Video Encryption Exploiting Non-Standard 3D Data Arrangements

Stefan A. Kramatsch[1,*] Herbert Stögner[1] and Andreas Uhl[1,2,+]

[1]School of Telematics and Network Engineering, Carinthia Tech Institute, Austria
[2]Department of Computer Sciences, Salzburg University, Austria
[+]Corresponding author e-mail: uhl@cosy.sbg.ac.at

Keywords: video encryption. (M)JPEG2000, lossless video compression, scan order

Abstract – Video data is not necessarily interpreted as a sequence of frames ordered in time. We investigate different strategies how to scan and order the data with respect to their respective suitedness for lossless video compression and subsequent encryption. Selective video encryption schemes with lower computational demand but offering equal security as compared to classical data arrangement strategies are obtained.

1 Introduction

Medical imaging is the main application field for lossless video coding. In these applications, most techniques employ lossless image coding techniques like lossless JPEG, JPEG-LS, or lossless JPEG2000 on a per-frame basis. Of course, temporal redundancy is ignored in such schemes which results in limited compression performance. On the other hand, three dimensional transforms (e.g. 3D DWT [3]) are used to decorrelate the video data but these techniques suffer from high complexity and high memory requirements. Motion compensation techniques as employed in lossy schemes are seldom used in lossless environments.

In this work, we view and process video data in a different manner as compared to classical approaches. We interpret a video as a 3D block of data which can be viewed and processed in any arbitrary order, in particular we do not consider the temporal direction as being necessarily of special nature.

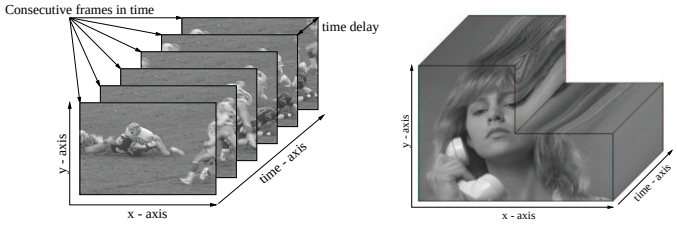
Image and video encryption schemes [10] have been mostly discussed in the context of digital rights management (DRM) systems where the emphasis lies on lossy schemes. In this field, different strategies apply partial encryption either on a per frame basis (where selected coefficients or VLC codewords are protected only [2]) or on a per group of picture (GOP) basis (where selected frames – I-frames – or selected macroblocks – I-blocks are protected only [1]). Obviously, medical imaging is also an application field where privacy and confidentiality are important aims. We discuss privacy schemes for lossless video. Since our underlying coding scheme does not use GOPs, we employ a frame based encryption scheme in our approach.

In section 2, we discuss and visualize various techniques how to scan and compress video data. Section 3 introduces corresponding video encryption schemes based on JPEG2000 and provides experimental results showing significantly superior performance as compared to classical JPEG2000 encryption. Section 4 concludes this paper.

*This artificial name represents a group of students working on this paper in the framework of the System Security lab in summerterm 2005: Agnes Gruber, Alexander Krapesch, Stefan Matschitsch, Thomas Mayerdorfer, Stefan Miedl, Stefan Moser, Martin Tschinder, and Stefan Zorn-Pauli.

2 Non-standard frame structure

The classical view of video data is depicted in Fig. 1.a. The video consists of a set of spatial frames which are temporally ordered still images. As can be seen from the figure, these frames are very similar, even in the presence of strong motion. The aim of inter-frame coding techniques is to exploit this similarity, either by employing block-matching motion compensation [4] among frames (motion compensated hybrid coding) or by applying a transform stage in the temporal direction as well (3D techniques, e.g. [3]). This latter 3D interpretation of video is shown in Fig. 1.b where the frames are accumulated to form a three dimensional (3D) data block of visual data.



