symmetry

MDPI

Article

Inter-Frame Based Interpolation for Top–Bottom Packed Frame of 3D Video

Phan Van Duc [1], Phu Tran Tin [2], Anh Vu Le [3,*], Nguyen Huu Khanh Nhan [3] and Mohan Rajesh Elara [4]

[1] Faculty of Automobile Technology, Van Lang University, Ho Chi Minh City 700000, Vietnam; duc.pv@vlu.edu.vn (P.V.D.)
[2] Faculty of Electronics Technology, Industrial University of Ho Chi Minh City, Ho Chi Minh City 700000, Vietnam; phutrantin@iuh.edu.vn (P.T.T.)
[3] Optoelectronics Research Group, Faculty of Electrical and Electronics Engineering, Ton Duc Thang University, Ho Chi Minh City 700000, Vietnam; nguyenhuukhanhnhan@tdtu.edu.vn (N.H.K.N.)
[4] ROAR Lab, Engineering Product Development, Singapore University of Technology and Design, Singapore 487372, Singapore; rajeshelara@sutd.edu.sg (M.R.E.)
* Correspondence: leanhvu@tdtu.edu.vn

Abstract: The frame-compatible packing for 3D contents is the feasible approach to archive the compatibility with the existing monocular broadcasting system. To perceive better 3D quality, the packed 3D frames are expanded to the full size at the decoder. In this paper, an interpolation technique enhancing and comparing the quality of enlarged halt vertical left and right stereo video in the top–bottom frame-compatible packing is presented. To this end, the appropriate interpolation modes from fourteen available modes for each row segment, which exploit the correlation between left and right stereoscopic as well as current and adjacent frames of individual view, are estimated at the encoder. Based on the information received from the encoder, at the decoder, the interpolation scheme can select the most appropriate available original data to find the missing values of to-be-discarded row segments. The proposed method outperformed than the state-of-the-art interpolation methods in terms of subjective visualization and numerical PSNRs and SSMI about 11%, with an execution time of about 12% comparisons.

Keywords: 3D frame packing; interpolation; 3D compression; 3D video broadcasting

Citation: Van Duc, P.; Tin, P.T.; Le, A.V.; Nhan, N.H.K.; Elara, M.R. Inter-Frame Based Interpolation for Top–Bottom Packed Frame of 3D Video. *Symmetry* **2021**, *13*, 702. https://doi.org/10.3390/sym13040702

Academic Editor: Dumitru Baleanu

Received: 22 March 2021
Accepted: 14 April 2021
Published: 16 April 2021

1. Introduction

Recently, with the increasing popularity of 3D contents, 3DTV broadcasting services becomes a reality. However, compared with convention digital TV (DTV) broadcasting services, 3DTV needs double the transmission bandwidth for the same video resolution since the left and right frame of 3D video are required to perceive 3D contents [1,2]. To utilize existing transmission infrastructure including teh H.264/AVC compression scheme, the compatibility of 3D video with conventional digital television (DTV) transformation format is strongly required [3,4]. The existing broadcasting network can be used by converting the 3D video format into a single DTV frame. For example, a top–bottom packing after the horizontal line sub-sampling can be treated as a single frame of the existing DTV. Other packing methods include side by side, interleaved formats [5]. The existing 3D compatible frame packing formats are shown in Figure 1.

With the results of utilizing the packing methods, we can utilize the existing transmission infrastructure for the stereoscopic 3DTV similar to [6]. However, the reduction in the spatial resolution due to the decimation of frame-compatible packing is a critical problem when the user wants to experience the video's full quality. One way to solve this problem is that one suitable interpolation method should be applied to recover the decoder's original resolution.

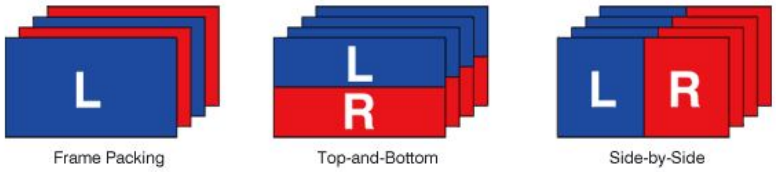

Figure 1. Frame packing for stereo video.

Interpolation is the problem of image processing for many years. To yield the unknown pixels of the upsampled image, the conventional interpolation methods can be used. The simple bilinear and bicubic interpolations use the linear and non-linear correlations of known surrounding neighbor pixels, and the sophisticated iterative methods use higher-order correlations such as methods [7,8]. NEDI6 [9] and the updated version of NEDI4 [10], which uses six original pixel values in the low-resolution image to estimate coefficients. Computation time is the main advantage of real-time applications. The idea of sending side information based on the horizontal and vertical line to the decoder is considered for an adaptive linear interpolation for the line-pruned image in [11] which proved good for the horizontal and vertical edges. Regarding the stereo images, the relation between the left and right images can play an essential cue for interpolation to full resolution, which has not been considered by conventional methods. The pixel-based stereo matching method [12] produces the horizontal disparities between left image pixels with their right image corresponding pixels and similar for right and left frames case. Because all pixels should be estimated, the disparity value by searching the match point, the burden of computation time and the accuracy of this method lead to motivation to find new methods.

