IMPROVING THE VISUAL PERFORMANCE OF S/DISCUS

Gabriel Domínguez-Conde[1] *and Deepa Kundur*[2]

1. Signal Theory and Communications Department, Universidade de Vigo, Vigo 36310
2. Department of Electrical & Computer Engineering, Texas A&M University, College Station, TX 77843-3128 USA

ABSTRACT

Wireless multimedia sensor networks (WMSNs) emerge as a solution in wireless unattended video surveillance scenarios. The G-E-M methodology was recently introduced by Czarlinska *et al.* to address the problem providing protection to visual wireless surveillance systems in the presence of another hostile sensing system. This work builds upon the G-E-M framework by introducing pre- and post-processing stages that reduce decoding errors and visual noise as well as allowing the effective control of bitrate.

Index Terms— Keyless security, distributed source coding, low-complexity video coding, wireless multimedia sensor networks, Slepian-Wolf coding

1. INTRODUCTION

Boosted by the reduction of the cost of the multimedia hardware, wireless multimedia sensor networks (WMSN) have recently emerged as a new branch in the wireless sensor networks area [1]. The advanced capabilities of WMSNs enable its application to a wide variety of problems ranging from health care monitoring, autonomous surveillance and industrial process control where visual information is critical for situational awareness.

WMSNs are uniquely characterized by their scale often consisting of hundreds or thousands of densely populated nodes. In many practical scenarios, WMSNs are unattendly deployed in a vast territory in order to carry out their monitoring mission. Due to these special network features, the sensors of WMSNs are subjected to two main design constraints: low cost and low power consumption. The first constraint is mandatory because of the scale of the networks, while the second constraint is required because of the unattended nature of the application making the task of battery replacement or recharging either infeasible or painstaking. It is well known that together, these constraints severely restrict the individual capabilities of the nodes making it necessary to design sensing, compression, communications and security algorithms that are cost-effective.

A common aspect of WMSNs is the high correlation among the information streamed by neighboring wireless nodes; thus, an appropriate source coding technique should be used to avoid the transmission of that redundant information. A naive approach to such processing would require joint source coding whereby neighboring nodes coordinate amongst one another. This requires internodal communication subsequently spending energy and bandwidth.

This solution has been widely rejected by the research community, which is instead focused on applying the Slepian-Wolf theorem [2] in such a wireless sensor networks scenario. The Slepian-Wolf theorem indicates that a pair of correlated random variables representing sensor readings whose correlation model is known a priori can be independently coded without loss of performance compared to the jointly coded case. This paradigm promotes an asymmetrical distribution of algorithm complexity and alleviates processing at the lightweight source nodes by trading off with decoder complexity where more resources are available. It is worth pointing out that although the aforementioned theorem was introduced in the 1970s, there were not implementations until the seminal work of Pradhan and Ramchandran [3]. In fact, most of the subsequently appeared implementations are based on their DISCUS algorithm.

Among other scenarios, the WMSNs are specially suited for autonomous video surveillance. To address the issue of security of such systems in the presence of a hostile adversary Czarlinska *et al.* introduced the G-E-M methodology in [4]. This scheme can be divided into three parts: data-gathering, encryption, and energy drain mitigation. Among other features, G-E-M uses S/DISCUS (Secure DISCUS) [5] to distributedly code and encipher the video streams captured from the visual sensor even if the secret keys are temporally compromised. S/DISCUS accounts for practical limitations of multimedia wireless sensor systems. However, much room exists to improve the system performance, specifically, to increase quality of image reconstruction.

In this manuscript, the previously presented S/DISCUS

Dr. William Luh is kindly acknowledged for providing a preliminary version of the S/DISCUS code, which was subsequently modified and updated by the authors for the simulations in the paper.

This work was partially supported by Xunta de Galicia under Projects 10PXIB322231PR (SCALLOPS), 2010/85 (Consolidation of Research Units), and by the Spanish Government and the European Regional Development Fund (ERDF) under projects DYNACS (TEC2010-21245-C02-02/TCM) and COMONSENS (CONSOLIDER-INGENIO 2010 CSD2008-00010).

