Global Motion Compensation and Spectral Entropy Bit Allocation for Low Complexity Video Coding

Malavika Bhaskaranand and Jerry D. Gibson
Department of Electrical and Computer Engineering
University of California, Santa Barbara, CA - 93106
Email: {malavika, gibson}@ece.ucsb.edu

Abstract—**Most standard video compression schemes such as H.264/AVC involve a high complexity encoder with block motion estimation (ME) engine. However, applications such as video reconnaissance and surveillance using unmanned aerial vehicles (UAVs) require a low complexity video encoder. Additionally, in such applications, the motion in the video is primarily global and due to the known movement of the camera platform. Therefore in this work, we propose and investigate a low complexity encoder with global motion based frame prediction and no block ME. We show that for videos with mostly global motion, this encoder performs better than a baseline H.264 encoder with ME block size restricted to 8×8. Furthermore, the quality degradation of this encoder with decreasing bit rate is more gradual than that of the baseline H.264 encoder since it does not need to allocate bits across motion vectors (MVs) and residue data. We also incorporate a spectral entropy based coefficient selection and quantizer design scheme that entails latency and demonstrate that it helps achieve more consistent frame quality across the video sequence.**

I. INTRODUCTION

Traditional video compression schemes such as MPEG-2 and H.264/AVC use a highly complex encoder with a block motion estimation (ME) scheme and typically target applications such as entertainment video storage and playback (DVDs), video streaming, video conferencing, and entertainment broadcasting. In contrast, unmanned aerial vehicles (UAVs) present video compression requirements different from the typical video compression applications. It is desirable to have low complexity video encoders in UAV payloads due to the tight constraints on space, weight, and power of UAVs with the desire for longer endurance, greater functionality, and lower fuel consumption [1]. Additionally, the motion in the video sequences is primarily global and due to the known motion of the UAV and the camera mounts. Hence video compression in UAVs has constraints and degrees of freedom not addressed by traditional video compression schemes.

Previous research on low complexity video encoders has been dominated by low complexity versions of the H.264 standard and methods based on the Wyner-Ziv distributed source coding theory. Low complexity H.264 encoders [2], [3] discard certain complex toolsets such as CABAC entropy coding or pursue short-cuts in mode decisions, resulting in poorer performance. Wyner-Ziv video codecs [4], [5] outperform H.264 intra and sometimes even H.264 'zero-motion' coding. Nevertheless, they require a feedback channel and instantaneous decoding at the receiver [6] and modifications proposed to avoid the feedback channel [7], [8] increase the complexity of the encoder and result in a loss of quality.

Research has been done on incorporating global motion in standard codecs such as MPEG-4 and H.264/AVC. Global motion compensated frames are used as reference frames for block motion estimation (ME) in [9], [10]. Block prediction based on global motion is used in addition to regular H.264 intra and inter modes in [11]–[13]. However, these techniques have been used within a block ME framework for standard video sequences but have not been used on entire frames for video sequences where the motion is almost entirely global.

Video compression schemes tailored for UAV applications that exploit the available global motion information have also been proposed. Gong *et.al* [14] use a homography to model the global motion, merge the first intra frame and subsequent inter frame residues in a frame group into a single "big image", and code it using JPEG2000. However, the creation of the "big image" is not highly conducive to JPEG2000 compression since one portion is the intra frame and the rest are residues while JPEG2000 is primarily designed for natural images. Rodriguez *et.al* [15] use the available global motion information to simplify block ME in a MPEG-4 encoder. While their approach reduces the complexity of a standard video encoder, transmitting the motion vectors is redundant since the motion parameters of the UAV and camera mounts is readily available at the receiver.

In this work, we propose a low complexity encoder that utilizes the global motion information available for global motion compensated frame prediction. Results demonstrate that this encoder has better rate-distortion performance than a H.264 encoder with ME block size constrained to 8×8 that has complexity in the same order as the proposed encoder. The quality degradation with decreasing bit rate is also more graceful for the encoder with global motion since bits need not be allocated across motion vectors and residue data. We also integrate a spectral entropy based coefficient selection and bit allocation scheme that examines future frames and show that it helps make the quality of frames more consistent across the video sequence. The performance of the encoders is

---

This research has been supported by Raytheon Applied Signal Technology, Inc.

evaluated in terms of the average bit rate required to achieve a given quality level and standard deviation of the quality across frames. The frame quality is quantified using peak signal-to-noise ratio (PSNR) and Structural SIMilarity (SSIM) [16] measures.

