# Image Confidentiality Using Progressive JPEG

Thomas Stütz and Andreas Uhl
Department of Scientific Computing
Salzburg University, Austria
E-mail: {tstuetz,uhl}@cosy.sbg.ac.at

***Abstract*— Baseline JPEG is not well suited for designing efficient encryption schemes. Selective encryption technology can be applied much better to visual data in scalable representation. We use the three progressive modes as defined in the JPEG extended system for providing confidentiality to visual data by encrypting selected scans. Whereas the hierarchical progressive and spectral selection modes turn out to be highly insecure when encryption effort is decreased, reasonable security can be maintained using successive approximation.**

***Keywords*—progressive JPEG, image encryption, layered security**

## I. Introduction

Encryption schemes for multimedia data need to be specifically designed to protect multimedia content and fulfil the application requirements for a particular multimedia environment [1].

For example, real-time encryption of visual data 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 [2]). In this context, several selective or partial encryption schemes have been proposed recently which do not strive for maximum security, but trade off security for computational complexity by restricting the encryption to the perceptually most relevant parts of the data.

The (historically) first and most numerous attempts have been made to secure DCT-based multimedia representations, among them the selective encryption of MPEG streams [3], [4] has attracted the most attention. This has been accomplished by encrypting I-frames (or I-encoded macroblocks) only [5], [6], by manipulating motion vector data [7], or by permuting coefficients [8], [9]. One of the most recent proposals [7] 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. Under such circumstances, selective encryption will not help to reduce the processing demands of the entire application [10].

Using the visual data in the form of scalable bitstreams is a possible solution to this problem. 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 [11], [12], [13], 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 [14] for this purpose.

An additional issue is resolution or quality adaptation in the encrypted domain. For many applications (e.g. transmission over heterogenous networks) visual data needs to be scaled with respect to its data rate, a process usually refered to as transcoding. In the general case when encryption is involved the data needs to be decrypted, transcoded, and re-encrypted which is fairly inefficient in terms of involved computations and key management. Visual data in scalable representation is much better suited for this type of environments since also encryption can be made scalable [15] and transcoding may be applied in the encrypted domain.

The baseline JPEG format does not correspond well to these requirements. For example, in order to selectively protect DC and large AC coefficients of a JPEG image (as discussed for example in [16], [13]), the file needs to be parsed for the EOB symbols 0x00 to identify the start of a new $8 \times 8$ 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) does not equal 0 there will be no 0x00 and the AC coefficients have to be counted). It is clear that selective encryption will not be helpful to reduce complexity under these circumstances. Also with respect to transcoding there is no way to design a scheme allowing this operation to be performed in the encrypted domain using baseline JPEG.

In this work we systematically investigate the different JPEG progressive modes as defined in the JPEG extended system [17] with respect to their usefulness for providing confidentiality to visual data in a flexible and efficient way. Section 2 reviews the three modes which are compared to the JPEG baseline system in terms of compression performance and data organisation in Section 3. Section 4 finally discusses the respective suitability in a selective encryption context, the paper is concluded in Section 5.

## II. Progressive JPEG Modes

The basic idea of DCT-based scalable coding [18] 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 three JPEG progressive modes are defined as follows (JPEG uses the term “scan” instead of layers, the first two

modes are often denoted as sequential progressive modes):

- Spectral selection: 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. A typical choice is to encode all DC coefficients into the first scan, subsequently groups of 6 and 7 AC coefficients are organized into one scan.
- Successive approximation: scans are organized according to the binary representation of the coefficients. The first 6 bit of a coefficient is the smallest fraction which the JPEG standard allows to specify. This fraction is coded as in baseline JPEG, while the following bits are emitted without coding. According to the standard, DC and AC coefficients have to be treated separately. A typical setting is to use 6 scans (first 6 bit of all DC coefficients Huffman coded, 1 bit more of DC coefficient data, 1 bit more of DC coefficient data, first 6 bit of all AC coefficients Huffman coded, 1 bit more of AC coefficient data, 1 bit more of AC coefficient data).
- Hierarchical progressive mode: 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, bi-linearly upsampled, and the difference to the next resolution level is computed and stored as JPEG with possibly different quantization strategy (similar to P and B frames in MPEG). This is repeated until the top level of the pyramid is reached.

The JPEG standard also allows to mix different modes. 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 – 6 sensible scans (given a $2^8 \times 2^8$ pixels image). Successive approximation is restricted to 6 scans (assuming 8 bpp grayscale data). Similar to the scalability profiles of MPEG-2, the JPEG progressive modes are not used very much and are poorly supported and documented in commercial software.