978-1-61284-350-6/11/$26.00 ©2011 IEEE

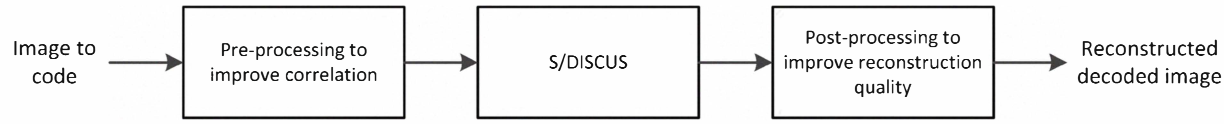

Fig. 1. Basic diagram of the scheme to improve quality of the reconstructed images encoded with S/DISCUS algorithm.

for image/video streaming is adapted in order to increase the quality of the reconstructed stream of images preserving its cryptographic properties. Additionally, these techniques can be used to effectively control the bitrate of the coded images.

The structure of the paper is as follows. A brief description of S/DISCUS applied to images is presented in Sect. 2. In Sect. 3 our dual approach to improve the quality of the image reconstruction is introduced presenting practical results, and we conclude in Sect. 4.

2. S/DISCUS APPLIED TO IMAGES

The S/DISCUS algorithm source codes and enciphers (without the use of traditional cryptographic keys) correlated sensor readings. In the spirit of Slepian-Wolf, no collaboration between sensor nodes is required. It is straightforward to realize that a lightweight distributed encipher and source coding algorithm like S/DISCUS shows significant advantages and is applicable to a broad set of sensor networking applications.

In order to explain S/DISCUS, let us assume that there exist m nodes generating k-length messages U_i^k (each element of these vector from Galois field GF(q)), $i = 1, \ldots, m$. S/DISCUS analysis indicates that in order to decode the distributedly encoded messages without errors, the following correlation condition between the messages has to be fulfilled

$$w(U_1^k + \ldots + U_m^k) < t, \tag{1}$$

where in the previous expression $w(\cdot)$ denotes the Hamming weight and the addition is over GF(q).

The messages U_i^k are independently encoded by each node obtaining their corresponding symbol $x_i^{n_i}$ as

$$X_i^{n_i} = \mathbf{H}_i U_i^k, \tag{2}$$

where $\mathbf{H}_i$ is a $n_i \times k$ matrix. At the decoder, if (1) holds, the original messages can be obtained from the received symbols $X_i^{n_i}$ of the m nodes.

S/DISCUS uses a supercode that can correct t errors and divides this supercode into subcodes imposing that the resulting subcodes must be maximum distance separable (MDS) codes to meet the secrecy requirements. The matrices used in (2) to encode the messages are the parity check matrices of these subcodes.

From (1), one may state that S/DISCUS requires a "hard" symbol correlation condition for encoding. This requirement is especially difficult to fulfill when the sources are images due to the complex nature of the associated luminance statistics. Thus, in order to at least partially meet the correlation model, the following scene post-processing scheme was introduced:

- Extract image background and send it to the receiver.
- Code the difference between the extracted background and each captured image. The black part of the difference image (which would correspond with the background) is maintained in order to increase the correlation of the pixels. Furthermore, the value of the pixels is quantized (removing the least significant bits of the pixels) to increase the correlation. Finally, the pixels of the resulting quantized and background subtracted image are interleaved to make uniform the correlation through the whole image.

By analyzing the PSNR curves in [4], one can see that the largest value of PSNR is around 20 dB, which can be considered insufficient in most of the practical scenarios. This distortion appears because the correlation condition described in (1) is not fulfilled; thus, errors messages are decoded and these errors are spread across the reconstructed images (because of the image interleaving). Due to the impulse-like nature of the noise, these error pixels can be accurately modelled as the well-known salt-and-pepper noise.