It is to be noted that in addition to being used in military applications, UAVs are being increasingly used in commercial applications such as monitoring areas that are unsafe for humans (like Japan's Fukushima Daiichi tsunami-damaged nuclear plant), plume tracking, and collecting bio-related data from large areas [17]. Also, video compression algorithms motivated by the UAV video surveillance can be employed in applications such as video surveillance of remote areas with little transient human activity expected, where low cost and low complexity video capture units are desirable and majority of the motion in videos is global and due to the known motion of the camera mounts. Therefore, the proposed encoder can be used in several non-military applications.

This paper is organized as follows. Section II describes the encoder architecture and briefly explains the spectral entropy based coefficient selection and bit allocation schemes. Section III presents the results and discusses them. Section IV summarizes the work and draws conclusions.

II. LOW COMPLEXITY VIDEO ENCODER

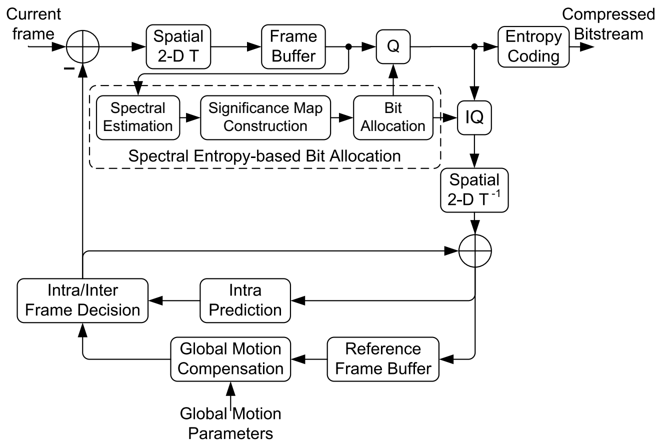


Fig. 1. Architecture for the proposed low complexity video encoder

The proposed encoder is shown in block diagram form in Fig. 1. The first frame in each group of pictures (GOP) is coded as an intra frame with intra prediction. The remaining frames in the GOP are inter predicted using global motion compensated reconstructed frames. The global motion parameters are derived from the known camera motion and are not estimated from the original frames using complex algorithms. Unlike in conventional block ME, this prediction is for the whole frame and not for individual blocks. The prediction residue is put through a spatial 2-D transform and the variances of the transform components are estimated using all the coefficients in the frame. The significant coefficients that are to be coded are chosen based on spectral entropy as proposed in [18] and bits are allocated to the chosen coefficients as per the scheme developed in [19]. The bits allocated are used to design scalar quantizers that operate independently on the transform components. The quantized coefficients are then entropy coded to generate the compressed bitstream.

We have implemented the proposed encoder using the H.264 integer transform [20] as the spatial 2-D transform. We also have adopted the H.264 quantization framework and design H.264 quantization matrices (QMs) using spectral entropy principles. The GOP size is set to 8 and the inter frames are predicted using two global motion compensated reconstructed frames: the preceding frame and the next available intra frame. The following subsections briefly describe the global motion model and the spectral entropy based coefficient selection and bit allocation schemes.

A. Global motion model

In the test sequences used in this work, the global motion consists mostly of rotation, scaling, and translation. Hence the RST global motion model is employed which is defined by

$$T_\theta(x,y) = \begin{bmatrix} \theta_3 & -\theta_4 \\ \theta_4 & \theta_3 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix}, \tag{1}$$

where (x,y) is the position of the pixel in the original frame, $T_\theta(x,y)$ the position of the pixel in the motion compensated frame, $[\theta_1,\theta_2]^T \in \mathbb{R}^2$ is the translation vector, and $(\theta_3,\theta_4) \in \mathbb{R}^2$ are the parameters describing the rotation and scaling. Higher order, more complex models such as homography could be easily used in place of the RST model, if the video sequence requires it.