To reduce the computation time of pixel-based stereo matching, in the patent [13], block-based stereo matching is proposed. Each pixel of the left image (or the right image in this block) is assigned the value of the corresponding pixel of another image for each matched block. This leads to the interpolated pixel being less accurate due to a block mismatch problem. Another issue of stereo matching is the accuracy of disparity estimation on the occlusion region where there are no matched regions between the right and left frames. For the top–bottom packing, the disparity estimation for each deleted line can be exploited to determine an appropriate interpolation mode [14]. Although this method significantly improves the video quality, it does not consider the inter-frame relations to improve the quality of interpolated 3D video sequences.

An effective interpolation technique addresing the unpacking top bottom format of stereo sequences was proposed and evaluated. Specificaaly, our proposal's primary concern is to use the available values of the previous and subsequent frames of the current view to preserve the quality of the interpolated stereo images. To achieve our goal, adding to the modes of paper [14] by the modes which include the previous and subsequent frame of the current frame is taken into account to exploit the high correlation. The best interpolation mode among the proposed interpolation modes is estimated at the encoder (sending side), before sending the decoders to yield the best interpolation results. The proposed method can archive better running time and subjective evaluation comparing to the evaluated interpolation methods.

The paper is organized as follows. The the frame-compatible top–bottom packing is presented in Section 2. The the proposed inter-frame based interpolation is detailed in Section 3. The optimal proposed interpolation strategy is evaluated in Section 4. The conclusion and future works are in the last Section 5.

2. Frame-Compatible Top–Bottom Packing

The proposed optimal mode based interpolation uses the top bottom packing scheme of 3D video during broadcasting 3D contents (see Figure 1). The top bottom packing scheme has a similar configuration with [14], but with the addition of the same packing

scheme for the current frame and adjacent frame (previous and subsequent frames) of one view of the stereo sequences. Specifically, each line of two consecutive horizontal rows will be discarded to squeeze the height by half right and left images, and a similar process is applied for the current and previous frames detailed, as shown in Figure 1. To utilize each image's remaining lines effectively, the horizontal downsized right image I_r and current frame of one view has a left image I_l with one line offset and a previous frame of one view, respectively. Then, the sub-sampled left and right views are combined as an individual frame within the stereoscopic stream for compression and transmission.

3. Proposed Inter-Frame Based Interpolation

In this paper, the idea of exploiting the correlations of inter-frames optimally at the encoder to leverage interpolating the top–bottom packed stereoviews at the decoder is implemented. The first motivation is that the need-to-be-interpolated pixels on the discarded horizontal line of one frame of the current view can be recovered by the corresponding pixels that are not discarded on the other view frame. This is the characteristic of the well-rectified stereo image as the result of epipolar constraint. To obtain the best perception of 3D effects, epipolar constraint states that the corresponding pixel in the right image to the considered pixel in the left image stay on the same row. The second motivation is that the inter-frame correlation is significantly high for frames of a single view. As a result, the current frame's information can exist in the adjacent frames, such as previous or subsequent frames. As a result, the current frame's information can exist in the adjacent frames, such as previous or subsequent frames. Figure 2 is an overview of the transmission system.

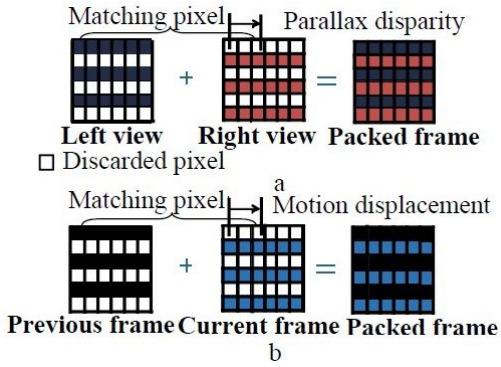

Figure 2. Frame-compatible top–bottom packing: (**a**) left and right frames; (**b**) previous and current frames of one view.

Conventional stereo matching, which wants to find the matched pixel (or matched block) by using some matching criteria like the sum of absolute differences (SAD), sum of squared differences (SSD) [15,16], to interpolate the right frame from the available left pixel on the left frame (similar for interpolating the right frame). In this paper, with the original available data from not to-be-discarded pixels of the corresponding left pixels and the pixels of previous and after right frames, the combinations of this available information can be estimated at the encoder and exploited to the decoder as shown in Figure 3. To this end, using the biliear, firstly the half vertical resolution right and left views are interpolated to full resolution at the prepossessing step. Each deleted row is divided into S parts with the length L similar to the method [14]. Each part will have the same interpolation modes. To find the best mode for each horizontal part, the mode will be found by the estimation scheme at the encoder. Note that among the total 14 modes in Table 1, modes 1 is the equation of parallax added left view I_l data $(x^j_{i,k*} \rightarrow x^j_{i,k*+L})$ and modes 2 and 3 are the function of adjacent right data $(x^{r-1}_{i,l*} \rightarrow x^{r-1}_{i,l*+L})$, $(x^{r+1}_{i,g*} \rightarrow x^{r+1}_{i,g*+L})$ and mode 4 are the formulas which combine the parallax-compensated left image with the previous and next frame motion estimation. Note that the left x^l_i and right x^r_i are the full resolution frame expanded by bilinear interpolation method and right x^{r-1}, right x^{r+1} are the full

resolution expanded by the bilinear interpolation of the previous and subsequent right frames, respectively. The horizontal parallax $k*$ in interpolation modes can be found as Equation (1).

