Depth-Assisted Temporal Error Concealment for Intra Frame Slices in 3-D Video

Meng Yang, *Student Member, IEEE,* Xuguang Lan, *Member, IEEE,* Nanning Zheng, *Fellow, IEEE,* and Pamela Cosman, *Fellow, IEEE*

***Abstract*—We propose a depth-assisted error concealment method for slice losses in intra frames of 2-D+depth encoded 3-D video. Intra frames in the depth sequence are offset from intra frames in the 2-D view sequence, such that the corresponding frames in the other sequence are not also intra mode. For slice losses in an intra frame of the 2-D sequence, motion vector (MV) candidates to conceal the lost macroblocks come from the co-located MVs in the depth frame as well as from MVs in the previous 2-D frame chosen based on motion information from the depth frame. Then the temporal smoothness of the depth sequence is used to filter out MV candidates. Finally, an enhanced distortion criterion is used for MV selection, based on contour information in the depth frame and boundary matching. Experimental results show that the proposed method provides improved performance over existing methods.**

***Index Terms*—Error concealment, intra frame, slice loss, 2-D+depth 3-D video.**

I. Introduction

2-D+DEPTH encoding of 3-D video shows promise due to its low bit rate compared with multi-view coding (MVC) techniques. In 2-D+depth 3-D video, a depth sequence associated with the 2-D video sequence is captured or computed, compressed, and then transmitted independently. At the decoder, depth-image-based rendering (DIBR) [1], [2] and other smart postprocessing techniques [3] like that can be used to reconstruct the 3-D video based on the 2-D video stream and the associated depth stream.

Similar to 2-D video techniques, all wired/wireless 3-D video applications over error-prone networks suffer packet losses during transmission. This can significantly degrade video quality. In the case of point-to-point communications, packet losses can often be handled by ARQ retransmissions, however for broadcast applications where there are many receivers, a lost packet cannot be re-transmitted, and therefore error concealment methods at the decoders are of prime importance. When the coded depth stream and 2-D stream of 2-D+depth 3-D video are transmitted independently, the packet loss problem for them can be solved separately with error concealment techniques of 2-D video. The dominant solutions at the decoder include spatial error concealment (SEC) which is mainly for intra mode (I frames of the H.26x standard), and temporal error concealment (TEC) which is mainly for inter modes (P, B frames of the H.264x standard). In TEC, the key point is to recover the motion information of the lost region to exploit the correlation between the current frame and the reference frame [4]. A well-known TEC method is the boundary matching algorithm (BMA) [5], which first estimates several candidate motion vectors (MVs) of the lost region from the surrounding regions, then adopts the optimal MV using a boundary matching (BM) distortion criterion [5]. The concealment is then done by implementing motion compensation (MC) from the reference frame, using the chosen MV. BM is originally designed for loss/erasure of isolated macroblock(MB). In real transmission, the packet/slice size is always more than one macroblock, for example, one MB row. The disadvantage of MB-based solutions for packet/slice error concealment is that the estimated MVs of the lost MBs are often not reliable [6]. So some BM-based solutions are designed specially for the consecutive MB loss problem [7]. All these BM-based TEC methods use the MV information of the current frame, which is only suitable for inter frames. As motion estimation is not implemented in intra frames, the dominant solution for packet losses in intra frames is SEC, such as spatial interpolation [8], [9]. However, most SEC methods perform less well than TEC, especially for consecutive MB losses in natural video sequences, because spatial interpolation methods induce blur. Some edge-based SEC methods [10], [11] can reduce blur; however, they include motion estimation (ME) at the decoder to recover the lost edges, which increases the computational complexity of the decoder. In this paper, we focus on a TEC solution for slice losses in intra frames.

For intra mode, the simplest TEC solution is frame copy which means to copy the co-located block from the previous frame. This only performs well for video sequences with small movement. A dominant TEC solution is to use BMA [5] with the co-located MVs in the previous frame as the MV candidates. However, for natural video, the motion information of the lost region may be more similar with that of spatial neighbor regions rather than that of temporal neighbor regions due to scene changes and object movement. Another dominant TEC method is to re-implement ME at the decoder.

Manuscript received August 15, 2012; revised May 5, 2013; accepted July 4, 2013. Date of current version June 4, 2014. This work was supported in part by the National Science Foundation under Grant CCF-1160832 in U.S., in part by the Program 973 under Grant 2012CB316400 and Grant 2010CB327902, and in part by NSFC under Grant 61175010 in China.

M. Yang, X. Lan, and N. Zheng are with the Institute of Artificial Intelligence and Robotics, Xi'an Jiaotong University, Xi'an 710049, China (e-mail: myang59@gmail.com; xglan@mail.xjtu.edu.cn; nnzheng@mail.xjtu.edu.cn).

P. Cosman is with the Department of Electrical and Computer Engineering, University of California at San Diego, La Jolla, CA 92093-0407 USA (e-mail: pcosman@ece.ucsd.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TBC.2014.2321676

However, the computational complexity of ME is generally considered to be unacceptable for decoding, even though it can be optimized as with the decoder motion vector estimation (DMVE) [12]. Another efficient TEC method is motion data hiding [13], which executes ME for consecutive intra frames at the encoder and then imperceptibly embeds MVs into the coefficients of other MBs in the same intra frame. If MB loss occurs in an intra frame at the decoder, the MV of this MB may be extracted from the current intra frame. However, this method relies on both the decoder and the encoder, and it requires extra computational complexity at the encoder and extra bits.

The loss problem in intra frames can be reconsidered in 2-D+depth 3-D video transmission, because there is a close relation of the motion information between the 2-D sequence and the depth sequence [14]. We previously checked the similarity of the MVs between the 2-D and depth sequences in [15], and reported that the correlation of MVs between a 2-D sequence and its corresponding depth sequence is usually strong with the correlation coefficient larger than 0.5. When the 2-D sequence and depth sequence are compressed and transmitted independently, the redundancy between the MVs of the 2 sequences can be utilized efficiently. This was originally considered for efficient 3-D video compression [16]. In recent years, it was utilized for error concealment of the frame/packet loss problem in 2-D + depth video and MVC+depth video, in which 2-D+depth 3-D technique is combined with MVC 3-D technique. It is reported that the packet loss problem of the depth sequence can be easily solved by utilizing either the information extracted from itself [17] or the associated information in the 2-D sequence [18], because the depth sequence only simply represents the basic structure of the corresponding 2-D sequence. However, depth information does not perform well when it is directly used for packet losses of the 2-D sequence in 2-D+depth video. Yan [19] first used the motion information of the depth sequence to optimize the conventional BMA method for the packet loss problem in inter frames with good results. Then he considered the frame loss problem in 2-D+depth video transmission [20]. Liu *et al.* [17] further considered the temporal smoothness of the depth sequence to improve the BM criterion for the lost region in inter frames. Chung [21] *et al.* extended the solution from 2-D+depth video to MVC+depth video. However, all these works focus on inter modes (P and B frames), and do not work for intra mode (I frame), because there is no MV information in either the 2-D sequence or the depth sequence for the intra frames.

