Seafloor Video Compression Using Adaptive Hybrid Wavelets and Directional Filter Banks

Yang Zhang
Dept. of Ocean Eng.
Ocean University of China
Qingdao, China 266100
Email: zzephyyr@gmail.com

Qingzhong Li
Dept. of Automation Eng.
Ocean University of China
Qingdao, China 266100
Email: liqingzhong@ouc.edu.cn

S. Negahdaripour
ECE Department
University of Miami
Coral Gables, FL 33146
Email: shahriar@miami.edu

***Abstract*—In this paper, a new video compression technique based on adaptive hybrid wavelets and directional filter banks is proposed to achieve both high coding efficiency and good reconstruction quality at very low-bit rates. A key application is the real-time transmission of video from an autonomous underwater vehicle to a surface station, e.g., for man-in-the-loop monitoring and inspection operations, through acoustic channels that have limited bandwidth. The proposed method can maintain details in texture regions at relatively low bit rates, while overcoming the ringing artifacts within smooth regions, for intra-frame coding. For inter-frame coding, improved efficiency is achieved by making use of: 1) a new spatio-temporal just-noticeable-distortion model to remove perceptual redundancy; 2) motion interpolation to reduce bit rate; 3) variable-precision in quantizing the residual error; and 4) block inter-leaver to reduce transmission errors. Experiments with underwater video sequences are presented to assess the effectiveness of the proposed approach, in comparison to traditional wavelet-based techniques.**

***Index Terms*—underwater imagery; video compression; wavelet; directional filter banks; just-noticeable distortion; low-bit-rate coding.**

I. Introduction

Autonomous Underwater Vehicles (AUVs) are commonly deployed for ocean exploration, coral reef and terrain surveys and monitoring, sampling and surveillance, etc. The communication from the AUV to the surface for human-in-the-loop missions is typically through acoustic channels with very limited bandwidth. Where video is to be transmitted in real time, very high data compression has to be achieved. For example, even with a relatively high 16 kbps bandwidth, the transmission of relatively low-resolution 352 [pix]$\times$288 [pix] color images[1] at 30 frame per second (fps) requires a compression ratio (CR) of about 1000:1.

Underwater images typically have low contrast and color variations, although seafloor natural objects (e.g., coral reefs) can have rich texture with contours at arbitrary orientations. The conventional MPEG-4 and H.264 coding schemes, designed for images of terrestrial environments, do not generally offer good compression performance, because of the variable characteristics of underwater video. Effective methods for underwater video compression need to produce a visually-pleasing decoding quality for different types of video imagery, e.g., low-contrast deep-sea images, high-detailed frames of coral reef and other natural seafloor objects, as well as varying and (or) non-uniform illumination.

Some earlier works have proposed underwater video compression techniques based on discrete wavelet transform (DWT) [1]-[3], but the compression rates are inadequate. Moreover, the DWT does not achieve satisfactory decoding quality for high-detail seafloor imagery, containing large areas of texture and varying edge orientation, due to limited directional representation [5]. To handle texture and contour structures for still image compression more effectively, we recently proposed a technique based on hybrid wavelets and directional filter banks (HWD) [4]. However, ringing artifacts arise in low-contrast images due to frequency aliasing between wavelets and directional filter banks (DFBs).

To overcome this problem, this paper describes a technique based on an adaptive HWD transform (AHWD), which is also used for the intra-frame coding of the reference image in video coding. We achieve high compression efciency and visually-pleasing reconstruction by preserving details and minimizing compression artifacts. Here, we divide the high-frequency of DWT into textured and untextured subblocks, and apply the DFBs only to textured subblocks to produce directional subbands adaptively. For inter-frame video coding, we propose a novel DWT-based spatio-temporal just-noticeable-distortion (JND) model to remove perceptual redundancy. Moreover, the residual error images are quantized and encoded by variable precision, where a block interleaving scheme is introduced to effectively reduce the impact of transmission error on video reconstruction. The motion vectors and radiometric tansformations are coded losslessly with fixed length.

Our AHWD-based compression scheme provides three main contributions:

- We overcome ringing artifacts encountered in conventional HWD-based compression algorithms. According to image content, the AHWD can provide rich and flexible directional decomposition for sparse signal representation. Consequently, using AHWD transform, we can achieve visually-pleasing reconstructed quality for both low-contrast and high-detail video imagery.
- We present a motion-compensated residual preprocessing method in DWT domain based on spatio-temporal JND

[1]This is the resolution of images used in our experiments.

978-0-933957-43-5 ©2015 MTS

model. The reduction of variances of DWT coefficients leads to lower residual signal distortion and improved residual coding efficiency [6]. The novel spatio-temporal JND model suppresses transform coefficient values of residues in the DWT domain before quantization, and accounts for the temporal impact by incorporating a modulation factor that depends on motion vectors and radiometric variations between consecutive frames.

- We have devised an efficient video coding scheme to reduce the computational complexity and bit rate by grouping every sixteen consecutive frames – one I-frame, five P-frames and ten B-frame – and implementing the motion estimation and compensation of P-frames in the low-frequency subband of one-level DWT. Moreover, the motion vectors and radiometric components of B-frame are predicted directly from the adjacent P- or I-frame at the decoder, avoiding the need to estimate and encode optical flow information of B-frames at the encoder.