B. Spectral entropy based coefficient selection [18]

In transform based compression schemes where the bandwidth is limited, it is not possible to transmit all transform coefficients and hence some coefficients need to be discarded. Therefore it is important to choose or sample the transform coefficients that best represent a signal (significant coefficients) and code them with high fidelity. Assuming zero mean, coefficients with magnitudes greater than a threshold are retained and the remaining are discarded. In many video encoders, the thresholds used to determine the coded coefficients are chosen heuristically or based on perceptual tests. Yang and Gibson [18], [21], [22] developed a method to choose significant coefficients with theoretical underpinnings based on Campbell's coefficient rate [23].

Consider a zero-mean stationary continuous-time random process $X(t)$ whose K-L coefficients in the time interval $[0,T]$ are $\{C_1, C_2, \ldots, C_M\}$ with C_i's being uncorrelated random variables with $\mathbf{E}[C_i] = 0$ and $\mathbf{E}[C_i^2] = \lambda_i$. Let $y(t_1, t_2, \ldots, t_N)$ be a product of N independent sample functions of $X(t)$. Yang and Gibson [18] showed that for large N, the number of occurrences of λ_i in the high energy terms of the energy of $y(t_1, t_2, \ldots, t_N)$ is proportional to λ_i and given by

$$n_i = \frac{\lambda_i}{\sigma^2} N, \qquad i = 1, 2, \ldots, M, \tag{2}$$

where $\sigma^2 = \sum_{i=1}^{M} \lambda_i$ is the total average energy of the $X(t)$.

Therefore if transform coefficients from N blocks of data are available, (2) suggests that the number of coefficients coded in each transform component should be proportional to the energy in that component. Once the number of coefficients to be coded in each transform component are determined, the coefficients across all the N blocks are examined and those with the largest magnitudes are chosen and coded. This method requires latency to compute the component energies and to examine the coefficients across all blocks. In contrast to the classical method, this coefficient selection mechanism has been shown to achieve better SNR (with no rate control) [21] and subjective quality [22].

C. Spectral entropy based bit allocation

Let there be N sampling functions (blocks/frames) each with M transform components. Out of the total $M \times N$ coefficients, let L coefficients be coded. Then the spectral entropy based coefficient selection derived in the previous subsection II-B dictates that the number of coefficients n_i coded in each component be proportional to the energy λ_i of that component *i.e.* $n_i = \frac{\lambda_i}{\sigma^2} L$.

If $b_i^{(S)}$ is the average number of bits spent to code a significant coefficient of component i, the total number of bits spent is $B^{(S)} = \sum_{i=1}^{M} n_i b_i^{(S)} + B(\text{sig.map})$ where $B(\text{sig.map})$ is the number of bits required to code the binary significance map that indicates the significant coefficients. The coding distortion is generated by two sources: quantization and discarding coefficients. Hence the expected value of the distortion of the ith component can be written as

$$\begin{aligned} d_i^{(S)} &= n_i \times \mathrm{E}(\text{quantization error}) + \\ &\qquad (N - n_i) \times \mathrm{E}(\text{energy of discarded coefficients}) \\ &= n_i \times h_i \lambda_i 2^{-2b_i^{(S)}} + (N - n_i) \times \lambda_i \end{aligned} \tag{3}$$

Here, the quantization error is computed assuming that the overload distortion is negligible and the high-resolution approximation holds and h_i is a constant determined by the distribution of the normalized random variable $C_i/\sqrt{\lambda_i}$ [24].

Setting up the bit allocation problem as finding $b_i^{(S)}$ for $i = 1, 2, \ldots, M$ so as to minimize the overall bit consumption $B^{(S)}$ subject to the constraint that $D = \sum_{i=1}^{M} d_i^{(S)}$, the optimal $b_i^{(S)}$ can be shown to be

$$b_i^{(S)} = \frac{1}{2} \log_2 \left(\frac{\lambda_i h_i}{[D - \sum_{i=1}^{M} (N - n_i) \lambda_i]/L} \right). \tag{4}$$

Spectral entropy bit allocation results in each transform component having the same average quantization distortion per coded coefficient. The bit allocation expression in (4) has a form similar to the classical bit allocation result $b_i^{(C)} = \frac{1}{2} \log_2 \left(\frac{\lambda_i h_i}{D/(MN)} \right)$. The denominator within the logarithm in (4) averages only the quantization distortion over only the significant coefficients while that in the classical result averages the total distortion budget over all coefficients. The spectral entropy bit allocation scheme hence can be interpreted as dividing the total distortion budget between quantizing and discarding coefficients.