In this paper, we consider slice losses for intra frames of the 2-D sequence in 2-D+depth encoded 3-D video transmission. We propose a TEC solution, named depth-assisted TEC, which avoids implementing ME at the decoder. Three kinds of information in the depth sequence are extracted and used for slice loss problem in intra frames of the 2-D sequence in 2-D+depth video, including MVs, contours, and context. Part of this paper was presented in [15], in which only the motion information of the depth sequence is used to solve the slice loss problem of 2-D sequence in 2-D+depth video. The novelties of our method compared with existing methods are:

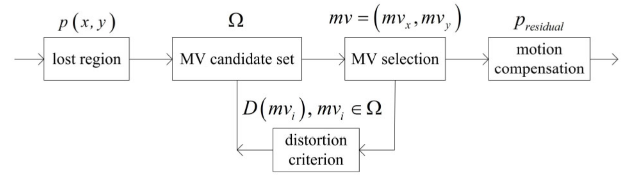

Fig. 1. Framework of TEC method.

1. we focus on slice losses in intra frames of 2-D+depth video. As far as we know, no existing work considers packet losses of intra frames, considering the correlation of MVs between the 2-D and depth sequences. [17], [19] mainly considered the loss problem in inter frames, which does not work for intra frames. 2. Unlike the MV candidate generating method of conventional BMA [5], we propose an offset MV candidate searching method based on the depth information, which is more effective especially for loss regions with a scene change or fast object movement. 3. Moreover, we design an enhanced distortion criterion for MV selection from the MV candidates based on BM [5] and DMVE [12], considering the unreliability of conventional BM method for slice losses.

The rest of this paper is organized as follows. In Section II, the proposed depth-assisted TEC is introduced. In Section III, we evaluate the proposed method, and conclude the paper in Section IV.

II. Depth-Assisted Temporal Error Concealment for Intra Frames

Fig. 1 shows the common framework of TEC, mainly including three parts: MV candidate generation for the lost region, MV selection from the candidate set and its distortion criterion, and motion compensation. For the lost region of an intra frame, the MV candidate set consists of the MVs of the co-located MBs in the previous frame in [5], or a continuous motion region in the previous frame in [12]. Then the optimal MV is adopted based on a distortion criterion. A well-known distortion measurement is BM [5], which is defined as the absolute difference between the external boundary of the lost region and the internal boundary of the predicting region in the previous frame. Finally, MC is implemented to recover the lost region. The principle is formulated in (1).

$$p(x, y) = p_r(x + mv_x, y + mv_y) + p_{residual}. \tag{1}$$

$p(x, y)$ denotes the pixel in the lost region and $p_r(x, y)$ denotes the pixel in the reference frame, where (x, y) is the top-left pixel coordinates of the lost region. $mv = \left(mv_x, mv_y\right)$ denotes the selected MV from the MV candidate set Ω by the distortion function $D(mv_i)$ $mv_i \in \Omega$, that is $mv = \arg\min_{mv_i\in\Omega} D(mv_i)$. $p_{residual}$ denotes the residual for the lost pixel. If the residuals for the lost region are also lost, the residuals when the co-located block in the previous frame is encoded can simply be used. In our scheme, we suppose the residuals are lost together with the MVs of the lost region, and will not consider the residual recovery. So the MV candidate set generation and the distortion criterion are the two key steps for TEC. The proposed depth-assisted TEC for intra slices of the 2-D sequence in 2-D+depth video tries to improve these two steps separately by the information extracted from the depth sequence in 2-D+depth video.

A. MV Candidate Generation

MV candidates are used to recover the lost region of the current 2-D frame from the previous 2-D frame. For slice losses in the intra frame of the 2-D sequence, MV candidates can be extracted from both the corresponding depth frame and the previous 2-D frame.

1) Depth-Offset for MV Extraction from the Depth Frame: MVs in the depth sequence can be utilized to recover the MV of lost region in the 2-D sequence because of the similarity of the motion information between the 2-D and depth sequences. We have reported the similarity of the MVs between the 2 sequences for some test 3-D sequences in [15]. However, for an intra frame in the 2-D sequence, typically the corresponding frame in the depth sequence is also intra mode, so there is no MV information from it that can be used. We impose a temporal offset to the intra frames of the depth sequence. We use G to denote the GOP length and G_{offset} to denote the offset $\left(0 < G_{offset} < G\right)$. The offset is imposed on the depth sequence starting from the 2nd GOP. That means the size of the 1st GOP in the depth sequence is $G - G_{offset}$, and the size of the other GOPs is G. Then, except for the first frame, each frame in the depth sequence that corresponds to an intra frame in the 2-D sequence is inter coded. In our experiments, the GOP size of the sequence is set to $G = 4$ to generate more intra frames in the sequence for the experiment. The offset parameter is $G_{offset}= 1$. That means the format of the 2-D sequence is "IPPPIPPPIPPP…" and the format of the depth sequence is "IPPIPPPIPPPI…".

For an MB loss in the 2-D sequence, all the co-located and neighboring MVs of the corresponding depth frame are exploited as MV candidates. For a macroblock MB_{lost}^{2D} in the lost slice, the co-located MB and its eight neighbors in the inter coded frame n of the depth sequence are denoted MB_i^{depth} $(i = 0, 1, \ldots, 8)$. Their MVs are extracted and denoted mv_i^{depth} $(i = 0, 1, \ldots, 8)$. All of them are used as MV candidates for the lost MB in the 2-D sequence.