## III. Compression Performance

Although providing much better functionality for transmission based applications, the compression performance could be expected to decrease using JPEG progressive modes. This would of course not favour the use of these techniques for providing security. As we shall see, compression performance equivalent and even exceeding the baseline system may be achieved provided the coding options are chosen carefully. The tests concerning the sequential progressive modes were conducted using the IJG's (Independent JPEG Group) reference library. The default configuration for progressive JPEG is a mixture between spectral selection and successive approximation (first 7 bit DC coefficients Huffman coded, first 6 bit of first 5 AC coefficients Huffman coded, first 6 bit of the remaining AC coefficients Huffman coded, next bit of all AC coefficients, lowest bit of DC coefficients, lowest bit of all AC coefficients). Fig. 1 clearly shows the superior rate-distortion performance of the progressive variant. All tests in this work employ the Lena image with $512^2$ pixels and 8bpp.

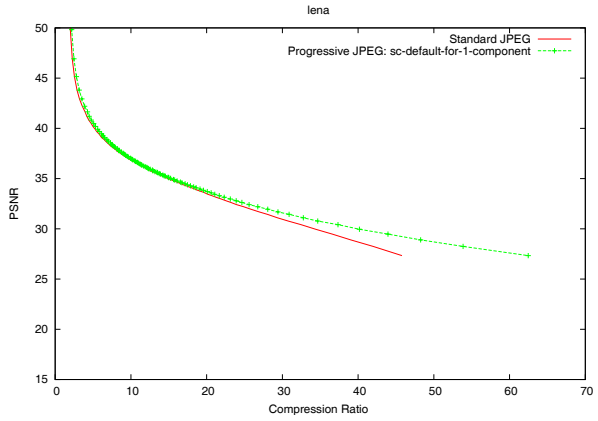


Fig. 1. Compression performance of baseline JPEG vs. sequential progressive JPEG.

Also the "pure" sequential progressive algorithms provide superior compression performance as compared to the baseline scheme.

As the IJG's reference implementation does not include a hierarchical mode, we implemented a corresponding coding option. As suggested by the standard, the downsampling for constructing the image pyramid is done by weighted averaging involving three neighbouring pixels (which is done in horizontal and vertical direction) and the upsampling simply uses bi-linear interpolation. The visual data in the pyramid levels are encoded as JPEG files where we removed redundant header informations. Since hierarchical JPEG is not widely used, little is known about good parameter choices – however, a large variety of coding choices may be specified: in addition to the number of levels used in the image pyramid, the quantization tables and the corresponding scaling factors for each level in the pyramid need to be specified. Beside the default quantization table we use "uniform" quantization tables for residual data similar to those used in MPEG for P- and B-frames. Fig. 2 shows the large variety of coding performance achieved when varying these settings.

The results turn out to be somewhat disapointing when considering the target applications of hierarchical JPEG: the best results are achieved with a low number of pyramid levels (i.e. two scans: one base layer, one enhancement layer) and the results are better for configurations when the base layer is encoded with low quality. Both properties are not desirable for progressive applications of course. However, it also turns out that we may achieve compression performance close to baseline JPEG using such a scheme (the scheme depicted in Fig. 2 uses 15% quality for the lowest pyramid level and uniform quantization matrices for higher levels).

It is also possible to combine hierarchical JPEG with the sequential progressive modes. This leads to a significant improvement of the coding performance, even superior to that of baseline JPEG in some cases. Fig. 3 gives an example for three pyramid levels.

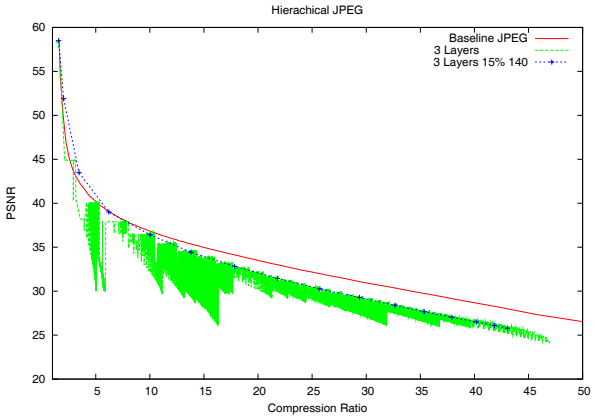


Fig. 2. Compression performance of baseline JPEG vs. hierarchical progressive JPEG.

