grey base layer results in significantly better rate-distortion performance (up to 5 dB) for medium and high



**Figure 9: Rate-distortion plot for SVC with two dyadic spatial layers and slice groups. Different base layers (depicted in grey and black) result in significantly different enhancement layer Y-PSNR.**

bit rates.

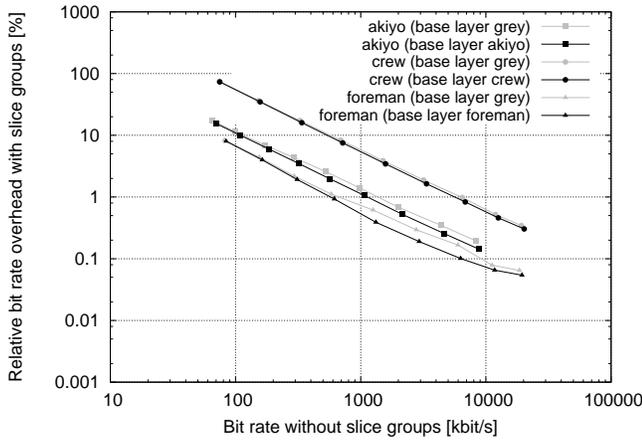
Since an all-grey base layer greatly improves rate-distortion performance avoiding the need for additional drift compensation due to encryption in the base layer, it can be considered a better solution than a classical base layer for this use case. As the overhead due to slice groups is similar in both, the all-grey and the classical base layer scenario (see above), this is also true for other potential use cases in which the base layer does not have to be the downsampled input sequence.

Note that an all-grey base layer in a scenario with two spatial layers defies the purpose of scalable video coding, since one of the two layers becomes unusable for content. However, it allows establishing a baseline for comparison in terms of overhead and allows assessing the usefulness of the concept. In order for all-grey base layers to be practically useful, a scenario with three spatial layers has to be considered so that two spatial layers remain for actual content.

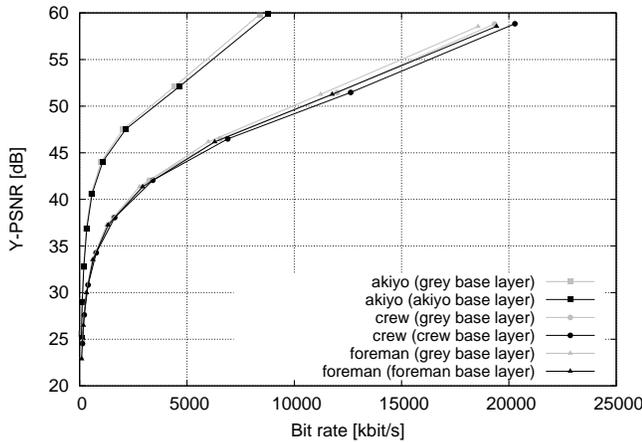
When increasing the number of spatial layers to the maximum of three (see section 3), the overhead due to slice groups increases, as depicted in Figure 10. The overall overhead is significantly higher than in the two-layer case (see Figure 8) for low to medium bit rates. This is due to the fact that slice groups introduce prediction borders which reduce coding efficiency and the three-layer case (with two enhancement layers with slice groups) uses double the amount of slice groups than the two-layer case (with one enhancement layer with slice groups). However, for high bit rates, the overhead is still relatively small and therefore practically negligible for most use cases.

Compared to the two-layer case, the all-grey base layer configuration in the three-layer case allows for an overhead which is approximately as low as the overhead in the classical base layer configuration. Although the all-grey base layer configuration exhibits a higher overhead for medium-to-high bit rates, the actual overhead is only insignificantly higher.

However, in the three-layer case the rate-distortion performance improvement of the all-grey base layer is only very small, as depicted in Figure 11. Although there are still



