PERCEPTUALLY LOSSLESS COMPRESSION WITH ERROR CONCEALMENT FOR PERISCOPE AND SONAR VIDEOS

Chiman Kwan[1], Jude Larkin[1], Bence Budavari[1], Eric Shang[1], and Trac D. Tran[2]

[1]Applied Research LLC, Rockville, Maryland, USA
[2]The Johns Hopkins University, Baltimore, Maryland, USA

ABSTRACT

We present a video compression framework that has two key features. First, we aim at achieving perceptually lossless compression for low frame rate videos (6 fps). Four well-known video codecs in the literature have been evaluated and the performance was assessed using four well-known performance metrics. Second, we investigated the impact of error concealment algorithms for handling corrupted pixels due to transmission errors in communication channels. Extensive experiments using actual videos have been performed to demonstrate the proposed framework.

KEYWORDS

Perceptually lossless compression; error recovery; maritime and sonar videos

1. INTRODUCTION

Video compression has been widely used in many commercial and military applications [1][2]. Well-known video codecs include X264 [3] and X265 [4].In some applications such as security monitoring videos where video quality is of prime importance, people are still using lossless image compression algorithms such as JPEG [5] and JPEG-2000 [6] for taking snapshots for storage. Some recent algorithms such as X264 [3] and X265 [4] also provide lossless compression options. This may be too conservative, as lossless compression can only achieve two to four times of compression.

JPEG, X264, and X265 are discrete cosine transform (DCT) based algorithms and JPEG-2000 is wavelet based [7]. About 15 years ago, there were some development in DCT based algorithms where overlapped blocks known as lapped transforms (LT) were used to further improve the compression [8]. In the past few years, researchers at Xiph have incorporated LT [8] into an open source video codec known as Daala [9].

Transmission of compressed videos over a wireless channel is susceptible to natural or manmade interferences. A lost bit may damage a whole macro block. Although there are mechanisms such as flexible macro block ordering (FMO) in video codecs that can deal with lost packets to certain extent, it is still not enough, as errors tend to propagate to future frames [10]. Another common way is to insert extra bits to protect the bit stream. However, this will limit the amount of information that can be transmitted over bandwidth constrained channels.

DOI: 10.5121/sipij.2019.10201

In our earlier study [11], we proposed an error concealment scheme for still image compression. It was observed that lost information can be recovered via error concealment algorithms. No additional bandwidth will be used to protect the bit streams. The error concealment is done at the receiving end, which usually has more computational power.

Originally, our sponsor required us to achieve perceptually lossless compression with 8:1 compression ratio for videos at frame rate of 6 frames per second. In our companion paper [11], we have demonstrated that 10 to 1 perceptually lossless compression can be achieved for still images. Hence, we aim at 20 to 1 compression for videos in this research, as videos have more redundancy than still images. We propose to apply four state-of-the-art video codecs (J2K, X264, X265 and Daala) to compress videos. Four performance metrics were used in our study. Moreover, since some applications use wireless channels to transmit compressed videos back to a control center, there are corrupted pixels due to channel errors. We also investigate the possibility of using error concealment techniques to repair the corrupted pixels.

A shortened version of this paper was presented in a conference [12]. We have significantly expand our paper. First, we included more compression studies. One periscope and one sonar videos were used in our studies, as they are both of interest to our sponsor. In each study, we have included new tables summarizing the performance metrics of the compression algorithms at 20 to 1 compression or equivalently at a compression ratio of 0.05. Second, we have included more error concealment studies. Two algorithms were compared and objective and visual comparisons were included.

Although the compression codecs are not new, the key contribution of our project is to integrate three components (compression, error concealment, and perceptually lossless evaluation) in the video compression system into a single framework. Our paper is organized as follows. Section 2 summarizes the technical approach and its components. Section 3 summarizes all the experiments using actual images that are of interest to our customer. Finally, concluding remarks will be given in Section 4.

2. TECHNICAL APPROACH

2.1. Proposed Video Compression and Error Recovery Framework

In this research, we will perform objective evaluations of different compression algorithms in the literature. This is to ensure that we will deliver the best approach to our customer. Our overall technical approach can be summarized as follows. First, we will briefly review the literature to determine the compression algorithms available in the market. At the same time, we will describe four performance metrics for algorithm evaluation. The focus will be on metrics that can better model human perception. We will also mention the error resilient algorithms and error concealment techniques. Second, we will obtain realistic periscope and sonar images for algorithm evaluation. Third, we will apply the various compression algorithms to the collected images and generate various performance metrics. Finally, we will also apply advanced algorithms to deal with the corrupted pixels due to channel errors.

Our proposed approach is shown in the following diagram.

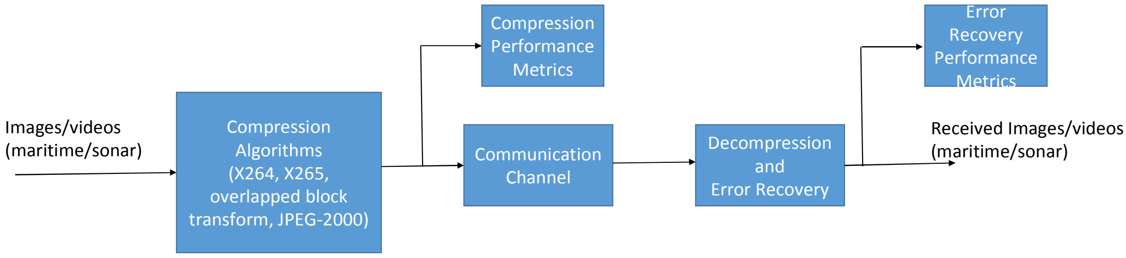


Figure 1. Proposed approach to evaluating different compression and error concealment algorithms for videos.

2.2. Brief Review of Video Compression Algorithms

Here, we will briefly review video codecs in the market.