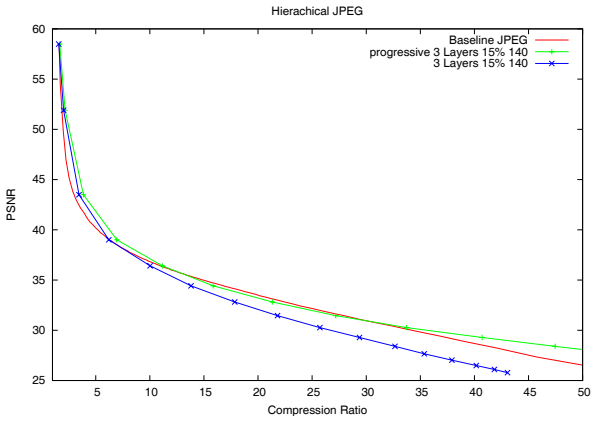


Fig. 3. Compression performance baseline JPEG vs. a mixed mode progressive JPEG.

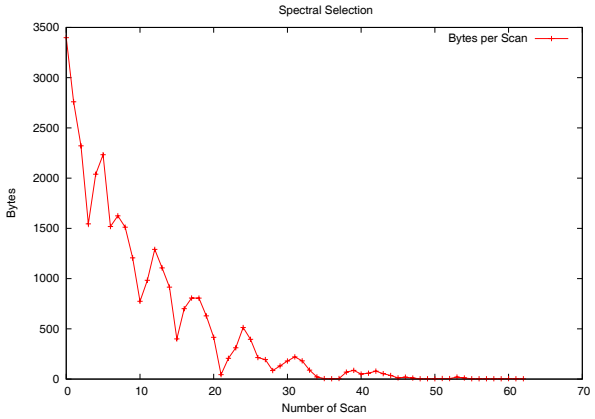


Fig. 4. Data distribution across different scans for spectral selection.

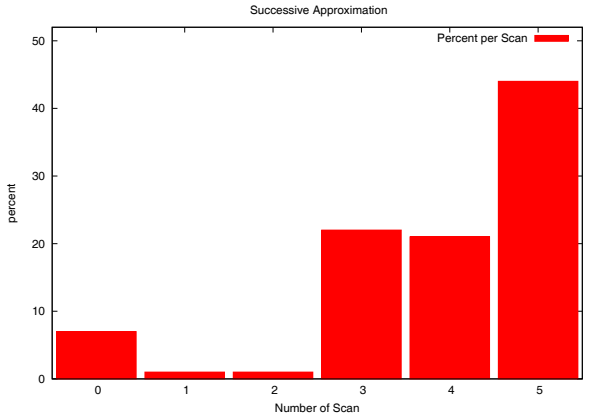


Fig. 5. Data distribution across different scans for successive approximation.

Finally, Figs. 4 and 5 show how the different scans contribute to the overall file size. This is important knowledge for subsequent selective encryption since we want to be able to estimate the reduction of computational load in case encryption is restricted to certain scans. In the spectral selection case (Fig. 4) each scan contains one coefficient from each block and as it is expected, the size of the scan decreases for increasing coefficient frequency.

In our example, successive approximation uses the scan configuration as used as an example in section 2. We realize that the two scans containing the single DC coefficient bits do not contribute much, whereas the three scans corresponding to single AC coefficient bits contribute with 21% and 44% of the overall data, which makes them interesting targets to be left unencrypted.

## IV. Confidentiality

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.

Decoding a partially encrypted image by treating the encrypted data as being unencrypted leads to images severely degraded by noise type patterns (which originate from the encrypted parts). 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” [7] or “replacement attack” [19] in the literature. The IJG software ignores empty scans, for conducting an attack therefore scans affected by encryption simply can be set to zero. In the hierarchical JPEG case we simply set residual pyramid levels to zero if affectd by encryption, the base layer is replaced by uniform gray value 128. See also [18] for these attacks against DCT-based coding/encryption schemes. In order to assess the quality of the visual material after reconstruction in addition to visual inspection we use PSNR and ESS (Edge Similarity Score [20]), the latter measuring the similarity of dominating edges on a block basis in the range $[0, 1]$.

Considering a JPEG file using spectral selection, first of all we have to encrypt the DC coefficient scan, otherwise a

subsampled/low-pass filtered version of the image is readily available to the attacker. Furthermore, Fig. 6 shows that even the last 50% of AC data contain enough image information to draw concrete conclusions of the image content. Hence strictly more than 55% have to be encrypted in order to guarantee at least some confidentiality which is still not satisfactory. Therefore, spectral selection can be categorized as being not suited for a confidentiality focused scheme.

