Enhancing the Error Detection Capabilities of DCT Based Codecs using Compressed Domain Dissimilarity Metrics

Reuben A. Farrugia and Carl J. Debono
Department of Communications and Computer Engineering, University of Malta, Msida MSD2080, Malta
Email: *rrfarr@eng.um.edu.mt; cjdebo@eng.um.edu.mt*

***Abstract*— Video compression standards are implemented in wireless data transmission technologies to provide multimedia services efficiently. These compression standards generally utilize the Discrete Cosine Transform (DCT) in conjunction with variable length codes (VLC) in order to achieve the required high compression ratios. While providing the necessary high data rates, this technique has the disadvantage of making the system more susceptible to transmission errors. The standard decoders do not manage to detect a large number of corrupted macroblocks, 40.54% not detected for H.263+, contributing to a significant reduction in the end-to-end video quality as perceived by the end-user.**

This paper presents three dissimilarity metrics which contain both color and texture information and that can be extracted directly from the compressed DCT coefficients. These metrics can be used to enhance the error-detection capabilities of standard DCT based codecs. Simulation results show that the proposed algorithm increases the error detection rate by 54.06% with a gain in peak signal-to-noise ratio (PSNR) of 3.21 dB. This improvement in performance is superior to other solutions found in literature.

***Keywords*— Error detection coding, Quality assurance, Video codecs, Video signal processing.**

I. Introduction

Nowadays, a huge number of multimedia services are available; these include videoconferencing, videophones, digital TV broadcasting, and image databases, to mention a few. Such services usually adopt video compression standards [1] – [5] in order to reduce the transmission bit-rate and storage space requirements. Most of these image and video compression standards employ the Discrete Cosine Transform (DCT) to remove the spatial redundancies present in each block. The resulting DCT coefficients are quantized and entropy encoded in order to achieve high compression ratios.

The resulting compressed video sequence is more susceptible to transmission errors, since a transmission error in a codeword will not only effect this codeword, but may also effect subsequent codewords, resulting in major distortions in the reconstructed sequence [6]. In order to reduce the synchronization problem, these codecs usually adopt a fixed-length synchronization marker in every group of block (GOB) and Picture header. Although these synchronization markers limit the propagation of errors to the following GOBs, the synchronization problem will still effect a number of macroblocks (MBs) embedded within the corrupted GOB.

A number of error concealment algorithms [7] – [9] were presented in the past to provide high quality video even in the presence of transmission errors. However, error concealment algorithms require that the corrupted macroblocks are accurately detected and localized to operate appropriately. The standard decoders usually adopt syntax and semantic violations in order to detect transmission errors in the sequence. This approach does not manage to accurately localize the corrupted macroblocks. Taking as an example the case of the H.263+ codec, 40% ~ 60% of the corrupted macroblocks are not accurately detected [10].

Another bit-level error detection approach was suggested in [11], in which reversible VLC (RVLC) was used. The RVLC algorithm allows the compressed stream to be decoded in both forward and backward directions. Although this approach manages to improve the error detection capabilities of the decoder, it is still based on syntax and semantic violations and therefore not all corrupted macroblocks are detected. Moreover, some errors have no impact on the quality of the video, and it would be a waste of resources to try to correct them [12]. For this reason, error detection at picture level was considered in this work.

A number of pixel-level error detection algorithms are present in literature [6], [12] – [17]. However, these algorithms only manage to detect 20% – 40% of the corrupted macroblocks. Moreover, these algorithms provide a number of false detections which will reduce the quality of the reconstructed video sequence even when transmitted over an error-free channel. Additionally, most of these pixel-level features require that the whole frame is decoded which results in the introduction of an additional delay in the system.

An iterative error detection and concealment approach was presented in [12], where a combination of dissimilarity metrics were adopted by the authors, each of which requires the whole frame to be decoded. Although this algorithm achieves a significant gain in peak signal-to-noise ratio (PSNR) of 4 – 6 dB, this is achieved at the expense of a significant increase in computational time required by the decoder.

In order to reduce the computational complexity, the authors in [18] applied error detection in the DCT transform domain. However, the algorithm presented was tested on just one video sequence and therefore the algorithm may be sequence dependent.