$$k* = \underset{k\in\Phi}{argmin}\left(\sum_{n=1}^{L}|x^{r}_{i,\Delta+n} - x^{l}_{i,k+n}|\right) \quad (1)$$

Moreover, the motion displacement l * between the segment of the current right frame and previous right frame can be found as Equation (2):

$$l* = \underset{l\in\Phi}{argmin}\left(\sum_{n=1}^{L}|x^{r}_{i,\Delta+n} - x^{r-1}_{i,k+n}|\right) \quad (2)$$

Similarly, Equation (3) is used to find the motion displacement $g*$ between the segment of the right frame and the next right frame:

$$g* = \underset{g\in\Phi}{argmin}\left(\sum_{n=1}^{L}|x^{r}_{i,\Delta+n} - x^{r+1}_{i,k+n}|\right) \quad (3)$$

where the parameter $\Phi = [\Delta - m_d, \Delta + m_d]$ is a set of a horizontals shift, and Δ is the first pixel of the considered segment, and m_d is the maximum horizontal disparity. Based on the original pixels and the pre-calculated candidate interpolation modes, the best interpolation mode $b^{*}_{r}(i,s)$ of sequences of deleted row, this can be calculated as Equation (4):

$$b^{*}_{r}(i,s) = \underset{m=(1toM)}{argmin}\left(\sum_{n=1}^{S}|x^{r}_{i,\Delta+j}(m) - x^{r}_{i,\Delta+j}|\right) \quad (4)$$

where right $x^{r}_{i,\Delta+j}$ is the pixel in the original right frame at time t and right $x^{r}_{i,\Delta+j}(m)$ is the interpolated pixel by the mode number m of the proposed interpolated modes. Note that after being estimated, the optimal mode is sent to the decoder to improve the accuracy of the interpolation.

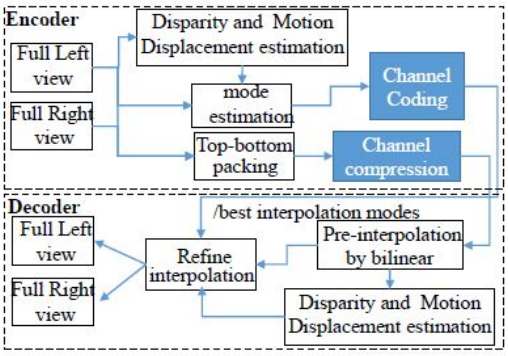

Figure 3. Block diagram of proposed method.

Table 1. Interpolation modes.

Mode	Interpolation Condition	Mode	Interpolation Mode
1	$\hat{x}_{i,j}^{right}(1) = x_{i+1,j}^{right}$	2	$\hat{x}_{i,j}^{right}(2) = x_{i-1,j}^{right}$
3	$\hat{x}_{i,j}^{right}(3) = \frac{1}{2}(x_{i+1,j}^{right} + x_{i-1,j}^{right})$	4	$\hat{x}_{i,j}^{right}(4) = \frac{1}{2}(x_{i-1,j+1}^{right} + x_{i+1,j-1}^{right})$
5	$\hat{x}_{i,j}^{right}(5) = \frac{1}{2}(x_{i-1,j-1}^{right} + x_{i+11,j+1}^{right})$	6	$\hat{x}_{i,j}^{right}(6) = \frac{1}{4}(x_{i-1,j-1}^{right} + x_{i+1,j+1}^{right} + x_{i-1,j}^{right} + x_{i+1,j}^{right})$
7	$\hat{x}_{i,j}^{right}(7) = x_{i,j*}^{right-1}$	8	$\hat{x}_{i,j}^{right}(8) = x_{i,g*}^{right+1}$
9	$\hat{x}_{i,j}^{right}(9) = \frac{1}{2}(x_{i,j*}^{right-1} + x_{i,g*}^{right+1})$	10	$\hat{x}_{i,j}^{right}(10) = x_{i,k*}^{left}$
11	$\hat{x}_{i,j}^{right}(11) = \frac{1}{2}(x_{i,g*}^{right+1} + x_{i,k*}^{left})$	12	$\hat{x}_{i,j}^{right}(12) = \frac{1}{2}(x_{i,j*}^{right-1} + x_{i,k*}^{left})$
13	$\hat{x}_{i,j}^{right}(13) = \frac{1}{3}(x_{i,j*}^{right-1} + x_{i,g*}^{right+1} + x_{i,k*}^{left})$	14	$\hat{x}_{i,j}^{right}(14) = \frac{1}{5}(x_{i-1,j}^{right} + x_{i,j+1}^{right} + x_{i,j*}^{right-1} + x_{i,g*}^{right+1} + x_{i,k*}^{left})$

4. Experimental Results

The experiments were done on MATLAB version R2016a environment installed on the machine with the configuration: Intel®core™i5CPU 4G Ram Windows 10 64 bit operating system. Six stereoscopic sequences named Book, Car, Door, Horse, Moabit, Bullinger are downloaded from the open source database [17] and used to evaluate the efficiencies of the proposed method. For a fair comparison, we selected the 3D sequences as same as the 3D-based frame-compatible interpolation in reference [14]. There are 150 frames for each tested sequence. The 3D frames of tested sequences are down-sampled for the top–bottom packing. The conventional interpolation methods including the Bilinear, NEDI6 [9], pixel-based matching method [12], block-based matching patent [13], the method [14] are implemented to compare with the proposed method in terms of subjective and numerical data.

