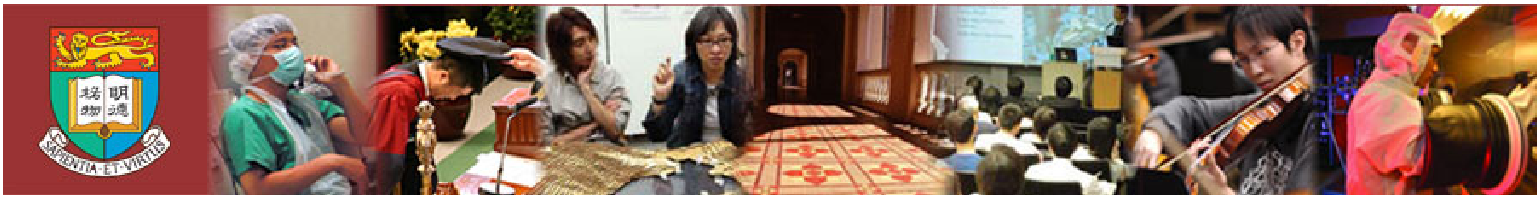


Title	On a region-of-interest based approach to robust wireless video transmission
Author(s)	Wong, ACW; Kwok, YK
Citation	Proceedings Of The International Symposium On Parallel Architectures, Algorithms And Networks, I-Span, 2004, p. 385-390
Issued Date	2004
URL	http://hdl.handle.net/10722/46463
Rights	

On a Region-of-Interest Based Approach to Robust Wireless Video Transmission

Alec Chi-Wah Wong and Yu-Kwong Kwok
Department of Electrical and Electronic Engineering
The University of Hong Kong, Pokfulam Road, Hong Kong
Corresponding Author: Yu-Kwong Kwok (email: *ykwok@hku.hk*)

Abstract— **This paper presents a scheme aiming at transmitting real-time video to wireless channel with vigorously varying quality, which is in practice the norm rather than the exception. Region of Interest (ROI) is an efficient approach to making the video more adaptive to the wireless channel because ROI is the region that human eyes tend to put more attention to than the Remainder Region (RM) . In our proposed scheme, we will adopt this feature. The real-time source video stream is divided into two regions, the ROI and the RM regions. The two regions will be encoded using H.263 standard codec such that the video transmission is adaptive to the current channel state, which is characterized by the effective data rate that varies from tens of kilobits per second to hundreds of kilobits per second. Channel state parameters are fed back to the source coder to adjust the compression ratio as well as the intra/inter options of the encoders. Results including frame loss probability, compression characteristics, Peak Signal the Noise Ratio (PSNR) against channel states are given, indicating that the resulting adaptive video codec can respond judiciously to time-varying channel quality. Our scheme is evaluated together with a ROI-enabled moving picture coding standard JPEG2000. Using the features provided in JPEG2000, we have made the JPEG2000 codec adaptive to the vigorously varying wireless channel and then compared it with the H.263 scheme. Our technique is suitable for a broad area of applications including real-time news reporting and video conferencing.**
Keywords: simulations, network measurements, region of interest (ROI), remainder (RM), channel adaptation, real-time video compression, H.263 (DCT), JPEG2000 (wavelet), PSNR.

I. Introduction

It is known that human perceives the outside world in a rather complicated manner [17]. The human visual system (HVS) is dramatically different from camera sensing. Specifically, our visual stimuli do not behave linearly to the physical properties of the stimuli such as intensity and color. Some information in the scene is inherently more important than other. For instance, when viewing a scene, human eye will focus on an area while letting the other areas blurred. Video conferencing is a case in point—people will focus more on the participants rather than the background. Thus, the area on focus, the Region of Interest (ROI), should be allocated more resources, so that the perceived quality would be better than the area that is not in focus. This approach could also utilize the bandwidth usage. We believe that this approach performs better than using fixed quality throughout the whole scene.

This research was supported by a grant from the Hong Kong Research Grants Council under project number HKU 7162/03E.