In the remainder, we introduce the proposed AHWD transform in Section II. We describe the DWT-based spatio-temporal JND model for perceptual redundancy removal in Section III. We give the implementation details of the proposed AHWD-based video coding algorithm in Section IV, and the experimental results with seafloor video in Section V. We present a summary and conclusions from this investigation in Section VI.

II. Adaptive Hybrid Wavelets and Directional Filter Banks

A. Ringing Artifacts of HWD

For still image compression, we have developed an HWD-based coding method that maintains texture and contour structures more effectively than earlier DWT-based algorithms [4]. However, ringing artifacts are introduced in smooth regions, where some of the small HWD transform coefficients are coded as zero, for the following reasons:

- Smooth regions generally have nonzero transform coefficients in the coarse levels of DWT. Since the smooth signals are best represented by wavelet basis functions, it is unnecessary to decompose these signals further by DFBs. The human vision system (HVS) is more sensitive to the change in coarser levels of DWT due to the low frequency content in these subbands. Employing DFBs to the entire wavelet high-frequency subband deteriorates the representation of smooth signals and results in ringing artifacts that can be readily observed by HVS.
- As large-size fan filters are used to construct DFBs, the size of directional basis functions may become larger than the size of wavelet subbands when employing DFBs in the coarser DWT levels. In order to obtain the filtered signals, the input signals have to be amplified by concatenation to match the size of directional filter [5], thus distorting the output. Although the size of directional filters is the same as the size of wavelet subbands in the finer levels, the distortions still exist due to the low-frequency leakage. Fig. 1(a) shows the frequency partitioning of one-level DWT. The two green regions denote the horizontal high-frequency subbands. Fig. 1(b) shows how the directional subbands are produced using full-tree and half-tree DFBs, respectively. It should be noted that the DWT cannot perfectly separate low frequency (white part in subband H_1) from high frequency in a high-frequency subband. Since the low-frequency content of wavelet subband is the input to the directional decomposition by both full-tree and half-tree DFBs, frequency scrambling still exists in the fine DWT level, which gives rise to artifacts.

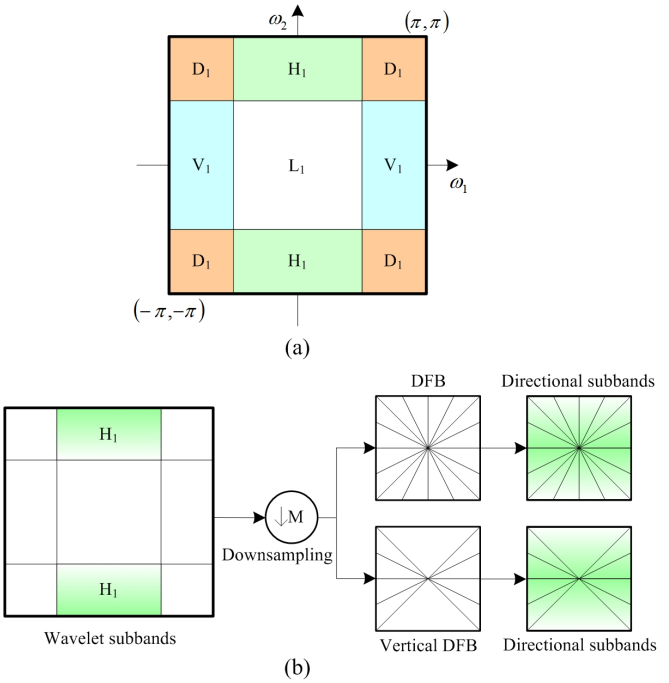


Fig. 1. (a) Frequency partitioning of one-level DWT; (b) Schematic plot of directional decomposition for a wavelet subband in the HWD transform.

B. Implementation of Adaptive HWD

Although frequency scrambling occurs in all levels of wavelet high-frequency channels, it is better to reduce the size of DFBs to avoid low-frequency leakage. Traditional HWD transform designed two schemes to adapt to the different image types [5], but manual intervention is needed to decide which scheme should be used. Neither scheme makes full use of image content, so the traditional HWD transform is non-adaptive. We propose an adaptive HWD transform to overcome the above problem. Here, we separate the texture and smooth regions in the wavelet domain. DFBs are only applied to texture subblocks to avoid ringing artifacts. Since the coarser levels of wavelet high-frequency subbands are sensitive to directional decomposition, we only apply DFBs to the first three DWT levels. Each wavelet high-frequency subband in the first three levels is divided into non-overlapping subblocks. The width and height of each subblocks are defined as

$$d_l = \begin{cases} D \cdot 2^{-(l+1)} & D \leq 256 \\ D \cdot 2^{-l}/3 & D > 256 \end{cases} \quad (1)$$

where $D = \{W, H\}$ is the $\{width, height\}$ of the image, and d_l is the corresponding width/height of subblock in the DWT level $l (l = 1, 2, 3)$.

We label a subblock as textured or untextured using the following criterion:

$$t_{k,l} = \frac{\sigma_{k,l}}{\sigma_{max}} \quad (2)$$

where $\sigma_{k,l}$ is the standard deviation of transform coefficients of subblock k in the DWT level l, and σ_{max} denotes the maximum of standard deviations. A subblock is labelled as texture/smooth region if $t_{k,l}$ is above/below threshold $T_{texture}$. The suitable value of $T_{texture}$ has been determined empirically to be 0.65, for experiments reported in this paper. To reduce computational complexity, we compute the texture measure at the finest DWT level ($l = 1$) only, and then apply to the other two DWT levels. If a subblock is textured, the DFB is applied to decompose this subblock into directional subbands. If not, we keep the subblock unchanged.