The reader should note that in order to objectively compare the results of this paper with those of the original visual S/DISCUS [4], we employ the same pair of test images and the same family of Reed Solomon supercodes (i.e., $(15, k)$).

3. DUAL ENHANCEMENT APPROACH TO IMPROVE THE QUALITY OF S/DISCUS CODED IMAGES

As discussed in the previous section, the application of S/DISCUS to images does not exhibit a level of reconstruction quality sufficient for many practical applications. To address this issue, we focus on performance enhancement by applying both a pre- and post-processing stage. A summary of our dual enhancement approach is presented in Figure 1.

In this way, we can employ S/DISCUS as a "black box" such that the preprocessing aids in shaping the image data to exhibit the statistical properties needed for better reconstruction. Furthermore, other non-spatial domains such as the 8×8 DCT or 4×4 integer DCT for compatibility with JPEG and h.264 can be employed. Transform domains have the advantage of both controlling the correlation of the symbols to be coded by S/DISCUS and enabling bitrate adjustment of the enciphered stream (e.g., by selectively transmitting subbands of the relevant visual bands).

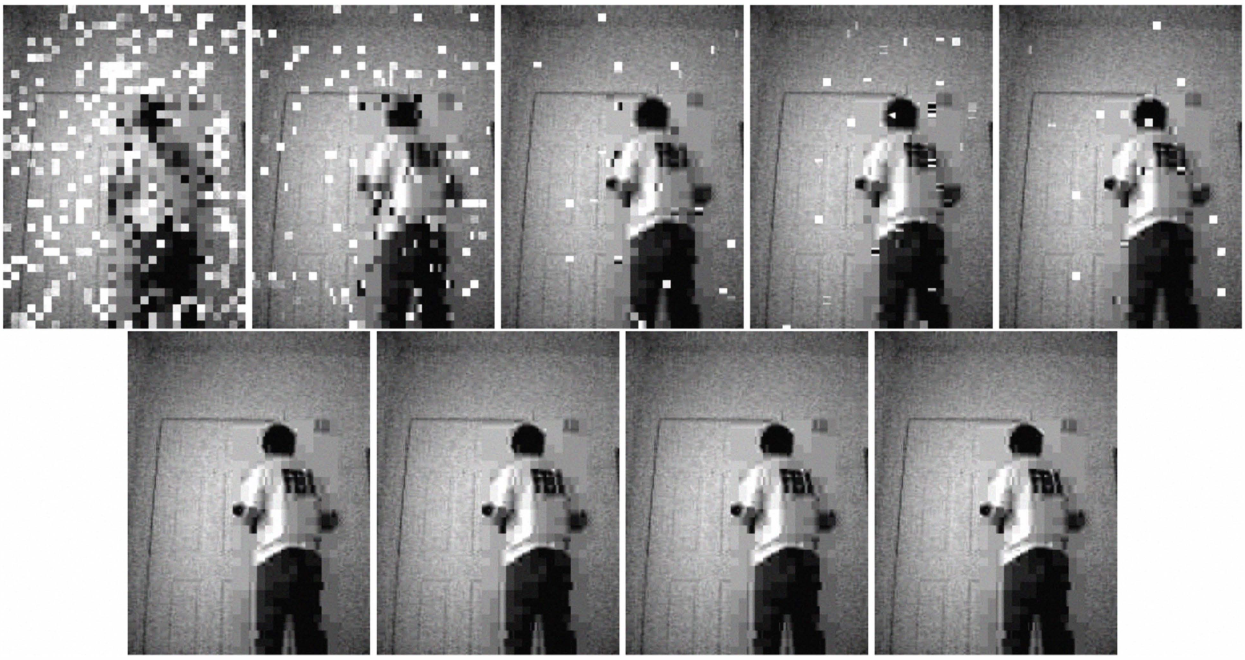

Fig. 3. Examples of S/DISCUS decoded images. In these cases, S/DISCUS encodes the quantized 8×8 DCT coefficients of the images (with an equivalent JPEG compression factor of 70). The number of coded coefficients of each block ranges from 1 to 9, where the corresponding images are sequentially ordered from top-left to bottom-right. The quantized coefficients of different frequencies are mixed and encoded together.