In our implementation, the total number of coefficients to be coded in each frame is decided based on the relative energy of the frame within that GOP and the target average distortion. Then for each frame, the number of significant coefficients in each transform component is computed to be proportional to the empirical energy of that component computed over the entire frame. Finally, coefficients with the largest magnitudes are chosen and bits are allocated to them so as to meet the target distortion. Therefore, the use of the spectral entropy based coefficient selection and bit allocation scheme requires a latency equal to number of frames in a GOP.

III. RESULTS AND DISCUSSION

In this section, we analyze the contributions of the global motion compensation and spectral entropy based quantizer design by comparing the performance of encoders with the following different combinations of toolsets.

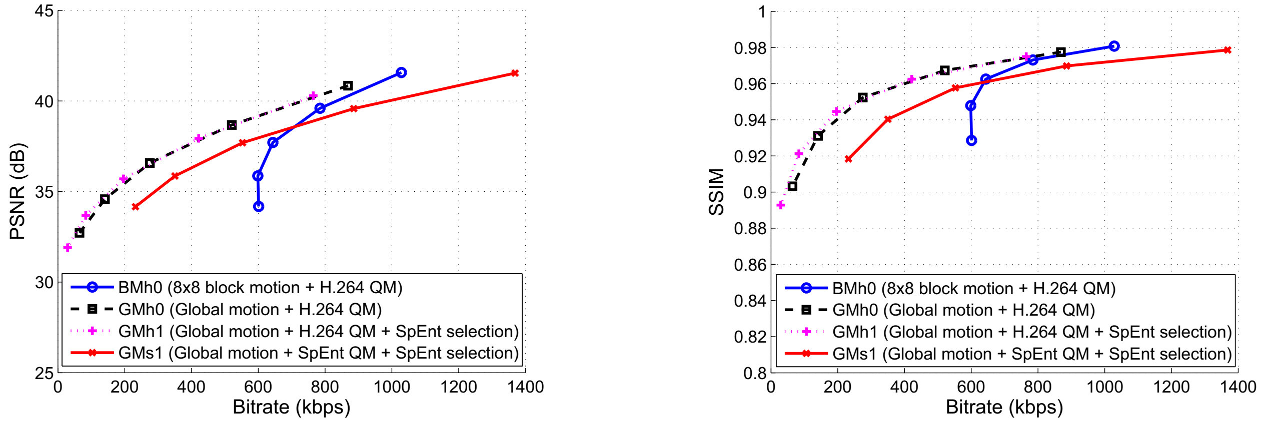
- BMh0 - Block motion estimation with 8×8 blocks with default H.264 QM. (This is the baseline H.264 encoder with no deblocking filter.)
- GMh0 - Global motion compensation with default H.264 QM.
- GMh1 - Global motion compensation with default H.264 QM and spectral entropy coefficient selection
- GMs1 - Global motion compensation with spectral entropy QM and spectral entropy coefficient selection

The notation used here is as follows: ‘BM’ refers to block motion while ‘GM’ refers to global motion; ‘h’ indicates default H.264 QM use while ‘s’ indicates spectral entropy QM; and ‘1’ denotes spectral entropy coefficient selection while ‘0’ denotes no spectral entropy coefficient selection (all coefficients that are non-zero after quantization are transmitted).

The baseline H.264 encoder (BMh0) is chosen such that its complexity is in the same order as, but higher than that of the GMs1 encoder. Complexity analysis of these two encoders in terms of storage requirements, memory accesses, and computations has been done in [25].

