SELECTIVE ENCRYPTION OF VISUAL DATA

Classification of Application Scenarios and Comparison of Techniques for Lossless Environments

Champskud J. Skrepth*
Carinthia Tech Institute, Austria
9956@edu.fh-kaernten.ac.at

Andreas Uhl†
Department of Scientific Computing, Salzburg University
and Carinthia Tech Institute, Austria
uhl@cosy.sbg.ac.at

Abstract We discuss techniques for selective encryption of visual data. After introducing a classification of application scenarios for selective encryption we derive conditions for the sensible employment of such an approach in each scenario. Finally, we propose and evaluate experimentally several selective encryption approaches for a specific lossless application scenario.

Keywords: Selective image encryption, partial or soft encryption

Introduction

In the area of multimedia security, the term "soft encryption" is sometimes used as opposed to classical "hard" encryption schemes like AES. Such schemes do not strive for maximum security and trade off security for computational complexity. They are designed to protect multimedia content and fulfill the security requirements for a particular multimedia application. For example, real-time encryption for an entire video stream using classical ciphers requires much computation time due to the large

*This artificial name represents a group of students working on this project in the framework of the multimedia 1 laboratory (winterterm 2001/2002).

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

amounts of data involved, on the other hand many multimedia applications require security on a much lower level (e.g. TV broadcasting [6]) or should protect their data just for a short period of time (e.g. news broadcast). Therefore, the search for fast encryption procedures specifically tailored to the target environment is mandatory for multimedia security applications.

Selective or partial encryption (SE) of visual data is an example for such an approach. Here, application specific data structures are exploited to create more efficient encryption systems (see e.g. SE of MPEG video streams [1, 5, 10, 11, 12, 15], of wavelet-based encoded imagery [2, 4, 8, 9, 13, 15], and of quadtree decomposed images [2]). Consequently, SE only protects (i.e. encrypts) the visually most important parts of an image or video representation relying on a secure but slow "classical" cipher.

In this work we discuss selective image encryption techniques for lossless environments. Section 1 provides a classification and discussion of four different application scenarios for SE. We derive conditions for the sensible employment of SE in each scenario and give concrete sample applications. In section 2 we present several different SE approaches for a specific lossless application scenario which are experimentally compared in section 3. In the conclusion we summarize the main results and give an outlook to further work in this direction.

1. Application scenarios for selective image and video encryption

Intuitively, SE seems to be a good idea in any case since it is always desirable to reduce the computational demand involved in signal processing applications. However, the security of such schemes is always lower as compared to full encryption. The only reason to accept this drawback are *significant* savings in terms of processing time or power. Therefore, the environment in which SE should be applied needs to be investigated thoroughly in order to decide whether its use is sensible or not.

In the following we discuss a classification of application scenarios for SE of images and videos. The first classification criterion is whether the application operates in a lossless or lossy environment. There exist several reasons why a lossy representation may not be acceptable or necessary:

- Due to requirements of the application a loss of image data is not acceptable (e.g., in medical applications because of reasons related to legal aspects and diagnosis accuracy [14]).

- Due to the low processing power of the involved hardware encoding or decoding of visual data is not possible (e.g. mobile clients).
- Due to the high bandwidth available at the communication channel lossy compression is not necessary.

Note that no matter if lossy or lossless, compression has always to be performed prior to encryption since the statistical properties of encrypted data prevent compression from being applied successfully. Moreover, the reduced amount of data after compression decreases the computational demand of the subsequent encryption stage.

The second classification criterion is whether the image data is given as plain image data or in form of a bitstream resulting from prior compression. For example, in on-line applications the plain image data may be accessed directly after being captured by a digitizer before being compressed. On the other hand, as soon as visual data has been stored or transmitted it has been compressed in some way which is true for most off-line applications. The following table introduces four types of possible SE application scenarios based on the criteria introduced so far which will be discussed subsequently.