Post-processing filtering is applied on the S/DISCUS decoded images in order taking into account the distribution of the noise provoked by an error at the receiver. A median filter is widely recommended to deal with this noise family.

3.1. Pre-processing to improve the correlation

3.1.1. 8×8 *Discrete Cosine Transform - JPEG*

To meet the correlation requirements set in Equation (1), we consider transforming the image using the 8×8 DCT - JPEG. Our reasoning is that we can exploit the energy compaction properties of DCT subbands in order to better control the correlation. Application of this transform also enables the classification of image coefficients as a function of their visual significance.

We assert that the popularity of the 8×8 DCT in standards such as JPEG makes the integration of G-E-M and its component S/DISCUS, in particular, easier in popular image and video applications such as video streaming and image gathering. Furthermore, we can exploit the many techniques that have been studied on improving the efficiency of DCT processing. Essentially JPEG performs the DCT of non-overlapping 8×8 blocks of the image followed by coefficient quantization. The degree quantization is used to control the compression ratio. The associated matrix is a function of the quality factor $\in [0, 100]$, such that the associated quantization level of a given coefficient takes into account the "relevance" of a particular DCT band. As in the JPEG DCT, we have arranged the quantized coefficients of each block following the well-known zig-zag order.

In Fig. 2, two curves (denoted by the suffix "indep") are shown employing the JPEG DCT (fixing the quality factor 70, i.e., medium quality), where S/DISCUS is independently applied to each subband, i.e., applying S/DISCUS on DC, then to first AC subband, second AC subband and so on, the results do not show good performance. This is because an error of the DC coefficient induces an error of a whole 8×8 image block.

From the example described above, one can conclude that the independent use of S/DISCUS on each subband does not show an appropriate behavior to deal with real scenarios. As in the spatial case, the decoding errors are mainly produced by the lack of correlation of the quantized coefficients of each subband. Since quantized coefficients of higher subbands usually take on lower values compared to the quantized coefficients of the lower subbands (because a quantization matrix based on the visual human system is used), one can think of use these higher subband coefficients as increasing the correlation at the encoder by encoding the quantized coefficients of different blocks and subbands together. In order to implement this, the subbands of the transform of the image blocks to be transmitted must be firstly selected (e.g., the DC subband and the two lowest AC subbands of the transform of each image block). A vector composed by the quantized coefficients corresponding to the selected subbands of each image block of the image is defined. A pseudorandom permutation

Fig. 4. Examples of S/DISCUS decoded images. In these cases, S/DISCUS encodes the quantized 4×4 integer DCT coefficients of the images (with an equivalent H.264 Quantization Parameter of 20). The number of coded coefficients of each block ranges from 1 to 9, where the corresponding images are sequentially ordered from top-left to bottom-right. The quantized coefficients of different frequencies are mixed and encoded together.

is carried out on this vector, and its elements are sequentially coded with S/DISCUS. The perceptual improvement of this technique is shown in Fig. 2, where the two PSNR curves corresponding to this method (denoted by the suffix "mix") are shown with the PSNR curves using the method described in the previous paragraph, showing an increase of approximately 12 dB; thus, revealing the suitability of this second technique. In addition, these curves indicate that coding more than 10 of each block does not make sense in this specific experiment because the PSNR values do not increase using more coefficients, i.e., a maximum is reached. A visual example of this second pre-processing technique can be shown in Fig. 3.