Wavelet based algorithm

J2K for Video
J2K is a wavelet based compression algorithm for still images. Motion J2K was developed in 2001. It does not have inter-frame coding and hence the coding efficiency is not as good as other videos codecs such as X-264, X265, etc.

DCT based algorithms

- JPEG [5] :
JPEG is the very first image compression standard. The video counterparts are the MPEG-1 and MPEG-2 standards.

- VP8 and VP9 [13] [14]:
These video compression algorithms are owned by Google. The performance is somewhat close to X-264. However, it is not as popular as X264.

- X-264 [3] :
X264 is the current state-of-the-art in video compression. Youtube uses X264. It has good still image compression.

- X-265 [4] :
This is the next-generation video codec. However, the computational complexity is much more than that of X264. In general, X265 has the same basic structure as previous standards. There are some incremental improvements in X265 as compared to X264, including:
    - Flexible partitioning
    - Flexibility in prediction modes and transform block sizes
    - Sophisticated interpolation and deblocking filters
    - Sophisticated prediction and signalling of modes and motion vectors

X264 and X265 are optimized versions of H264 and H265, respectively.

- Daala [9]
Recently, there is a parallel activity at xiph.org foundation, which implements a compression codec called Daala [9]. It is based on DCT. There are pre- and post-filters to increase energy compaction and remove block artifacts. Daala borrows ideas from [8], which was written by one of us (T. Tran).

The block-coding framework in Daala can be illustrated in Figure 2.

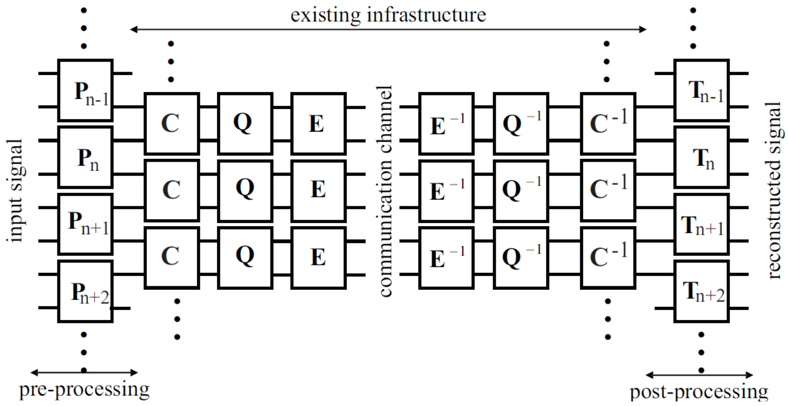


Figure 2. Daala codec for block-based image coding systems.

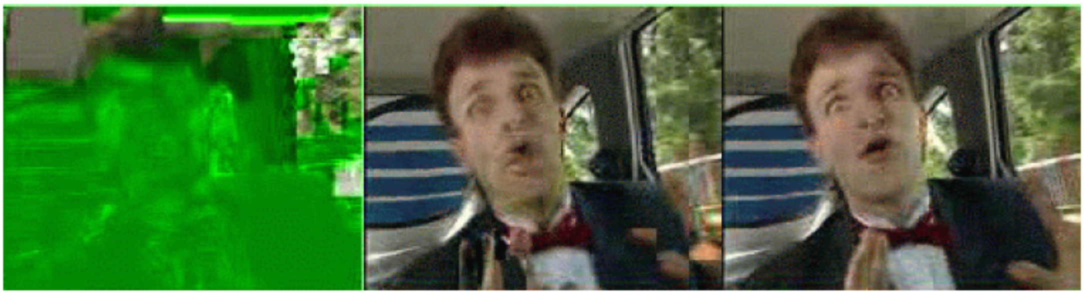
In this study, we have mainly compressed videos using Daala, X264, X265, and JPEG-2000.

2.3. Error Resilient/Error Concealment

In [15], several coding schemes were evaluated. From
Figure, it can be seen that there is a limit to which the channel errors can be corrected.

It should be noted that there are some built-in error concealment mechanisms in codecs such as H264. Flexible Macro Ordering (FMO) is one of them. In one of our papers, we have evaluated some of these mechanisms and found that they are only effect to certain extent [10].

In practical applications, we believe a combination of the built-in error concealment and post-decompression error concealment is needed.



Unprotected (SNR 7 dB AWGN) RS+convolutional coded Turbo coded

One way decoding Two way decoding

Figure 3. Comparison of several error resilient coding schemes [15]. There are still errors that cannot be fixed.

2.4. Principle of Video Error Concealment [16]

Similar to still images, error resilient coding incurs more overhead in bandwidth and may not be able to deal with severe channel interferences. It is therefore necessary to apply error concealment techniques to recover lost pixels in the video.

Our general approach for these problems is based on the observation that for every small block in a frame, there always exist some motion estimations (ME) from the previous or future frames. Each patch with missing or corrupted values is grouped with similar patches from the partial information of the patch and stacked into columns of a matrix. All the matched motion estimations should have similar underlying image structures and the completed version of these patches should lie in a low dimensional subspace. Therefore, the constructed matrix becomes a very low-rank structure [17]-[26] and can be applied in low-rank techniques for reconstruction.