On the other hand, by gathering channel state information (CSI) for the current channel characteristics [10], [15], [16], a more judicious resources allocation can be achieved since we have a better understanding on the available bandwidth in the channel. CSI can be collected in the physical layer of the mobile device or through indirect inferences (e.g., monitoring the ARQ activities) at the link layer. Recently, there have been quite a few research attempts on the idea of channel adaptive wireless transmission. Bhagwat et al. [1] developed a downlink channel adaptive approach for enhancing wireless TCP performance. There has also been other work done on adaptive scheduling on the downlink and uplink directions. For instance, Kwok et al. [9] proposed a novel channel adaptive uplink access control protocol for nomadic computing. However, these previous contributions are not targeted at the application level, as in the wireless video considered in this paper. Indeed, link adaptation, albeit not a new idea, has not been applied together with the above mentioned ROI concept in adaptive wireless video.

Through joint source-channel coding, we could have a better allocation of resources to source and channel coding. There are many novel joint-source channel video coding approaches suggested in the literature [2], [3], [5], [6], [7], [8]. In [4], the rate-shaping joint source-channel coding that allows a portion of the video stream to be dropped before sending so that more resources can be reserved for channel coding. There are other approaches for joint source-channel coding like distortion-based. Kondi el al. [5], [6], [7], [8] have developed numerous effective models on joint source-channel coding and have conducted studies of those models using multi-path fading channels. However, none of these previous work are based on the ROI concept.

In this paper, a ROI-based channel adaptive source coding scheme for real-time video is introduced. Our proposed scheme is effective in that given a particular amount of source coding resources under the vigorously changing channel states, the compression quality is dynamically adjusted such that the source coding can still provide a high quality presentation for the ROI. We also use the ROI and layer-progressive organization features in JPEG2000 to create an adaptive rate video transmission scheme for comparison. In Section II, we provide some information concerning CSI. In Section III, DCT-based channel adaptive video transmission will be introduced with our proposed scheme. In Section IV, Wavelet-based channel adaptive video transmission will be introduced. In Section V, extensive simulation results and interpretations will be offered. The last section provides some concluding remarks.

II. Channel State Information (CSI)

In a realistic environment, the quality of the wireless link is characterized by two components, namely the fast fading component and the long-term shadowing component. Fast fading is caused by the superposition of multipath components and is therefore fluctuating in a very fast manner (on the order of a few msec). Long-term shadowing is caused by terrain configuration or obstacles and is fluctuating only in a relatively much slower manner (on the order of one to two seconds). The combined effect is that the channel quality varies continuously. Specifically, the channel quality stays constant for a short amount of time (called the coherence time) and then changes to a different quality level (e.g., the goodput can degrade from 1 Mbps to 750 Kbps within 20 msec for a wireless LAN connection). Using channel adaptive physical layer (i.e., multiple-mode channel coder and modulator), the channel state information (CSI) can be obtained and used by the upper layers (e.g., the application).

III. DCT-based Channel Adaptive Video Transmission

Figure 1 shows the block diagram of our proposed channel adaptive transmission scheme, which consists of three major components—splitting and encoding, channel adaptive transmitter, and feedback control.

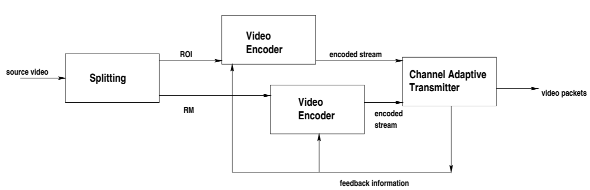


Fig. 1. The proposed channel adaptive video transmission scheme.

A. Splitting and Encoding

In this component, the source video data is split into two regions, the Region of Interest (ROI), and the Remainder Region (RM). The two regions will be separately encoded using two independent video encoders. The compression ratio of an encoder is defined as the ratio between the encoded data size and the corresponding encoded data size at maximum quality (i.e., ratio one). For instance, the data size of an image sequence at ratio n should be n times smaller then that of the same video sequence compressed at ratio one. Moreover, when feedback information is unavailable, the encoder will generate one intra-frame followed by four inter-frames. This is defined as one cycle.

