LAYERED ENCRYPTION TECHNIQUES FOR DCT-CODED VISUAL DATA*

Mark M. Fisch[1,+]*, Herbert Stögner*[1]*, and Andreas Uhl*[1,2]

[1]School of Telecommunications & Network Engineering, Carinthia Tech Institute
Primoschgasse 8, A-9020 Klagenfurt, AUSTRIA

[2]Department of Scientific Computing, Salzburg University
Jakob-Haringerstr.2, A-5020 Salzburg, AUSTRIA
e-mail: uhl@cosy.sbg.ac.at

ABSTRACT

Selective encryption technology can be applied efficiently to visual data in scalable representation. In this work we experimentally compare different ways to represent DCT-encoded visual data in a scalable way in terms of their suitability for partial encryption. We find that MPEG-2 SNR scalability is superior to several other approaches.

1. INTRODUCTION

Encryption schemes for multimedia data need to be specifically designed to protect multimedia content and fulfil the security requirements for a particular multimedia application. For example, real-time encryption of an entire video stream using classical ciphers requires heavy computation due to the large amounts of data involved, but many multimedia applications require security on a much lower level (e.g. TV news broadcasting [6]). In this context, several selective encryption schemes have been proposed recently which do not strive for maximum security, but trade off security for computational complexity. The (historically) first and most numerous attempts have been made to secure DCT-based multimedia representations, among them the selective encryption of MPEG streams [2, 10] has attracted the most attention. This has been accomplished by encrypting I-frames (or I-encoded macroblocks) only [1], by manipulating motion vector data [12, 14], or by manipulating coefficients: [11] proposes coefficient permutation, [2, 14] suggest to scramble coefficient data. One of the most recent proposals [12] has been made in the context of MPEG-4 IPMP and cleary shows that selectively encrypting MPEG data while maintaining bitstream compliance implies a significant processing overhead. In case a selective encryption process requires a multimedia bitstream to be parsed in order to identify the parts to be subjected to encryption, the problem of high processing overhead occurs in general. For example, in order to selectively protect DC and large AC coefficients of a JPEG image (as suggested by some authors), the file needs to be parsed for the EOB symbols 0x00 to identify the start of a new 8×8 pixels block (with two exceptions: if 0xFF is followed by 0x00, 0x00 is used as a stuffbit and has to be ignored and if AC63 (the last AC-Coefficient) not equals 0 there will be no 0x00 and the AC coefficients have to be counted). Under such circumstances, selective encryption will not help to reduce the processing demands of the entire application [9].

A possible solution to this problem is to use the visual data in the form of scalable bitstreams. In such bitstreams the data is already organized in layers according to its visual importance and the bitstreams do not have to be parsed to identify the parts that should be protected by the encryption process. In previous work [3, 4, 5], several suggestions have been made to exploit the base and enhancement layer structure of the MPEG-2 scalable profiles as well as to use the MPEG-4 FGS [13] for this purpose. However, there exist several possibilities how to organize MPEG data into base and enhancement layers and it is not clear which variant is most suited for the selective encryption application.

In this work we systematically investigate the different possibilities how to organize DCT-coded visual data into several quality layers (Section 2) and we experimentally compare the respective applicability to the selective encryption application in Section 3. Section 4 concludes the paper and provides an outlook to further work in this direction.

2. LAYERED RESPRESENTATION OF DCT-CODED VISUAL DATA

The basic idea of DCT-based scalable coding is to organize the data into a base layer which contains a low quality approximation to the original data and several enhancement layers which, if combined with the base layer, successively improve the quality. The MPEG-2 scalability profile provides three types of scalability:

- SNR Scalability: the base layer contains a full resolution but strongly quantized version of the video, the enhancement layers consist of DCT coefficient differences to weaker quantized versions of the data.
- Resolution Scalability: the base layer is a low resolution version of the video (usually generated by repeated weighted averaging and subsequent downsampling), the enhancement layers contain the difference between different resolutions of the data.
- Temporal Scalability: the base layer is a version of the video with reduced frame rate, the enhancement layers simply contain the frames required to achieve higher frame rates.