Table 1. Application scenarios for selective encryption

	Lossy	Lossless
Bitstream	Scenario A	Scenario B
Image	Scenario C	Scenario D

In the following, t denotes the time required to perform an operation, E is the encryption function, SE the selective encryption function, C is compression, P is the preprocessing involved in the selective encryption scheme (where P means the extraction of relevant features), and $>>$ means significantly larger. Please note that the processing time t is not equivalent to computational complexity: for example, if compression is performed in hardware and encryption in software, the time required for compression will be considerably lower as for encryption, contrasting to the relation if both operations are performed in software (compare scenario C).

- **Scenarios A and B:** Given the bitstream B resulting from prior compression, the following condition must be fulfilled in order to justify the use of SE:

$$t(E(B)) >> t(P) + t(SE(B)) \tag{1}$$

In this case, P is the identification of relevant features in the compressed bitstream. Depending on the type of bitstream, $t(P)$ may range from negligibly small to a considerably large amount of time. If the bitstream is embedded or is composed of several quality layers, the identification of parts subjected to SE is straightforward ($t(P) = 0$) – the first part of the embedded bitstream or the base layer is encrypted only. In case of a less structured bitstream it might be necessary to partially decode or at least parse the bitstream to identify significant features (e.g. DC or large AC coefficients of a JPEG encoded image), thereby causing $t(P)$ to increase linearly with $t(SE(B))$. However, given an embedded bitstream the condition $t(E(B)) >> t(SE(B))$ and therefore equation (1) may be satisfied easily and SE definitely makes sense. Concrete applications are video on demand for scenario A and retrieval of medical images from a database for scenario B.

- **Scenario C**: Given the raw image data I, the following condition must be fulfilled in order to justify the use of SE:

  $$t(C(I)) + t(E(C(I))) >> t(C(I)) + t(P) + t(SE(C(I))) \qquad (2)$$

  As in scenarios A and B, P is again the identification of relevant features in the compressed bitstream and the same considerations about its execution time apply here. In any case, even if $t(P) = 0$, equation (2) is very hard to satisfy since $t(C(I)) >> t(E(C(I)))$ holds for most lossy coding schemes and symmetrical ciphers if both schemes are executed in software. Therefore, the difference between $t(E(C(I)))$ and $t(SE(C(I)))$ often does not matter in practice. This effect is even more pronounced for high compression ratios of course (since the resulting bitstream after compression is already rather small). Consequently, given the raw image data, the decrease in terms of security often does not justify the marginal savings in processing time as achieved by SE in a software based system. Concrete applications for this scenario are videoconferencing and on-line surveillance.

- **Scenario D**: Given the raw image data I, the following condition must be fulfilled in order to justify the use of SE:

  $$t(E(I)) >> t(P) + t(SE(I)) \qquad (3)$$

  In contrast to scenarios A - C, encryption is applied directly to the raw image data. Note that if $t(C(I)) + t(E(C(I))) > t(E(I))$ does not hold, the image data is compressed using a lossless codec and the considerations of scenario B apply. Usually, this is not the case since $t(C(I)) >> t(E(I))$ is valid also for almost all lossless codecs and symmetrical ciphers. Additionally, the data reduction of lossless schemes is much lower as compared to lossy ones making the contribution of $t(E(C(I)))$ significant as well. When applying encryption to the raw image data, P denotes the identification of relevant features in the raw image data which may be done in various ways (see next section for some examples). However, it is crucial that $t(P)$ is not too large to satisfy equation (3). If $t(P)$ can be made small, SE is a sensible approach in this setting. A concrete sample application for this scenario is teleradiology with mobile clients to enable fast and exact on-site diagnosis.

Most SE schemes discussed in literature involve lossy coding schemes and are therefore categorized into scenarios A or C. Regarding the analysis given above, it makes a big difference with respect to the usefulness of SE whether the raw image data or a compressed bitstream is given – this fact is usually ignored which is an obvious shortcoming in many papers on SE. Consequently, it is important to analyse the concrete application setting in which SE should be used in order to judge its appropriateness. In the following section we discuss several approaches to extract significant image features (i.e. preprocessing approaches) for SE schemes in application scenario D.