2) Offset MV Candidates in the Previous 2-D Frame: Because the depth sequence only represents the basic structure of the corresponding 2-D sequence,the MV of the lost region can not be recovered well only by the MV candidates from the depth sequence. In this case, intra slice concealment of the 2-D video can also be done using the MVs of the co-located MBs in the previous frame as MV candidates. However, due to a scene change or object movement, the MVs of the co-located MBs may not be optimal. Instead of using these ones, we try to find the offset region in the previous frame which corresponds to the lost region in the current frame using the depth motion information. For a lost macroblock MB_{lost}^{2D} whose coordinate is (x_{lost}, y_{lost}), the MV of the co-located macroblock MB^{depth} in the corresponding depth frame is denoted mv^{depth}. Due to the motion similarity between the 2-D and depth sequences, mv^{depth} is likely similar to the MV of MB_{lost}^{2D}. So we use mv^{depth} to find the MB (denoted MB_{offset}^{2D}) in the previous frame

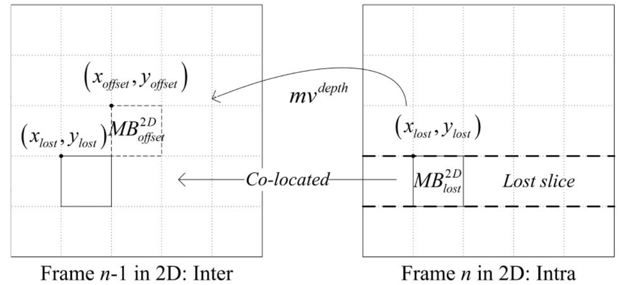

Fig. 2. Offset MV candidate generation.

that corresponds to MB_{lost}^{2D}. The coordinates (x_{offset}, y_{offset}) of MB_{offset}^{2D} are calculated as follows:

$$\begin{cases} x_{offset} = \left\lfloor \frac{x_{lost}+mv_x^{depth}+N/2}{N} \right\rfloor \cdot N, & mv_x^{depth} \geq 0 \\ x_{offset} = \left\lfloor \frac{x_{lost}+mv_x^{depth}-N/2}{N} \right\rfloor \cdot N, & mv_x^{depth} < 0 \end{cases} \tag{2}$$

$$\begin{cases} y_{offset} = \left\lfloor \frac{y_{lost}+mv_y^{depth}+N/2}{N} \right\rfloor \cdot N, & mv_y^{depth} \geq 0 \\ y_{offset} = \left\lfloor \frac{y_{lost}+mv_y^{depth}-N/2}{N} \right\rfloor \cdot N, & mv_y^{depth} < 0 \end{cases} \tag{3}$$

where $\lfloor \cdot \rfloor$ is the floor function. This is depicted in Fig. 2, where the size of an MB is $N \times N$. Then the MVs of MB_{offset}^{2D} and its eight neighbors in the previous 2-D frame are used as MV candidates instead of the co-located ones. These MVs are denoted $mv_{offset,i}^{2D}$ $(i = 0, 1, \ldots, 8)$, where i indexes the MB and its 8 neighbors. Finally, the MV candidate set Ω for MB_{lost}^{2D} is

$$\begin{aligned} &\Omega\left(MB_{lost}^{2D}\right) \\ &\triangleq \left\{ mv_i^{depth},\ mv_{offset,i}^{2D},\ (i = 0, 1, \ldots, 8)\,;\ (0, 0) \right\} \end{aligned} \tag{4}$$

where $(0, 0)$ denotes the zero motion case.

B. Region Size Selection

TEC technique in H.26x is commonly implemented based on macroblocks of size 16×16. However, this will not be optimal for some cases, such as the lost regions around object contours, adjacent slice losses, etc.

1) Pixel-Level MB Segmentation Based on Depth Contours: For regions around object contours that are not homogeneous, different part of the region may have different movement. So it is necessary to segment them as foreground and background. In [17], MBs in inter frames are segmented based on subblock (8×8) size according to the depth values. We segment the lost MBs in the intra frames at the pixel level, which is more precise, utilizing the contour information of the depth frame.

First, we extract contour information from the slices of the depth frame that correspond to the lost intra frame slices. For a pixel in the depth sequence corresponding to the lost region, we denote its variance over a 3×3 sliding window as v. If $v > v_0$, then the pixel is declared an edge pixel. The threshold was chosen to be $v_0 = 100$ in our experiment. The contour region consists of all pixels whose variance exceeds the threshold as well as all internal holes in such

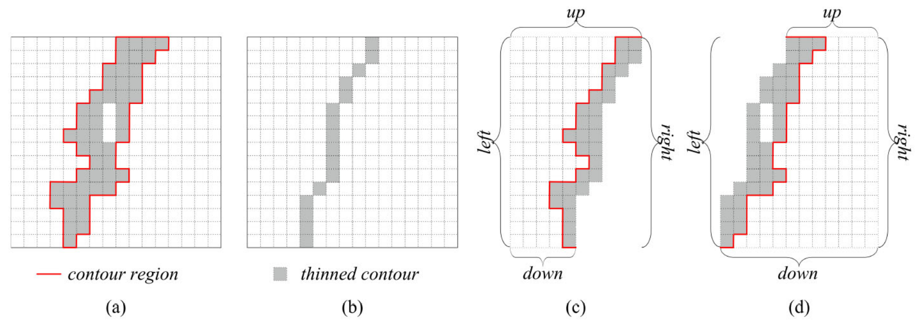

Fig. 3. Pixel-level MB segmentation. (a) 16 × 16 MB with contour region. (b) Thinned contour. (c) Segmented left region. (d) Segmented right region.

regions, as shown in Fig. 3(a). If some pixels in both top and bottom pixel rows of an MB are included in the contour region, then this MB is pronounced a left-right segmented one. Then, we thin the extracted contour regions to 1-pixel wide by taking their central pixels. Fig. 3 shows the left-right segmentation case. Fig. 3(b) shows the 16 x 16 MB, where a contour crosses from top to bottom. Fig. 3(c) shows the left segment, including the pixels of the thinned contour and all pixels to the left of it, and Fig. 3(d) shows the right segment, which also includes the thinned contour pixels and all the pixels to the right. For each segmented part, the whole region is denoted *region*, and the four inner boundary pixel sets are denoted *up*, *down*, *left*, and *right*. Note that the upper left pixel in Fig. 3(c) is included in set *left* and in set *up*, and the lower left pixel is included in set *left* and set *down*, etc. These pixel sets will be used in the boundary matching distortion calculation. Macroblocks crossed by a contour are selected for segmentation except for the following two cases.

Case 1: The MB is almost static comparing with the background. As the background in our test videos is always either static or moving only slightly, we simply check if the foreground MV is close to 0. Note that if the background is not static, the relative movement of the MB towards the background should be calculated instead. Using the MV of the co-located MB in the depth frame, we calculate the average values of the x and y components of the MVs of the four subblocks (in the JM software, ME is implemented on subblocks). We sum their absolute values. If this is lower than a predefined threshold 42, then we declare this a static MB, and do not segment it. The threshold was trained over 10 test videos. For our simulations, to reduce complexity, this threshold was not optimized for any particular video stream; it might be possible to improve the performance by training the threshold for a given video or establishing a model to adjust this parameter.