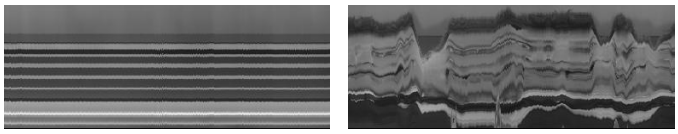
(a) Sequence of frames (b) 3D block

Figure 1: Different views of video data.

This block may be cut along different directions as shown in the figure providing alternative views of the visual and temporal content of the video. In the following, we consider the video as being given in this accumulated 3D manner.

Given the 3D block of video data, the classical way to view or process it is orthogonal to the spatial plane in temporal direction (labelled as "normal view"). However, different ways to view or process the data are possible. The "horizontal view" takes a sideview of the video data as shown in Fig. 1.b. Using this approach, we result in "horizontal frames" which have the original image height and the number of frames as their respective spatial dimensions. Stepping through the video composed of horizontal frames, the adjacent frames correspond to spatially adjacent slices consisting of temporally translated columns of the original frames. Therefore, the first horizontal frame is just the side-plane of the 3D video cube. In Fig. 2 we visualize two types of horizontal frames: the third horizontal frame (Fig. 2.a) is close to the edge of the video and consists mostly of background columns which do hardly change

over time. This results in almost straight lines at the border of these regions. On the other hand, the horizontal frame 095 consists of columns at the center of the original frames which frequently change in time. Fig. 2.b displays this frame which shows the movements of parts of the head giving an interesting visual impression.



(a) horizontal frame 003 (b) horizontal frame 095

Figure 2: "Carphone" video: 176 x 144 x 383 pixels

The "vertical view" looks at the video data from above. Using this approach, we result in "vertical frames" which have the original image width and the number of frames as their respective spatial dimensions. Stepping through the video composed of vertical frames, the adjacent frames correspond to spatially adjacent slices consisting of temporally translated lines of the original frames. As a consequence, the first vertical frame is just the upper plane of the 3D video cube.

In recent work we have shown that videos represented by horizontal and vertical frames may be compressed in lossless mode more efficiently under certain circumstances [6]. The following table lists compression ratios for the two test videos specified in Section 3 for lossless JPEG2000 applied on a frame basis.

Table 1: Compression ratios for JPEG2000 compression.

mode \\ video	Carphone	Claire
normal	1,877	2,727
vertical	2,122	3,908
horizontal	2,046	4,080

It can be clearly observed that the obtained compression ratios are higher for the two alternative scan orders. This effect is more pronounced for the low-motion video Claire.

3 Video encryption

Our aim is to exploit the alternative perspectives of video data as discussed in the last section for lossless video compression and encryption. The classical technique ("normal view") which we compare our results to corresponds to Motion JPEG2000 (MJPEG2000) encryption on a per frame basis (which is in fact JPEG2000 encryption).

Encryption of JPEG2000 bitstreams while maintaining format compliance with focus on confidentiality has been discussed in literature to some extent. Grosbois et al. [5] propose the first partial encryption scheme for JPEG2000 bitstreams – a pseudo random inversion of the bits in certain layers is suggested. In earlier work [8], we have investigated which JPEG 2000 coding options are most suited for subsequent partial encryption, and we have investigated how much packet data needs to be protected to provide reasonable confidentiality. In subsequent work [9] we discuss transparent encryption of JPEG2000 bitstreams for try-and-buy scenarios.

3.1 *Experimental Settings*

We use the following two testvideos (and give their spatial and temporal resolutions): Carphone (176 x 144 x 383) and Claire (176 x 144 x 494), which represent sequences with high motion content and low motion content, respectively.

We compress the data frame-by frame using lossless JPEG2000 (the JJ2000 JAVA implementation[1] is used for this purpose). Since the aim is to operate directly on the bitstream without any decoding we need to discriminate packet data from packet headers in the bitstream. This can be achieved by using two special JPEG 2000 optional markers which were originally defined to achieve transcoding capability, i.e. manipulation of the bitstream to a certain extent without the need to decode data. For technical details see [8]. The bitstream segments identified are then encrypted by AES in CFB mode taking care of marker emulation issues.