B. Adaptive Transmitter

In general, for image sequences composing a video, there exists temporal redundancy, and hence an intra-frame is usually larger than an inter-frame. Thus the following inequality holds:

$$\text{Intra ROI} + \text{Intra RM} > \text{Intra ROI} + \text{Inter RM} \qquad (1)$$

If intra ROI and intra RM are sent under the same frame period, the available bandwidth (obtained from CSI) may not be enough and that will lead to RM frame loss. To avoid this adverse effect, the intra/inter frames selections are independent for ROI and RM. For instance, under the same video timestamp, an intra-frame in ROI does not imply an intra-frame for RM and vice versa. Initially, the compression ratios of ROI and RM are set to maximum with poorest quality (in practice any starting ratio can be set). As human viewers typically focus more on ROI, ROI should have a higher priority than RM. Thus ROI is sent first under the available bandwidth for each frame. If ROI is sent successfully, the remaining available bandwidth is allocated to RM. This leads to three cases:

Case 1: Within each frame period, both ROI and RM can be sent. If the ROI frame is an intra-frame (intra frame is usually larger than inter frames), a GIFT (elaborated below) is given to try to make more efficient use of available bandwidth.

Case 2: Within each frame period, both ROI and RM cannot be sent due to insufficient available bandwidth. A PENALTY02 (elaborated below) is given in the hope that the next frame can be sent successfully.

Case 3: This is the last case. Only ROI can be sent. That means the RM frame is lost. In this case, a PENALTY03 (elaborated below) is given in the hope that the RM frame can be recovered in coming frames.

C. Feedback Control

There are two independent feedback paths from the channel adaptive transmitter to the video encoders. There are three types of feedback messages: GIFT, PENALTY02, and PENALTY03.

C.1 GIFT

When an intra-ROI frame and an intra/inter RM frame can be sent successfully within the frame period, a GIFT message will be fed back to the two encoders. The GIFT is a decrement in compression ratio (better quality) in ROI and RM such that a better use of channel bandwidth is achieved. Firstly, residue is defined as follows:

$$\text{residue} = \text{available bandwidth} - \text{ROI frame size} - \text{RM frame size} \qquad (2)$$

and the new ROI compression ratio is as follows:

$$\text{new ROI compression ratio} = \frac{(\text{ROI size} \times \text{ROI ratio})}{(\text{ROI size} + \text{residue})} + 1 \qquad (3)$$

The above equation is formed based on the inversely proportional characteristics between size of ROI frame and the compression ratio. The "+1" acts as a relaxation such that the new compression ratio will not exhaust the channel that may lead to subsequent frame loss. Since there is not much significance in RM, the scheme slightly decreases the compression ratio by 1. When a GIFT message is received at the encoders, the compression ratios will be updated in the next cycle.

C.2 PENALTY02

When both ROI and RM are failed to send within the frame period, a PENALTY02 message will be fed back to the two encoders. The PENALTY02 is an increment in compression ratio (worse quality) in both ROI and RM such that in the coming frame period the video transmission can be recovered. The encoders will also be reset immediately when a PENALTY02 message is received, i.e., a new cycle is forced to begin. That means the next frame to be encoded must be an intra-frame for both ROI and RM. This could avoid error propagation to the coming frames. The new compression ratio will be adjusted as follows:

$$\text{new ROI compression ratio} = \frac{(\text{ROI size} \times \text{ROI ratio})}{\text{available bandwidth}} + 1 \quad (4)$$

The increment of ROI compression ratio will be two or by the above formula, whatever larger. A fixed compression ratio increment of one unit will be enforced for RM to further lower the required bandwidth. This is done in the hope that the next frame can be transmitted successfully.

C.3 PENALTY03

When only ROI can be sent with RM lost within a frame period, a PENALTY03 message will be fed back to RM encoder. The RM encoder will be reset immediately when a PENALTY03 message is received, i.e., a new cycle is forced to begin. That means the next incoming RM frame to be encoded must be an intra-frame. As a result, there will be four possible combinations of inter/intra ROI/RM frames. For instance, when an intra-ROI can be sent with intra-RM lost, the next encoded frame will become an inter-ROI with an intra-RM. As the inter frames are usually small, the intra-RM frame could be sent hopefully following the inter-ROI frame in the same frame period. There is no adjustment for compression ratio for this case—since most likely frame loss is caused by a transmission of intra-ROI plus intra-RM. Resetting RM encoder will lead to sending intra-RM again in next frame, which will accompany with inter-ROI. By Equation (1), the resulted total size will be smaller and hence the probability of success will be much higher in general.