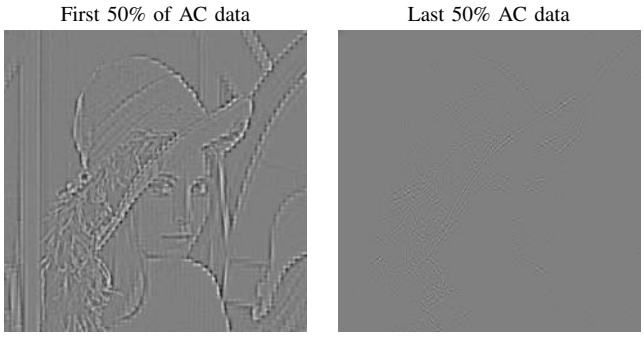


Fig. 6. Reconstructions based on AC coefficient data only

Also the application of selective encryption to hierarchical JPEG for confidential encryption is quite limited. In fact every enhancement layer contains sufficient edge and contour information to get at least an idea of the image content, the base layer itself is a low reolution approximation. Therefore, every layer needs to be encrypted to guarantee that no content revealing information can be reconstructed. This is visually confirmed by Fig. 7 where all but the finest resolution level residual is encrypted – edge information is still present with surprising quality which is also confirmed by the high ESS values.

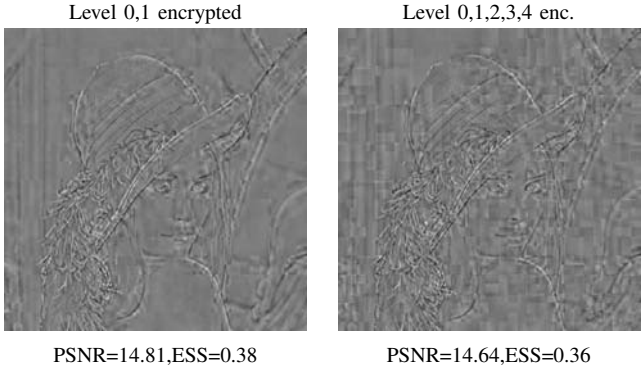


Fig. 7. Encryption of hierarchical JPEG with 3 and 6 pyramid levels

In the case of successive approximation, according to Fig. 8 the first 6 bit of DC and AC coefficients respectively have to encrypted since these data contain most of the image's luminance and edge information.

Applying this method (i.e. encrypting scans 0 and 3) to the Lena image leads to good results (see 9). In this example 29%

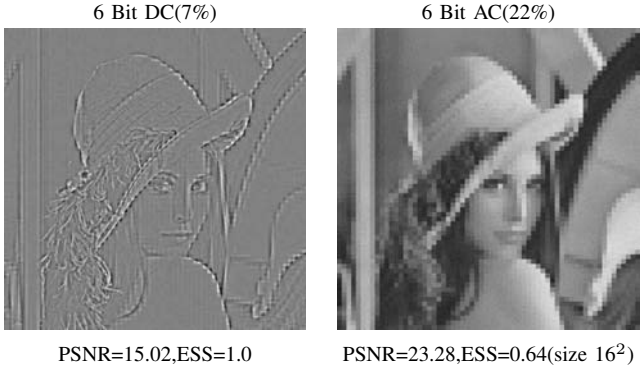


Fig. 8. Encryption of Successive Approximation Scans 0 and 3 only (reconstruction with replacement attack).

of the image data are encrypted and no visual information is left in the image (which is confirmed by the 0.0 value of ESS).

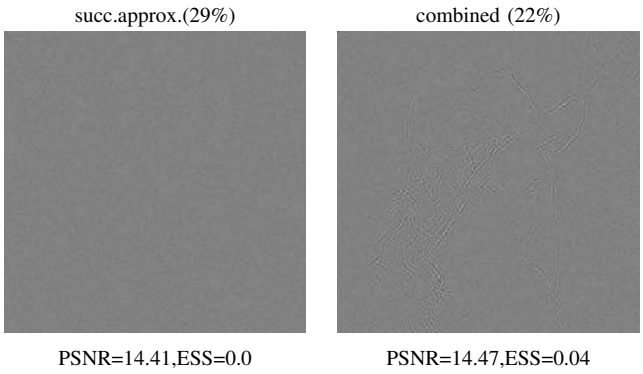


Fig. 9. Encryption of successive approximation JPEG mode.