Note that, in the bilinear method, the interpolated pixel p_i can be estimated by the Equation (5) where x_l, x_r, x_t, x_b are the left right top and bottom undiscarded pixels' data, and c_l, c_r, c_t, c_b are the corresponding distances from the interpolated pixel. In NEDI6 [9], Io^d is the downsample of the original image I_o which is expanded to I_i of size mxn with zero value row at $k_t h$ row of I_o, the row of Io after the downsample is linked to the odd rows of expanded frame I_i by $I_i(i, 2j-1) = I_o(i, j$, and then using the sixth order h^6, the discarded row is derived from the undiscarded row as Equation (7). The pixel-based matching method [12] and block-based matching patent [13] can be used to derive the unknown discarded pixels by searching for the minimal value of SAD values within the pixel order or block of pixels order, respectively. The SAD algorithm has the advantage of computational efficiency. The SAD, as Equation (6), finds the disparity $disp\ ab$ windows of corresponding pixels in the left image I_l and right image I_r. The optimal mode-based method [14] only considers the current left and right frame to derive the optimal interpolation modes:

$$p_i = \frac{x_l c_l + x_r c_r + x_t c_t + x_b c_b}{c_l + c_r + c_t + c_b} \tag{5}$$

$$SAD(i, j, disp) = \sum_{k=-a/2}^{a/2} \sum_{l=b/2}^{b/2} |I_r(i+k, j+l) - I_l(i+k-j+l+disp)| \tag{6}$$

$$I_i(i, 2j) = \sum_{k=-1}^{1} \sum_{l=0}^{1} (h_{3l+k+1}^{6} I_i(i+k, 2j+2l-1)) \tag{7}$$

We applied the strategy by calculating the average measure from one top–bottom frame, and then averaged all 150 frames per one tested video sequence. As a result, the peak signalnoise ratio (PSNR), structural similarity index measure (SSIM) and multiscale structural similarity index measure (MSSSIM) were added in Tables 2–4, respectively. Note that, PSNR measures the lossy interpolated images. On another hand, SSMI and MS-SSMI are the objective image quality measures that have a higher correlation with the human

visual system. Given original image I_o and interpolated image I_i with the size M, n and standard deviation μ and mean σ, Equations (8) and (10) were used to find the PSNR and MMSI, whilst the MS-SSIM is implemented by applied multiple scale SSIM:

$$PSNR(I_o, I_i) = \sum_{i=1}^{M} \sum_{j=1}^{N} (I_o(i,j) - I_i(i,j))^2 \tag{8}$$

where MSE is calculated as Equation (9):

$$MSE = \frac{\sum_{n=1}^{N} \sum_{m=1}^{M} (I_o(i,j) - I_i(i,j))^2}{MN} \tag{9}$$

$$SSIM(I_o, I_i) = \frac{(2\mu_{Io}\mu_{Ii} + c_1)(2\sigma_{IoIi} + c_2)}{((\mu_{Io}^2 + \mu_{Ii}^2 + c_1)(\sigma_{Io}^2 + \sigma_{Ii}^2 + c_2)} \tag{10}$$

where c_1, c_2 are constants.

Table 2. PSNR comparison (dB).

Method -	Bilinear (dB)	[9] dB	[12] Method dB	[13] dB	[14] dB	Proposed dB
Book	34.66	34.39	35.66	35.82	36.66	37.90
Car	38.93	38.91	39.12	38.87	39.68	41.38
Door	36.16	36.06	37.24	37.12	38.26	39.76
Horse	32.69	32.24	32.72	32.91	33.73	35.34
Moabit	31.93	31.94	33.12	33.01	34.22	36.24
Bullinger	43.33	43.49	43.51	43.42	44.37	45.68
Mean	**36.28**	**36.19**	**36.90**	**36.86**	**37.82**	**39.38**
Standard Deviation	**4.27**	**4.41**	**4.05**	**3.96**	**3.93**	**3.80**

Table 3. SSIM comparison.

Method -	Bilinear	[9]	[12]	[13]	[14]	Proposed Method
Book	0.872	0.881	0.887	0.891	0.925	0.943
Car	0.901	0.911	0.919	0.922	0.935	0.966
Door	0.891	0.901	0.908	0.921	0.932	0.951
Horse	0.881	0.885	0.896	0.907	0.912	0.933
Moabit	0.903	0.912	0.904	0.911	0.924	0.941
Bullinger	0.908	0.915	0.921	0.926	0.936	0.974
Mean	**0.892**	**0.901**	**0.906**	**0.913**	**0.928**	**0.951**
Standard Deviation	**0.0139**	**0.0146**	**0.0131**	**0.0134**	**0.0090**	**0.0157**

Table 4. MS-SSIM comparison.