Case 2: One of the segmented parts is too small for reliable boundary matching. If the number of pixels in sets $up + down$ shown in Fig. 3(c) is less than 8, or if the number of pixels in sets $up + down$ shown in Fig. 3(d) is less than 8, the MB will not be segmented.

In our experiment, the distortion criterion for the current segmented MB will be calculated from the segmented left and right parts separately. Then two optimal MBs, denoted MB_l and MB_r, in the previous 2-D frame will be chosen and

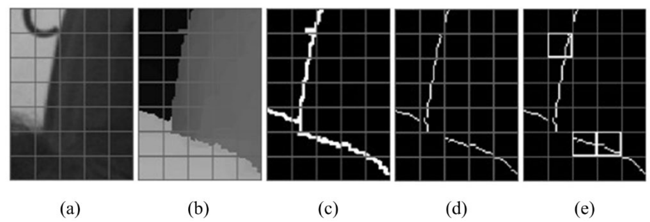

Fig. 4. Example of MB Segmentation. (a) 2-D frame. (b) Corresponding depth frame. (c) Extracted contour region. (d) Thinned contour. (e) Segmented MBs. In this example, the 2nd slice from the top and the 2nd slice from the bottom are lost.

used for motion compensation. The final concealed MB is calculated using

$$MB_{mc}(i,j) = MB_l(i,j) \cdot mask_l(i,j) + MB_r(i,j) \cdot mask_r(i,j) \tag{5}$$

where $i, j = 1, 2, \ldots, N$, and $mask_l$ is an N×N array defined in (6).

$$mask_l(i,j) = \begin{cases} 1, & (i,j) \in region/contour \\ 1/2, & (i,j) \in contour \\ 0, & others \end{cases} \tag{6}$$

$mask_r$ is similarly defined for the rightside segment. Similar to left-right segmentation, the MB will be segmented into top and bottom regions if it is crossed by a contour from left to right. The procedure is similar. Fig. 4 shows an example of our pixel-level MB segmentation, including 3 cases: left-right segmented MB, top-bottom segmented MBs, and non-segmented MB.

2) Adjacent Slice Losses: MBs which are not segmented are concealed based on a whole MB. In this case, *region* denotes the $N \times N$ MB, and *up*, *down*, *left*, and *right* denote the 4 inner N-pixel boundaries of the MB. It is possible that two adjacent slices in the same intra frame are lost. For the uppermost slice of the two, only the top boundary (*up*) is available for BM. For the lower slice, only the lower boundary (*down*) is available for BM. This makes it less reliable to select the optimal MV for the lost region from the MV candidate set Ω. In this case, we combine the two lost slices and conceal them together. Then both the *up* boundary and *down* boundary can be used for BM. All error concealment procedures are implemented based on the combined MB, such as the MB segmentation. If the combined MB is not segmented, *region* denotes the $2N \times N$ combined MB, *up* and *down* denote the upper and lower inner N-pixel boundaries of the combined MB, and *left* and *right* denote the left and right inner $2N$-pixel boundaries of the combined MB respectively. The MV candidate set of the combined MB is the union of the two MV candidate sets.

C. Distortion Criterion Enhancement

Given an MV candidate set Ω, the next step is to select the optimal MV from it using a distortion criterion. When a slice is taken to be a horizontal row of MBs, it may be unreliable to adopt the conventional BM distortion criterion, especially when the number of MV candidates is large or the spatial correlation is low. That is because only the upper and lower boundary pixels can be used to calculate the distortion, since the left and right boundaries of the lost MB are adjacent to other lost MBs. We try to enhance the distortion criterion based on BM and the information extracted from the depth frame.

1) MV Candidate Filtering Based on Depth Temporal Smoothness: We first consider the temporal smoothness of the depth sequence. One of the reasons why BM may not be reliable is that the boundary pixels do not represent the whole lost region well, when the spatial texture is complex. We use the depth to evaluate the temporal smoothness of the lost region in the 2-D sequence. For an MV candidate $mv \in \Omega$, the depth temporal smoothness *SAD* (sum of absolute differences) is defined as

$$SAD_{region}^{depth}(mv) = \frac{16}{size(region)} \cdot \sum_{(x,y)\in region} \left| P_n^{depth}(x,y) - P_{n-1}^{depth}\left(x+mv_x, y+mv_y\right) \right| \tag{7}$$

where $size\,(region)$ stands for the number of pixels in *region*, which is defined in Section II-B; P_n^{depth} and P_{n-1}^{depth} denote the pixel values of depth frames n and $n-1$ respectively.

We use it to filter out bad MV candidates from the original MV candidate set Ω. For a lost region, its motion information is likely to be similar to that of the same object or same part in an object. The depth value in the same object or the same part of an object is also likely to be similar. So the depth smoothness (6) can be used to filter out bad MV candidates from Ω. If $SAD_{region}^{depth}(mv)$ is above a threshold SAD_T^{depth}, the MV candidate mv will be rejected. The threshold is trained as $SAD_T^{depth} = 50$ in our experiment.

2) Enhanced BM-Based Distortion Function: The distortion function is established for each *region* separately. BM [5] was originally designed for loss of isolated MBs. For slice loss problem, it may still be unreliable to select the optimal MV from the MV candidate set by BM, because only the boundaries above and below can be used. That is, since the slice is a horizontal row of MBs, the neighbor MBs above and below are usually available for BM, but the left and right neighbors are generally not available (although if they were concealed already, then the concealment MBs are available for BM). Especially when the spatial correlation is low, or the MB is left-right segmented, the *up* and *down* pixel sets are much smaller for BM. In this section, we try to enhance the distortion function.

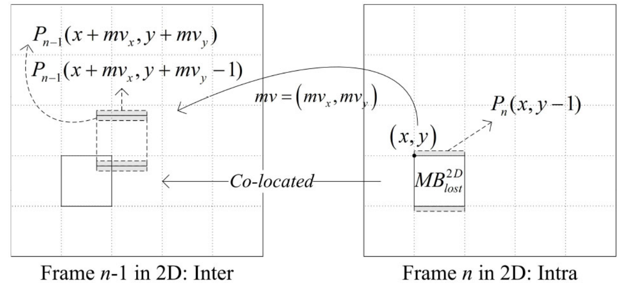

Fig. 5. Enhanced BM for UP and DOWN boundaries.