The DFBs with smaller size are applied to texture regions only, rather than to the entire wavelet subband to avoid frequency aliasing. Thus, the AHWD transform can suppress ringing artifacts significantly, while maintaining textures effectively. Moreover, the proposed AHWD transform makes use of the same preprocessing and coding schemes as in our previous work [4], for the reference frames in video compression.

III. Spatio-Temporal JND Model for Wavelet Domain

The key to inter-frame coding efficiency is the residual encoding. When the quantized residual data are entropy-coded, the MSE-based signal distortion is proportional to the variances of DWT coefcients of the residues [6]. It is expected that reduction of the variances of DWT coefcients leads to reduction of residual signal distortion and improvement in the residual coding effciency. In previous work on still image compression [4], we employed a HWD-based JND model to remove peceptual redundancy. Here, *we propose a novel spatio-temporal JND model to suppress transform coefficient values of residues in the DWT domain before quantization*. We account for the temporal impact by employing a modulation factor, defined in terms of the motion vectors and radiometric variations of consecutive frames.

The JND model exploits the perceptual limitations of the HVS by calculating a visibility threshold, below which a change cannot be perceived by the majority of human observers. That is, we aim to adaptively remove visually-insensitive transform coefcients to reduce the variances of residual data, thus achieving the goal of reducing redundancy. The spatio-temporal JND threshold can be expressed based on the contrast sensitivity function modulated by luminance adaption, contrast masking, and temporal masking effects:

$$T_{JND} = T_{base} \cdot A_l \cdot M_c \cdot M_t \quad (3)$$

explained in detail below, T_{JND} is the JND threshold of transform coefficient; T_{base} is the base JND threshold in the DWT domain generated by spatial contrast sensitivity function; A_l and M_c are luminance adaption and contrast masking functions, respectively, as used in our previous work [4]; M_t is the temporal modulation factor.

A. Contrast Sensitivity Function

The contrast sensitivity function measures the human sensitivity to various spatial frequencies of visual stimuli. The larger the contrast sensitivity becomes, the smaller the base JND threshold. Consequently, the base JND threshold T_{base} is proportional to the inverse of contrast sensitivity function CSF:

$$T_{base} = \begin{cases} 0; & \text{for low-freq. subband} \\ \frac{1}{r+(1-r)\beta^2} \cdot \frac{1}{CSF}; & \text{for high-freq. subbands} \end{cases} \quad (4)$$

where the scaling term $1/(r + (1 - r)\beta^2)$ accounts for the directionality. To elaborate, the HVS is more sensitive to the horizontal/vertical than the diagonal direction. We set $\beta = 1$ and $\beta = \sqrt{2}/2$ for vertical/horizontal and diagonal subbands, respectively, and r is empirically set to 0.6.

For the DWT coefcient at location (i, j) within subband d at level l, the contrast sensitivity function $CSF_{i,j,d,l}$ is set to [7], [8]:

$$CSF_{i,j,d,l} = 2.6 \times \left(0.0192 + \frac{f_{i,j,d,l}}{f_0}\right) \cdot \exp\left[-\left(\frac{f_{i,j,d,l}}{f_0}\right)^{1.1}\right] \quad (5)$$

where $f_{i,j,d,l}$ [cycles/degree] is the spatial frequency of a transform coefficient [7], [9], corresponding to the spatial frequency $f(u, v)$ in the image domain:

$$f(u,v) = \sqrt{\left(\frac{u}{2W_x \cdot \theta_y}\right)^2 + \left(\frac{v}{2W_y \cdot \theta_x}\right)^2}; \quad 0 \leq u \leq W_x,\ 0 \leq v \leq W_y \quad (6)$$

where W_x/W_y is the image width/height, respectively, and θ_x/θ_y is the horizontal/vertical visual angles of a pixel [9]:

$$\theta_\hbar = 2 \arctan\left(\frac{1}{2 \cdot R_\hbar \cdot P_\hbar}\right) \quad (7)$$

where $R_\hbar$ and $P_\hbar$ ($\hbar = \{x, y\}$) denote the ratio of viewing distance between the human eyes and monitor to image width/height, and the number of pixels in the horizontal/vertical directions, respectively.

In Eq. (5), f_0 represents the spatial frequency corresponding to the peak sensitivity. Since the HVS is more sensitive to low frequency, we set f_0 to the spatial frequency of DWT low-frequency subband

$$f_0 = f\left(W_x/2^L,\ W_y/2^L\right) \tag{8}$$

where L is the total number of DWT level.

B. Temporal Modulation Factor

An effective spatio-temporal JND model must account for the temporal characteristics of HVS, namely, the sensitivity of the HSV to variations in successive frames. The traditional spatio-temporal JND models produce higher temporal masking for larger inter-frame motion vectors, according to the insensitivity of the HVS to changes under fast motions. In contrast, slower motions lead to smaller temporal masking. However, traditional JND models consider the geometric transformations (also known as optical flow) to establish temporal masking. They neglect the often significant impact of radiometric variations in underwater applications [10].