Method -	Bilinear	[9]	[12]	[13]	[14]	Proposed Method
Book	0.842	0.873	0.879	0.882	0.915	0.955
Car	0.866	0.869	0.875	0.873	0.904	0.934
Door	0.887	0.891	0.904	0.911	0.920	0.931
Horse	0.871	0.879	0.882	0.887	0.911	0.926
Moabit	0.893	0.905	0.911	0.919	0.922	0.932
Bullinger	0.896	0.912	0.919	0.921	0.935	0.961
Mean	**0.874**	**0.889**	**0.895**	**0.899**	**0.918**	**0.939**
Standard Deviation	**0.0204**	**0.0140**	**0.0186**	**0.0206**	**0.0106**	**0.0144**

The proposed method proved that it archives the better quality when applied to the sequences with complicated textual and many corners like the Alt-moabit sequence.

For the sequences in which their left frames and right frames consist of occlusions, the proposed method derived the best results in terms of subjective evaluation than stereo-based matching methods [12,13].

We calculated the ANOVA Fisher's least significant difference (LSD) procedure the multicomparison post hoc statistical test to check which pairs of means have a significantly different variation from the mean PSNR, SSIM, and MS-SSIM of different interpolation methods. The numerical data are shown in Figure 4 for PSNR, Figure 5 for SSIM , and Figure 6 for MS-SSIM, respectively. As one can observe, the performance of the method [14], which is the second-best method, is considerably lower than proposed optimal mode based interpolation technique. Even though the proposed method's execution time is slightly higher than the Bilinear and [14], the proposed method outperformed all validated methods in PSNR, SSIM, and MS-SSIM with significant margins and comparable standard deviations.

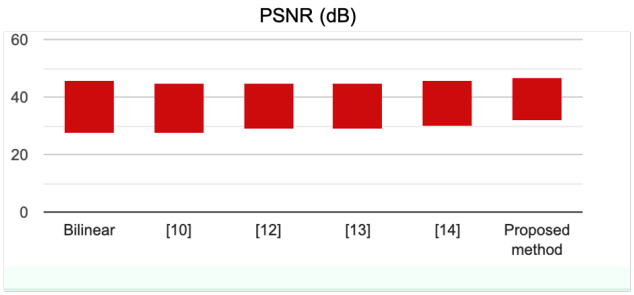

Figure 4. Variation of mean PSNR.

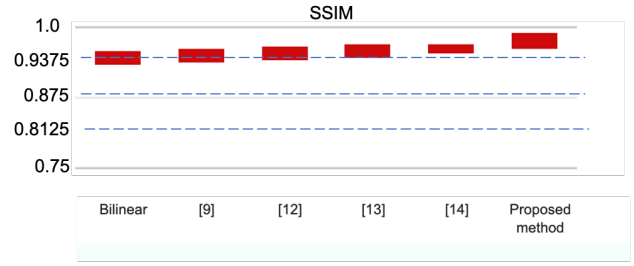

Figure 5. Variation of mean SSIM.

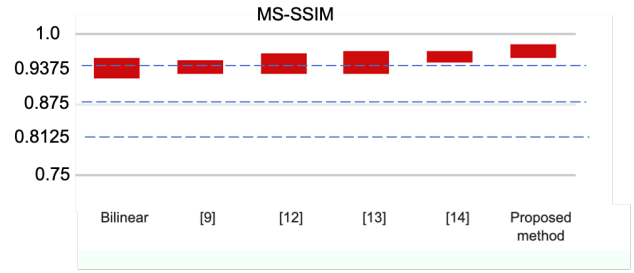

Figure 6. Variation of mean MS-SSIM.

We zoom in on the interpolated frame of the tested sequences for subjective comparisons in Figures 7–10, for the Moabit, Car, Horse, and Book-Arrive sequences, respectively. From the results, the proposed method archives outperforms than others, especially at thin horizontal lines. This better quality caused by deleted pixels on one view still exists in another view, and its value can be almost recovered after interpolation based on feature-based matching. Figure 11 shows the results of the tested method for the zoom in on the vertical edge. The proposed method fixes zigzag artifacts generated by incorrectly estimated disparity. The execution time comparison for encoder and decoder processes of all tested methods is shown in Table 5. Although the bilinear method yeilds the smallest time

values, its visual effects are worst among the tested methods. The proposed method takes more time than the bilinear and much less time than the pixel-based stereo matching and NEDI6. Note that the execution time to derive the optimal mode for considered segment at the encoder is method's primary consumption time [14] and the proposed method.