**Figure 10: Overhead with slice group coding for different CIF sequences when using three dyadic spatial layers**

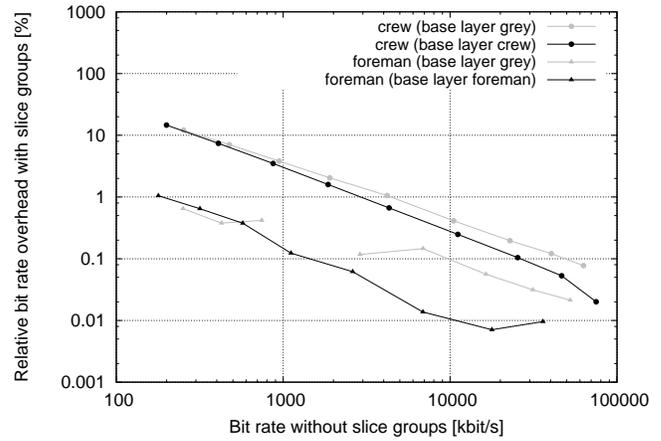


**Figure 11: Rate-distortion plot for SVC with three dyadic spatial layers and slice groups. Different base layers (depicted in grey and black) result in similar enhancement layer Y-PSNR.**

differences of up to 1 dB between an all-grey and a classical base layer in terms of enhancement layer Y-PSNR, but the performance improvement is nowhere near the improvements of the two-layer case (see above).

This is mainly due to the fact that there are two enhancement layers, which use most of the bit rate and the fact that the first enhancement layer can be used to predict parts of the second one through inter-layer prediction. This makes the three-layer case with an all-grey base layer similar to a two-layer case with an additional all-grey bit stream, which is very likely not used at all for inter-layer prediction. However, an all-grey base layer still has advantages compared to a classical base layer for the use case in this paper, since base layer encryption cannot rely on slice groups due to base layer limitations (see above). Thus, an all-grey base layer is still to be preferred over a classical base layer in the three-layer case.

### 4.3 Overhead (4CIF)



**Figure 12: Overhead with slice group coding for different 4CIF sequences when using two dyadic spatial layers**

In order to analyze the influence of spatial resolution on overhead, we repeat the experiments of the previous section with sequences in 4CIF resolution. Note that a 4CIF version of *akiyo* could not be obtained, which is why the following paragraphs only describe results for the *foreman* and the *crew* sequences.

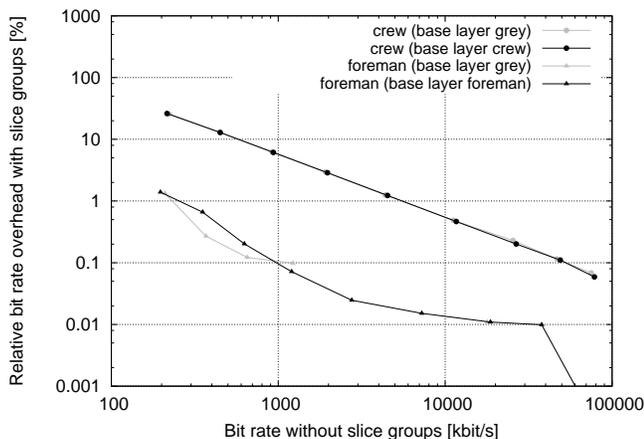
Figure 12 shows the overhead induced by moving slice groups with two spatial layers, like in the previous section. As expected, the decrease of the relative overhead with increasing bit rate is quasi identical, while the overhead values are mostly smaller. Due to the higher spatial resolution, the percentage of macroblocks which are affected by the slice-group-induced prediction borders is smaller, thereby increasing coding efficiency compared to the CIF case depicted in figure 8.

The overhead for the *foreman* sequence (triangles) is 1% or lower for all QP. Although the use of an all-grey base layer introduces a larger overhead than in the CIF case depicted in figure 8, it can still be considered insignificantly small for most QP.

The overhead for the *crew* sequence (circles) is about two to three times lower than in the CIF case depicted in figure 8 when comparing equal QP, and quasi identical when comparing equal bit rates. This is due to the high number of slice groups in the *crew* sequence which induce a significant number of prediction borders. At 4CIF resolution, these have a smaller effect than at CIF resolution, as described for the *foreman* sequence above.

When the *crew* sequence is encoded with an all-grey base layer (grey circles), the overhead is slightly higher than in the regular base layer case (black circles). Although this deviates from the behavior at CIF resolution, where both curves overlap quasi completely, the difference can still be considered to be insignificantly small.

Figure 13 shows the overhead induced by moving slice groups with three spatial layers. As in the two-layer case, the overhead for the *foreman* sequence (triangles) is about 1% or lower. The overhead for the *crew* sequence (circles) is about two to four times lower than in the CIF case depicted in figure 10 when comparing equal QP, and quasi identical when comparing equal bit rates.