2. Selective image encryption techniques in lossless environments

SE applied in application scenario D consists of two stages: preprocessing P which extracts significant features of the image and the selective encryption process $SE(I)$ itself which encrypts those features. Therefore, two components of the overall execution time of such a scheme need to be investigated: $t(P)$ and $t(SE(I))$ where the latter is directly proportional to the percentage of data encrypted in this stage (`perc`) relative to the original data amount. Following the IEEE standard data types we assume the image to be given in 8bit/pixel (bpp) precision (`char`), whereas `int` and `float` data are assumed to require 32 bpp.

2.1. Spatial domain techniques

2.1.1 Multiresolution pyramids. As a first step we construct a quater-sized version of the image ("approximation") using a 4-pixel average (AV), a 4-pixel median (ME), or subsampling by 2 in each direction (DS). Subsequently, we construct a full-sized version of the image ("prediction") by using an interpolated version of the approximation only (different types of interpolation are used: linear (default setting) and cubic (CI)). This prediction is subtracted from the original image resulting in the "residual". The approximation is encrypted and transmitted together with the residual in plaintext. The construction of the approximation is iterated to construct smaller versions. Whereas the bitdepth is not influenced by these operations (the residual can usually be represented as `signed char` requiring 8bpp as well) the overall amount of data to be transmitted is. After one iteration we result in 125% of the original data (residual plus quater-sized approximation), 106.25% after the second iteration, etc. There is only a small amount of sensible values attained for `perc`: 25, 6.25, 1.56, …, $t(P) \neq 0$ of course but it is reasonably small using our simple approach.

2.1.2 Bitplane encryption. We consider the 8bpp data in the form of 8 bitplanes, each bitplane associated with a position in the binary representation of the pixels. The SE approach is to encrypt a subset of the bitplanes only, starting with the bitplane containing the MSB of the pixels. Each possible subset of bitplanes may be chosen for SE, however, the minimal value for `perc` is 12.5 (when encrypting the MSB bitplane only) increasing in steps of 12.5. $t(P)$ is negligible using this approach which is denoted BP1. Alternatively, the Gray-code binary representation can be used (BP2) which causes $t(P) \neq 0$ due to the conversion operation which has a complexity order of magnitude similar to an addition operation for each pixel. The encrypted bitplanes are transmitted together with the remaining bitplanes in plain text.

2.2. Transform domain techniques

In contrast to spatial domain methods the operation P (i.e. the transform) increases the bitdepth significantly. No matter if `float` or `int` data are used in the transform, we result in 400% of the original data after the transform. Consequently, if we encrypt 25% of the coefficients for example, `perc` is 100. Additionally, $t(P) \neq 0$ and it is considerably large. Note that for all techniques discussed the amount of data to be transmitted is 400% of the original image data.

2.2.1 DCT. The DCT is well known to extract global image characteristics efficiently and is used for watermarking applications for these reasons (see e.g. Cox's scheme [3]). We use the DCT in two flavours: as full frame DCT (DCT1) and as DCT applied to 8 × 8 pixels blocks (DCT2) due to complexity reasons. Following the zig-zag scan (compare e.g. JPEG) we encrypt the first coefficients or the first coefficients from each block, respectively. The encrypted coefficients are transmitted together with the non-encrypted ones. Given a 512 × 512 pixels image and using DCT2, the lowest value for `perc` is 6.25 (i.e. the DC coefficient is encrypted only for each block) and increasing in steps of 6.25 per additional coefficient, whereas `perc` may be set almost arbitrarily with DCT1.

2.2.2 Wavelet transform. In many applications wavelet transforms (WT) compete with and even replace the DCT due to their improved localization properties (e.g., the WT is used in many watermarking schemes [7]). We use the Haar transform due to complexity reasons and investigate two different SE approaches. For both schemes the decomposition depth is a parameter, WT1 subsequently encrypts the approximation subband only, whereas WT2 encrypts both approximation subband and a number of additional significant coefficients (i.e. larger than a threshold) from other subbands. WT1 delivers only a small amount of sensible values for `perc` (25, 6.25, ...), whereas `perc` may be set almost arbitrarily with WT2.