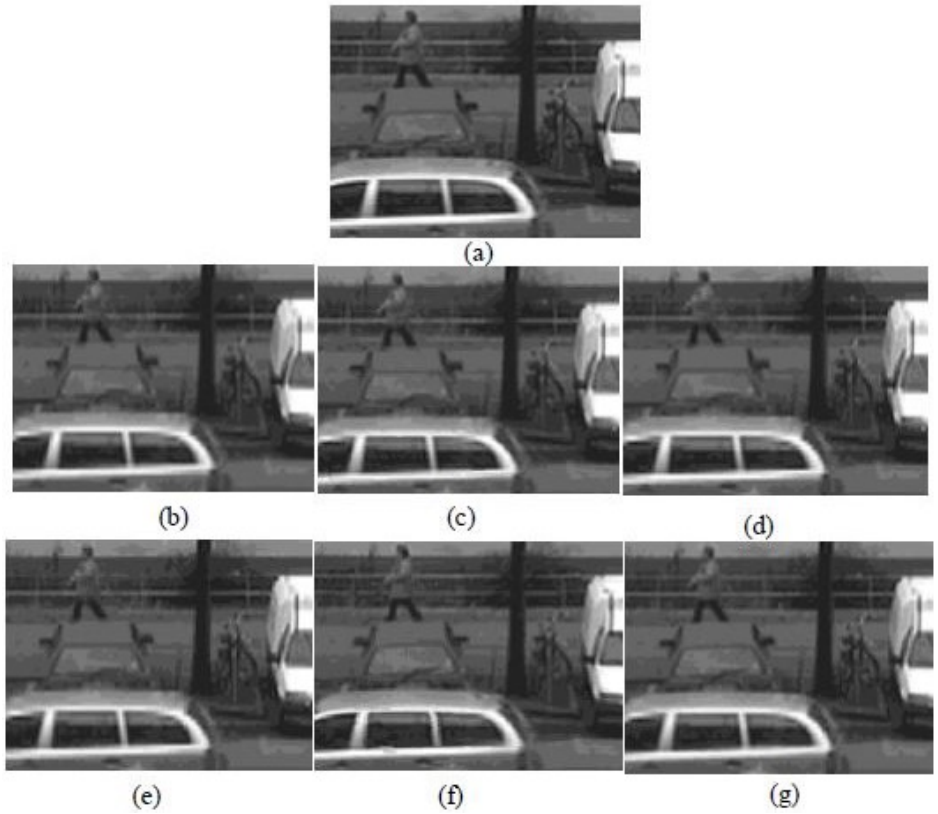

Figure 7. Visualized comparison after the interpolations for the Moabit sequence: (**a**) original image; (**b**) cropped original; (**c**) NEDI6 [9]; (**d**)pixel-based matching [12]; (**e**) patent [13]; (**f**) paper [14]; and (**g**) proposed method.

Table 5. Execution time comparison (s).

Method -	Bilinear	[9]	[12]	[13]	[14]	Proposed Method
Book	0.15	10.60	39.09	1.20	1.76	1.81
Car	0.18	5.10	28.78	0.65	1.05	1.12
Door	0.25	9.76	37.09	0.91	1.49	1.15
Horse	0.31	5.28	32.36	0.86	1.55	1.59
Moabit	0.21	10.30	40.77	1.10	2.42	2.62
Bullinger	0.16	2.94	13.82	0.31	0.60	0.69
Mean	**0.21**	**7.33**	**31.99**	**0.84**	**1.47**	**1.50**
Standard Deviation	**0.0609**	**3.2822**	**9.9408**	**0.3223**	**0.6206**	**0.6755**

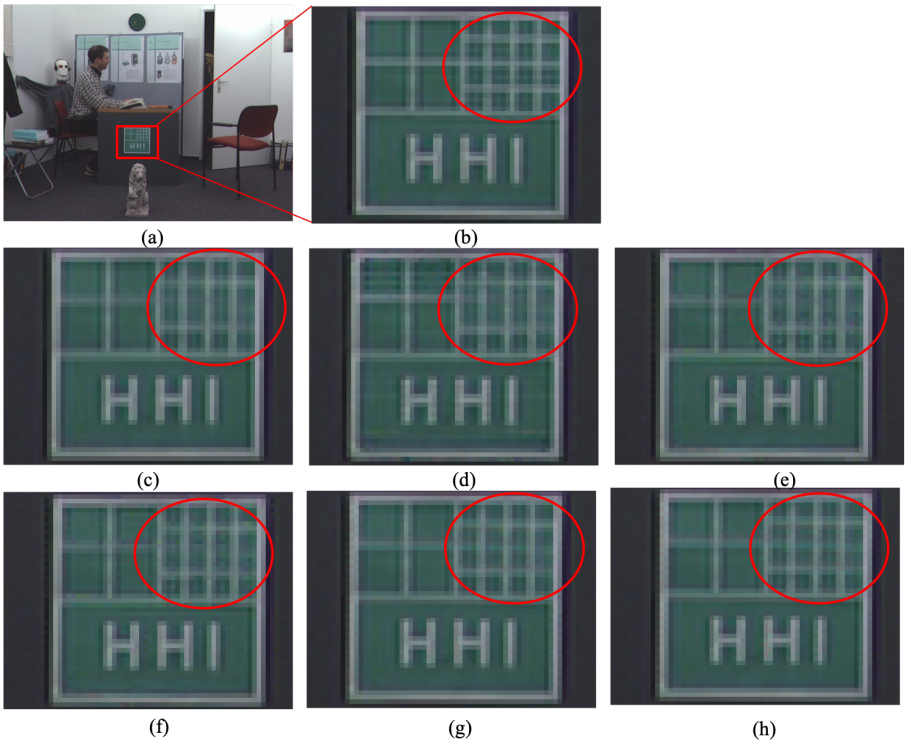

Figure 8. Zooming in for a visualized comparison between the interpolation methods for the Door sequence: (**a**) original image; (**b**) Bilinear; (**c**) cropped original; (**d**) NEDI6 [9]; (**e**)pixel-base matching [12]; (**f**) patent [13]; (**g**) paper [14]; and (**h**) proposed method.