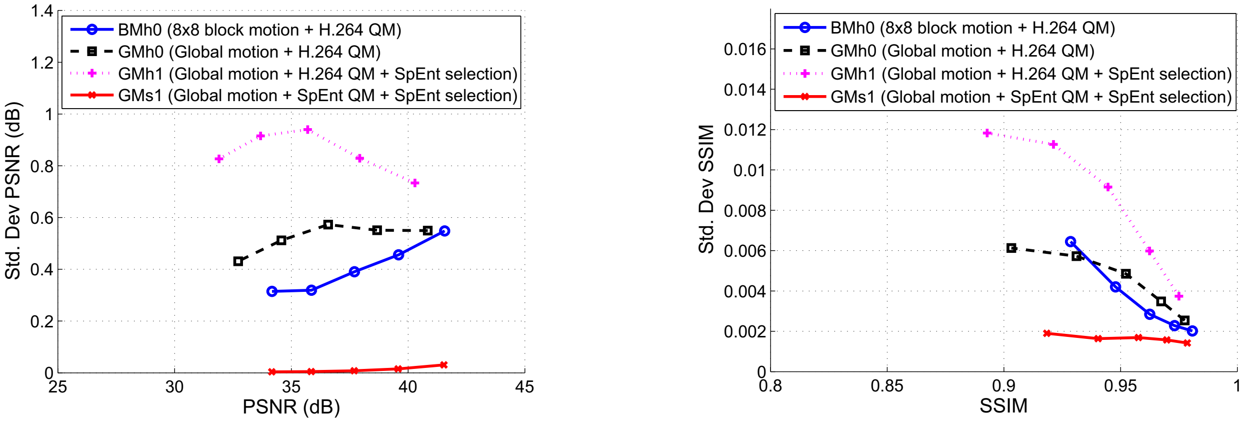
We have tested these encoders for video sequences where most of the motion is global and due to the motion of the camera platform. Here we present and discuss results only for the “aerial_beach1” sequence at 400×240 resolution which has been captured from a helicopter flying along a beach. The majority of the motion in the video is global and due to the movement of the helicopter and the panning and rotation of the camera. The results for the other test video sequences show trends similar to those for the “aerial_beach1” sequence and similar conclusions can be drawn.

The average quality in terms of PSNR and SSIM is plotted against bit rate of the various encoders for the “aerial_beach1” sequence in Figs. 2 (a) and (b). The corresponding standard deviation of quality is plotted against the average quality in Figs. 2 (c) and (d). Comparing the curves for BMh0 and GMh0 in Figs. 2(a) and (b), it can be seen that the quality degradation with bit rate of the GMh0 encoder at lower bit



(a) Average PSNR vs. bit rate

(b) Average SSIM vs. bit rate



(c) Standard deviation of PSNR vs. bit rate

(d) Standard deviation of SSIM vs. bit rate

Fig. 2. Comparison of performance for 400 × 240 “aerial_beach1” sequence.

rates is gradual while the quality for the BMh0 encoder drops sharply. This is due to the fact that the BMh0 encoder allocates bits across MVs and residue data, while the GMh0 encoder does not. The use of global motion instead of block motion also gives significant performance gains at lower bit rates. This is because at lower bit rates, motion vectors (MVs) consume a large fraction of the total bit rate when block motion is used and MVs could use more bits than the frame residues. The performance gains decrease at higher bit rates where the bit rate savings on MVs are less important and the savings on residue data are not very significant. Encoder bit rate savings in UAV payloads are also critical since fewer bits to be transmitted translates to lower power consumption.

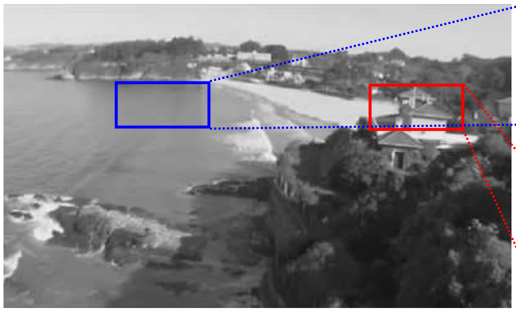
In Fig. 2, comparing the results of the GMh0 and GMs1 encoders, we see that using spectral entropy-based coefficient selection and quantization matrix (QM) design improves the constancy of quality across frames although it degrades the average rate-distortion performance. Also, the variation in quality for the GMs1 encoder is roughly constant across different bit rates as indicated by the flatness of the curves in Figs. 2 (c) and (d). This is because the spectral entropy QM design looks ahead at frames in the GOP and ensures constant quality.