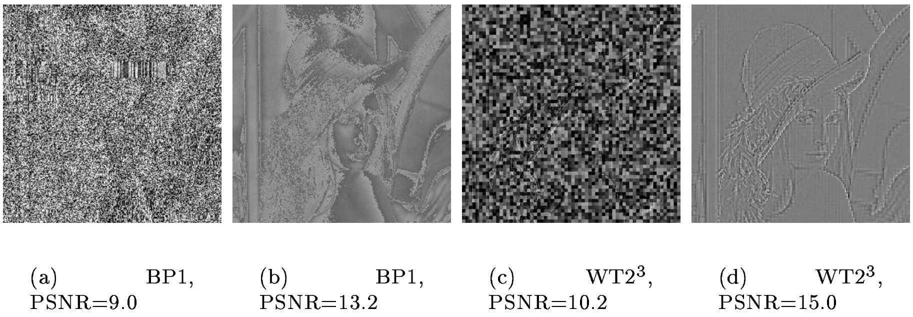
3. Experiments

The aim of the experimental section is twofold. First, we want to evaluate the effectiveness of the proposed SE schemes with respect to execution efficiency. Second, and most important, the security of these schemes is assessed and compared. Assuming the cipher in use is unbreakable we conduct a simple ciphertext-only attack by reconstructing the selectively encrypted images. The encrypted parts would introduce noise-type distortions in directly reconstructed images. Therefore, we replace the encrypted parts by artificial data mimicing typical images (see section 3.1 for details). Subsequently, reconstruction is performed as usual, treating the encrypted and replaced parts as being non-encrypted. A major shortcoming of many SE investigations is the lack of quantifying the quality of the visual data that can be obtained by attacks against SE. Mostly visual examples are provided only. Here, the quality of the obtained images is assessed using PSNR. Additionally, we relate the numerical values to visual examples.

3.1. Experimental settings

We use the classical 8bpp, 512 × 512 pixels Lena grayscale image as testimage. All SE schemes are implemented using MATLAB®, as cipher we use an AES implementation specifically developed for this particular purpose with blocksize 128 bit and a 128 bit key. The 128 bit block of AES is filled with data according to data types: a quater of a line when encrypting bitplanes, 16 consecutive pixels of a line when encrypting `char` and 4 values when encrypting `float` transform coefficients.

In order to conduct our ciphertext-only attacks the encrypted data needs to be replaced for reconstruction as explained in the previous section. In the case of multiresolution pyramids, the encrypted approximation is replaced by an approximation consisting of constant grayvalue 128. Similarly, the encrypted wavelet approximation subband is replaced by an approximation subband resulting from decomposing an image with constant grayvalue 128. The encrypted DCT coefficients are replaced by a linearly decreasing sequence of coefficients, starting from a DC coefficient again obtained from transforming an image with constant grayvalue 128, terminating at the first non-encrypted original coefficient. Bitplane encryption is attacked by replacing the encrypted bitplane with a constant 0 bitplane and additionally compensating for the decrease in average luminance by adding 64 to each pixel if only the MSB bitplane was encrypted, 96 if the MSB and next bitplane have been encrypted, and so on. The efficiency of these simple replacements is shown in Figure 4. The images 4.a and 4.c are directly reconstructed, whereas 4.b and 4.d are reconstructed using the replacement strategy.



(a) BP1, PSNR=9.0

(b) BP1, PSNR=13.2

(c) WT2^3, PSNR=10.2

(d) WT2^3, PSNR=15.0

Figure 1. Visual examples for the efficiency of the replacement operation, perc=25.

Obviously, not only the visual appearance but also the numerical PSNR values have been significantly improved by the replacement strategy. These effects are equivalently observed for multiresolution pyramid and DCT based methods.