This paper presents three dissimilarity metrics which can be employed to enhance the error detection capabilities of standard DCT based codecs. These metrics exploit both the color and the textural consistency in both the spatial and the temporal domains. This novel approach extracts the metrics directly from the DCT transformed domain, thus presenting a significant reduction in computational complexity. In addition, the proposed features can be extracted at decode time, and thus, in contrast to the other algorithms present in the literature, no additional delay is introduced in the system.

This paper is organized as follows. An overview of the DCT is provided in Section II followed by a discussion on the visual impairments caused by transmission errors in Section III. Section IV describes in some detail the DCT based dissimilarity metrics followed by the presentation and interpretation of the achieved simulation results in Section V. Finally comments and conclusion are provided in Section VI.

II. Discrete Cosine Transform

Block transform coding is widely adopted in image and video compression systems. The DCT successively operates on 8 x 8 image blocks, and is used in the JPEG [1], H.261 [2], H.263+ [3], MPEG-1 [4], and MPEG-2 [5] standards. The DCT makes use of the high degree of correlation between adjacent image pixels to provide energy compaction in the transformed domain. This results in a lossless representation of the original signal that is more suitable for quantization.

The DCT is defined on and N x N array of pixels [19]:

$$F_{u,v} = \frac{C_u C_v}{4} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \cos \frac{(2i+1)v\pi}{2N} \cos \frac{(2j+1)v\pi}{2N} f_{i,j} \quad (1)$$

with

$$C_u, C_v = \begin{cases} \frac{1}{\sqrt{2}} & \text{for u, v} = 0 \\ 1 & \text{otherwise} \end{cases} \quad (2)$$

where i and j are the horizontal and vertical indices of the N x N spatial array, and u and v are the horizontal and vertical indices of the N x N DCT coefficient array.

The distributions of the coefficients in the transformed block contain few large coefficients positioned in the upper-left hand corner of the transformed block and small coefficients elsewhere. Thus, the DCT transform considerably reduces the spatial redundancy of the block [19].

Each DCT coefficient is a linear combination of all pixel values within the block. From (1) it can be concluded that the DC coefficient of a block (u = 0, v = 0) represents the average pixel intensity of the block. The other AC coefficients reflect variation in pixel intensity in a certain direction at a certain rate [20]. For example, the upper horizontal coefficients and the left vertical coefficients in DCT transform domain represent some vertical and horizontal edge information, respectively [20 – 21].

III. Distortion Caused by Transmission Errors

As discussed in the previous sections, most of the image and video coding standards employ VLC codes. However, when VLC codes are transmitted over an error-prone channel, a single corrupted bit will desynchronize the bitstream until the next synchronization marker. The corrupted bitstream can still form valid entries in the VLC table, and therefore the decoder may not detect nor localize the errors [15]. In [10] it was found that for the H.263+ codec, around 40.54% of the corrupted macroblocks are not detected by the syntax and semantic violation test. In these cases, the decoder will continue decoding the unsynchronized sequence until a syntax violation, a semantic violation or a synchronization marker is encountered. Fig. 1 shows the resulting quality of a video sequence when the corrupted macroblocks are not precisely detected and concealed.

Fig. 1 A decoded frame of the Erik sequence representing the distortions caused by channel errors.

These visual impairments generally provide macroblocks whose color and texture does not fit in either the spatial or temporal domains. The proposed dissimilarity metrics exploit this property, and are used to accurately detect and localize the corrupted macroblocks that provide considerable visual impairments. The macroblocks which are detected as corrupted by these dissimilarity metrics are concealed. As will be confirmed through simulation results, this method achieves a significant improvement in the overall quality of the decoded video sequence.

IV. Dissimilarity Metrics

The corrupted macroblocks can be detected using the DCT coefficients, since these contain both color and textural information [20 – 21]. Fig. 2 illustrates the features that can be used to enhance the error detection capabilities of DCT based codecs. Each feature contains information about the color, together with (1) vertical, (2) diagonal and (3) horizontal edge information. These features are computed by summing the coefficients along the respective direction and are given by equations (3) – (5):

$$F_1 = \sum_{i=0}^{N-1} p(i,0) \quad (3)$$

$$F_2 = \sum_{i=0}^{N-1} p(i,i) \quad (4)$$

$$F_3 = \sum_{i=0}^{N-1} p(0,i) \quad (5)$$

where p is the DCT coefficient and N is the block size.