Generally, human subjects can readily detect the changes or additive noise in successive frames with small brightness variations, however, less/not the case with large brightness variations. Here, we propose a temporal masking function based on both geometric and radiometric variations in the video sequence. More precisely, we incorporate a modulation factor in the temporal masking function based on the frame-to-frame radiometric transformation:

$$M_t = \begin{cases} e^{0.05\delta I}; & f_{i,j,d,l} < 5 \text{ and } f_t < 10 \\ 1.07^{(f_t-10)} \cdot e^{0.05\delta I}; & f_{i,j,d,l} < 5 \text{ and } f_t \geq 10 \\ 1.07^{f_t} \cdot e^{0.05\delta I}; & f_{i,j,d,l} \geq 5 \end{cases} \tag{9}$$

where the radiometric transformation field δI is estimated by the optical flow method in [2], [10], devised based on the Generalized Dynamic Image Model (GDIM), and f_t [Hz] is the temporal frequency of a transform coefficient:

$$f_t = f_x v_x + f_y v_y \tag{10}$$

where f_x/f_y and v_x/v_y denote the horizontal/vertical components of spatial frequency $f_{i,j,d,l}$ and object velocities on the retina plane (horizontal/vertical components of optical flow), respectively [9].

C. Implementation of JND-Based Residue Preprocessor

The residual error images are generated by inter-frame motion estimation and compensation. A one-level DWT is applied to further decompose the residual image before inter-frame coding. In the JND-based residue preprocessor, the JND suppression by the spatial-temporal JND model in Eq. (3) is implemented by subtracting the JND threshold from the DWT coefficient value as

$$\left|c'_{i,j,d,l}\right| = \begin{cases} |c_{i,j,d,l}| - T_{JND}; & |c_{i,j,d,l}| > T_{JND} \\ 0; & \text{otherwise} \end{cases} \tag{11}$$

where $c_{i,j,d,l}$ and $c'_{i,j,d,l}$ denote the transform coefficients before and after JND suppression, respectively.

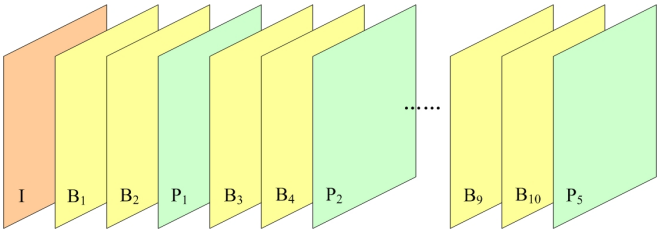


Fig. 2. 16-Frame structure in the proposed video coding scheme.

IV. Video Coding Scheme

In the proposed underwater video coding scheme, every 16 consecutive frames, comprising of one reference I-frame, five P-frames and ten B-frames, are grouped together; Fig. 2 depicts the 16-frame structure in image sequence. As stated, the reference I-frame is encoded based on the proposed AHWD transform, using the same coding algorithm from our previous work [4]. The P-frames and B-frames are coded by the estimation of inter-frame geometric (optical flow) and radiometric transformation δI, as detailed next.

A. Estimation and Coding for Motion Information

We employ the GDIM-based optical flow method [10] to estimate both geometric and radiometric transformations of P-frames in the underwater image sequence. The motion estimation and compensation are implemented in the low-frequency subband of a one-level DWT, to reduce computational complexity. Since the size of low-frequency subband of one-level DWT is a quarter of the original image, this yeilds a 75% bit-rate saving in the coding of both motion information and residual error.

In order to improve the coding efficiency, the motion information of B-frames are computed by motion interpolation from adjacent P- or I-frames. To elaborate, the motion and radiometric transformation fields between frames P_1 and P_2 in Fig. 2 are set to $(\delta x, \delta y)$ and δI, respectively. Exploiting motion continuity, we set the same between frame P_1 and B_3 frame to $(\delta x, \delta y)/3$ and $\delta I/3$, and $2(\delta x, \delta y)/3$ and $2\delta I/3$ between frames P_1 and B_4. There is no need to estimate and code the optical flow information for B-frames at the encoder side, thus yielding a remarkable bit-rate reduction. Since motion vectors and radiometric variations are very important for video reconstruction, we code them losslessly by fixed bit length.

B. Quantization and Block Interleaving for Residual Image

For residual error images generated by motion compensation, we use the JND-based preprocessor, described in Section III, to suppress the transform coefficients of a one-level DWT before quantization and coding. Referring to Fig. 3, the DWT coefficients of residues are quantized by variable-precision quantization intervals. Setting the maximum value of transform coefficients to C_{max}, coefficients larger than C_{max} are treated as significant for coding. As depicted, smaller/larger coefficients are assigned to smaller/larger subintervals, shown

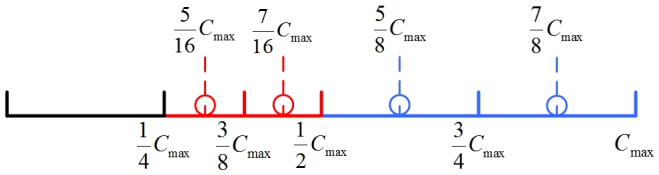


Fig. 3. Quantization interval for residual error image.