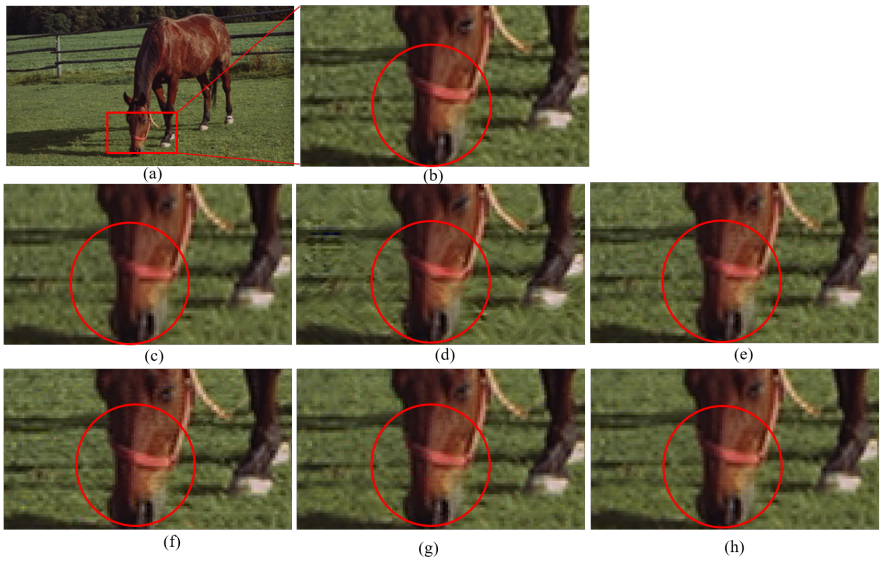

Figure 9. Zooming in for a visualized comparison between the interpolation methods for the Horse sequence (**a**) original image; (**b**) Bilinear; (**c**) cropped original; (**d**) NEDI6 [9]; (**e**)pixel-base matching [12]; (**f**) patent [13]; (**g**) paper [14]; and (**h**) proposed method.

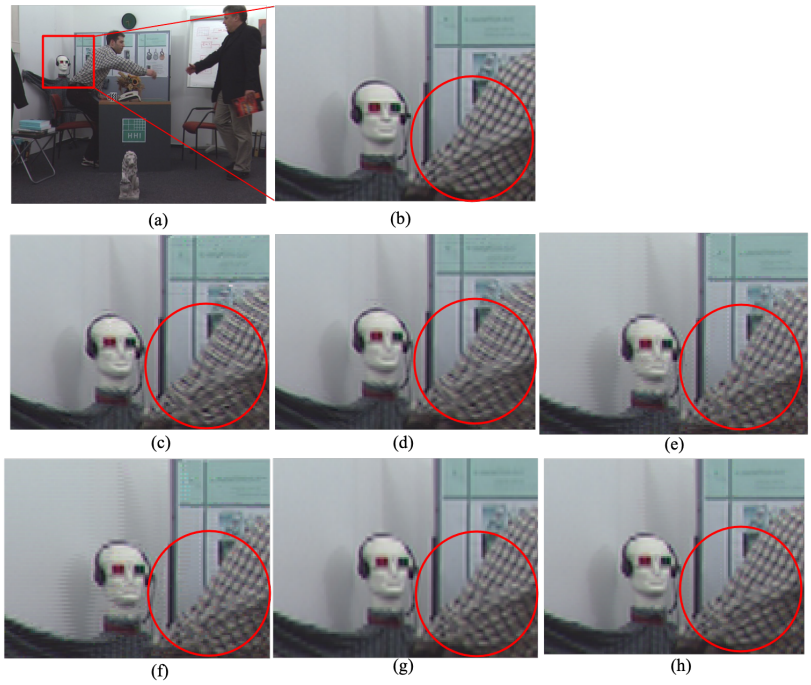

Figure 10. Zooming in for a visualized comparison between the interpolation methods for the Book sequence: (**a**) original image; (**b**) Bilinear; (**c**) cropped original; (**d**) NEDI6 [9]; (**e**)pixel-base matching [12]; (**f**) patent [13]; (**g**) paper [14]; and (**h**) proposed method.

Figure 11. Zigzag artifact at the vertical edge caused by an incorrect estimate disparity: (**a**) patent [13]; (**b**) paper [14]; and (**c**) proposed method after fixing artifact.

Furthermore, similar to [18], we add the depth image rendered by the proposed method using the interpolated left and right frame and compare with the method [4]; and as shown in Figure 12, one can observe that the proposed interpolation technique creates the less noise and higher PSRN for the rendered depth images from the interpolated left and right frames.

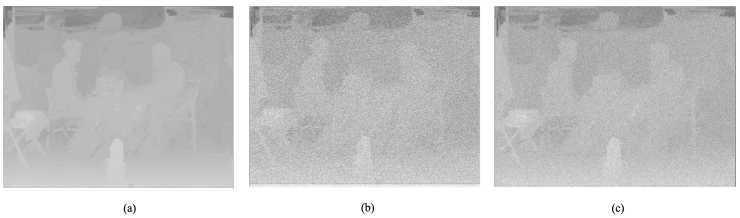

Figure 12. Depth images from the left and right interpolated frame: (**a**) ground true; (**b**) method [14] with PSNR 21.124 dB; and (**c**) the proposed method with PSNR 22.246.