As the upper and lower boundaries of the lost MBs are always reliable for BM, we enhance the distortion function for them. Derived from [5] and [12], we consider the outer boundary of the *region* in the current frame n, and both the outer and inner boundaries of the predicted region in the previous frame $n-1$, as shown in Fig. 5, where we take the MB for example. Then for a MV candidate $mv = \left(mv_x, mv_y\right) \in \Omega$, the distortions for the upper and lower boundaries are calculated as

$$\begin{aligned} SAD_{up}^{2D}(mv) = \sum_{(x,y)\in up} \Big[& \left| P_n(x, y-1) - P_{n-1}\left(x+mv_x, y+mv_y\right) \right| \\ & + \left| P_n(x, y-1) - P_{n-1}\left(x+mv_x, y+mv_y-1\right) \right| \Big] \end{aligned} \tag{8}$$

$$\begin{aligned} SAD_{down}^{2D}(mv) = \sum_{(x,y)\in down} \Big[& \left| P_n(x, y+1) - P_{n-1}\left(x+mv_x, y+mv_y\right) \right| \\ & + \left| P_n(x, y+1) - P_{n-1}\left(x+mv_x, y+mv_y+1\right) \right| \Big] \end{aligned} \tag{9}$$

where *up* and *down* are the pixel sets defined in Section II-B, and P_n and P_{n-1} are the pixel values of the 2-D frames n and $n-1$ respectively.

For the region *region*, the left and right neighboring MBs may be lost or concealed. If they are concealed, we use the conventional BM [5] directly. That is

$$SAD_{left}^{2D}(mv) = \sum_{(x,y)\in left} \left| P_n(x-1, y) - P_{n-1}(x+mv_x, y+mv_y) \right| \tag{10}$$

$$SAD_{right}^{2D}(mv) = \sum_{(x,y)\in right} \left| P_n(x+1, y) - P_{n-1}(x+mv_x, y+mv_y) \right| \tag{11}$$

where *left* and *right* are the defined pixel boundaries of *region* in Section II-B.

Finally, for a MV candidate $mv \in \Omega$, the distortion criterion is defined as follows

$$\begin{aligned} SAD(mv) = {} & \alpha SAD_{up}^{2D}(mv) + \beta SAD_{down}^{2D}(mv) \\ & + \gamma SAD_{left}^{2D}(mv) + \delta SAD_{right}^{2D}(mv) \end{aligned} \tag{12}$$

The parameters α, β, γ, and δ are related to *region*, and determined like this:

a) If the MB is top-bottom segmented, for the upper part $\alpha = 1$ and $\beta = 0$, and for the lower part $\alpha = 0$ and $\beta = 1$.

b) If the left neighbor of the current MB is not concealed, γ is set to 0. If the right neighbor of the current MB is not concealed, δ is set to 0.

c) For other cases, all parameters are set to 1.

III. Experiments

We implement the proposed scheme based on H.264/AVC codec reference software (JM16.2). The motion search range is 32 and fast full motion search mode in JM16.2 is adopted. The quantization parameter (QP) is set to 28 and 40, which correspond to higher and lower bit rate cases respectively. The slice is an MB row. Six 200-frame 2-D+depth sequences are tested: Cafe (960 $\times$ 512), Newspaper (1024 $\times$ 768), Mobile (720 $\times$ 512), Balloons (1024 $\times$ 768), Bookarrival (1024 $\times$ 768), and Orbi (720 $\times$ 512). Here "2-D" sequence refers to either the left view or the right view sequence of the rendered stereoscopic 3-D video and all the "depth" sequences used are manually generated. Then GOP size is set as 4 to generate more intra frames for simulation. There are 50 intra frames in each test sequence. Random packet losses in intra frames are simulated by modified "rtp-loss" tool of the JM16.2 software. The slice is defined as an MB row. In this scenario, it is also possible that 2 or more adjacent slices are lost together. It is reported in [17] that the result of TEC without utilizing any information of 2-D video is very good if packet losses happen in the depth sequence, because the context of depth video is much simpler than that of 2-D video. So we only focus on the packet losses of the 2-D sequence in 2-D+depth video.

As far as we know, no existing work considers exploiting the correlation between the 2-D sequence and the depth sequence for the problem of intra slice losses in 2-D+depth video. The dominant solutions for this problem are those for 2-D video, including spatial interpolation [9], frame copy, motion re-estimation at the decoder [12], and boundary matching [5]. The proposed method is compared against those 4 solutions. The detailed settings are as follows. 1) Spatial interpolation (denoted "Spatial"): We recover each pixel in the lost slice by linear interpolation using the nearest pixels in the top and bottom neighbor MBs. 2) Frame copy (denoted "Copy"): The co-located MBs of the previous decoded frame replace the MBs of the lost slice. 3) Decoder ME combined with BMA (denoted "ME+BMA"): For lost MB, implement motion estimation at the decoder for the neighbor MBs above and below to determine MV candidates, and then conceal the lost region by BMA. This is similar to DMVE [12]. 4) BMA: Use the MV candidates of the co-located MB and its 8 neighbors in the previous frame. 5) DMVE: Redo motion estimation at the decoder for the lost region [12]. The upper pixel line and the lower pixel line of the lost region are used for distortion measurement. We provide the PSNR result of the Y component (denoted "SNRY") averaged over all the intra frames of the 2-D sequence (left/right view in rendered stereoscopic video) only, the PSNR result of the Y component averaged over both the intra frames and the inter frames of the 2-D sequence, and the decoding time of the 2-D video sequence. The packet loss rate (PLR) in all the intra frames is set to 5%, 10%, 15%, 20%, and 25%, in which the first IDR frame of the 2-D sequence is error-free. Each method is averaged over 20 realizations.

TABLE I
Average Decoding Times (ms) for PLR = 10% Case

	Spatial	Copy	ME+BMA	BMA	DMVE	Proposed
TIME	140	126	2182	174	1349	277

We first take the result of all intra frames of the input sequences. One important issue of video decoding is the decoding time, which is highly related with the computational complexity of the decoder. The computational complexity of the proposed method is expected to be similar with that of "Spatial", "Copy", and "BMA", all of which do not implement motion re-estimation at the decoder, while the computational complexity of "ME+BMA" and "DMVE" might be much higher because of motion re-estimation at the decoder. We measured the software decoding times. The average decoding times of those methods for PLR=10% for the CAFE sequence are shown in Table I. For other PLR cases, the results are similar. The "SNRY" results for QP=28 case are shown in Table II. Compared with "Spatial" and "Copy", the SNRY result of the proposed method is much better for all the test sequences, and the decoding time of the proposed method is 1.9-2.2 times higher. Compared with "ME+BMA" and "DMVE", the SNRY result of the proposed method is better for most test sequences. One reason for "DMVE" is that only the upper pixel line and lower pixel line of the lost region can be used for distortion measurement in slice loss problem. The result for the Balloons sequence is an exception, where the result of "ME+BMA" and "DMVE" is comparable with or slightly better than the proposed method. This is because the spatial correlation is higher than the temporal correlation in this sequence. However, the decoding time of "ME+BMA" and "DMVE" that implements motion estimation at the decoder is an order of magnitude higher than that of the proposed method. Compared with "BMA", which is the dominant TEC solution for slice losses in intra frames, the proposed method improved the result by up to 0.8 dB for PLR=25%. The decoding time of the proposed method is 50% higher than that of "BMA". We additionally show the "SNRY" results for QP=40 case in Table III, whose bit rate is lower than that for QP=28. It is shown that the improvement of the proposed method is slightly lower than that when QP=28, which is up to 0.6 dB for PLR=25%.