This technique may be applied directly to the temporal (i.e. "normal" in Fig. 1), horizontal, and vertical frames of the videos as they are generated. Note that a JPEG2000 bitstream which is selectively encrypted in the described way is fully compliant to the standard.

Subsequently, we decode and view the video data. In order to assess the quality of the visual material after reconstruction in addition to visual inspection we measure the quality of the reconstructed frames using PSNR and ESS (Edge Similarity Score [7]), the latter measuring the similarity of dominating edges on a block basis in the range $[0, 1]$.

In order to additionally assess the security of the scheme, we exploit a built-in error resilience functionality in JJ2000 in order to conduct an "error concealment attack" (see [8] for technical details).

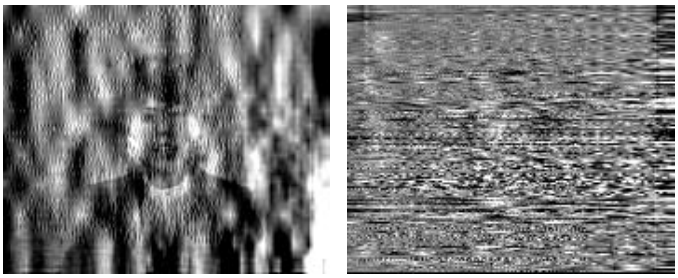
3.2 *Results*

All displayed result show the 18th ("normal") frame of the corresponding test videos, numerical values represent average ESS and/or PSNR values for the entire sequence (in normal view) considered. Apart from Fig. 3, all results have been generated with error-concealment attack mounted.

The first result shown in Fig. 3 displays the 18th frame of the Claire video comparing normal and vertical frame based compression and encryption, when 3% of the original data amount (i.e. 3% of the MJEPG2000 compressed frame, encryption starts at the beginning of the bitstream) is encrypted. It is clearly visible that the quality of the normal view still allows to recognize the content of the frame whereas this is not possible in the vertical view where only noise remains visible. The only information left in the image are some subtle edges corresponding in part to the shape of the person but this is probably not sufficient to recognize that there is a person present in the plain image. Note also that this visual impression does not correspond at all to the numerical quality results since both PSNR and ESS indicate that the images are similar in quality which is obviously not the case.

Fig. 4 shows the numerical results for ESS where an error concealment attack has been mounted, depending on the

[1] http://jj2000.epfl.ch/



(a) normal: ESS 0.54, 10.63 dB (b) vertical: ESS 0.55, 10.67 dB

Figure 3: Claire video: 3% of the original data size is encrypted (direct reconstruction, no attack performed).

amount of data encrypted. When comparing these values to the results shown in the previous figure, it is clear that the attack increases the quality of the reconstructed results significantly, at least in the numerical evaluation.

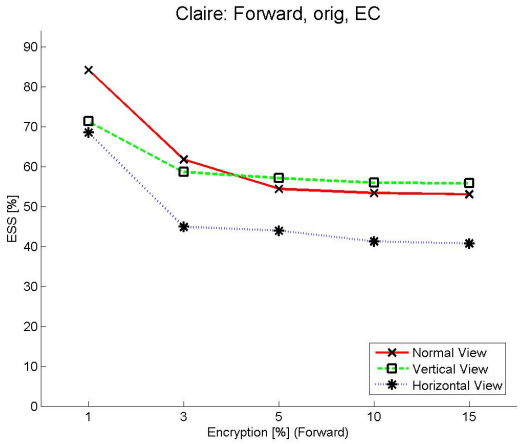
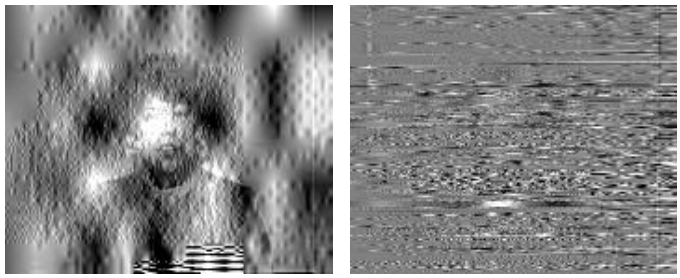


Figure 4: Claire video: ESS vs. encryption amount in %.