3.2. Experimental results

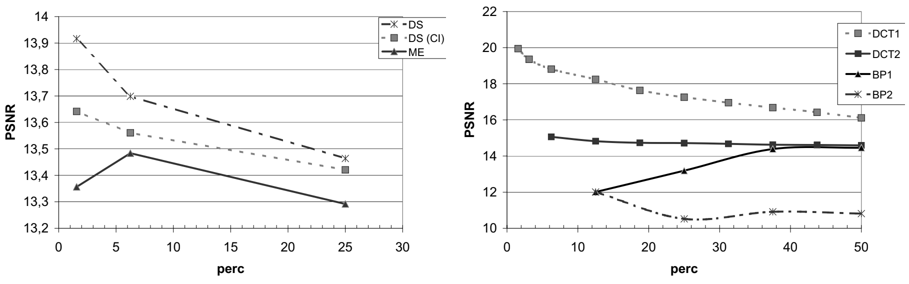
Equation 4 relates the time demand of the preprocessing stage $t(P)$ of the various SE schemes introduced in our implementation. For schemes based on multiresolution pyramids or WT, superscripts denote the number of decomposition iterations performed and CI denotes cubic interpolation to generate the prediction in the case of multiresolution pyramids.

$$\begin{aligned} 0 = BP1 < BP2 << DCT2 < WT1^i < WT2^i < \\ DS^i < AV^i < ME^i < ME^i(CI) << DCT1 \end{aligned} \tag{4}$$

Note that the computational effort to encrypt the entire image with AES is about two times as high as for performing a wavelet transform and is comparable to the full frame DCT for the image size considered. This fact makes the approach DCT1 useless for practical applications unless the DCT is performed in hardware and AES in software. For all techniques but the bitplane encryption methods, sensible performance gains compared to full encryption can consequently be expected for `perc` ≤ 25 only. Bitplane encryption is more efficient than full AES encryption even for `perc` $= 87.5$ due to its low complexity preprocessing stage P.

$t(SE(I))$ is entirely independent of $t(P)$ and is assessed in the following by directly relating the amount of data encrypted in this stage (expressed in `perc`, see section 2 for a definition) to the PSNR of the reconstructed images. Note that "better" methods have lower PSNR values !

In the case of multiresolution pyramids Figure 2.a shows the extremal curves for DS and ME. All other techniques fall within this range. Cubic interpolation CI enhances the result of simple downsampling DS significantly which is not true for the median approximation ME.

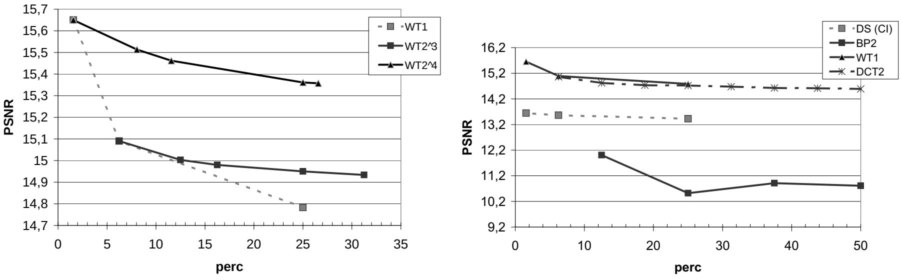


(a) Multiresolution pyramids

(b) Bitplane encryption and DCT

Figure 2. Recovering selectively encrypted images

Figure 2.b compares the two flavors of bitplane encryption and DCT, respectively. The Gray-code representation shows better results as compared to the classical binary code. Surprisingly, both approaches do not exhibit monotonically decreasing PSNR values for increasing `perc`, BP1 even has increasing ones. The DCT based methods are clearly inferior to bitplane encryption, DCT1 which is based on global DCT is 2 dB and more inferior to DCT2. Finally we compare the wavelet transform techniques in Figure 3.a. WT1 which simply encrypts the approximation subband is superior to both techniques relying on encrypting approximation data plus significant coefficients.