Fig. 4. Schematic diagram of block interleaving with 8 groups.

in red/blue color. For inverse quantization the four subintervals are set to values $5C_{max}/16$, $7C_{max}/16$, $5C_{max}/8$, and $7C_{max}/8$, and assigned code words 00, 01, 10, and 11.

After quantization, we adopt a block interleaving technique to effectively reduce the impact of transmission error on video reconstruction. Fig. 4 shows a schematic diagram of the block interleaving with 8 groups. Every coding subblock is an 8×8 region. The coding subblocks with the same color are allocated to the same group. For example, all the subblocks with gray color belong to group 8. The coding order is set on a group basis, from group 1 to group 8. Once all the subblocks in a group are finished encoding, the coding for next group starts. The transmission of code words is also implemented group by group. If a transmission error occurs in a certain group, the impact on decoding quality is weakened due to the dispersed coding locations in the group. This block interleaving scheme is used for the coding of both residual error and reference image.

V. Experimental Results

To demonstrate the three main contributions of this work, we present the results from three types of experiments with five underwater videos, each with 150 frames. The sample images in Fig. 5 have a resolution of 352 [pix]×288 [pix], and depict the characteristics and texture contents of the sequences.

A. Coding Performance of Reference Frame

The first two tests assesses the decoding quality of reference frame, by comparing the proposed AHWD-based algorithm with three wavelet-based coding techniques, namely, JPEG 2000, SPIHT [11], and WTWDR [12], as well as our previous HWD-based algorithm [4]. Our proposed method involves a four-level AHWD transform with two-level DFB, the JPEG 2000, SPIHT, and WTWDR algorithms employ a four-level

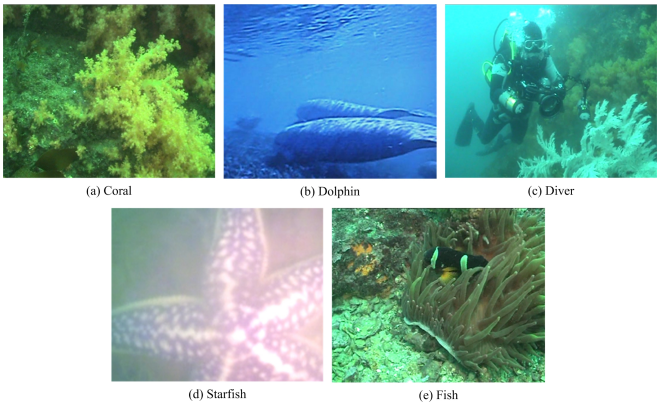


Fig. 5. Samples from five underwater video sequences used in experiments.

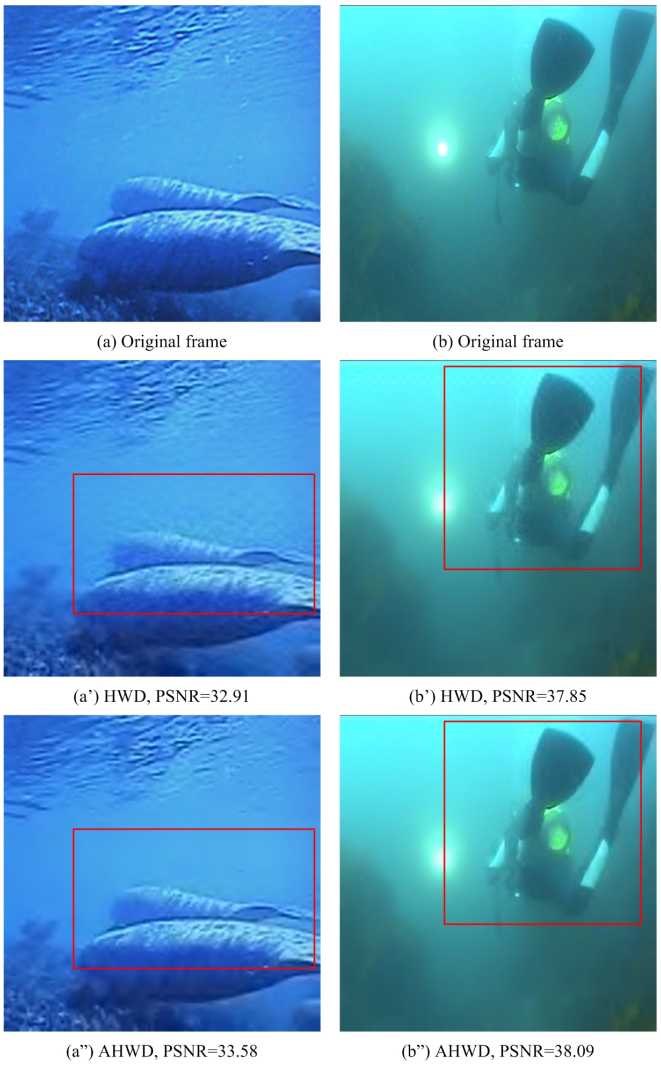
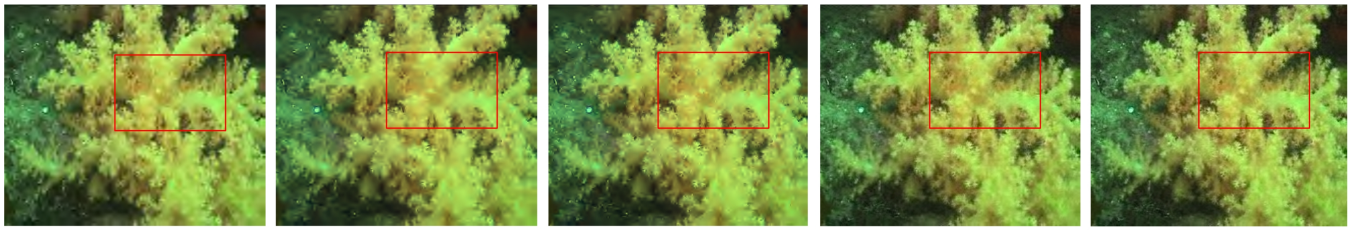