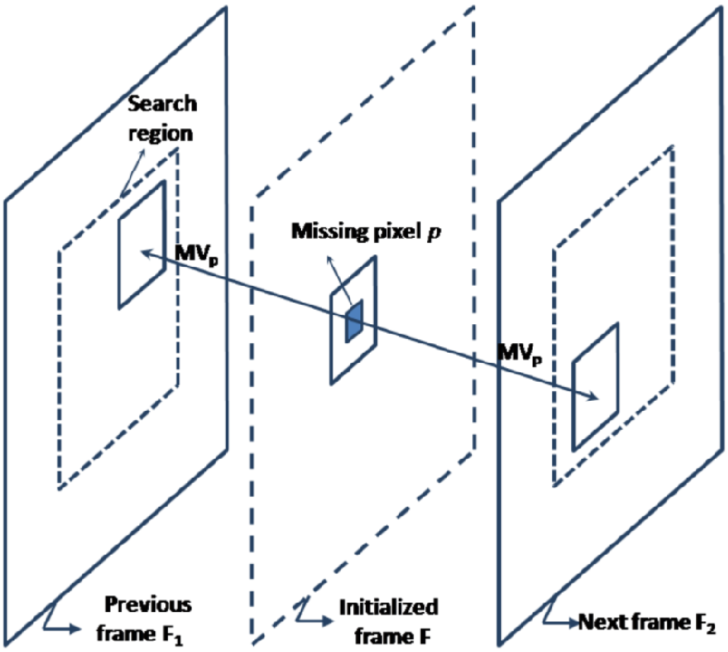
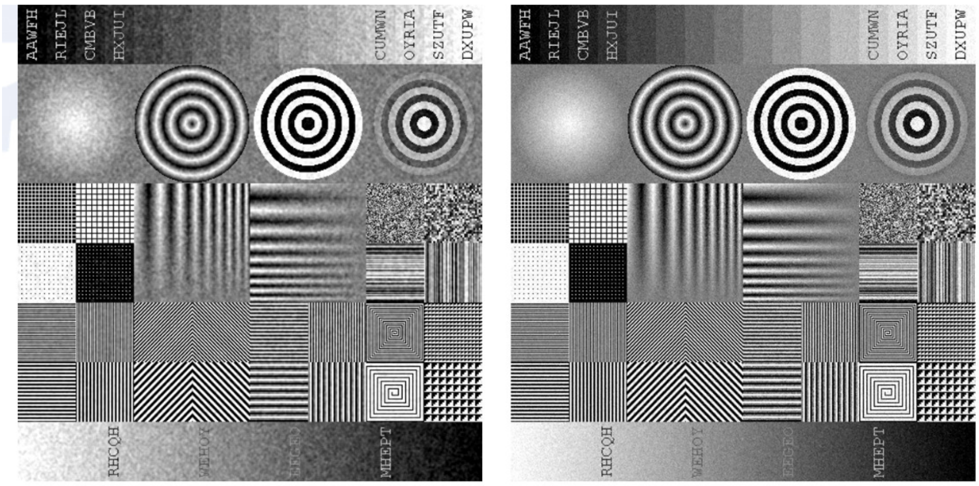


Figure 4. Bi-directional motion compensation approach to reconstruct corrupted pixels [16].

2.5. Performance Metrics

In many papers, researchers used peak signal-to-noise ratio (PSNR) or structural similarity (SSIM) to algorithm evaluation. Given a fixed compression ratio, algorithms that yield higher PSNR or SSIM will be regarded as better algorithms. However, PSNR or SSIM do not correlate well with human perception. Recently, a group of researchers investigated a number of different performance metrics [28]. Extensive experiments were performed to investigate the correlation between human perception with various performance metrics. According to the results found in [28], it was determined that two performance metrics known as human visual system (HVS) and human visual system with masking (HVSm) correlate well with human perception. A summary of the findings in [28] can be seen in Figure 5. It can be seen that PSNR-HVS (HVS in short) and PSNR-HVS-M (HVSm in short) have high correlation with human subjective evaluation results. For completeness, we include Table 1. It can be seen that HVSm and HVS have much higher correlation with human perception than PSNR and SSIM in terms of Spearman and Kendall correlation coefficients.

Hence, in addition to PSNR and SSIM, we also used HVS and HVSm for assessing perceptually lossless compression.



SSIM = 0.80, PSNR-HVS-M = 25.50 dB
SSIM says: "This is better!"

SSIM = 0.79, PSNR-HVS-M = 31.29 dB
PSNR-HVS-M says: "This is better!"

Figure 5. Comparison of SSIM and HVSM. HVSM has better correlation with human perception [28].

Table 1. Correlation of different metrics to human's visual perception [28].

Measure	Reference	Spearman correlation	Kendall correlation
PSNR-HVS-M	N. Ponomarenko, F. Silvestri, K.Egiazarian, M. Carli, V. Lukin, On Between-Coefficient Contrast Masking of DCT Basis Functions, *CD-ROM proceedings of Third International Workshop on Video Processing and Quality Metrics for Consumer Electronics VPQM-07*, January, 2007, 4p	0.984	0.948
PSNR-HVS	Egiazarian K., Astola J., Ponomarenko N., Lukin V., Battisti F., Carli M. New full-reference quality metrics based on HVS, *CD-ROM Proceedings of the Second International Workshop on Video Processing and Quality Metrics*, Scottsdale, USA, 2006, 4 p	0.895	0.712
NQM	Damera-Venkata N., Kite T., Geisler W., Evans B. and Bovik A., Image Quality Assessment Based on a Degradation Model, *IEEE Transactions on Image Processing*, Vol. 9, 2000, pp. 636-650	0.857	0.673
DCTune	Solomon J. A., Watson A. B., and Ahumada A., Visibility of DCT basis functions: Effects of contrast masking, *Proceedings of Data Compression Conference*, 1994, pp. 361-370 http://vision.arc.nasa.gov/dctune/ - DCTune 2.0 page	0.829	0.712
UQI	Wang Z., Bovik A., A universal image quality index, *IEEE Signal Processing Letters*, vol. 9, March, 2002, pp. 81-84	0.550	0.438
PSNR	Peak Signal to Noise Ratio	0.537	0.359
VQM	Xiao F., DCT-based Video Quality Evaluation, *Final Project for EE392J*, 2000	0.441	0.281
SSIM	Wang Z., Bovik A., Sheikh H., Simoncelli E., Image quality assessment: from error visibility to structural similarity, *IEEE Transactions on Image Processing*, vol.13, 2004, pp.600-612	0.406	0.358
VIF	Sheikh H. R. and Bovik A. C., Image Information and Visual Quality, *IEEE Transactions on Image Processing*, vol. 15, February, 2006, pp. 430-444	0.377	0.255
PQS	Miyahara, M., Kotani, K., Algazi, V.R., Objective picture quality scale (PQS) for image coding, *IEEE Transactions on Communications*, vol. 46, issue 9, 1998, pp. 1215-1226	0.302	0.242

3. Experimental Results

3.1. Data

Our sponsor is interested in low frame rate periscope and sonar videos. We searched the Internet and found some public videos: one periscope video and one sonar video.