IV. Wavelet-based Channel Adaptive Video Transmission

A. *Evaluations on Features of JPEG2000*

In our evaluation, kakadu implementation of JPEG2000 codec will be used [13]. Each picture will be encoded as a JPEG2000 picture with one tile.

A.1 Layer-Progressive Organization (LPO)

This feature will divide the JPEG2000 picture into layers. By adding one more layer, the picture quality will be improved. In channel adaptive video transmission, it is desirable to be able to transmit frames at different sizes and JPEG2000 has this feature—you can truncate the file size into any length (of course it needs to have a minimum length for the decoder to decode the minimum information). Without layer-progressive organization, the truncated picture will degrade very rapidly as more portion is truncated.

TABLE I
Characteristics of a sample picture.

Sample Picture A	PSNR ROI/RM (dB)
Encoded at 1 bit/pixel	24.699/17.673
1 bit/pixel → 0.25 bit/pixel (no LPO)	19.111/17.616
1 bit/pixel → 0.25 bit/pixel (LPO)	20.49/17.627
Encoded at 0.25 bit/pixel	20.538/17.624

As can be seen from Table I and Figure 2, the PSNR of those truncated graphs with LOP enabled shows a better performance than that of those without LOP at the same bit/pixel. The performance of truncated graph with LOP enabled is very near to those pictures encoded at the truncated rate.

A.2 Max-Shift Method vs. Cost Function for ROI Processing

Details of Max-shift Method are described in [13], [14]. One drawback of this method is that the shifting coefficient cannot be too small in order for the decoder to recognize [11], [12]. For instance, for a reversible compression of a 8-bit color images requires a shift value of at least 12, i.e., the scale factor of ROI becomes $2^12 = 4096$. That means the quality of ROI will be improved all the way to loseless when adding quality layers while the RM will remain poor. Thus, there has been another method in which a selectable scaling factor or Cost Function is used to allocate code-block contributions to quality layers in accordance with ROI.

TABLE II
Characteristics of the max-shift and cost function.

Sample Picture A	PSNR ROI/RM (dB)
1 bit/pixel using Max-Shift	24.699/17.673
1 bit/pixel using Cost Function	23.949/18.806

As can be seen from Table II and Figure 3, the Max-Shift Method provides a better ROI PSNR but a poorer RM PSNR. It seems the Cost Function method makes the transition too long from clear to blur, even though the ROI is sent at right half only. In our case, Max-Shift method is chosen since ROI has much more significance and hence the maximum possible resources should be allocated to ROI before RM.

(a) sample picture A—without LPO

(b) sample picture A—with LPO

Fig. 2. Two sample pictures.

(a) sample picture A—max-shift

(b) sample picture A—cost function

Fig. 3. Two sample pictures using different methods.

B. Scheme Used for Creating Motion JPEG2000

In our case, a four quality layered JPEG2000 encoder with Max-Shift ROI processing is used. The same set of encoding parameters is used throughout the whole real-time video transmission. Under a frame-period, packets will be sent from the beginning of the encoded picture whenever a CSI is available. This will lead to a case that the transmitted size of different frames will be different.

V. Simulation Results and Interpretation

In simulation, a 131 frame real-time video will be captured and processed. The video has 5 frames per second or with a frame period of 200 ms. Due to space limitations, we can only present a small set of results here. For the complete set of results, the reader is referred to [18].