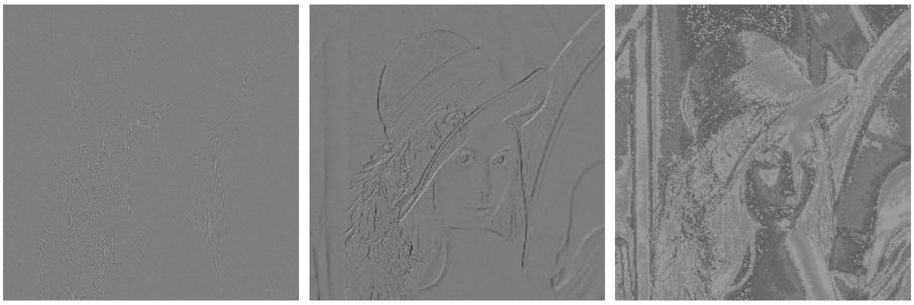
(a) Wavelet transform

(b) Overall comparison

Figure 3. Recovering selectively encrypted images

For an overall comparison, we have selected the best performing SE algorithm from each group (i.e. that with lowest PSNR), except for the multiresolution pyramid techniques since it has turned out that the visual performance of DS(CI) is significantly better (i.e. worse image quality) as compared to ME (see Figures 4.a and 4.b). In Figure 3.b we notice that the transform based techniques can not compete with the spatial domain techniques. This seems to be surprising at first – note that this is partially due to the data expansion caused by the transforms. However, from the plot we see that this is definitely not the only reason since even compensating the factor 4 expansion does not make WT1 and DCT2 competitive to BP2 and DS(CI). BP2 outperforms DS(CI) with PSNR values up to 2 dB lower. Consequently, in terms of pure PSNR performance, bitplane encryption using Gray-code representation is the best SE technique of the proposed ones. In the following we will relate these findings to visual perception.

Figure 4 compares the visual appearance of recovered images which have been subject to spatial domain selective encryption at a rate of `perc`=25 (i.e. one quater of the original data is encrypted). We face a severe mismatch between PSNR and perceived quality – whereas almost no information is visible at 13.4 dB with DS(CI) (Figure 4.a), we clearly

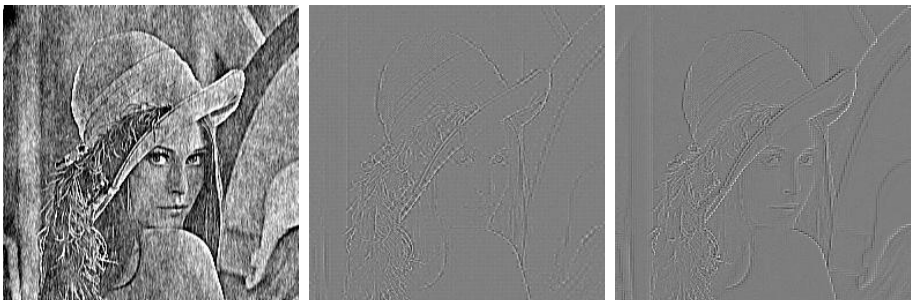


(a) DS(CI), PSNR=13.4 (b) ME, PSNR=13.3 (c) BP2, PSNR=10.2

Figure 4. Visual examples of recovered images after attack, perc=25.

see some edges of the Lena image at 13.3 dB with ME (Figure 4.b). The extremely low PSNR values of BP2 (10.2 dB in this case, Figure 4.c) do not correspond with the visual perception at all. Whereas the luminance appearance is significantly alienated, the objects in the image are perceived clearly.

The visual appearance of recovered images which have been subject to transform domain selective encryption is compared in Figure 5.



(a) DCT1, PSNR=16.7 (b) DCT2, PSNR=14.8 (c) WT1, PSNR=14.8

Figure 5. Visual examples of recovered images after attack, perc=25.

The poor performance of DCT1 is visually confirmed, also the similar performance of DCT2 and WT1 in terms of PSNR corresponds well to perception. However, the 1.5dB difference between ME and DCT2/WT1 can not be confirmed visually, all three reconstructions show an almost equal amount of edge information. Note also that the PSNR computed between the image Lena and its entirely AES encrypted version is 9.2 dB whereas PSNR between Lena and an image with constant grayvalue 128 is 14.5 dB ! Both images do not carry any structural information related to Lena, however, the PSNR values differ more than 5 dB. Also, the value 14.5 dB is higher (i.e. more similar to Lena) as compared to the values corresponding to Figures 4.a and 4.b where at least some edge information is visible.