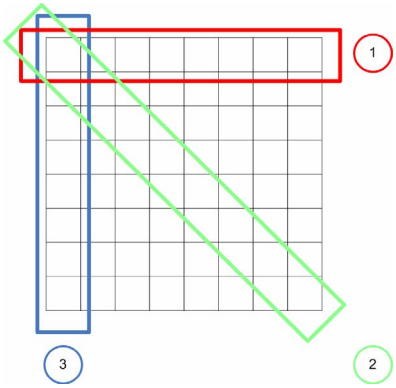

Fig. 2 DCT features containing both color and textural information.

A. Spatial Dissimilarity Metric

The spatial dissimilarity metric is based on the fact that there is a correlation between the DCT coefficients of the current block being decoded, and the DCT coefficients of the neighboring blocks in space. Therefore, the DCT coefficients of the current macroblock can be approximated from the surrounding blocks. If the color and textural information of the predicted DCT block is significantly different from the color and texture of the current block, then the block is probably corrupted.

DCT block prediction is performed using data from the same luminance or chrominance components (Y, C_B, or C_R). As illustrated in Fig. 3, the DCT block prediction algorithm adopts both the block above and the block to the left of the current macroblock being tested, similar to the one adopted in [3] and [22]. Since these macroblocks are already available, this method can be applied at decode time. The DC coefficient is derived using the average DC value from the above and left macroblock. The vertical AC coefficients are derived from the first row of the above macroblock while the horizontal AC coefficients are derived from the first column of the block to the left.

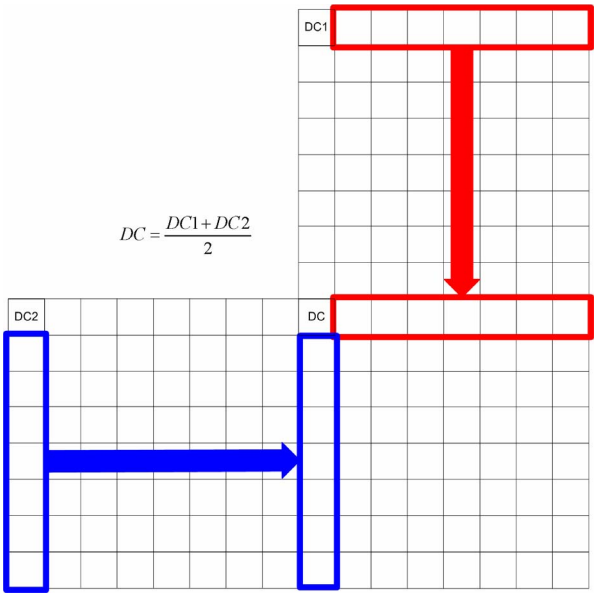

Fig. 3 Neighboring blocks used by the DCT block prediction algorithm.

The DCT block prediction algorithm derives only the results of the first row and column of each block. Therefore, only the F_1 and F_3 features can be extracted from the predicted DCT block. For each color component, the features F_{1p} and F_{3p} are extracted from the predicted macroblock, while F_{1t} and F_{3t} are extracted from the current macroblock under test. The dissimilarity metrics d_1 and d_2 can then be computed using equation (6):

$$d_n = \sqrt{\left\{\frac{1}{4}\sum_{i=0}^{3}\left|F_{npY(i)} - F_{ntY(i)}\right|\right\}^2 + \left|F_{npCb} - F_{ntCb}\right|^2 + \left|F_{npCr} - F_{ntCr}\right|^2} \quad (6)$$

Finally, the spatial dissimilarity metric $d_{spatial}$ is computed by averaging the two dissimilarity metrics d_1 and d_2. This dissimilarity metric will provide large values for corrupted macroblocks and small values for uncorrupted macroblocks, thus, by applying an appropriate threshold, errors can be easily detected.

B. Temporal Dissimilarity Metric

The temporal dissimilarity metric is based on the fact that the difference between the color and textural information of the current block and the corresponding block in the previous frame is usually small when the block is not corrupted. On the other hand, this difference would be significantly large when the block is corrupted.

For each color component, the features F_{1t-1}, F_{2t-1} and F_{3t-1} are extracted from the previous macroblock. Similarly, features F_{1t}, F_{2t} and F_{3t} are extracted from the current macroblock. The dissimilarity metrics d_1, d_2 and d_3 are then computed using (6). The temporal dissimilarity metric $d_{temporal}$ is derived from the average of the dissimilarity metrics d_1, d_2 and d_3.