**Figure 13: Overhead with slice group coding for different 4CIF sequences when using three dyadic spatial layers**

When using an all-grey base layer (grey circles), the overhead is quasi indistinguishable from the regular base layer case (black circles). Consequently, using all-grey base layers at higher resolutions is recommended for two and three layers, as in the CIF resolution case.

Regarding the overhead results for higher resolutions in general, it can be concluded that equal bit rates yield quasi equal overhead values. Since the overhead decreases with bit rate, the relative overhead decreases with increasing resolution. Thus, at higher resolutions than 4CIF, it is to be expected that the overhead with moving slice groups becomes so small that it can, for most use cases, be ignored.

## 5. FUTURE WORK

This paper shows that slice groups help containing drift in SVC. Although this result can be extended to (non-scalable) H.264 for the most part (since SVC is built upon H.264), a detailed analysis of the overhead induced by slice groups in (non-scalable) H.264 bit streams is desired. Furthermore, the use of B frames and other GOP structures on both, the overhead and the ability to contain drift has to be investigated.

In addition, the detailed effects of SNR scalability have to be studied. Although SNR scalability can be considered as a special case of spatial scalability where width and height remain the same, the overhead of slice groups in SNR layers may be significantly lower due to the more restricted inter-layer prediction mechanisms. This would make SVC encryption yet more feasible, since SNR layers are identical to spatial layers in terms of drift as analyzed in this paper.

## 6. CONCLUSION

We showed the impact of slice group coding on post-compression encryption for a typical surveillance use case. We analyzed the slice-group-induced bit rate overhead as well as the usefulness of slice groups for the containment of drift. For medium and high bit rates, configurations with two and three spatial layers can be used to reduce drift with slice groups with relatively low overhead. For low bit rates, the overhead is too large for practical use at CIF resolution, but moderate at 4CIF and higher resolutions since the relative

overhead decreases with increasing resolution. Furthermore, we introduced the concept of all-grey base layers which simplifies encryption significantly in the two- and three-layer case, albeit at the cost of losing one spatial scalability layer.

## 7. ACKNOWLEDGMENTS

This work is supported by FFG Bridge project 832082.

## 8. REFERENCES

- [1] T. M. Bae, T. C. Thang, D. Y. Kim, J. W. K. Yong Man Ro, and J. G. Kim. Multiple Region-of-Interest Support in Scalable Video Coding. *ETRI Journal*, 28(2):239–242, Apr. 2006.
- [2] T. E. Boulton. PICO: Privacy through invertible cryptographic obscuration. In *IEEE/NFS Workshop on Computer Vision for Interactive and Intelligent Environments*, pages 27–38, Lexington, KY, USA, Nov. 2005.
- [3] P. Carrillo, H. Kalva, and S. Magliveras. Compression Independent Reversible Encryption for Privacy in Video Surveillance. *EURASIP Journal on Information Security*, 2009:1–13, Jan. 2009.
- [4] F. Dufaux and T. Ebrahimi. Video surveillance using JPEG 2000. In *Proceedings of the SPIE Applications of Digital Image Processing XXVII*, volume 5588, pages 268–275, Aug. 2004.
- [5] F. Dufaux and T. Ebrahimi. H.264/AVC video scrambling for privacy protection. In *Proceedings of the IEEE International Conference on Image Processing, ICIP '08*, pages 47–49, San Diego, CA, USA, Oct. 2008. IEEE.
- [6] F. Dufaux and T. Ebrahimi. Scrambling for privacy protection in video surveillance systems. *IEEE Transactions on Circuits and Systems for Video Technology*, 18(8):1168–1174, 2008.
- [7] F. Dufaux and T. Ebrahimi. A framework for the validation of privacy protection solutions in video surveillance. In *Proceedings of the IEEE International Conference on Multimedia & Expo, ICME '10*, pages 66–71, Singapore, July 2010. IEEE.
- [8] D. Grois, E. Kaminsky, and O. Hadar. Roi adaptive scalable video coding for limited bandwidth wireless networks. In *2010 IFIP Wireless Days (WD)*, pages 1–5, Oct. 2010.
- [9] ITU-T H.264. Advanced video coding for generic audiovisual services, Nov. 2007. <http://www.itu.int/rec/T-REC-H.264-200711-I/en>.
- [10] Y. Kim, S. Jin, and Y. Ro. Scalable Security and Conditional Access Control for Multiple Regions of Interest in Scalable Video Coding. In Y. Shi, H.-J. Kim, and S. Katzenbeisser, editors, *International Workshop on Digital Watermarking 2007 (IWDW 2007)*, volume 5041, pages 71–86. Springer Berlin / Heidelberg, 2008.
- [11] Y. Kim, S. Yin, T. Bae, and Y. Ro. A selective video encryption for the region of interest in scalable video coding. In *Proceedings of the TENCON 2007 - IEEE Region 10 Conference*, pages 1–4, Taipei, Taiwan, Oct. 2007.
- [12] J.-H. Lee and C. Yoo. Scalable ROI algorithm for H.264/SVC-based video streaming. In *2011 IEEE*