Interestingly the graph suggests the images resulting from horizontal view to be of significantly lower quality as compared to vertical view. This cannot be confirmed at all by visual inspection (not shown) which shows vertical and horizontal view as being of comparable quality. This again emphasizes that the quality metrics currently available do not lead to satisfying results in case of low quality visuals.

Also when visual assessment is employed the effectiveness of the attack is confirmed. In case of the mounted attack, it is necessary to encrypt about 10% of the original data to result in an image quality comparable to Fig. 3.b where no image content is recognizable (recall that in this former case only 3% of the data have been encrypted). This is shown in Fig. 5.b (vertical view is used), whereas Fig. 5.a shows the same amount of data encrypted in normal view where clearly image details are visible. Here, the numerical values for PSNR suggest the opposite, which makes clear that PSNR is not suited at all to assess the quality of low quality images. ESS at least indicates a slight superiority (i.e. lower quality) for the vertical view which is very distinct in visual inspection.

The results change when transparent encryption [9] is considered, which is used in try-and-buy scenarios to attract the users' interest. The simplest way to implement this idea is to encrypt the data starting at the end of the



(a) normal: ESS 0.56, 9.81 dB (b) vertical: ESS 0.50, 11.43 dB

Figure 5: Claire video: 10% of the original data size is encrypted.

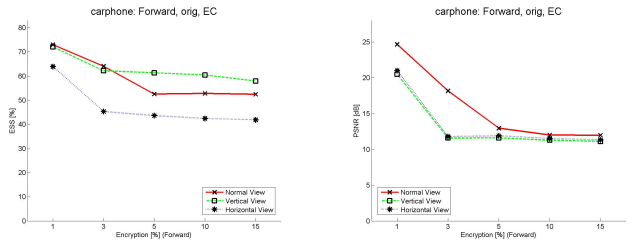
bitstream. Fig. 6 displays results when 10% of the original data size is encrypted in this way. It can be observed, that contrasting to the privacy/confidentiality focussed case the visual impression is very similar for both normal and vertical view. We may also notice that both quality metrics give comparable results which indicates that these measures are applicable as soon as the image quality is above a certain level. Therefore, our approach does not lead to improved results for transparent encryption.

(a) normal: ESS 0.78, 22.28 dB (b) vertical: ESS 0.80, 22.21 dB

Figure 6: Claire video (transparent encryption): 10% of the original data size is encrypted.

Coming back to the privacy/confidentiality focused application scenario, we additionally investigate the behaviour of a second test video. Fig. 7 shows the numerical results for an increasing percentage of encrypted data starting from the beginning of the bitstream. Again, a significant difference between the results corresponding to vertical and horizontal view is indicated in Fig. 7.a, which cannot be confirmed by visual inspection and is obviously caused by a directional bias in the computations of ESS. Fig. 7.b even indicates better PSNR quality for the normal view images, which is not at all the case as can be seen in Fig. 8. Based on numerical values, no clear conclusion could be drawn concerning the effects of our proposed approach.

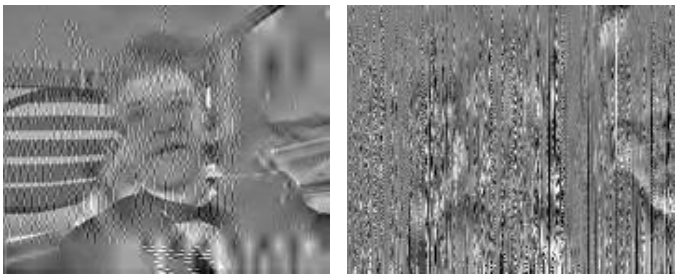
Visual inspection of Fig. 8 again shows that protection of alternative views requires a significantly lower encryption effort, this time we display the horizontal view results. However, contrasting to the case of the Claire test video, for Carphone even encrypting 15% of the data is not sufficient to provide sufficient protection for a privacy/confidentiality application scenario (see Fig. 8.b). The higher motion content of this video reduces the advantages of compression of alternative scan orders and this is propagated as well to the encryption amount, which has to be higher as compared to



(a) ESS (b) PSNR

Figure 7: Carphone video: video quality vs. encryption amount in %.