C. Spatio-Temporal Dissimilarity Metric

The combined effect of the spatial and temporal dissimilarity metric is termed as being the spatio-temporal dissimilarity metric and is derived formally as:

$$d_{spatio-temporal} = d_{spatial} + d_{temporal} \quad (7)$$

V. Simulation Results

The aim of the dissimilarity metrics described in the previous section is to provide small metrics for uncorrupted macroblocks and large metrics for corrupted macroblocks. However, the metrics of corrupted and uncorrupted macroblock classes usually overlap and therefore a macroblock cannot be easily classified as being corrupted or not. Another important factor is that the false detection rate should be kept at a minimum, since false detection will conceal a number of uncorrupted macroblocks thus reducing the quality of the video sequence even if no errors are present in the channel.

The performance of the proposed dissimilarity metrics was evaluated from the probability of successfully detecting corrupted macroblocks given a 0% false detection rate, p (bad, no false detection), as shown in Fig. 4. Large values of p (bad, no false detection) mean that the dissimilarity metric provides better separation between the classes, and therefore is more robust. Five video sequences were considered in this simulation, including videoconferencing like sequences “Erik”, “Silent” and “Akiyo”, the sports sequence “Football”, which contains fast moving objects, and “News” which contains an abrupt shot.

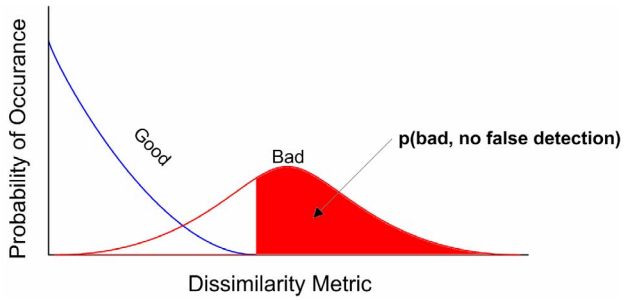

Fig. 4 Probability density function of Good (uncorrupted MBs) and Bad (corrupted MBs)

The video sequences under test were stored at CIF resolution and were compressed by an H.263+ Encoder. The resulting compressed bitstream was then corrupted with a BER of 1.3E-3. If a syntax or semantic violation is detected by the standard decoder, all the following macroblocks, until the next synchronization marker, are concealed. Otherwise, the spatial, temporal and spatio-temporal dissimilarity metrics discussed in the previous section are computed. If the dissimilarity metric is larger than a global threshold, it conceals this macroblock and the following macroblocks, until the next synchronization marker.

Table I summarizes the performance of the error detection algorithms using different dissimilarity metrics. These results show that the spatial dissimilarity metric achieves an average separation of about 41.81% and is the most stable metric of the three. It can also be observed that the performance of the temporal dissimilarity metric performs well for videoconferencing applications, with an average separation of 66.12%. However, this metric suffers in the presence of shots and fast moving objects present in the sequence. The spatio-temporal dissimilarity metric provides a compromise between the two algorithms.

TABLE I. DISSIMILARITY METRIC CLASS SEPARATION TEST

Diss Metric	Erik	Akiyo	Silent	News	Foot-ball
$d_{Spatial}$	41.53%	46.98%	34.45%	36.44%	49.65%
$d_{Temporal}$	71.19%	77.59%	49.58%	7.63%	37.06%
$d_{Spatio-Temporal}$	66.95%	63.79%	60.50%	35.59%	46.58%

The global thresholds (T), required by the decoder to decide whether a specific macroblock is corrupted or not, were derived heuristically in order to maximize the error rate while minimizing the false detection rate. The error concealment adopted by the system replaces each macroblock with the motion compensated macroblock from the previous frame whose motion vector was computed from the median of the motion vectors of the neighboring macroblocks.

Table II summarizes the performance of the spatial, temporal and spatio-temporal dissimilarity metrics in terms of P(ED) (overall error detection rate), P(FD) slow (false detection rate for sequences containing neither shots nor fast moving objects in the sequence) and P(FD) all (overall false detection rate). In order to derive more accurately the false detection rate, a larger collection of video sequences consisting of a total of 2000 frames encoded in CIF resolution was considered.