3.1.1 Periscope Video

The video frame size is 720 x 1280 with 6 fps. There are 42 frames. Two frames are shown below.

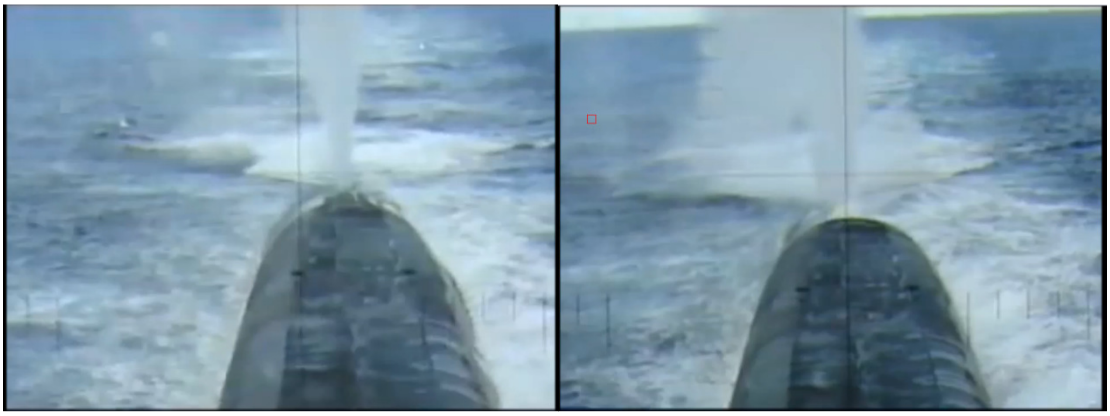


Figure 6. Two frames of a periscope video.

3.1.2 Sonar Video

The sonar video has a size of 720 x 1280 at 6 fps. There are 87 frames. Two frames are shown below.

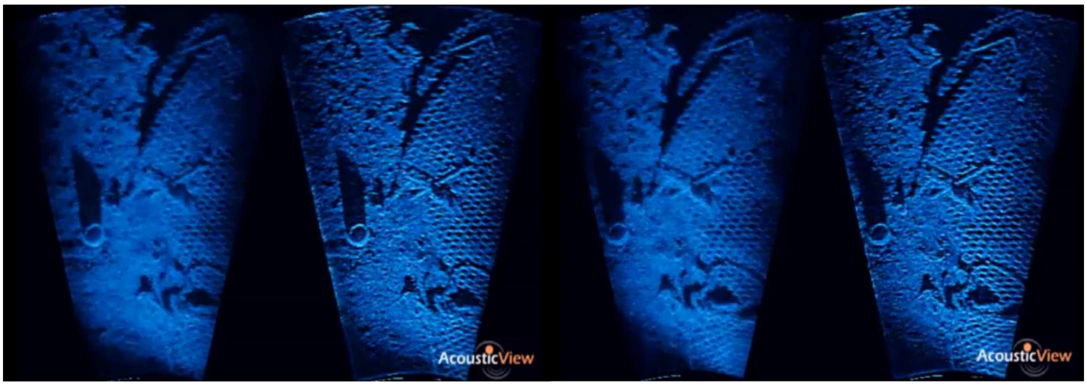


Figure 7. Two frames of a sonar video.

3.2. Video Compression Results

3.2.1 Periscope Video Compression

Four compression algorithms (J2K, Daala, X264, X265) were compared. Four performance metrics (PSNR, SSIM, HVS, HVSm) were used for evaluation. As shown in Figure, the performance metrics are very high and can be considered as perceptually lossless. Table summarizes the performance metrics of various codecs at a compression ratio of 0.05. It is worth to mention that X264 performed slightly better than X265. However, X265 performed better at lower compression ratio such as 0.01. J2K and Daala were not as good as X264 and X265. Since the computational burden of X264 is much less than X265, it is suitable for real-time applications.