Therefore the GMs1 encoder gives higher quality with lower quality variation than the baseline H.264 (BMh0) encoder at lower bit rates. Fig. 3 compares the visual quality of these two encoders for two magnified portions of frame 83 of the “aerial_beach1” sequence. From Figs. 3(b), (d), and (f), it can be seen that there are fewer blocking artifacts in the GMs1 encoder output. This is because of the absence of block ME and the better quantization. The improved quantization also helps retain more sharp edges as seen in Figs. 3(c), (e), and (g). In Fig. 4, the frame-wise PSNR is plotted for the BMh0 and GMs1 encoders operating at an average PSNR of 35.87dB. It can be observed that for the GMs1 encoder, the PSNR within a GOP remains relatively constant, while for the BMh0 encoder the PSNR could change by up to 2dB within 5 frames (frames 84-88).

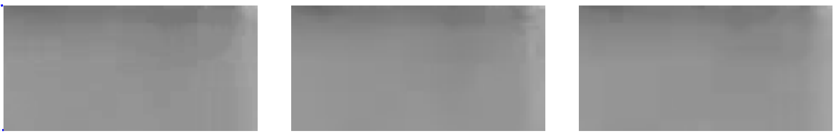
IV. CONCLUSION

UAV reconnaissance and video surveillance applications are different from traditional video applications in that they require low complexity encoders and most of the motion in the video is global. Motivated by this application, we have proposed a low complexity video encoder tailored for sequences with mostly global motion whose characteristics are

- no block motion estimation,
- frame-level global motion compensated prediction with the global motion parameters derived from the known camera platform movements, and
- spectral entropy based coefficient selection and quantizer design.



(a) Frame 83 of original sequence



(b) Crop 1 - BMh0 (44.04dB, 0.989)

(d) Crop 1 - Original

(f) Crop 1 - GMs1 (43.13dB, 0.990)

(c) Crop 2 - BMh0 (31.53dB, 0.909)

(e) Crop 2 - Original

(g) Crop 2 - GMs1 (31.70dB, 0.909)

Fig. 3. Comparison of frame 83 of 400 × 240 "aerial_beach1" sequence reconstructed using BMh0 and GMs1 encoders. (a) Original frame (b),(c) Crops of frame encoded using BMh0 (d),(e) Crops of original frame (f),(g) Crops of frame encoded using GMs1. PSNR and SSIM are indicated in parenthesis.

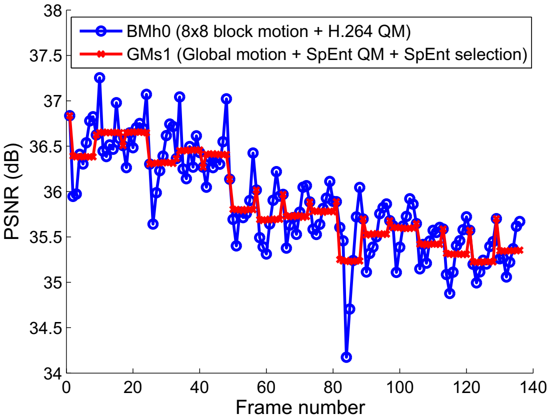


Fig. 4. Comparison of PSNR variation across frames at 35.87dB average PSNR for 400 × 240 "aerial_beach1" sequence.

We have compared its performance to a baseline H.264 encoder and separately analysed the performance gains due to the use of global motion compensation and spectral entropy based quantizer design. An encoder using global motion compensation instead of block motion estimation can achieve similar frame quality at significantly lower bit rates, reducing the transmission power required. Use of global motion also makes the drop in quality with bit rate less drastic, since bits are not allocated across motion vectors and residues. It also is a major contributor to the reduction of the encoder complexity. Incorporating spectral entropy based bit allocation in the encoder, entails latency but improves the constancy of quality across frames.

REFERENCES

[1] T. Klassen, "The UAV video problem: using streaming video with unmanned aerial vehicles," *Military and Aerospace Electronics*, vol. 20, no. 7, Jul. 2009.

[2] Y. Tan, W. Lee, J. Tham, and S. Rahardja, "Complexity-rate-distortion optimization for real-time H.264/AVC encoding," in *Proc. International Conf. on Computer Communications and Networks*, Aug. 2009, pp. 1–6.

[3] L. Su, Y. Lu, F. Wu, S. Li, and W. Gao, "Real-time video coding under power constraint based on H.264 codec," in *Proc. Visual Communications and Image Processing (VCIP)*, vol. 6508, Jan. 2007, pp. 1–12.