3.1.2. 4×4 *Integer Discrete Cosine Transform - H.264*

In the previous section, we showed how we can control image correlation for S/DISCUS by coding the correlated visual sources in the well-known 8×8 block DCT domain. However, WMSNs are significantly constraints in their computational resources (due to the low cost hardware and also the power consumption requirements). In this way, it may be possible that the 8x8 block DCT is not computationally appropriate. The roadmap for the evolution of video coding algorithms has followed a similar route. Thus, we next apply the H.264 integer version of 4×4 DCT. This transform was specifically designed to establish a good trade-off between compression and complexity. In contrast to the JPEG transform, the core of the H.264 transform can be implemented solely via additions, subtractions and shifts [6].

Fig. 5 shows the PSNR curves for independent coding of each subband. Note the range of values of the quality factor of the H.264 transform is between 0 and 51. In this case study, we consider that a quantization parameter $QP = 20$ (medium quality) in a similar scenario to JPEG compression with quality 70. The corresponding results are improved over the 8x8 DCT results. However, the PSNR quality is still insufficient for practical situations.

As discussed in the previous section, we hypothesize that coding all the subbands together instead of coding them independently could increase the correlation in Equation (1); thus, reducing the probability of a decoding error. Fig. 5 shows the PSNR versus the number of coefficients for the situation of coding together. The PSNR increase is significant making this version suitable for more realistic.

A visual example of this latter pre-processing technique is shown in Fig. 4 demonstrating its performance when the number of coded coefficients is increased.

3.2. Post-processing for improved reconstruction quality

3.2.1. *Median Filtering*

In order to better recover the original image in the presence of decoding error, we approximate the error as salt-and-pepper noise. It is well-known that the preferred filter to tackle this kind of noise is the 2D median filter. In Fig. 6 the PSNR vs S/DISCUS parameter k (i.e., the dimension of the used Reed-Solomon code to implement S/DISCUS) curves are depicted

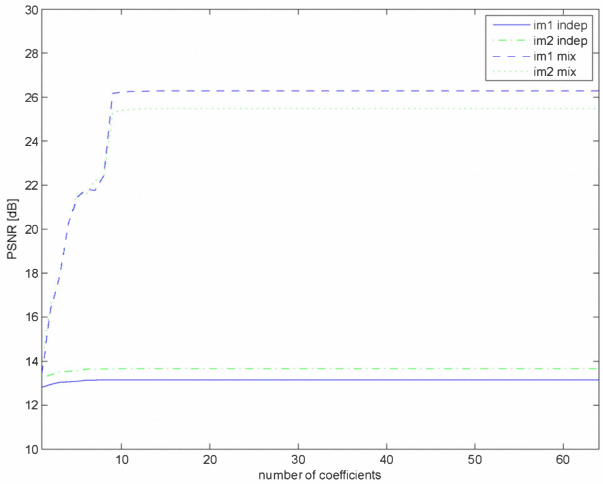

Fig. 2. PSNR vs number of coefficients curves for im1 (image 1) and im2 (image 2) for k = 2, 8 × 8 DCT and JPEG quality factor 70 independently encoding each subband (denoted by "indep") and encoding all the subbands together (denoted by "mix").

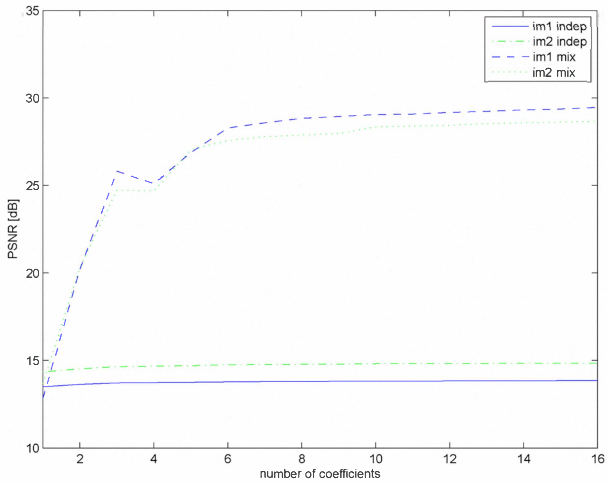

Fig. 5. PSNR vs number of coefficients curves for im1 (image 1) and im2 (image 2) for k = 2, 4 × 4 integer DCT and H.264 QP = 20 independently encoding each subband (denoted by "indep") and encoding all the subbands together (denoted by "mix").