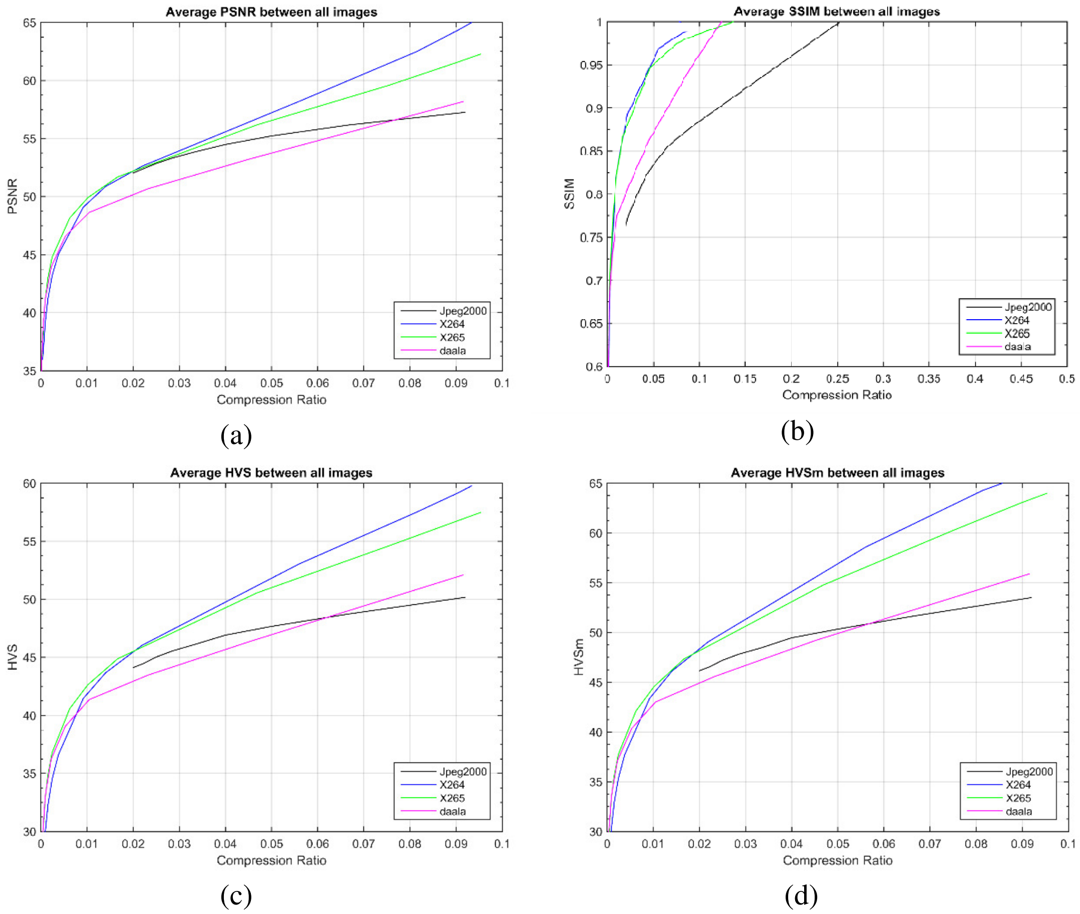


Figure 8. Performance metrics for periscope video compression: (a) Averaged PSNR in dB of all frames versus compression ratio; (b) Averaged SSIM of all frames versus compression ratio; (c) Averaged HVS in dB of all frames versus compression ratio; (d) Averaged HVSm in dB of all frames versus compression ratio.

Table 2. Performance metrics of four codecs at 0.05 compression ratio for a periscope video. Bold numbers indicate the best performing method.

	PSNR (dB)	SSIM	HVS (dB)	HVSm (dB)
J2K	55.25	0.83	47.5	50.4
X264	**57**	**0.96**	**51.5**	**56.8**
X265	56.5	0.95	51	55.3
Daala	53.75	0.87	46.8	49.8

3.2.2 Sonar Video Compression

Sonar videos are somewhat harder to compress as there are a lot noises and textures in the frames. It is well known that noisy frames are hard to compress. Here, we also applied four compression algorithms and four performance metrics to evaluate the results. Figure shows the performance metrics. It can be seen that all the compression algorithms work well in achieving high numbers.

Table summarizes the metrics of various codecs at a compression ratio of 0.05. Similar to the periscope video, we also observe that X264 performs better than x265. J2K is the worst in this case.

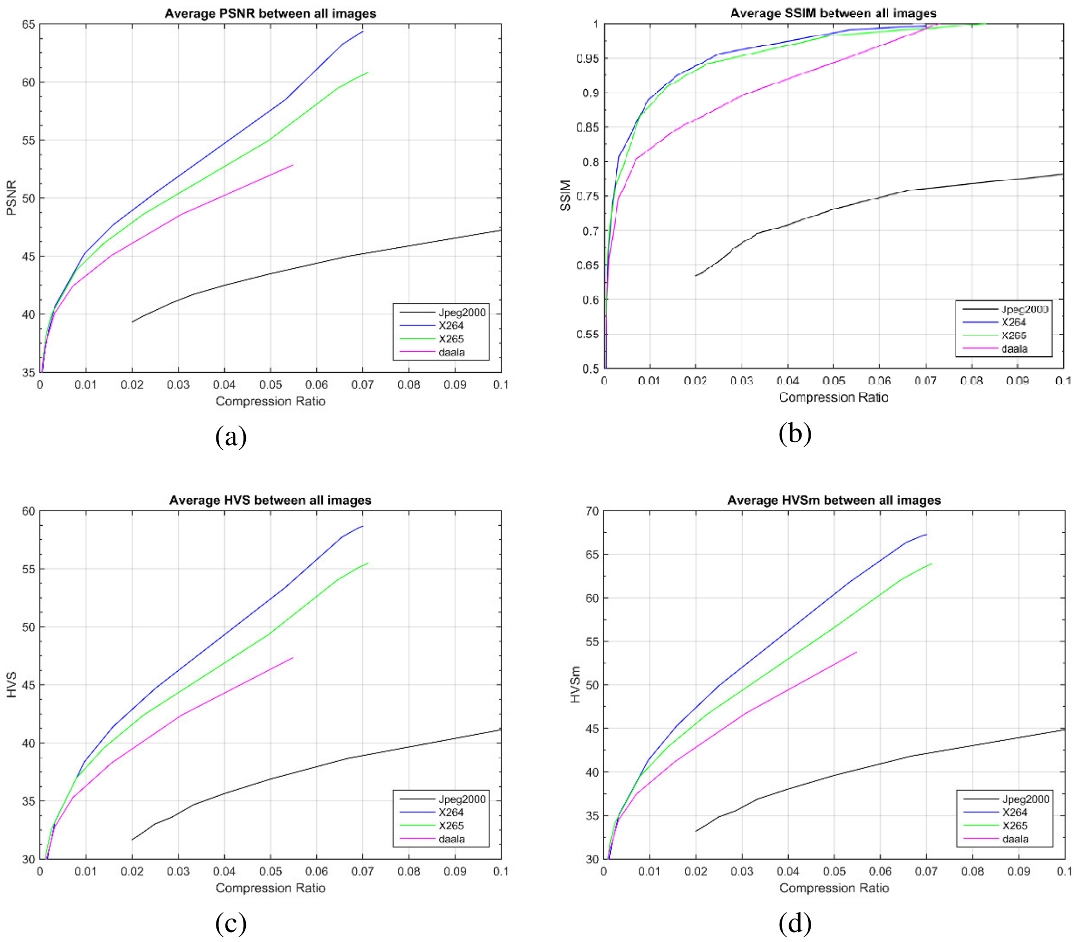


Figure 9. Performance metrics for sonar video compression: (**a**) Averaged PSNR in dB of all frames versus compression ratio; (**b**) Averaged SSIM of all frames versus compression ratio; (**c**) Averaged HVS in dB of all frames versus compression ratio; (**d**) Averaged HVSm in dB of all frames versus compression ratio.