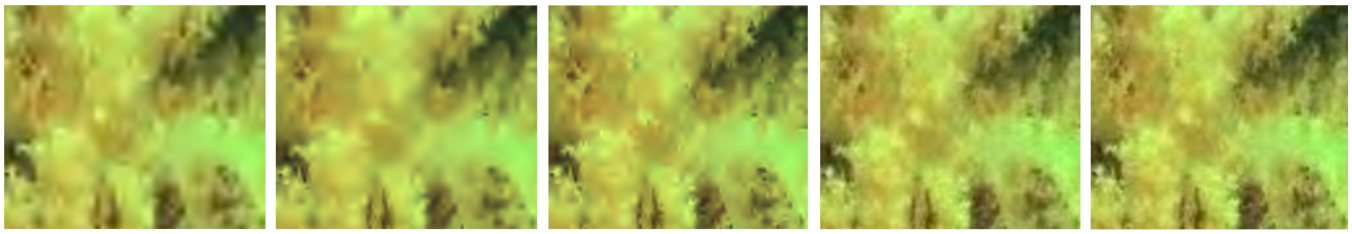
Fig. 6. Reconstructed results of two reference frames by HWD- and AHWD-based algorithms at 0.1 bpp rate.

DWT, and our previous HWD-based coding algorithm incorporates a four-level HWD transform. For a fair comparison, these all apply the same preprocessing method before compression.

We first demonstrate the improvement over our previous



(a) JPEG 2000, PSNR=30.44 (a') SPIHT, PSNR=31.58 (a") WTWDR, PSNR=31.82 (a"') HWD, PSNR=32.64 (a"") AHWD, PSNR=32.81

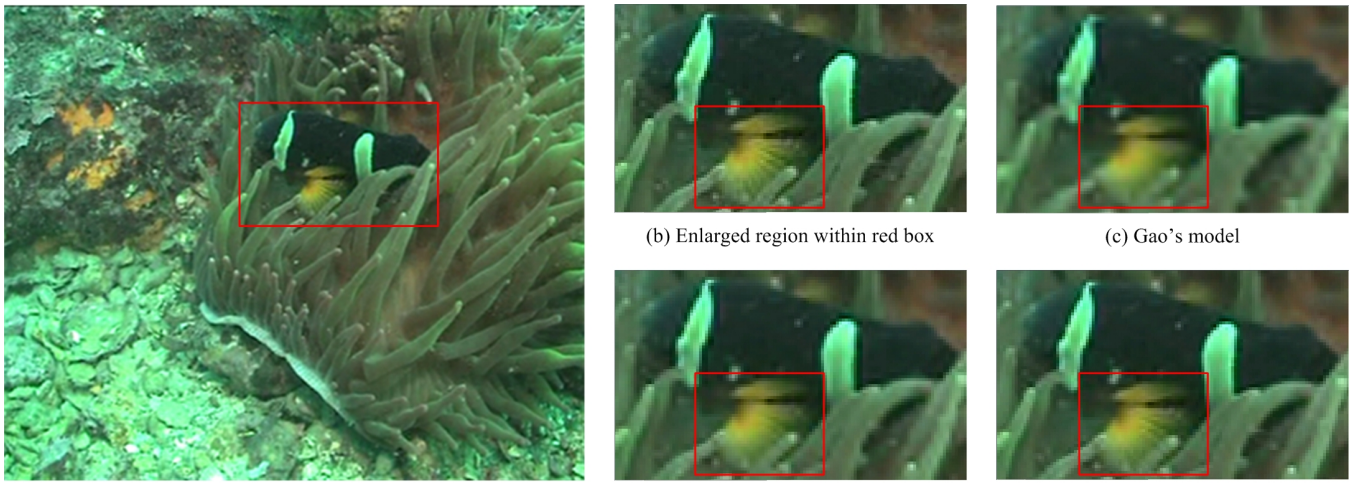


(b) JPEG 2000 (b') SPIHT (b") WTWDR (b"') HWD (b"") AHWD

Fig. 7. Reconstructed results of a reference frame with large texture regions by five algorithms at 0.1 bpp rate.