Another important issue for video coding is the error propagation to the subsequent inter frames if the slice losses occur in an intra frame. Minimizing the average distortion of only intra frames does not always lead to an optimized quality of the whole sequence, so we show the average "SNRY" of the whole sequence for QP=28 case in Table IV. The trends are similar to those in Table I.

Note that not all the sequences have high quality depth information. It is known that the quality of Cafe, Mobile and Orbi depth sequences is better than that of Newspaper, Balloons and Bookarrival. So the proposed method for Cafe, Mobile and Orbi is more effective than for the other three. This result

TABLE II
AVERAGE SNRY (dB) RESULT OF ONLY INTRA FRAMES (QP = 28)

	Cafe						Newspaper					
PLR	Spatial	Copy	ME+BMA	BMA	DMVE	Proposed	Spatial	Copy	ME+BMA	BMA	DMVE	Proposed
0%	42.27						39.92					
5%	38.12	40.49	41.54	41.53	41.49	**41.71**	35.47	37.76	39.31	39.33	39.28	**39.44**
10%	36.06	39.23	40.87	40.94	40.86	**41.23**	33.46	36.30	38.75	38.77	38.71	**39.02**
15%	34.76	38.21	40.19	40.33	40.17	**40.66**	32.09	35.13	38.30	38.33	38.31	**38.58**
20%	33.79	37.52	39.75	39.93	39.77	**40.37**	30.89	34.17	37.76	37.90	37.80	**38.18**
25%	31.78	36.40	39.11	39.14	39.15	**39.85**	28.83	33.23	37.35	37.36	37.33	**37.67**
	Mobile						Balloons					
PLR	Spatial	Copy	ME+BMA	BMA	DMVE	Proposed	Spatial	Copy	ME+BMA	BMA	DMVE	Proposed
0%	39.95						42.45					
5%	35.01	38.79	39.14	39.25	39.16	**39.44**	38.66	40.60	41.51	41.36	41.50	**41.57**
10%	32.36	37.82	38.58	38.57	38.60	**38.92**	36.44	39.07	40.74	40.40	**40.76**	40.66
15%	30.47	36.99	37.96	38.06	37.99	**38.44**	35.05	37.98	40.00	39.57	**40.05**	39.97
20%	29.49	36.26	37.40	37.46	37.45	**38.02**	34.29	37.21	**39.52**	39.09	39.49	39.47
25%	27.51	35.46	36.89	36.86	36.93	**37.70**	32.42	36.10	38.66	38.09	**38.69**	38.60
	Bookarrival						Orbi					
PLR	Spatial	Copy	ME+BMA	BMA	DMVE	Proposed	Spatial	Copy	ME+BMA	BMA	DMVE	Proposed
0%	39.56						39.86					
5%	35.02	37.53	38.79	38.82	38.76	**38.98**	34.89	37.47	38.61	38.57	38.64	**38.82**
10%	34.17	37.09	38.31	38.35	38.29	**38.6**	34.06	36.84	38.26	38.19	38.27	**38.63**
15%	32.95	36.19	37.68	37.69	37.66	**38.11**	33.11	35.97	37.49	37.36	37.51	**38.06**
20%	31.84	35.08	37.14	37.17	37.13	**37.76**	31.73	34.89	37.1	36.94	37.13	**37.85**
25%	31.06	34.75	36.55	36.59	36.50	**37.28**	30.99	34.52	36.61	36.5	36.65	**37.48**

TABLE III
AVERAGE SNRY (dB) RESULT OF ONLY INTRA FRAMES (QP = 40)

	Cafe						Newspaper					
PLR	Spatial	Copy	ME+BMA	BMA	DMVE	Proposed	Spatial	Copy	ME+BMA	BMA	DMVE	Proposed
0%	36.25						33.28					
5%	33.47	34.54	35.38	35.40	35.36	**35.68**	30.79	31.52	32.55	32.58	32.53	**32.73**
10%	31.38	33.27	34.64	34.66	34.63	**34.93**	29.68	30.67	32.18	32.24	32.20	**32.52**
15%	30.12	32.18	34.05	34.13	34.01	**34.51**	27.93	29.48	31.76	31.87	31.77	**32.21**
20%	28.95	31.5	33.62	33.78	33.57	**34.3**	25.65	28.32	31.41	31.52	31.38	**31.9**
25%	27.64	30.89	32.9	32.95	32.86	**33.57**	24.21	27.39	30.98	31.03	30.96	**31.47**
	Mobile						Balloons					
PLR	Spatial	Copy	ME+BMA	BMA	DMVE	Proposed	Spatial	Copy	ME+BMA	BMA	DMVE	Proposed
0%	31.6						35.25					
5%	28.14	30.1	30.92	31.05	30.89	**31.2**	32.75	33.67	34.57	34.43	34.56	**34.62**
10%	26.85	29.85	30.28	30.27	30.26	**30.51**	31.63	32.51	34.05	33.86	33.98	**34.1**
15%	25.66	28.79	29.76	29.88	29.71	**30.25**	29.8	31.84	**33.61**	33.2	33.55	33.58
20%	23.87	28.21	29.1	29.21	29.02	**29.84**	27.91	30.97	32.93	32.51	**32.95**	32.94
25%	22.94	27.59	28.79	28.84	28.69	**29.52**	26.52	30.03	32.47	31.97	32.38	**32.5**
	Bookarrival						Orbi					
PLR	Spatial	Copy	ME+BMA	BMA	DMVE	Proposed	Spatial	Copy	ME+BMA	BMA	DMVE	Proposed
0%	33.53						33.42					
5%	30.17	31.44	32.65	32.69	32.64	**32.85**	29.73	31.24	32.41	32.37	32.44	**32.69**
10%	29.53	30.7	32.07	32	32.04	**32.34**	28.65	30.58	31.82	31.74	31.79	**32.18**
15%	28.5	29.89	31.46	31.61	31.41	**32.03**	28.07	29.87	31.26	31.25	31.23	**31.77**
20%	27.92	29.01	31.03	31.14	31.04	**31.71**	27.11	29.46	30.79	30.7	30.76	**31.3**
25%	26.98	28.12	30.38	30.47	30.41	**31.08**	26.62	28.89	30.32	30.29	30.37	**31.04**

can also be predicted from the motion similarity between 2-D and depth sequences in [15]. It is shown that those sequences that perform best always have better motion similarity between color and depth sequences. This suggests that, if the quality of depth sequences is improved in the future, the proposed method can be more effective.