Table 3. Performance metrics of four codecs at 0.05 compression ratio for a sonar video. Bold numbers indicate the best performing method.

	PSNR (dB)	SSIM	HVS (dB)	HVSm (dB)
J2K	43	0.785	36.85	39.5
X264	**57.5**	**0.981**	**52.5**	**60.5**
X265	55	0.98	49.3	56.5
Daala	50.6	0.94	46.25	52.4

3.3 Video Error Concealment Results

As mentioned in the literature review section, error resilient coding and error concealment in the codecs can only protect the compressed bit stream to certain extent. If the interference in the communication channel is very severe due to either natural causes or enemy's intentional jamming, the bit stream will be heavily corrupted as a result. Hence, it is essential and absolutely necessary to have some high performance error concealment algorithms to repair the damaged packets. It is worth to mention that error concealment does not incur any bandwidth as compared to error resilient coding. Of course, some computational resource is needed at the receiving station, which should not be a problem in practice.

In this section, we applied our video error concealment approach and another algorithm in the literature to periscope and sonar videos. The theory has been described in an earlier section. So we focus on presenting the results.

3.3.1 Periscope Video

The video is 720x1280 with 42 frames. The frame rate is only 6. Here we introduced a large corrupted block of size 400x400 on frame 6.
Figure summarizes the original frame 6, frame 6 with a large missing block due to corrupted packets during transmission, and a recovered frame 6. For the 400x400 block, the PSNR is 34.52 dBs for the reconstructed image.

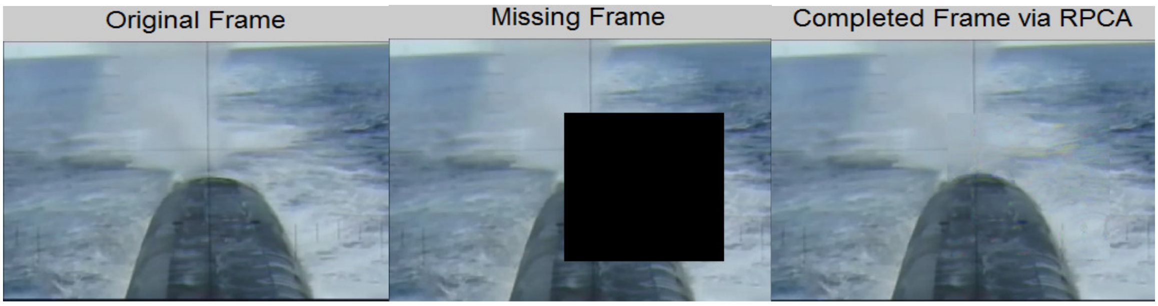


Figure 10. Summary of error recovery in a periscope video.

Here, we compared our algorithm with another one called BMA (Boundary Matching Algorithm) [22]. It should also be noted that all frames have corrupted pixels. The following figures show that our error recovery performance (34.52 dB) is 6.5 dB better than BMA (28.02 dB). This is very significant. One can also visually inspect the recovered images and notice that our performance is perceptually lossless.

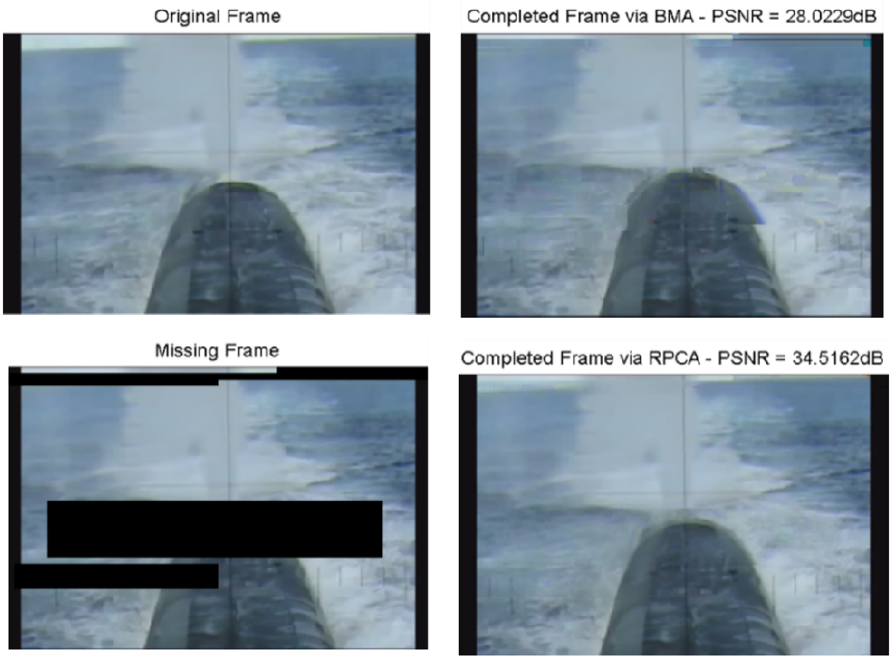


Figure 11. Video completion results for frame 6 – Periscope sequence

3.3.2 Sonar Videos

The video is 720x1280 with 87 frames. The frame rate is only 6. Here we introduced a large corrupted block of size 400x400 on frame 5.

Figure summarizes the original frame 5, frame 5 with a large missing block due to corrupted packets during transmission, and a recovered frame 5. For the 400x400 block, the PSNR is 40.365 dB for the reconstructed image using our algorithm.