Additionally, in the context of DVB there exists a way to partition MPEG-2 data into more and less important parts in order to enable unequal error protection functionality, where leading DCT coefficients and motion vector data constitute

[+] Mark M. Fisch is an artificial name representing a group of students working on this project in the framework of the Multimedia I laboratory (winterterm 2003/2004): H. Fischer, C. Gattringer, M. Mauritsch, M. Oberwasserlechner, C. Probst, M. Schauer, F. Schmidt, M. Schuster, C. Sturmer.

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

the important part and high frequency DCT coefficients the less important part. However, special MPEG units supporting this functionality are required.

MPEG-2 scalbility profiles have not found wide acceptance due to several reasons, reduced coding efficiency in case of using a high number of enhancement layers among them. The obvious advantages in the context of confidential video transmission might change this in the future. Additionally, the MPEG committee has recently launched a call for proposals for a scalable video codec which should overcome the problems of MPEG-2.

Since no MPEG software is publicly available which implements all scalability modes, we use the progressive JPEG modes from the JPEG extended system [7]. As we shall see, the different progressive JPEG modes perfectly simulate the types of MPEG scalability. In JPEG, the terminology is changed from layers to scans.

- Hierarchical progressive mode (HP): an image pyramid is constructed by repeated weighted averaging and downsampling. The lowest resolution approximation is stored as JPEG (i.e. the first scan), reconstructed, bilinearly upsampled, and the difference to the next resolution level is computed and stored as JPEG with different quantization strategy (similar to P and B frames in MPEG). This is repeated until the top level of the pyramid is reached. This mode corresponds well to MPEG-2 resolution scalability.
- Sequential progressive modes
  - Spectral selection (SS): the first scan contains the DC coefficients from each block of the image, subsequent scans may consist of a varying number of AC coefficients, always taking an equal number from each block. This mode is very similar to the abovementioned DVB/MPEG-2 data partitioning scheme.
  - Successive approximation (SA): the most significant bits of all coefficients are organized in the first scan, the second scan contains the next bit corresponding to the binary representation of the coefficients, and so on. Since quantization is highly related to reducing the bit depth of coefficents, this mode behaves similarly to SNR scalability.

The JPEG standard also allows to mix different modes – an important example is to use the DC coefficient as first scan, the subsequent scans contain the binary representation of the AC coefficients as defined by successive approximation (we denote this mode as mixed (MM)). The three modes allow a different amount of scans. Whereas spectral selection offers a maximum of 64 scans, the hierarchical progressive mode is restricted to 5 or 6 sensible scans (given a $2^8 \times 2^8$ pixels image). Successive approximation mostly uses a maximum of 10 scans (depending on the data type used for coefficient representation). Similar to the scalability profile of MPEG-2, the JPEG progressive modes are not used very much and are poorly supported and documented in commercial software. Although providing much better functionality for transmission based applications, the compression performance could be expected to decrease using JPEG progressive modes. As a matter of fact, compression performance is at least as good as for the baseline system and often better (Fig. 1 shows the rate distortion performance of the Photoshop baseline and progressive JPEG versions). However, the computational demand for encoding and decoding is of course higher.

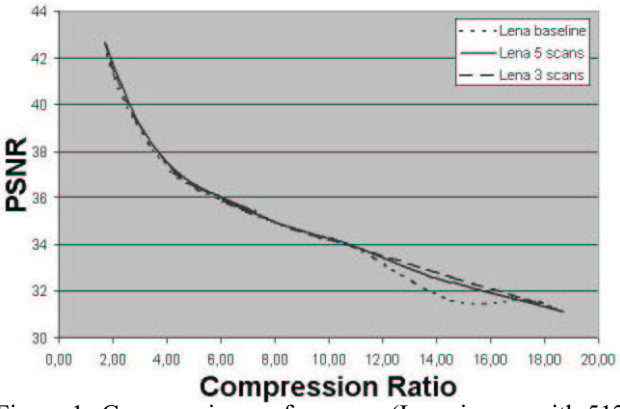


Figure 1: Compression performance (Lena image with 512^2 pixels and 8bpp) of Photoshops' baseline JPEG and progressive JPEG (with 3 and 5 scans).

This also serves as an excellent example how poorly documented the progressive JPEG modes are – there is no hint in the Photoshop documentation what type of progressive mode is employed. All subsequently used images are in 8bpp 512^2 pixels format.