TABLE I
PERFORMANCE COMPARISON FOR THE THREE JND MODELS IN TERMS OF STANDARD DEVIATION, AVERAGE PSNR AND BIT-RATE REDUCTION

Test Sequences	Standard Deviation				Average PSNR				Bitrate Reduction (%)		
	Original	Gao's	Liu's	Proposed	Original	Gao's	Liu's	Proposed	Gao's	Liu's	Proposed
Coral	3.91	3.11	2.05	1.68	31.97	32.14	32.21	32.23	5.51	8.71	15.99
Dolphin	2.94	2.19	2.23	1.98	35.94	36.10	36.35	36.39	8.34	7.22	11.91
Diver	4.38	3.45	3.17	2.76	34.58	34.65	34.70	34.72	8.62	10.56	17.70
Starfish	2.55	2.06	1.99	1.67	35.10	35.09	35.17	35.26	3.58	4.93	9.12
Fish	6.01	4.89	4.50	3.01	30.22	30.26	30.40	30.41	12.51	19.71	23.18
Average	3.96	3.14	2.79	2.22	33.56	33.65	33.76	33.80	7.71	10.23	15.58



(a) Originial frame (b) Enlarged region within red box (c) Gao's model (d) Liu's model (e) Proposed JND model

Fig. 8. Comparision among Subjective qualities of reconstructed frames in the *Fish* sequence by three JND models at 0.8 bpp.

HWD-based coding algorithm, shown to be more effective than the wavelet-based algorithms in terms of texture representation, but suffers from ringing artifacts within the smooth regions of low-contrast underwater images [4]. Fig. 6 depicts the reconstruction of two reference frames with large smooth regions at 0.1 bit/pixel (bpp). Applying the previous HWD transform for coding leads to severe ringing artifacts within smooth regions, e.g., regions within red boxes in Figs. 6(a')

and (b'). This is overcome in the proposed AHWD-based coding algorithm as depicted in Fig. 6(a") and (b"), thus achieving a better visual quality. This is also reflected in the slight improvements in the PSNR values by 0.67 dB and 0.24 dB, respectively.

The AHWD-based method performs well also for textured images. As an example, Fig. 7 shows the reconstruction of a reference frame by the five algorithms at 0.1 bpp rate. Comparing the enlarged areas in Fig. 7(b) to (b""), the AHWD-based algorithm is comparable to the HWD-based coding algorithm, and maintains more texture details relative to the JPEG 2000, SPIHT and WTWDR algorithms.

B. Verification of JND-Based Residue Preprocessor

In order to verify the effectiveness of the proposed JND-based residue preprocessor, we compare our proposed spatio-temporal JND model with two DWT-based JND models of Gao et al. [8] and Liu et al. [13], in terms of average PSNR, bit-rate reduction and standard deviation of residual image. As the baseline, we also give results without applying the JND. Note that the PSNR values are computed using the images before quantization and after decoding, to measure the quantization distortion. For fair comparison, all models are implemented in a one-level DWT domain.

Table I shows the performance comparison for the three JND models. Producing a higher JND threshold, the residue preprocessor of our JND model yields the smallest standard deviation with an average of 2.22. With a smaller standard deviation of transform coefficients, the quantized data can be more closed to the original values and reconstructed accurately. Referring to Fig. 3, the proposed JND model ensures that more transform coefficients of the residual image are put into the smaller quantization bins, during the variable-precision quantization process, which reduces the quantization distortion for decoding. This is also reflected in the slightly higher average PSNR values yield by the proposed JND model as shown in Table I. Compared to the original coding scheme without JND model, the average bit-rate reduction is 7.71% for Gaos model, 10.23% for Lius model and 15.58% for the proposed model (see Table I). That is, the coding scheme with the proposed JND model outperforms Gaos and Lius models, with higher bit-rate reductions by 7.87% and 5.35%, respectively.

Fig. 8 shows some image regions of the reconstructed frames using the Gaos, Lius and proposed JND models, as the residue preprocessor in the video coding scheme. Considering the pelvic fin within the enlarged areas in Figs. 8(b), (c), (d) and (e), the proposed spatio-temporal JND model produces less visual distortion, relative to the other two JND models.

C. Video Coding Performance

In the last experiment, we compare the proposed AHWD-based video coding algorithm with the best of three DWT-based technique [3], and an HWD-based method that uses the same coding scheme as our AHWD-based method but HWD (instead of AHWD) transform for reference-frame coding.

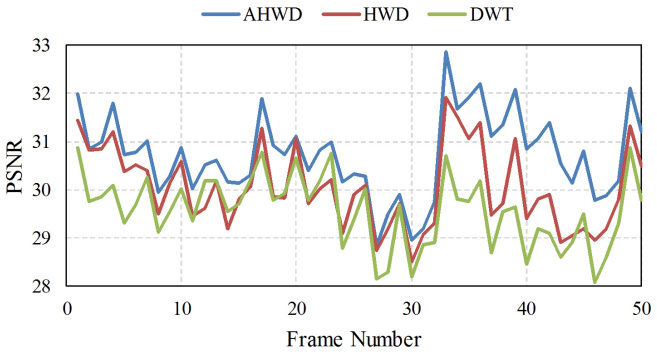


Fig. 9. PSNR comparison of the *Fish* sequence compressed by three coding methods

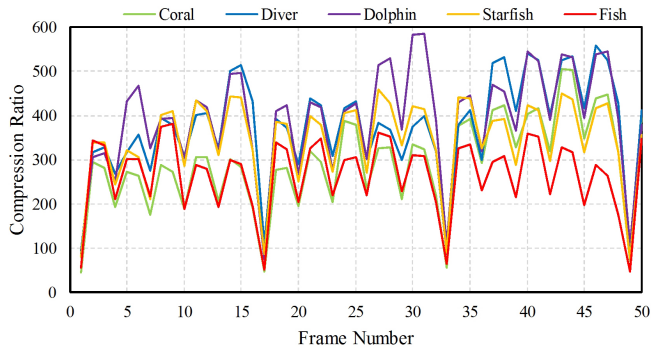


Fig. 10. Compression ratios of proposed video coding algorithm for five video sequences.