Consequently, we may state that PSNR and visual perception do not correspond at all at low quality. Therefore, PSNR is not suited to evaluate SE schemes where the aim is to achieve the lowest image quality possible. Only SE schemes not suited at all for SE (e.g. DCT1) may be identified correctly using PSNR.

When trying to combine both aspects, visual perception and PSNR, in order to identify the best SE technique, we find the multiresolution pyramid technique DS(CI) to be the winner from a pure security point of view. However, this approach is not very scalable since already the second most secure mode (`perc`=6.25) already reveals a certain amount of edge information. In cases where severe alienation is sufficient to protect the data, bitplane encryption based on Gray-code representation is the method of choice, since this technique is very fast as well.

4. Conclusion

After introducing a classification for selective encryption application scenarios we have investigated several techniques suited for lossless environments with low computing capacity (e.g. mobile clients). From a pure security point of view a specific multiresolution technique in the spatial domain has turned out to be the best method, if alienation is sufficient and speed important, bitplane encryption based on Gray-code representation is a possible choice as well. We have found that PSNR is hardly useful in order to rate SE schemes, consequently empirical methods involving human viewers are required for this aim at present. In future work we will accomplish this to assess the proposed schemes more accurately and we will investigate alternative numerical measures (e.g. picture quality scale) to evaluate low quality visual data consistently. Additionally, we will develop more scalable multiresolution pyramid methods (especially focusing on DS(CI)) by allowing approximations of arbitrary size.

References

[1] A. M. Alattar, G. I. Al-Regib, and S. A. Al-Semari. Improved selective encryption techniques for secure transmission of MPEG video bit-streams. In

Proceedings of the 1999 IEEE International Conference on Image Processing (ICIP'99). IEEE Signal Processing Society, 1999.

[2] H. Cheng and X. Li. Partial encryption of compressed images and videos. *IEEE Transactions on Signal Processing*, 48(8):2439–2451, 2000.

[3] Ingemar J. Cox, Joe Kilian, Tom Leighton, and Talal G. Shamoon. Secure spread spectrum watermarking for multimedia. In *Proceedings of the IEEE International Conference on Image Processing, ICIP '97*, volume 6, pages 1673–1687, Santa Barbara, California, USA, October 1997.

[4] 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*, San Diego, CA, USA, July 2001.

[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] P. Meerwald and A. Uhl. A survey of wavelet-domain watermarking algorithms. In Ping Wah Wong and Edward J. Delp, editors, *Proceedings of SPIE, Electronic Imaging, Security and Watermarking of Multimedia Contents III*, volume 4314, San Jose, CA, USA, January 2001. SPIE.

[8] A. Pommer and A. Uhl. Wavelet packet methods for multimedia compression and encryption. In *Proceedings of the 2001 IEEE Pacific Rim Conference on Communications, Computers and Signal Processing*, pages 1–4, Victoria, Canada, August 2001. IEEE Signal Processing Society.

[9] A. Pommer and A. Uhl. Selective encryption of wavelet packet subband structures for obscured transmission of visual data. In *Proceedings of the 3rd IEEE Benelux Signal Processing Symposium (SPS 2002)*, pages 25–28, Leuven, Belgium, March 2002. IEEE Benelux Signal Processing Chapter.

[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] C. Shi and B. Bhargava. A fast MPEG video encryption algorithm. In *Proceedings of the ACM Multimedia 1998*, pages 81–88, Boston, USA, 1998.

[12] 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.

[13] T. Uehara, R. Safavi-Naini, and P. Ogunbona. Securing wavelet compression with random permutations. In *Proceedings of the 2000 IEEE Pacific Rim Conference on Multimedia*, pages 332–335, Sydney, December 2000. IEEE Signal Processing Society.

[14] S. Wong, L. Zaremba, D. Gooden, and H.K. Huang. Radiologic image compression – a review. *Proceedings of the IEEE*, 83(2):194–219, 1995.

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