for both images considering and not considering the use of this non linear filter (median filter with size 3 × 3 pixels). The difference of PSNR between the cases of using and not using the median filter is significant, decreasing their distance as the S/DISCUS parameter k is increased. Note that given the length of the used RS code, the larger k, the larger the compression. Focusing on the case k = 2 the PSNR gain of using the median filter is approximately 10 dB. Note that a value of PSNR larger than 25 dB can be considered acceptable to practical scenarios. Thus, the use of a 2D median filter is justified in improving performance. Furthermore, it is worthy pointing out that this filter is implemented at the receiver often positioned at a WMSN base station so we do not increase the computation complexity at the visual sensor side.

3.3. Complete system to improve the quality of S/DISCUS coded images

Fig. 7 and Fig. 8 show the curves of PSNR versus the size of the image median filter applied after decoding the received images, where the images were encoded using the 4 × 4 integer DCT (the first figure) and using the 8 × 8 DCT (the second figure) for different cases with different numbers of quantized transform coded coefficients.

According to these figures, filtering the decoded images with a median filter does not always improve the performance (i.e., PSNR). For example in the 4 × 4 integer DCT case, if five or six transform quantized coefficients are used to encode the image, the restored images will show the same or worse quality compared to the unfiltered case (note that the filter 1 × 1 is the identity filter). This decrease of the PSNR can also be shown in the 8 × 8 DCT case when 9 or 10 coefficients are used and a 9 × 9 median filter is applied. In order to understand this decrease in PSNR with the size of the median filter, it is worth pointing out the advantages and the drawbacks of the median filters. Median filters are successful in reducing salt-and-pepper noise, but at the same time distort valuable image detail. The larger the median filter size, the greater the opportunity to reduce noise, but the higher the chance of image artefacts. In typical scenarios, there is a median filter size that provides an appropriate compromise. However, in our tests there are scenarios where the noise is absent (e.g., when the conditions of the correlation are fulfilled), in this case the use of the median filters does not make sense. An example of this phenomenon can be shown in Fig. 9.

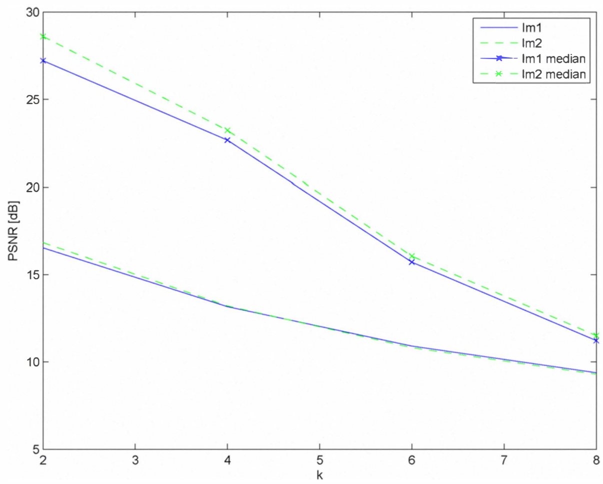

Fig. 6. PSNR vs k curves for im1 (image 1) and im2 (image 2) for both cases: using 3 × 3 median filter (denoted with "median") and without median filter.