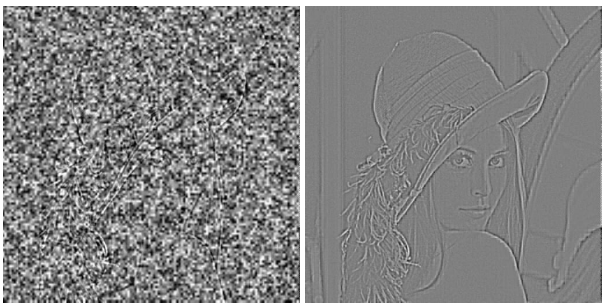
3. SELECTIVE ENCRYPTION USING LAYERED REPRESENTATION

The basic idea of selectively encrypting visual data in layered representation is to simply encrypt the base layer or the scans containing the perceptually most relevant information. In this case, the enhancement layers or remaining scans may be expected to contain data which is useless on its own although given in plaintext. Of course, this is not true in case of temporal scalability since the enhancement layer contains entire frames. As a consequence, temporal scalability can not be used for layered encryption.

Decoding a partially encrypted image by treating the encrypted data as being unencrypted leads to images severely degraded be noise type patterns (which originate form the encrypted parts, see Figs. 2.a and 3.a). Using these images to judge the security of the system leads to misinterpretations since a hostile attacker can do much better. In particular, an attacker could simply ignore the encrypted parts (which can be easily identified by statistical means) or replace them by typical non-noisy data. This kind of attack is called "error-concealment" [12] or "replacement attack" [8] in the literature.

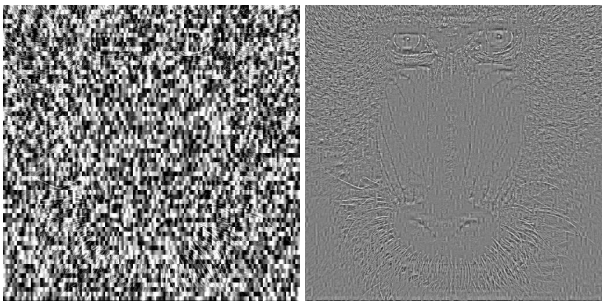
Figs. 2.b and 3.b clearly show that there can be still information left in the unencrypted parts of the data after selective encryption has been applied – in case of direct reconstruction this is hidden by the high frequency noise pattern. As a consequence, in order to facilitate a sound evaluation and comparison of the four modes to be considered, they are evaluated after a replacement attack has been mounted.

In order to be able to compare the different JPEG progressive modes for their suitability to follow the selective encryption approach, we set the amount of data to be encrypted to approximately 10 and 30%, respectivly. Since we use a 10 bit representation for quantized DCT coefficients, the percentages can be exactly achived in SA mode by encrypting the corresponding number of bitplanes. For HP, we get 31,25% of the original data encrypted by building a three level pyramid and encrypting the lowest resolution plus the first residual, and 8,3% by building a six level pyramid and



(a) direct reconstruction (b) replacement attack

Figure 2: Lena image; a three level pyramid in HP mode is used with the lowest resolution encrypted



(a) direct reconstruction (b) replacement attack

Figure 3: Mandrill image; SS mode is used with DC and first AC coefiicient encrypted

encrypting the lowest resolution plus the three next residuals. SS facilitates protection of 29,7% and 9,3% of the data by encrypting 19 or 6 coefficients, respectively. Finally, we achieve encryption of 31,09% and 11,4% in the case of MM by scrambling the DC coefficients and one bitplane or three bitplanes, respectively.

The replacement attack is conducted as follows: for HP, the encrypted first scan is replaced by an equally sized image with constant gray value and eventually encrypted residuals are replaced by constant zero residuals. For SS an encrypted bitplane is replaced by a constant 0 bitplane, and for SA the encrypted coefficients are replaced by zeros.

	HP	SS	SA	MM
Lena, 10% enc.	14.8	14.6	7.0	6.8
Lena, 30% enc.	14.7	14.5	6.2	6.4
Mandrill, 10% enc.	17.5	16.8	7.5	7.3
Mandrill, 30% enc.	17.0	16.2	6.4	6.4

Table 1: Objective quality (PSNR in dB) of reconstructed iamges

Table 1 shows the PSNR values of the different techniques applied to the Lena and Mandrill image. Note that in contrast to a compression application, a method exhibiting low PSNR values is most desirable (since this implies low image quality and therefore good resistance against the replacement attack). HP and SS show very similar results (at about 14.5 - 17.5 dB depending on the image and percentage of encryption) as well as do SA and MM at a much lower level (at 6.4 - 7.5 dB). However, it is interesting to note that there is not much numerical difference between the encryption of 10% and 30% of the data. As a consequence, we expect to perform SA and MM much better in terms of security as compared to HP and SS. In Fig. 4 we visually compare the reconstructed images underlying the numerical data of Table 1.