Fig. 9 demonstrates the improved PSNR of the proposed AHWD-based coding scheme over the 50 frames of the *Fish* sequence (note that these methods use the same inter-frame coding scheme). This confirms the significance of effective reference-frame coding in video compression. It is noted that the PSNR values of B-frames are lower than those of adjacent P-frames because we use the motion interpolation for B-frames to reduce bit rate at the expense of estimation accuracy. Fig. 10 illustrates the compression ratios of five video sequences compressed by the proposed coding scheme. As expected, large CR jumps arise when the coding algorithm is reset at every reference frame (after each 16 frames). Based on Fig. 10, the maximum CR by the proposed scheme amounts to 580:1 for the *Dolphin* sequence due to the low-contrast image content. An average CR of roughly 300:1 suggests that the bandwidth constraint of the acoustic channel can be met by transmitting at roughly 1/3 standard video rate.

VI. Conclusion

An AHWD-based compression technique algorithm has been designed for underwater video transmssion. In the intra-frame coding, an AHWD transform is proposed to overcome the ringing artifacts of HWD transform introduced by frequency aliasing. Using adaptive directional decomposition, the AHWD transform can maintain important texture structures for high-detail images, thus preserving the advantage of HWD

transform. For inter-frame coding, a spatio-temporal JND-based residue preprocessor is proposed for increased bit-rate savings. Incorporating a temporal modulation factor yields a higher visual threshold in order to remove more perceptual redundancy. The motion interpolation of B-frames lowers the computation burden and yields reduced bit rates. Additional enhancements are realized by variable-precision quantization and block interleaver. The results from various experiments with underwater image sequences have demonstrated that the proposed AHWD-based video compression technique can achieve both higher coding efciency and enhanced decoding quality at very low bit rates, compared to the wavelet-based coding algorithms. Future work will be focused on improved error resiliency to overcome the impact of inevitable noise in underwater acoustic telemetry.

ACKNOWLEDGMENT

Yang Zhang is a Ph.D. student at Ocean University of China. He is currently sponsored by China Scholarship Council (CSC) during appointment at University of Miami as a visiting scholar.

REFERENCES

[1] D. F. Hoag, V. K. Ingle, and R. J. Gaudette, "Low-bit-rate coding of underwater video using wavelet-based compression algorithms," *IEEE Journal of Oceanic Engineering*, vol. 22, no. 2, pp. 393-400, Apr. 1997.

[2] S. Negahdaripour and A. Khamene, "Motion-based compression of underwater video imagery for operations of unmanned submersible vehicle," *Computer Vision and Image Understanding*, vol. 79, no. 1, pp. 162-183, Jul. 2000.

[3] Q.-Z. Li, B. Wang, and W.-J. Wang, "An efficient underwater video compression algorithm for underwater acoustic channel transmission," *in Proc. CMC'09*, Yunnan, China, Jan. 2009.

[4] Y. Zhang, Q.-Z. Li, and S. Negahdaripour, "Seafloor image compression based on hybrid wavelets and directional filter banks," *in Proc. Oceans'15*, Genova, May 2015.

[5] R. Eslami and H. Radha, "A new family of nonredundant transforms using hybrid wavelets and directional filter banks," *IEEE Trans. Image Processing*, vol. 16, no. 4, pp. 1152-1167, Apr. 2007.

[6] X.-K. Yang, W.-S. Lin, Z.-K. Lu, E. Ong, and S.-S. Yao, "Motion-compensated residue preprocessing in video coding based on just-noticeable-distortion profile," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 6, pp. 742-752, Jun. 2005.

[7] C.-H. Chou and K.-C. Liu, "A perceptually tuned watermarking scheme for color images," *IEEE Trans. Image Processing*, vol. 19, no. 11, pp. 2966-2982, Nov. 2010.

[8] X. Gao, W. Lu, D. Tao, and X. Li, "Image quality assessment based on multiscale geometric analysis," *IEEE Trans. Image Processing*, vol. 18, no. 7, pp. 1409-1423, Jul. 2009.

[9] Z. Wei and K. N. Ngan, "Spatio-temporal just noticeable distortion profile for grey scale image/video in DCT domain," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 19, no. 3, pp. 337-346, Mar. 2009.

[10] S. Negahdaripour, "Revised interpretation of optical ow: Integration of geometric and radiometric transformations for dynamic scene analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 9, pp. 961-979, Sept. 1998.

[11] A. Said and W. A. Pearlman, "A new fast and efcient image codec based on set partitioning in hierarchical trees," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 6, no. 3, pp. 243-250, Jun. 1996.

[12] Q.-Z. Li and W.-J. Wang, "Low-bit-rate coding of underwater color image using improved wavelet difference reduction," *Visual Communication and Image Representation*, vol. 21, no.7, pp. 762-769, May 2010.

[13] Z. Liu, L. J. Karam, and A. B. Watson, "JPEG2000 encoding with perceptual distortion control," *IEEE Trans. Image Processing*, vol. 15, no. 7, pp. 1763-1778, Jul. 2006.