A. Frame Loss Probability

Figure 4 shows the ROI frame loss rate in a frame sequence of 131 frames. Loss probability for ROI is lower than that of RM due to higher priority of ROI frames. In general, the loss rate increases as the time-coherence increases. For instance, if the time coherence is 1000ms, 5 frames will be sent within that coherent time. It is because in high time coherence, multiple frames are sent within the length of time coherence. If the channel state in that particular time is poor, multiple frames will be lost in one burst. In 50 ms time coherence, the loss probability is high due to a long period of low bandwidth. In general, small time coherence will likely mitigate the effect of burst loss because during a frame period (1/frame frequency), channel state has changed several times. For instance, if the time coherence is 10 ms, the channel state will have changed 20 times after a frame is sent. That means low and high bandwidth are likely to happen within that frame period and hence the overall bandwidth for that frame is averaged.

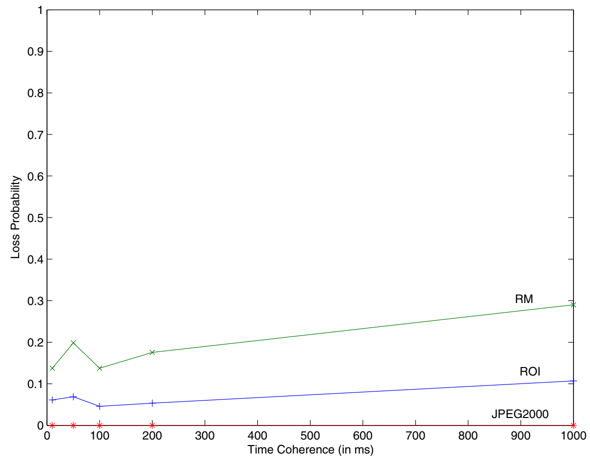


Fig. 4. ROI/RM/JPEG2000 frame loss probability against time coherence.

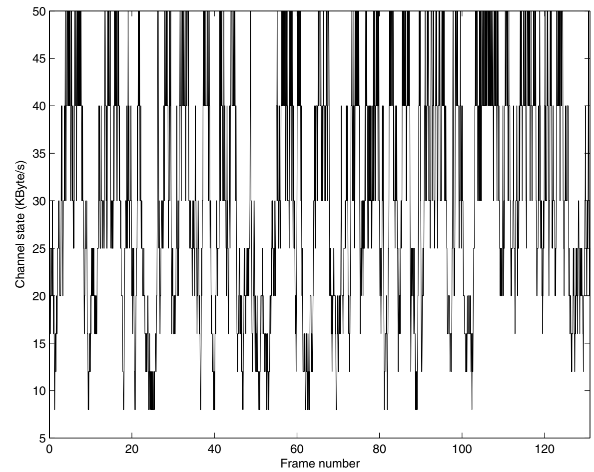
B. *Compression Ratio*

Figure 5 shows the compression ratio results of our proposed scheme at a coherence time of 10 ms (for results on other coherence times, please refer to [18]).

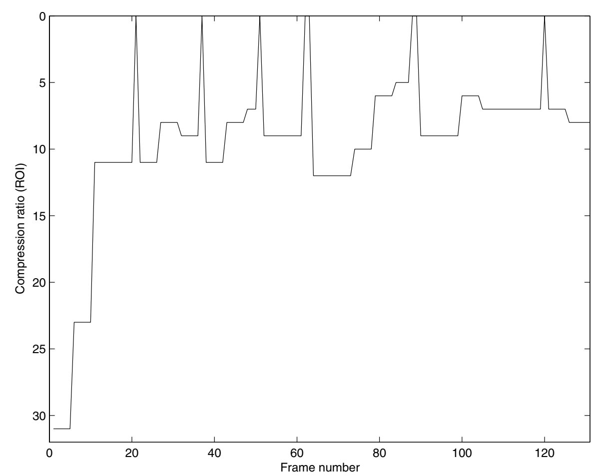
Interpretation on some instances of the compression ratio graphs with time coherence 10 ms is given as follows:

- Initially, the compression ratio of both ROI and RM are set to 31 (although in practice it is recommended to set the ratio to 1 which will be explained later). As the bandwidth is enough, a GIFT is fed back to the encoder and the compression ratio is decreased. So in next cycle, the ROI compression ratio becomes 23 and the RM compression becomes 30 (in frame 6).
- The good condition continues and the compression ratio is further reduced.
- In frame 11, when the compression ratio of ROI is 11 while that of RM is 29, the RM frame drops. PENALTY03 is fed back and the compression ratio of RM is increased to 30 in frame 12.
- In frame 21, both ROI and RM are lost due to insufficient bandwidth. PENALTY02 is fed back and the ROI and RM encoders are reset with new increased compression ratio (in this case the result of ROI is the same as before while the ratio of RM is increased by one).
- In frame 31, although there is a GIFT sent, the result of the suggested ratio is higher than before, it shows that the bandwidth is tight and so the relaxation relieves it.

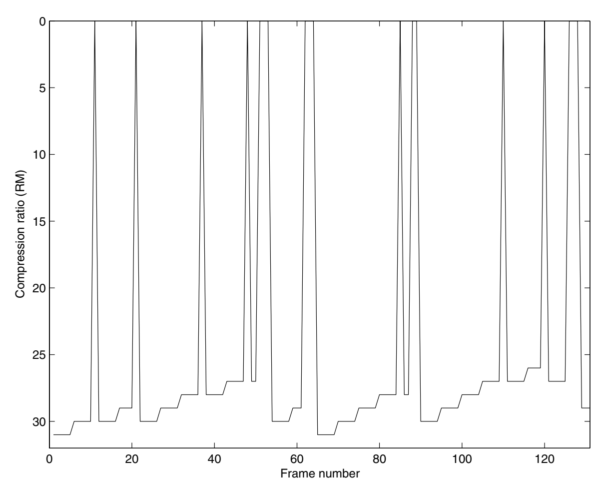
At the beginning, the ratio is set to 31 for both ROI and RM, this will maximize the probability for a successful transmission. You may use a more optimistic way by setting the initial ROI to 1 and RM to 31. This will likely introduce frame loss, but in the second frame it would likely to recover by the PENALTY02 feedback message. The other compression ratio graphs show similar characteristics as in above description.



(a) instantaneous channel data rate



(b) ROI compression ratio



(c) RM compression ratio

Fig. 5. Comparison of compression ratios (coherence time = 10 ms).

C. Compressed Video Size

We have also obtained results on compressed video sizes generated by different approaches [18]. These results indicate that the overall compressed size of H.263 is smaller than that of JPEG2000. It is due to two reasons:

- Motion JPEG2000 does not explore temporal redundancies and hence all are in fact intra-frames. On the other hand, H.263 has inter-frame coding and hence under a particular compression ratio, the intra frame may be as high as that of JPEG2000 while the inter-frame is much smaller.
- The transmitted data size of JPEG2000 is the total amount of bandwidth available and the final frame size sent could be of any value. On the other hand, within a frame cycle (unless a PENALTY02 is called in which everything will be reset), all frames are encoded in the same compression ratio and the intra-frame size is bounded by discrete compression ratio values. The relaxation will further lower the frame size although it is necessary since if the frame size is too tight, frame loss is more likely to occur as the channel state changes.

D. PSNR

The performance of different approaches in terms of PSNR are also obtained [18]. We find that the PSNR of JPEG2000 is higher than that of H.263. This is due to the smaller encoded size of H.263 plus the fact that JPEG2000 uses wavelet transform which performs better than DCT-based H.263.

VI. Conclusions

In this paper, we proposed a scheme for region-based channel adaptive source coding scheme. Channel state information (CSI) is obtained through the physical layer of the mobile device and then proper resources are allocated between source and channel coder. Our scheme focuses on how the source coding reacts with known channel information. The channel state is accurately simulated by using realistic physical models with channel bandwidth varying at different speeds, modeling different user mobility. Based on the available channel bandwidth for source coding, Region-of-Interest (ROI) and Remainder Region (RM) are sent or dropped accordingly with compression quality varied in each of the regions. Feedback messages are sent from the channel adaptive transmitter back to the encoders of the two regions to adjust the encoding parameters so that the resources can be used more efficiently. The scheme is compared with current standard JPEG2000 with ROI processing capability. The JPEG2000 codec is made channel adaptive and then compared with H.263 channel adaptive scheme. Results show that the flexibility and quality (PSNR) of JPEG2000 is higher than that of H.263 with the sacrifice of ignoring temporal redundancy which can greatly reduce the video stream size. The loss rate of JPEG2000 is zero while the loss rate of H.263 is not high at all.