*International Conference on Consumer Electronics (ICCE)*, pages 201–202, Jan. 2011.

- [13] C. Li, X. Zhou, and Y. Zhong. NAL level encryption for scalable video coding. In *Advances in Multimedia Information Processing, PCM'08*, pages 496–505. Springer-Verlag, Dec. 2008.
- [14] A. Massoudi, F. Lefèbvre, C. D. Vleeschouwer, B. Macq, and J.-J. Quisquater. Overview on selective encryption of image and video, challenges and perspectives. *EURASIP Journal on Information Security*, 2008(Article ID 179290):doi:10.1155/2008/179290, 18 pages, 2008.
- [15] Y. Ou, C. Sur, and K. H. Rhee. Region-based selective encryption for medical imaging. In *Proceedings of the International Conference on Frontiers in Algorithmics (FAW'07)*, Lecture Notes in Computer Science, pages 62–73, Lanzhou, China, Aug. 2007. Springer-Verlag.
- [16] H. Schwarz, T. Hinz, D. Marpe, and T. Wiegand. Constrained inter-layer prediction for single-loop decoding in spatial scalability. In *IEEE International Conference on Image Processing (ICIP) 2005*, volume 2, pages II–870–873, Sept. 2005.
- [17] H. Schwarz, D. Marpe, and T. Wiegand. Overview of the scalable H.264/MPEG4-AVC extension. In *Proceedings of the IEEE International Conference on Image Processing, ICIP '06*, pages 161–164, Atlanta, GA, USA, Oct. 2006. IEEE.
- [18] C. A. Segall and G. J. Sullivan. Spatial scalability within the H.264/AVC scalable video coding extension. *IEEE Transactions on Circuits and Systems for Video Technology*, 17(9):1121–1135, Sept. 2007.
- [19] S. S. F. Shah and E. A. Edirisinghe. Evolving Roi Coding in H.264 SVC. In *VISAPP 2008: Proceedings of the Third International Conference on Computer Vision Theory and Applications – Volume 1*, pages 13–19, 2008.
- [20] Z. Shahid, M. Chaumont, and W. Puech. Selective and scalable encryption of enhancement layers for dyadic scalable H.264/AVC by scrambling of scan patterns. In *16th IEEE International Conference on Image Processing*, pages 1273–1276, Cairo, Egypt, Nov. 2009.
- [21] H. Sohn, E. Anzaku, W. D. Neve, Y. M. Ro, and K. Plataniotis. Privacy protection in video surveillance systems using scalable video coding. In *Proceedings of the Sixth IEEE International Conference on Advanced Video and Signal Based Surveillance*, pages 424–429, Genova, Italy, Sept. 2009.
- [22] T. C. Thang, T. M. Bae, Y. J. Jung, Y. M. Ro, J.-G. Kim, H. Choi, and J.-W. Hong. Spatial Scalability of Multiple ROIs in Surveillance Video. [http://wftp3.itu.int/av-arch/jvt-site/2005\\_04\\_Busan/JVT-O037.doc](http://wftp3.itu.int/av-arch/jvt-site/2005_04_Busan/JVT-O037.doc), Jan. 2005.
- [23] L. Tong, F. Dai, Y. Zhang, and J. Li. Prediction restricted H.264/AVC video scrambling for privacy protection. *Electronic Letters*, 46(1):47–49, Jan. 2010.
- [24] T.-L. Wu and S. F. Wu. Selective encryption and watermarking of MPEG video (extended abstract). In H. R. Arabnia, editor, *Proceedings of the International Conference on Image Science, Systems, and Technology, CISST '97*, Las Vegas, USA, Feb. 1997.