TABLE II. ERROR DETECTION CAPABILITIES

Diss Metric	T	P(ED) (%)	P(FD) slow (%)	P(FD) all (%)
$d_{Spatial}$	964.47	27.80	0.00	0.00
$d_{Temporal}$	636.03	53.59	10.8E-3	61.99E-3
$d_{Spatio-Temporal}$	655.43	54.06	0.00	11.81E-3

From these results it can be concluded that the spatial dissimilarity metric is the most stable dissimilarity metric with an overall error detection rate of 27.80% and an overall false detection rate of 0.00%. However, the spatio-temporal dissimilarity metric performs well with video sequences which do not contain shots or fast moving objects in the sequence. This can also be seen in Table III, where the spatio-temporal dissimilarity metric has obtained the best quality video sequence, with an overall gain in PSNR of 3.21 dB. The gain in subjective quality of the video sequence is illustrated in Fig. 5.

TABLE III. PSNR GAIN USING (DB)

Diss Metric	Erik	Akiyo	Silent	News	Foot-ball
$d_{Spatial}$	1.8954	2.7273	2.3175	2.4352	0.5644
$d_{Temporal}$	2.8872	5.0020	4.1710	3.3325	-0.4288
$d_{Spatio-Temporal}$	2.8589	4.7551	4.1449	3.5729	0.6946

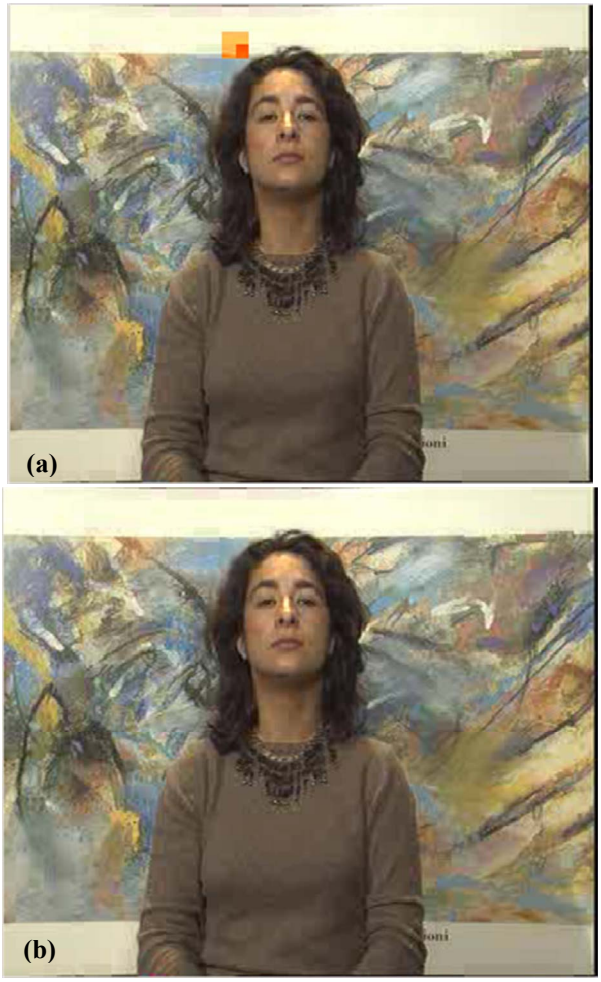

Fig. 5 Frame 4 of a corrupted Silent Sequence using (a) Standard H.263+ Decoder (b) Enhanced error detection capabilities using the spatio-temporal dissimilarity metric.

VI. Comments and Conclusion

This paper has presented three novel dissimilarity metrics that can be used to enhance the error detection capabilities of standard DCT-based video codecs. The proposed metrics exploit both the color and textural information contained within the DCT coefficients. These features can be extracted in the compressed domain, thus significantly reducing the computational time required by the algorithm to detect and conceal errors.

These metrics were tested on a wide range of video sequences including videoconferencing like sequences and sports sequences. From these tests it was found that the temporal dissimilarity metric performs well when dealing with videoconferencing like sequences but it suffers in the presence of abrupt shots or moving objects present in the sequence, since these generally generate false detections. The spatio-temporal dissimilarity metric achieves the best results with an overall gain in PSNR of 3.21 dB, which is comparable to the pixel domain dissimilarity metrics solution presented in [23], and superior to other solutions found in literature.

References

[1] ISO/IEC 10918, "Information technology – Digital compression and coding of continuous tone still images," 1994.