References

[1] P. Bhagwat, P. Bhattacharya, A. Krishna, and S. K. Tripathi, "Using Channel State Dependent Packet Scheduling to Improve TCP Throughput over Wireless LANs," *ACM/Baltzer Wireless Networks*, vol. 3, no. 1, pp. 91–102, 1997.

[2] J. Cai and C. W. Chen, "A High-Performance and Low-Complexity Video Transcoding Scheme for Video Streaming over Wireless Links," *Proc. IEEE WCNC 2002*, Mar. 2002.

[3] N. Chaddha and S. Diggavi, "A Framework for Joint Source-Channel Coding of Images over Time-Varying Wireless Channels," *Proc. International Conference on Image Processing*, pp. 89–92, Sept. 1996.

[4] T. P.-C. Chen and T. Chen, "Adaptive Joint Source-Channel Coding Using Rate-Shaping," *Proc. Of ICASSP*, May 2002.

[5] L. P. Kondi, F. Ishtiaq, and A. K. Katsaggelous, "Joint Source-Channel Coding for Scalable Video," *Proc. SPIE Conference on Image and Video Communications and Processing*, pp. 324–335, 2000.

[6] L. P. Kondi, S. N. Batalama, D. A. Pados, and A. K. Katsaggelos, "Joint Source/Channel Coding for Scalable Video Over DS-CDMA Multipath Fading Channels," *Proc. IEEE International Conference on Image Processing*, vol. I, pp. 994–997, Oct. 2001.

[7] L. P. Kondi and A. K. Katsaggelos, "Joint Source/Channel Coding for Scalable Video Using Models of Rate-Distortion Functions," *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. III, pp. 1377–1380, May 2001.

[8] L. P. Kondi, F. Ishtiaq, and A. K. Katsaggelos, "Joint Source-Channel Coding for Motion Compensated DCT Based SNR Scalable Video," *IEEE Trans. Image Processing*, vol. 11, no. 9, pp. 1043–1052, Sept. 2002.

[9] Y.-K. Kwok and V. K. N. Lau, "A Novel Channel-Adaptive Uplink Access Control Protocol for Nomadic Computing," *IEEE Trans. Parallel and Distributed Systems*, vol. 13, no. 11, Nov. 2002.

[10] V. K. N. Lau, "Performance Analysis of Variable Rate Symbol-by-Symbol Adaptive Bit Interleaved Coded Modulation for Rayleigh Fading Channels," *IEEE Trans. Vehicular Technology*, vol. 51, no. 3, pp. 537–550, May 2002.

[11] D. S. Taubman, UNSW, *kakadu Survey Documentation*, June 2003.

[12] Kakadu Implementation on JPEG2000 license.

[13] JPEG2000 Part I Final Committee Draft Version 1.0, March 2000.

[14] D. S. Taubman and M. W. Marcellin, *JPEG2000: Image Compression Fundamentals, Standards and Practice*, 2002.

[15] J. D. Parsons, *The Mobile Radio Propagation Channel*, 2nd Edition, Wiley, 2000.

[16] H. V. Poor and G. W. Wornwell, *Wireless Communications: Signal Processing Perspectives*, Prentice Hall, Signal Processing Series, 1998.

[17] Y. Q. Shi and H. Sun, *Image and Video Compression for Multimedia Engineering: Fundamentals, Algorithms, and Standards*, CRC Press, 2000.

[18] A. C.-W. Wong, *Real-Time Video Codec with Wireless Link Adaptation and Region-of-Interest Considerations*, M.Phil. thesis, Department of Electrical and Electronic Engineering, The University of Hong Kong, Aug. 2004.

Proceedings of the 7th International Symposium on Parallel Architectures, Algorithms and Networks (ISPAN'04)
1087-4089/04 $ 20.00 © 2004 IEEE