If the goal is to encrypt even less, one might think of combining this approach with spectral selection, but it is very likely that edges and contours will show up. Figure 9 shows the result if the last third of the 6 bit AC coefficient data is not encrypted (this only makes 7% of the encoded coeffiecient data). One can clearly recognize some of the most important edges of the original picture.

## V. 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 much more than 50% of the original image data. This makes these schemes useless for selective encryption. Successive approximation delivers much better results in terms of security and reduction of encryption effort where encrypting 30% of the data leads to reasonable security results. Considering these findings in the MPEG context, this means that PSNR scalbility is best suited to design progressive encryption schemes with MPEGs scalability profiles. In future work we will investigate the usefulness of the JPEG progressive modes for transparent encryption schemes.

## ACKNOWLEDGEMENT

The support of the Austrian Grid project is gratefully acknowledged. This work has been also partially supported by the Austrain Science Fund, project no. 15170.

## REFERENCES

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

[2] B. M. Macq and J.-J. Quisquater, "Cryptology for digital TV broadcasting," *Proceedings of the IEEE*, vol. 83, pp. 944–957, June 1995.

[3] B. Bhargava, C. Shi, and Y. Wang, "MPEG video encryption algorithms," *Multimedia Tools and Applications*, vol. 24, no. 1, pp. 57–79, 2004.

[4] L. Qiao and K. Nahrstedt, "Comparison of MPEG encryption algorithms," *International Journal on Computers and Graphics (Special Issue on Data Security in Image Communication and Networks)*, vol. 22, no. 3, pp. 437–444, 1998.

[5] I. Agi and L. Gong, "An empirical study of secure MPEG video transmissions," in *ISOC Symposium on Network and Distributed Systems Security*, (San Diego, California), pp. 137–144, 1996.

[6] G. Spanos and T. Maples, "Performance study of a selective encryption scheme for the security of networked real-time video," in *Proceedings of the 4th International Conference on Computer Communications and Networks (ICCCN'95)*, (Las Vegas, NV), 1995.

[7] J. Wen, M. Severa, W. Zeng, M. Luttrell, and W. Jin, "A format-compliant configurable encryption framework for access control of video," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 12, pp. 545–557, June 2002.

[8] S. Shin, K. Sim, and K. Rhee, "A secrecy scheme for MPEG video data using the joint of compression and encryption," in *Proceedings of the 1999 Information Security Workshop (ISW'99)*, vol. 1729 of *Lecture Notes on Computer Science*, (Kuala Lumpur), pp. 191–201, Springer-Verlag, Nov. 1999.

[9] L. Tang, "Methods for encrypting and decrypting MPEG video data efficiently," in *Proceedings of the ACM Multimedia 1996*, (Boston, USA), pp. 219–229, Nov. 1996.

[10] A. Pommer and A. Uhl, "Application scenarios for selective encryption of visual data," in *Multimedia and Security Workshop, ACM Multimedia* (J. Dittmann, J. Fridrich, and P. Wohlmacher, eds.), (Juan-les-Pins, France), pp. 71–74, Dec. 2002.

[11] J. Dittmann and R. Steinmetz, "Enabling technology for the trading of MPEG-encoded video," in *Information Security and Privacy: Second Australasian Conference, ACISP '97*, vol. 1270, pp. 314–324, July 1997.

[12] A. Eskicioglu and E. J. Delp, "An integrated approach to encrypting scalable video," in *Proceedings of the IEEE International Conference on Multimedia and Expo, ICME '02*, (Laussanne, Switzerland), Aug. 2002.

[13] T. Kunkelmann, "Applying encryption to video communication," in *Proceedings of the Multimedia and Security Workshop at ACM Multimedia '98*, (Bristol, England), pp. 41–47, Sept. 1998.

[14] 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.

[15] H. Yu, *Streaming media encryption*, pp. 197–217. CRC Press, 2005.

[16] H. Cheng and X. Li, "On the application of image decomposition to image compression and encryption," in *Communications and Multimedia Security II, IFIP TC6/TC11 Second Joint Working Conference on Communications and Multimedia Security, CMS '96* (P. Horster, ed.), (Essen, Germany), pp. 116–127, Chapman & Hall, Sept. 1996.

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

[18] M. M. Fisch, H. Stögner, and A. Uhl, "Layered encryption techniques for DCT-coded visual data," in *Proceedings (CD-ROM) of the European Signal Processing Conference, EUSIPCO '04*, (Vienna, Austria), Sept. 2004. paper cr1361.

[19] 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), IEEE Norway Section, Oct. 2002. file cr1037.pdf.

[20] 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), IEEE Signal Processing Society, Oct. 2004.