[2] ITU-T Rec. H.261, "Video Codec for audiovisual services at p x 64 bit/s," 1993.

[3] ITU-T Rec. H.263, "Video Coding for Low Bit-Rate Communication," 2005.

[4] ISO/IEC 11172, "Information technology – Coding of moving pictures and associated audio for digital storage media at up to 1.5 Mbit/s," 1993.

[5] ISO/IEC 13818, "Information technology – Generic coding of moving pictures and associated audio information," 2000.

[6] W. Chu and J. Leou, "Detection and Concealment of Transmission Errors in H.261 Images," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 8, no. 1, pp. 74-84, Feb. 1998.

[7] Y. Wang and Q. Zhu, "Error Control and Concealment for Video Communication: A Review," *Proc. IEEE*, vol. 86, no. 5, pp. 974-997, May 1998.

[8] J. Suh and Y. Ho, "Error Concealment Techniques for Digital TV," *IEEE. Trans. Broadcast.*, vol. 48, no. 4, pp.299-306, Dec. 2002.

[9] S. Tsekeridou and I. Pitas, "MPEG-2 Error Concealment Based on Block-Matching Principles," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 10, no. 4, pp. 646-658, Jun. 2000.

[10] W. Park, and B. Jeon, "Error Detection and recovery by Hiding Information into Video Bitstream using Fragile Watermarking," *Proc. SPIE Visual Communications and Image Processing 2002*, vol. 4671, pp. 1-10, Jan. 2002.

[11] J. Wen and J.D. Villasenor, "Reversible Variable Length Codes for Robust Image and Video Transmission," *Proc. 31st Asilomar Conf. on Signals, Systems, and Computers*, vol. 2, pp. 973-979, Nov. 1997.

[12] E. Khan, S. Lehmann, H. Gargi, and M. Ghanbari, "Iterative Error Detection and Correction of H.263 Coded Video for Wireless Networks," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 12, pp. 1294-1307, Dec. 2004.

[13] H. Shyu, and J. Leou, "Detection and Concealment of Transmission Errors in MPEG-2 Images – A Genetic Algorithm Approach," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 9, no. 6, pp. 937-948, Sept. 1999.

[14] Y. Han and J. Leou, "Detection and Correction of Transmission Errors in JPEG Images," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 8, no. 2, pp. 221-231, Apr. 1998.

[15] M.R. Pickering, M.R. Frater and J.F. Arnold, "A Statistical Error Detection Technique for Low Bit-Rate Video," *IEEE Proc. of TENCON Conf.,* vol. 2, pp. 773-776, Dec. 1997.

[16] S. Ye, X. Lin and Q. Sun, "Content Based Error Detection and Concealment for image Transmission over Wireless Channel," *IEEE Proc. of ISCAS Conf.*, vol. 2, pp. 368-371, May 2003.

[17] O. Lehtoranta, T.D. Hamalainen, and V. Lappalainen, "Detecting Corrupted Intra Macroblocks in H.263 Video," *IEEE Workshop on Multimedia Signal Processing*, pp. 33-36, Dec. 2002.

[18] K. Bhattacharyya and H.S. Jamadagni, "DCT Coefficient-Based Error Detection Technique for Compressed Video," *IEEE Proc. of ICME Conf.*, vol. 3, pp. 1483-1486, Aug. 2000.

[19] A. Sadka, *Compressed Video Communications*, pp. 25-54. John Wiley and Sons, UK, 2002.

[20] B. Shen and I.K. Sethi, "Direct feature extraction from compressed images," *Proc. SPIE*, vol. 2670, pp. 404-414, Storage and Retrieval for Image and Video Databases IV, 1996.

[21] H. Bae and S. Jung, "Image Retrieval using Texture Based on DCT," *IEEE Proc. of ICICS Conf.*, vol. 2, pp. 1065-1068, Sep. 1997.

[22] G. Côté, B. Erol, M. Gallant, and F. Kossentin, "H.263+: Video Coding at Low Bit Rate," *IEEE Trans. Circuits Syst. Video Technol*, vol. 8, no. 7, pp. 849-866, Nov. 1998.

[23] R.A. Farrugia and C.J. Debono, "Enhancing the Error Detection Capabilities of the Standard Video Decoder using Pixel Domain Dissimilarity Metrics," *Proc. of EUROCON 2007 Conf.,* Sept. 2007.