[4] A. Aaron, R. Zhang, and B. Girod, "Wyner-Ziv coding of motion video," in *Thirty-Sixth Asilomar Conference on Signals, Systems and Computers*, vol. 1, Nov. 2002, pp. 240–244.

[5] X. Artigas, J. Ascenso, M. Dalai, S. Klomp, D. Kubasov, and M. Ouaret, "The DISCOVER codec: Architecture, techniques and evaluation," in *Picture Coding Symposium*, 2007.

[6] C. Brites, J. Ascenso, and F. Pereira, "Feedback channel in pixel domain Wyner-Ziv video coding: Myths and realities," in *Proc. of 14th European Signal Processing Conference*, Sep. 2006.

[7] C. Brites and F. Pereira, "Encoder rate control for transform domain Wyner-Ziv video coding," in *IEEE International Conf. Image Process.*, vol. 2, Oct. 2007, pp. 5–8.

[8] C. Yaacoub, J. Farah, and B. Pesquet-Popescu, "Feedback channel suppression in distributed video coding with adaptive rate allocation and quantization for multiuser applications," *EURASIP Journal on Wireless Communications and Networking*, 2008.

[9] C. Morimoto, P. Burlina, and R. Chellappa, "Video coding using hybrid motion compensation," in *IEEE International Conf. Image Process.*, vol. 1, Oct. 1997, pp. 89–92.

[10] E. Steinbach, T. Wiegand, and B. Girod, "Using multiple global motion models for improved block-based video coding," in *IEEE International Conf. Image Process.*, vol. 2, 1999, pp. 56–60.

[11] A. Glantz, A. Krutz, and T. Sikora, "Adaptive global motion temporal prediction for video coding," in *Picture Coding Symposium*, Dec. 2010.

[12] X. Li, J. R. Jackson, A. K. Katsaggelos, and R. M. Merserau, "Multiple global affine motion model for H.264 video coding with low bit rate," *Proc. SPIE*, vol. 5685, no. 1, pp. 185–194, 2005.

[13] A. Smolic, Y. Vatis, H. Schwarz, and T. Wiegand, "Long-term global motion compensation for advanced video coding," in *10. ITG-Fachtagung Dortmunder Fernsehseminar*, Sep. 2003.

[14] J. Gong, C. Zheng, J. Tian, and D. Wu, "An image-sequence compressing algorithm based on homography transformation for unmanned aerial vehicle," in *International Symposium on Intelligence Information Processing and Trusted Computing (IPTC)*, Oct. 2010, pp. 37–40.

[15] A. F. Rodríguez, B. B. Ready, and C. N. Taylor, "Using telemetry data for video compression on unmanned air vehicles," AIAA Guidance, Navigation, and Control Conference and Exhibit, Aug. 2006.

[16] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[17] R. Schneiderman, "Unmanned drones are flying high in the military/aerospace sector [Special reports]," *IEEE Signal Process. Mag.*, vol. 29, no. 1, pp. 8 –11, 2012.

[18] W. Yang, J. D. Gibson, and T. He, "Coefficient rate and lossy source coding," *IEEE Trans. Inf. Theory*, vol. 51, no. 1, pp. 381–386, Jan. 2005.

[19] M. Bhaskaranand and J. D. Gibson, "Spectral entropy-based bit allocation," in *International Symposium on Inf. Theory and its Applications*, Oct. 2010, pp. 243–248.

[20] H. S. Malvar, A. Hallapuro, M. Karczewicz, and L. Kerofsky, "Low-complexity transform and quantization in H.264/AVC," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 598–603, Jul. 2003.

[21] W. Yang and J. D. Gibson, "Coefficient rate and significance maps in transform coding," in *Thirty-First Asilomar Conference on Signals, Systems and Computers*, vol. 2, Nov. 1997, pp. 1373–1377.

[22] ——, "Distortion analysis of the coding scheme based on coefficient rate," in *IEEE International Symposium on Inf. Theory*, 1998, p. 224.

[23] L. L. Campbell, "Minimum coefficient rate for stationary random processes," *Information and Control*, vol. 3, no. 4, pp. 360–371, 1960.

[24] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*. Kluwer Academic Publishers, 1991.

[25] M. Bhaskaranand and J. D. Gibson, "Low-complexity video encoding for UAV reconnaissance and surveillance," in *Military Communications Conference (MILCOM)*, 2011, pp. 1633 –1638.