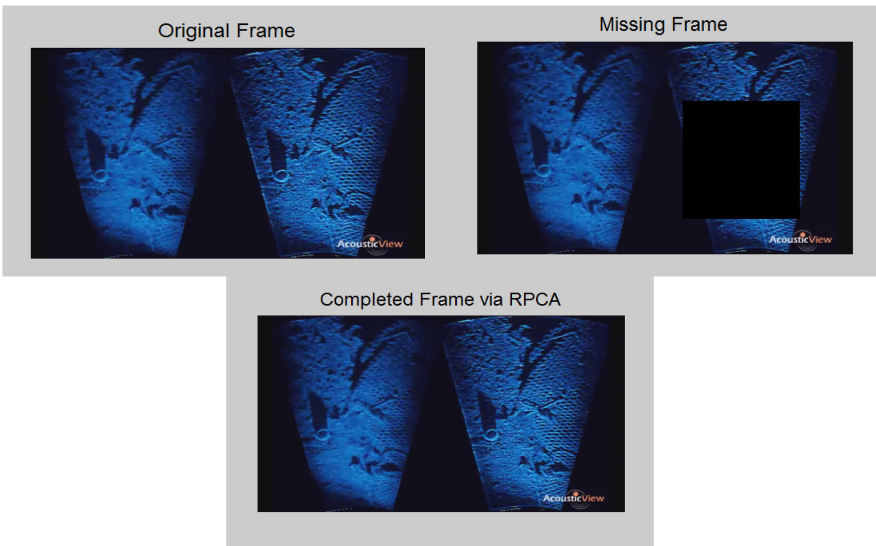


Figure 12. Summary of error recovery in a sonar video.

For the sonar video, we compared our algorithm with another one called BMA (Boundary Matching Algorithm) [27]. It should also be noted that all frames have corrupted pixels. The following figures show that our error recovery performance (40.36 dB) is 7.5 dB better than BMA (32.88 dB). This is very significant. From Figure 13, one can also visually inspect the recovered images and notice that our performance is perceptually lossless.

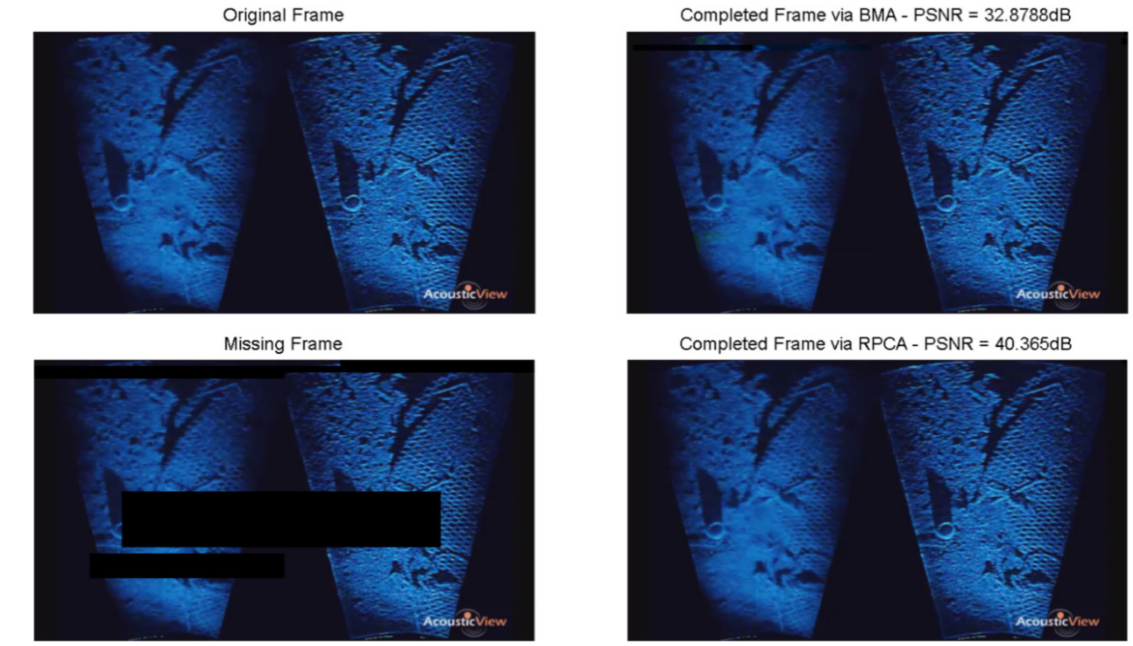


Figure 13. Video completion results for frame 5 – Sonar sequence

4. Conclusions

In this research, we were asked by our customer to achieve perceptually lossless compression at 8:1 compression for videos at 6 fps originally. After some investigations, we have decided to modify the objective to 20 to 1 compression. We have clearly achieved our objective. The best video compression codec has been determined to be X264 via experiments using four performance metrics. Moreover, we also applied two error concealment algorithms to repair those corrupted pixels due to transmission errors.

One future direction is to work with our sponsor and produce an integrated and fast video codec that has both compression and error concealment at the post-decompression stage.

Acknowledgements

This research was supported in part by the US Navy under contract N00024-14-P-4560.The views, opinions and/or findings expressed are those of the author(s) and should not be interpreted as representing the official views or policies of the U.S. Government.

References

[1] Strang, G. and Nguyen, T, (1997), "Wavelets and filter banks", Wellesley-Cambridge Press.

[2] Kwan, C.; Li, B.; Xu, R.; Tran, T. and Nguyen, T., (2001), "Very Low-Bit-Rate Video Compression Using Wavelets", Wavelet Applications VIII, 4391, 176-180.

[3] Kwan, C., Larkin, J., Budavari, B. and Chou, B., (2019), "Compression algorithm selection for multispectral Mastcam images," Signal & Image Processing: An International Journal.

[4] Kwan, C. and Larkin, J., (2018), "Perceptually Lossless Compression for Mastcam Images," IEEE Ubiquitous Computing, Electronics & Mobile Communication Conference, New York City,

[5] Pennebaker, W. B. and Mitchell, J. L., (1993), JPEG–Still image data compression standard, Van Nostrand Reinhold.