Another small advantage of the proposed method lies in the bit rate stability of the coded streams. The output bit rate of coded 2-D+depth sequences is the sum of the rates of the two separate sequences. For video coding, the total bit rates fluctuate sharply between the intra and inter frames, because the bit rates for intra frames are much higher than for inter frames. In the proposed method, intra frames in the 2-D sequence are offset from intra frames in the depth sequence. This results in shifting part of the bit rate from the peak to the trough. The experimental results are shown in Fig. 6, where "0 offset" stands for conventional solution without offset between the 2-D sequence and the depth sequence, "1 offset" and "2 offset" stand for our solution with offset $G_{offset} = 1$ and $G_{offset} = 2$ between the depth sequence and 2-D sequence respectively. It is shown that the proposed method benefits the stability of the output stream, reducing the peak bit rate by 15%-25%.

TABLE IV
AVERAGE SNRY (dB) RESULT OF BOTH INTRA AND INTER FRAMES (QP = 28)

	Cafe						Newspaper					
PLR	Spatial	Copy	ME+BMA	BMA	DMVE	Proposed	Spatial	Copy	ME+BMA	BMA	DMVE	Proposed
0%	41.90						39.61					
5%	37.98	40.14	41.13	41.17	41.09	**41.38**	35.36	37.45	38.93	38.98	38.92	**39.06**
10%	35.97	38.87	40.42	40.60	40.42	**40.92**	33.40	35.96	38.30	38.38	38.29	**38.58**
15%	34.72	37.83	39.73	39.97	39.74	**40.35**	32.06	34.80	37.82	37.93	37.81	**38.15**
20%	33.76	37.18	39.29	39.60	39.28	**40.07**	30.87	33.82	37.24	37.48	37.29	**37.75**
25%	31.77	36.10	38.65	38.89	38.70	**39.59**	28.81	32.91	36.76	36.85	36.77	**37.22**
	Mobile						Balloons					
PLR	Spatial	Copy	ME+BMA	BMA	DMVE	Proposed	Spatial	Copy	ME+BMA	BMA	DMVE	Proposed
0%	39.76						42.04					
5%	34.94	38.47	38.90	39.02	38.89	**39.22**	38.47	40.29	41.14	41.00	41.17	**41.22**
10%	32.33	37.41	38.25	38.29	38.26	**38.69**	36.36	38.84	**40.41**	40.10	40.40	40.39
15%	30.47	36.61	37.65	37.80	37.67	**38.25**	35.01	37.80	39.70	39.30	39.74	**39.73**
20%	29.49	35.79	37.05	37.15	37.09	**37.80**	34.27	37.08	**39.27**	38.86	39.2	**39.27**
25%	27.52	35.02	36.55	36.61	36.60	**37.43**	32.44	36.01	**38.49**	37.90	38.47	38.48
	Bookarrival						Orbi					
PLR	Spatial	Copy	ME+BMA	BMA	DMVE	Proposed	Spatial	Copy	ME+BMA	BMA	DMVE	Proposed
0%	39.27						39.57					
5%	34.98	37.54	38.55	38.59	38.51	**38.79**	34.78	37.46	38.37	38.34	38.36	**38.59**
10%	34.15	36.97	38.12	38.17	38.13	**38.43**	33.96	36.78	37.92	37.86	37.89	**38.26**
15%	32.88	36.17	37.45	37.47	37.42	**37.94**	33.02	35.9	37.25	37.15	37.22	**37.77**
20%	31.67	35.07	36.93	36.91	36.96	**37.62**	31.69	34.76	36.93	36.81	36.96	**37.62**
25%	30.92	34.69	36.46	36.5	36.40	**37.14**	30.89	34.51	36.4	36.3	36.44	**37.25**

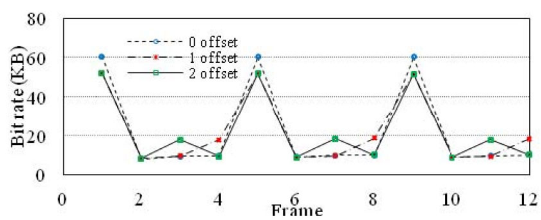

Fig. 6. Bit rate stability of the coded 2-D+depth streams (Newspaper).

IV. CONCLUSION

In this paper, we proposed a depth-assisted temporal error concealment method for slice loss in intra frames of 2-D+depth video. The correlation between the 2-D sequence and the depth sequence is exploited to assist error concealment based on the conventional TEC method. Experimental results show that the proposed method has improved PSNR, reduced computational complexity at the decoder, and more stable output bit rate.

REFERENCES

[1] C. Fehn, "Depth-image-based rendering (DIBR), compression and transmission for a new approach on 3DTV," in *Proc. 11th SPIE Stereoscopic Displays Virtual Reality Syst.*, San Jose, CA, USA, Jan. 2004, pp. 93–104.

[2] C. Zhu, Y. Zhao, L. Yu, and M. Tanimoto, *3D-TV System with Depth-Image-Based Rendering: Architecture, Techniques and Challenges*. New York, NY, USA: Springer, 2013.

[3] Y. Zhao, C. Zhu, Z. Chen, D. Tian, and L. Yu, "Boundary artifact reduction in view synthesis of 3D Video: From perspective of texture-depth alignment," *IEEE Trans. Broadcast.*, vol. 57, no. 2, pp. 510–522, Jun. 2011.

[4] J. Zhou, B. Yan, and H. Gharavi, "Efficient motion vector interpolation for error concealment of H.264/AVC," *IEEE Trans. Broadcast.*, vol. 57, no. 1, pp. 75–80, Mar. 2011.

[5] W. Lam, A. Reibman, and B. Liu, "Recovery of lost or erroneously received motion vectors," in *Proc. ICASSP*, Minneapolis, MN, USA, Apr. 1993, pp. 417–420.

[6] H. Gao, J. Tham, W. Lee, and K. Goh, "Slice error concealment based on size-adaptive SSIM matching and motion vector outlier rejection," in *Proc. ICASSP*, Prague, Czech Republic, May 2011, pp. 1809–1812.

[7] X. Qian, G. Liu, and H. Wang, "Recovering connected error region based on adaptive error concealment order determination," *IEEE Trans. Multimedia*, vol. 11, no. 4, pp. 683–695, Jun. 2009.