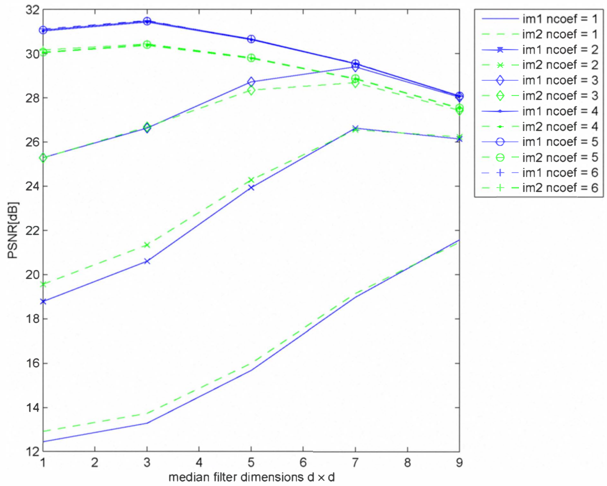

Fig. 7. PSNR vs size of the image median filter for different number of 4×4 integer DCT coded coefficients.

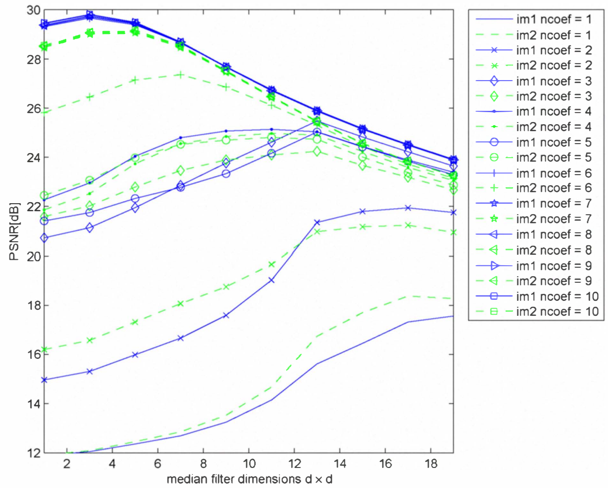

Fig. 8. PSNR vs size of the image median filter for different number of 8×8 DCT coded coefficients.

Fig. 9. Visual example before (left) and after (right) applying 5×5 image median filter to the decoded image with five coefficients of the 4×4 integer DCT.

4. CONCLUSIONS

An analysis of the performance of distributed source coding in more realistic scenarios is introduced. First, pre-preprocessing techniques are proposed in order to improve the correlation between the coefficients used to code the image in such a way that the error decoding probability is reduced. Next, a post-processing technique is employed in order to reduce the difference between the original image and the reconstructed image by taking into account the special statistics of the noise introduced when a decoding error appears.

Without significantly increasing encoder complexity, our algorithm shows an increase in the quality of the reconstructed images using S/DISCUS transmitting fewer bits per pixel compared to the spatial domain version especially when employing a 4×4 integer DCT.

5. REFERENCES

[1] Ian F. Akyildiz, Tommaso Melodia, and Kaushik R. Chowdhury, "A Survey on Wireless Multimedia Sensor Networks," *Computer Networks*, vol. 51, pp. 921–960, 2007.

[2] D. Slepian and J. Wolf, "Noiseless Coding of Correlated Information Sources," *IEEE Trans. Inform. Theory,*, vol. 19, no. 4, pp. 471 – 480, July 1973.

[3] S. Sandeep Pradhan, Eep Pradhan, and Kannan Ramchandran, "Distributed Source Coding Using Syndromes (DISCUS): Design and Construction," *IEEE Trans. Inform. Theory*, vol. 49, pp. 626–643, 1999.

[4] A. Czarlinska, W. Luh, and D. Kundur, "G-E-M Sensor Networks for Mission Critical Surveillance in Hostile Environments," in *Proc. IEEE INFOCOM Workshop on Mission Critical Networks*, Phoenix, Arizona, April 2008, pp. 1–6.

[5] W. Luh and D. Kundur, "Secure Distributed Source Coding with Side-Information," *IEEE Communication Letters*, vol. 12, no. 4, pp. 310–312, April 2008.

[6] Iain E. Richardson, *H.264 and MPEG-4 Video Compression: Video Coding for Next Generation Multimedia*, Wiley, August 2003.