[6] Marpe, D., George, G., Cycon, H. L., and Barthel, K. U., (2004) "Performance evaluation of Motion-JPEG2000 in comparison with H.264/AVC operated in pure intracoding mode," Proc. SPIE 5266, Wavelet Applications in Industrial Processing.

[7] Kwan, C.,Li, B., Xu, R., Tran, T., and Nguyen, T., (2001), "SAR image compression using wavelets," Wavelet Applications VIII, Proc. SPIE (vol. 4391).

[8] Tran, T. D., Liang, J., Tu, C., (2003)"Lapped transform via time-domain pre-and post-filtering," IEEE Transactions on Signal Processing.

[9] Valin, J.-M. and Terriberry, T. B., (2015), "Perceptual Vector Quantization for Video Coding," Proceedings of SPIE Visual Information Processing and Communication Conference.

[10] Kwan, C., Shi, E., Um, Y.,(2018),"High performance video codec with error concealment", Data CompressionConference.

[11] Kwan, C., Larkin, J., Budavari, B., Chou, B., Shang, E., Tran, T. D., (2019), "A Comparison of Compression Codecs for Maritime and Sonar Images in Bandwidth Constrained Applications," Computers.

[12] Kwan, C., Shang, E. and Tran, T., (2018), "Perceptually lossless video compression with error concealment", 2nd International Conference on Vision, Image and Signal Processing, Las Vegas, NV, USA.

[13] Ozer, J., (2010), "VP8 vs. H.264," Available online.

[14] Ozer, J., (2016), "What is VP9," Available online.

[15] Dogan, S., Sadka, A. H., Kongoz, A. M., (2005), "Error Resilient Techniques for Video Transmission Over Wireless Channels," Center for Communications System Research, U. Surrey, UK.

[16] Nguyen, D., Dao, M., Tran, T. D., (2011), "Error concealment via 3-mode tensor approximation", IEEE Int. Conf. on ImageProcessing (ICIP), Brussels, Sep. 2011.

[17] Kwan, C., Budavari, B., Dao, M., Zhou, J.,(2017),"New Sparsity Based Pansharpening Algorithm for Hyperspectral Images,"IEEE Ubiquitous Computing, Electronics & Mobile Communication Conference, p 88-93.

[18] Dao, M., Kwan, C., Ayhan, B., Tran, T.,(2016),"Burn Scar Detection Using Cloudy MODIS Images via Low-rank and Sparsity-based Models,"IEEE Global Conference on Signal and Information Processing, p 177 – 181.

[19] Wang, W., Li, S., Qi, H., Ayhan, B., Kwan, C., Vance, S.,(2015),"Identify Anomaly Component by Sparsity and Low Rank", IEEE Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensor (WHISPERS).

[20] Wang, W., Li, S., Qi, H., Ayhan, B., Kwan, C., Vance, S., (2014), "Revisiting the Preprocessing Procedures for Elemental Concentration Estimation based on CHEMCAM LIBS on MARS Rover", 6th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS).

[21] Zhou, J., Kwan, C.,(2018),“High Performance Image Completion using Sparsity based Algorithms”, SPIE Commercial + Scientific Sensing and Imaging Conference.

[22] Zhou, J., Ayhan B., Kwan, C., Tran, T.,(2018),“ATR Performance Improvement Using Images with Corrupted or Missing Pixels”, SPIE Defense + Security Conference.

[23] Kwan, C., Luk, Y.,(2018),“Hybrid sensor network data compression with error resiliency,”Data Compression Conference.

[24] Zhou, J., Kwan, C., (2018),“Missing Link Prediction in Social Networks,”15th International Symposium on Neural Networks.

[25] Kwan, C., Zhou, J.,(2015), Method for Image Denoising, Patent #9,159,121.

[26] Elad, M.,(2010), Sparse and Redundant Representations, Springer New York.

[27] Chen, Y., Hu, Y., Au, O. C., Li, H., Chen, C. W.,(2008), “Video error concealment using spatio-temporal boundary matching and partial differential equation,” IEEE Trans. on Multimedia, vol. 10, no. 1, pp. 2-15, 2008.

[28] Ponomarenko, N., Silvestri, F., Egiazarian, K., Carli, M., Astola, J., Lukin, V., (2007), “On between-coefficient contrast masking of DCT basis functions”, Proc. Third International Workshop on Video Processing and Quality Metrics for Consumer Electronics, Scottsdale, AZ, USA.

AUTHORS

Chiman Kwan received his Ph.D. degree in electrical engineering from the University of Texas at Arlington in1993. He has 15 issued and pending patents, 55 invention disclosures, 325 technical papers in journals and conferences, and more than 500 proprietary technical reports. He has been the principal investigator and program manager of over 115 diverse and competitively selected projects with total funding exceeding 36 million dollars from private and government agencies. He is the Chief Technology Officer of Signal Processing, Inc. and Applied Research LLC, working on robust control, fault diagnostics and prognostics, compressive sensing, data compression, target tracking, speech processing, deep learning, and remote sensing projects.

Jude Larkin received his B.S. in Computer Science from Franciscan University of Steubenville in 2015. He is a software engineer at ARLLC.

Bence Budavari received his B.S. in Audio Engineering from BelmontUniversity in 2015. He is a software developer at ARLLC.

Eric Shang received his B.S. in Electrical Engineering from the University of Maryland College Park in 2014. He was a software engineer at ARLLC and is now with Leidos.

Trac D. Tran received his B.S. from MIT and Ph.D. from U. Wisconsin Madison. He is a Professor in the Department of Electrical and Computer Engineering, The Johns Hopkins University. His pioneering research on integer-coefficient transforms and pre-/post-filtering operators has been adopted as critical components of Microsoft Windows Media Video 9 and JPEG XR – a recent international still-image compression standard ISO/IEC 29199-2. Moreover, the open-source video codec, Daala, also incorporates the pre-/post-filtering operators mentioned above. He was the recipient of the NSF Career Award and is a Fellow of IEEE for his contributions to multi-rate and sparse signal processing.