Claire. For carphone, about 20% - 25% of the original data need to be protected in horizontal or vertical view mode to achieve satisfying confidentiality.



(a) normal: ESS 0.46, 12.45 dB (b) horizontal: ESS 0.45, 11.48 dB

Figure 8: Carphone video: 15% of the original data size is encrypted.

4 Conclusion

We have found that the advances with respect to compression performance caused by alternative interpretation and scan order of video data do carry over to the encryption of JPEG2000 based corresponding frames. Privacy focused applications require a significantly lower amount of encryption effort when applying alternative scan orders as compared to classical MJPEG2000 video. We have also found that this is not the case for transparent encryption scenarios.

Acknowledgements

This work has been partially supported by the Austrian Science Fund, project no. 15170.

References

[1] A. Alattar G. Al-Regib. Evaluation of selective encryption techniques for secure transmission of MPEG-compressed bit-streams. In *Proceedings of the IEEE International Symposium on Circuits and Systems*, pages IV–340–IV–343, 1999.

[2] B. Bhargava, C. Shi, and Y. Wang. MPEG video encryption algorithms. *Multimedia Tools and Applications*, 24(1):57–79, 2004.

[3] S. Cho, D. Kim, and W. A. Pearlman. Lossless compression of volumetric medical images with improved 3-D SPIHT algorithm. *Journal of Digital Imaging*, 17(1):57–63, 2004.

[4] B. Furht, J. Greenberg, and R. Westwater. *Motion estimation algorithms for video compression*. Kluwer Academic Publishers Group, Norwell, MA, USA, and Dordrecht, The Netherlands, 1997.

[5] Raphaël Grosbois, Pierre Gerbelot, and Touradj Ebrahimi. Authentication and access control in the JPEG 2000 compressed domain. In A.G. Tescher, editor, *Applications of Digital Image Processing XXIV*, volume 4472 of *Proceedings of SPIE*, pages 95–104, San Diego, CA, USA, July 2001.

[6] S. Kramatsch, H. Stögner, and A. Uhl. Experimental study on scan order and motion compensation in lossless video coding. In P. Podhradsky et al., editors, *Proceedings EC-SIP-M 2005 (5th EURASIP Conference focused on Speech and Image Processing, Multimedia Communications and Services)*, pages 292–297, Smolenice, Slovak Republic, 2005.

[7] Y. Mao and M. Wu. Security evaluation for communication-friendly encryption of multimedia. In *Proceedings of the IEEE International Conference on Image Processing (ICIP'04)*, Singapore, October 2004. IEEE Signal Processing Society.

[8] Roland Norcen and Andreas Uhl. Selective encryption of the JPEG2000 bitstream. In A. Lioy and D. Mazzocchi, editors, *Communications and Multimedia Security. Proceedings of the IFIP TC6/TC11 Sixth Joint Working Conference on Communications and Multimedia Security, CMS '03*, volume 2828 of *Lecture Notes on Computer Science*, pages 194 – 204, Turin, Italy, October 2003. Springer-Verlag.

[9] A. Uhl and Ch. Obermair. Transparent encryption of JPEG2000 bitstreams. In P. Podhradsky et al., editors, *Proceedings EC-SIP-M 2005 (5th EURASIP Conference focused on Speech and Image Processing, Multimedia Communications and Services)*, pages 322–327, Smolenice, Slovak Republic, 2005.

[10] A. Uhl and A. Pommer. *Image and Video Encryption. From Digital Rights Management to Secured Personal Communication*, volume 15 of *Advances in Information Security*. Springer-Verlag, 2005.