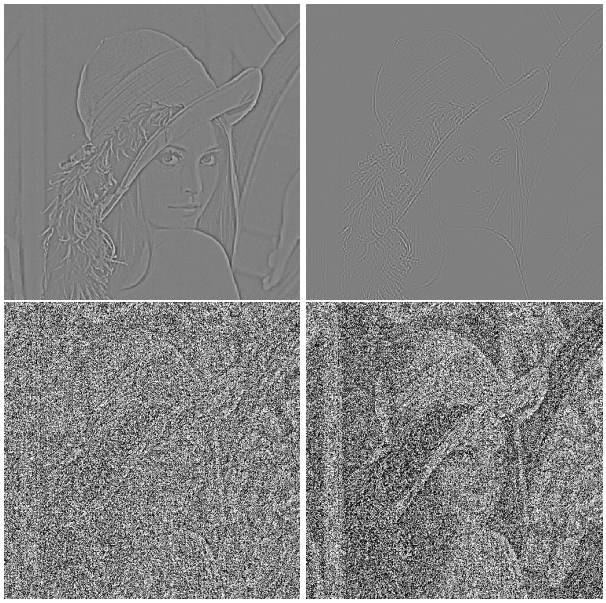
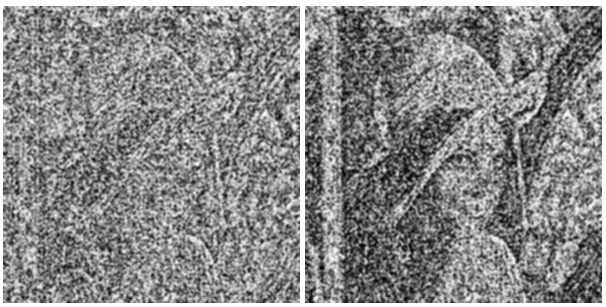


Figure 4: Subjective quality of reconstructed Lena image, 10% of the data encrypted (HP,SS,SA,MM, in clockwise direction starting at the upper left image).

The numerical results are clearly confirmed by visual inspection. Whereas HP and SS clearly exhibit still remaining high frequency information (which are much clearer in the HP case), almost no information is visible for SA and MM where the images are dominated by noise. This noise comes from the fact that on average 50% of the coefficients (no matter if high or low frequency) have been altered at their MSB in the binary representation which results in those randomly looking images. Note that the replacement attack is not effective in the case of SA and MM since no matter if directly reconstructed or under the replacement attack always on average 50% of the coefficents are altered at their MSB position. Although the results of SA and MM look rather satisfying from a security point of view, there is still visual information related to the original image left. Fig. 5 shows that this remaining information may be enhanced using simple image processing operations which leads to the conclusion that obviously MM is the most secure variant of our investigated selective encryption schemes and is the only one that can be securly operated at a level of encrypting 10% of the data. The additionally encrypted DC coefficient makes MM much more resistant against reconstruction as compared to SA.

Increasing the amount of encrypted data up to 30% does not leave any perceptaully relevant information in the re-



(a) MM (b) SA

Figure 5: Images from Fig. 4 median filtered (3x3 kernel) und blurred (5x5 filter).

maining data in the case of SA and MM. Little information is left in case of SS applied to the Lena image, HP still reveals some edge and texture information. The Mandrill image contains much high frequency information which is still visible after encrypting 30% for both, the HP and SS modes (see Fig. 6).