[8] S. Huang and S. Kuo, "Optimization of hybridized error concealment for H.264," *IEEE Trans. Broadcast.*, vol. 54, no. 3, pp. 49–516, Sep. 2008.

[9] B. Yan, H. Gharavi, and B. Hu, "Pixel interlacing based video transmission for low-complexity intra-frame error concealment," *IEEE Trans. Broadcast.*, vol. 57, no. 2, pp. 253–257, Jun. 2011.

[10] M. Ma, O. Au, S. Gary, and M. Sun, "Edge-directed error concealment," *IEEE Trans. Circuits Syst. Video. Technol.*, vol. 20, no. 3, pp. 382–395, Mar. 2010.

[11] S. Hisa,"An edge-oriented spatial interpolation for consecutive block error concealment," *IEEE Signal Process. Lett.*, vol. 11, no. 6, pp. 577–580, Jun. 2004.

[12] J. Zhang, J. Arnold, and M. Frater, "A cell-loss concealment technique for MPEG-2 coded video," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 10, no. 6, pp. 659–665, Jun. 2000.

[13] S. Chen and H. Leung, "A temporal approach for improving intra-frame concealment performance in H.264/AVC," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 19, no. 3, pp. 422–426, Mar. 2009.

[14] C. Hewage, S. Worrall, S. Dogan, and A. Kondoz, "A novel frame concealment method for depth maps using corresponding colour motion vectors," in *Proc. 3DTV-CON*, Istanbul, Turkey, May 2008, pp. 149–152.

[15] M. Yang, Y. Yang, and P. Cosman, "Depth-assisted error concealment for intra frame slices in 3D video," in *Proc. ICIP*, Orlando, FL, USA, Sep.–Oct. 2012.

[16] S. Grewntsch and E. Miiller, "Sharing of motion vectors in 3D video coding," in *Proc. ICIP*, Singapore, Oct. 2004.

[17] Y. Liu, J. Wang, and H. Zhang, "Depth image-based temporal error concealment for 3-D video transmission," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 4, pp. 600–604, Apr. 2010.

[18] C. Hewage and M. Martini, "Joint error concealment method for backward compatible 3D video transmission," in *Proc. IEEE VTC*, Budapest, Hungary, May 2011.

[19] B. Yan, "A novel H.264 based motion vector recovery method for 3D video transmission," *IEEE Trans. Consum. Electron.*, vol. 53, no. 4, pp. 1546–1552, Nov. 2007.

[20] B. Yan and J. Zhou, "Efficient frame concealment for depth image based 3D video transmission," *IEEE Trans. Multimedia*, vol. 14, no. 3, pp. 936–941, Jun. 2012.

[21] T. Chung, S. Sull, and C. Kim, "Frame loss concealment for stereoscopic video plus depth sequences," *IEEE Trans. Consum. Electron.*, vol. 57, no. 3, pp. 1336–1344, Aug. 2011.

Meng Yang (S'12) received the B.S. degree in electrical engineering from Xi'an Jiaotong University (XJTU), Xi'an, China, in 2008. He is currently pursuing the Ph.D. degree from the School of Electronic and Information Engineering, XJTU. He was a Visiting Scholar at the Department of Electrical and Computer Engineering, University of California at San Diego, La Jolla, CA, USA, from 2011 to 2012. His current research interests include video coding, processing, and communication.

Xuguang Lan (M'06) received the Ph.D. degree in pattern recognition and intelligent systems from Xi'an Jiaotong University (XJTU), Xi'an, China, in 2005. After visiting Ecole Centrale de Lyon for five months in France as a Visiting Researcher, he joined the Department of Computer Science and Technology to pursue the Post-Doctoral Research at XJTU, in 2005. He is currently a Professor with XJTU. His current research interests include image/video coding, processing, and communication, P2P technology, media content analysis, and VLSI design.

Nanning Zheng (SM'93-F'06) received the B.S. and M.S. degrees in information and control engineering from Xi'an Jiaotong University (XJTU), Xi'an, China, in 1975 and 1981, respectively, and the Ph.D. degree in electrical engineering from Keio University, Yokohama, Japan, in 1985. In 1975, he joined XJTU, where he is currently a Professor and the Director of the Institute of Artificial Intelligence and Robotics. His current research interests include computer vision, pattern recognition and image processing, and hardware implementation of intelligent systems. He became a member of the Chinese Academy of Engineering, in 1999, and is the Chinese Representative on the Governing Board of the International Association for Pattern Recognition. He also serves as the Executive Deputy Editor of the *Chinese Science Bulletin* and as an Associate Editor for the IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS.

Pamela Cosman (S'88-M'93-SM'00-F'08) received the B.S. degree with Honors in electrical engineering from the California Institute of Technology, Pasadena, CA, USA, in 1987, and the M.S. and Ph.D. degrees in electrical engineering from Stanford University, Stanford, CA, USA, in 1989 and 1993, respectively. She was an NSF Post-Doctoral Fellow at Stanford University and a Visiting Professor at the University of Minnesota, Minneapolis, MN, USA, during 1993–1995. In 1995, she joined the Faculty of the Department of Electrical and Computer Engineering (ECE) at the University of California, San Diego, CA, USA, where she is currently a Professor and Vice Chair. She was the Director of the Center for Wireless Communications from 2006 to 2008. Her current research interests include the areas of image and video compression and processing and wireless communications. She received the ECE Departmental Graduate Teaching Award, a Career Award from the National Science Foundation, a Powell Faculty Fellowship, the Globecom 2008 Best Paper Award, and the HISB 2012 Best Poster Award. She was a Guest Editor of the June 2000 special issue of the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS ON "ERROR-RESILIENT IMAGE AND VIDEO CODING," and was the Technical Program Chair of the 1998 Information Theory Workshop in San Diego. She has been a member of the Technical Program Committee or the Organizing Committee for numerous conferences, including ICIP 2008–2011, QOMEX 2010–2012, ICME 2011–2013, VCIP 2010, PacketVideo 2007–2013, WPMC 2006, ICISP 2003, ACIVS 2002-2012, ICC 2012, the Asilomar Conference on Signals, Systems, and Computers 2003, and EUSIPCO 1998. She was an Associate Editor for the IEEE COMMUNICATIONS LETTERS, from 1998 to 2001, and for the IEEE SIGNAL PROCESSING LETTERS from 2001 to 2005. She was the Editor-in-Chief from 2006 to 2009. Since 2010, she has been the Senior Editor of the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS, a position she also held from 2003 to 2005. She is a member of Tau Beta Pi and Sigma Xi.