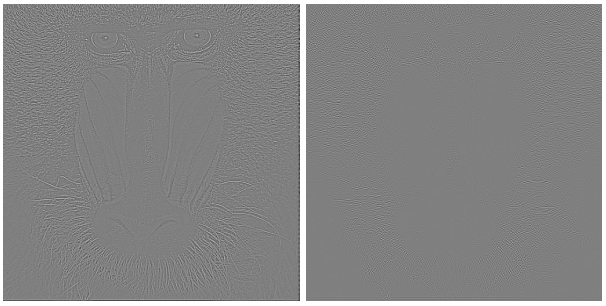


Figure 6: Subjective quality of the reconstructed Mandrill image, 30% of the data encrypted (comparison of HP and SS modes).

4. CONCLUSIONS AND FUTURE WORK

We have seen that selective encryption using the hierarchical progressive and spectral selection JPEG modes still leaves perceptually relevant information in the remaining data after encrypting 30% of the original image data. Successive approximation and especially a hybrid variant which additionally protects the DC coefficient deliver much better results in terms of security. Relating these results to the MPEG case, SNR scalability will be most suited to apply selective encryption to scalable video data. In future work we will investigate the resistance of SNR scalability based selective encryption against a reconstruction attack [8] and we will additionally use the GOP structure of MPEG video to additionally lower the encryption effort (i.e. a lower percentage of data is encrypted for P and B frames as compared to I frames).

REFERENCES

[1] I. Agi and L. Gong. An empirical study of secure MPEG video transmissions. In *ISOC Symposium on Network and Distributed Systems Security*, pages 137–144, San Diego, California, 1996.

[2] B. Bhargava, C. Shi, and Y. Wang. MPEG video encryption algorithms. *Multimedia Tools and Applications*, 2003. to appear.

[3] Jana Dittmann and Ralf Steinmetz. Enabling technology for the trading of MPEG-encoded video. In *Information Security and Privacy: Second Australasian Conference, ACISP '97*, volume 1270, pages 314–324, July 1997.

[4] Ahmet Eskicioglu and Edward J. Delp. An integrated approach to encrypting scalable video. In *Proceedings of the IEEE International Conference on Multimedia and Expo, ICME '02*, Laussanne, Switzerland, August 2002.

[5] Thomas Kunkelmann. Applying encryption to video communication. In *Proceedings of the Multimedia and Security Workshop at ACM Multimedia '98*, pages 41–47, Bristol, England, September 1998.

[6] Benoit M. Macq and Jean-Jacques Quisquater. Cryptology for digital TV broadcasting. *Proceedings of the IEEE*, 83(6):944–957, June 1995.

[7] W.B. Pennebaker and J.L. Mitchell. *JPEG – Still image compression standard*. Van Nostrand Reinhold, New York, 1993.

[8] M. Podesser, H.-P. Schmidt, and A. Uhl. Selective bitplane encryption for secure transmission of image data in mobile environments. In *CD-ROM Proceedings of the 5th IEEE Nordic Signal Processing Symposium (NORSIG 2002)*, Tromso-Trondheim, Norway, October 2002. IEEE Norway Section. file cr1037.pdf.

[9] A. Pommer and A. Uhl. Application scenarios for selective encryption of visual data. In J. Dittmann, J. Fridrich, and P. Wohlmacher, editors, *Multimedia and Security Workshop, ACM Multimedia*, pages 71–74, Juan-les-Pins, France, December 2002.

[10] Lintian Qiao and Klara Nahrstedt. Comparison of MPEG encryption algorithms. *International Journal on Computers and Graphics (Special Issue on Data Security in Image Communication and Networks)*, 22(3):437–444, 1998.

[11] L. Tang. Methods for encrypting and decrypting MPEG video data efficiently. In *Proceedings of the ACM Multimedia 1996*, pages 219–229, Boston, USA, November 1996.

[12] Jiangtao Wen, Mike Severa, Wenjun Zeng, Max Luttrell, and Weiyin Jin. A format-compliant configurable encryption framework for access control of video. *IEEE Transactions on Circuits and Systems for Video Technology*, 12(6):545–557, June 2002.

[13] C. Yuan, B.B. Zhu, Y. Wang, S. Li, and Y. Zhong. Efficient and fully scalable encryption for MPEG-4 FGS. In *IEEE International Symposium on Circuits and Systems (ISCAS'03)*, Bangkok, Thailand, May 2003.

[14] Wenjun Zeng and Shawmin Lei. Efficient frequency domain video scrambling for content access control. In *Proceedings of the seventh ACM International Multimedia Conference 1999*, pages 285–293, Orlando, FL, USA, November 1999.