5. Discussion and Conclusions

Since we need the fast interpolation method to obtain the intermediate full-size resolution at the decoder, we applied the bilinear method so that it could then be refined

by the optimal modes send by the encoder. The single image super-resolution such as the tested method yielded a lower quality than the proposed method. The deep learning-based method using the single image might provide better quality but it required heavy computation and advanced hardware to store and load the trained models which is not always available at the encoder of DTV.

In this paper, the matched region between the right and left frames and the adjacent frames of the same view was exploited for the problem of the interpolation frame-compatible top–bottom packing to full resolutions. The correlation of pixels between the view is the right cue to derive the appropriate value of to-be-interpolated pixels. The experimental section shows clearly that the proposed method yields higher PSNR by 1–2 dB and also better visualizes for interpolated images while comparing with the other methods. The future work will focus on (1) reduce mode options; (2) extending to online interpolation; (3) expanding the proposed approach into another frame-compatible packing system; (4) exploring the interpolation mode by AI-aided techniques, testing the method in real hardware devices; and (5) the images quality assessment, which must be studied intensively to provide the appropriate tool to evaluate the interpolated, restored images [19–23]. The method as mentioned in the recommendation ITU-R BT.500-14 [24] will consider as future works since it required setting up the evaluation environment with professional equipment and observers.

Author Contributions: Conceptualization, A.V.L.; data curation, N.H.K.N.; formal analysis, P.V.D., A.V.L.; methodology, P.T.T.; software, P.V.D., P.T.T.; project administration, M.R.E.; supervision, N.H.K.N. and M.R.E.; writing—original draft, A.V.L.; writing—review and editing, P.V.D., P.T.T. and N.H.K.N. and M.R.E. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the National Robotics Programme under its Robotics Enabling Capabilities and Technologies (Funding Agency Project No. 192 25 00051), National Robotics Programme under its Robot Domain Specific (Funding Agency Project No. 192 22 00058) and administered by the Agency for Science, Technology and Research and Industrial University of Ho Chi Minh City (IUH) under grant number 72/HD-DHCN.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare that there is no conflict of interest.

References

1. Vetro, A.; Lou, W.H.; Flynn, M. TV Architecture Supporting Multiple 3D Services. In Proceedings of the 2010 Digest of Technical Papers International Conference on Consumer Electronics (ICCE), Las Vegas, NV, USA, 9–13 January 2010; IEEE: New York, NY, USA, 2010; pp. 135–136.
2. Le, A.V.; Jung, S.W; Won, C.S. Directional joint bilateral filter for depth images. *Sensors* **2014**, *14*, 11362–11378. [CrossRef] [PubMed]
3. Nguyen, H.; Le, A.V.; Won, C.S. Fast selective interpolation for 3D depth images. In Proceedings of the IEEE International Symposium on Broadband Multimedia Systems and Broadcasting, Seoul, Korea, 27–29 June 2012; pp. 1–5.
4. Vetro, A.; Wiegand, T.; Sullivan, G.J. Overview of the stereo and multiview video coding extensions of the H.264/MPEG-4 AVC standard. *Proc. IEEE* **2011**, *99*, 626–642. [CrossRef]
5. Hur, N.; Lee, H.; Lee, G.S.; Lee, S.J.; Gotchev, A.; Park, S.I. 3DTV broadcasting and distribution systems. *IEEE Trans. Broadcast.* **2011**, *57*, 395–407. [CrossRef]
6. Le, A.V.; Won, C.S. Key-point based stereo matching and its application to interpolations. *Multidimens. Syst. Signal Process.* **2017**, *28*, 265–280. [CrossRef]
7. Siu, W.C.; Hung, K.W. Review of Image Interpolation and Super-Resolution. In Proceedings of the 2012 Asia Pacific Signal and Information Processing Association Annual Summit and Conference, Hollywood, CA, USA, 3–6 December 2012; pp. 1–10.
8. Giachetti, A.; Asuni, N. Real-time artifact-free image upscaling. *IEEE Trans. Image Process.* **2011**, *20*, 2760–2768. [CrossRef] [PubMed]
9. Vo, D.T.; Sole, J.; Yin, P.; Gomila, C.; Nguyen, T.Q. Selective data pruning-based compression using high-order edge-directed interpolation. *IEEE Trans. Image Process.* **2009**, *19*, 399–409. [CrossRef] [PubMed]
10. Li, X.; Orchard, M.T. New edge-directed interpolation. *IEEE Trans. Image Process.* **2001**, *10*, 1521–1527. [PubMed]

11. Le, A.V.; Kim, H.M.; Won, C.S. Fast interpolation for line-pruned images. *J. Electron. Imaging* **2011**, *20*, 033010.
12. Tourapis, A.; Pahalawatta, P.V.; Leontaris, A.; Stec, K.J. Encoding and Decoding Architectures for Format Compatible 3D Video Delivery. U.S. Patent 9,774,882, 26 September 2017.
13. Morita, C. Content Reproducing Apparatus and Method. U.S. Patent 12/502,318, 22 June 2010.
14. Won, C.S. Mode selective interpolation for stereoscopic 3D video in frame-compatible top-bottom packing. *Multidimens. Syst. Signal Process.* **2013**, *24*, 221–233. [CrossRef]
15. Di Stefano, L.; Marchionni, M.; Mattoccia, S. A fast area-based stereo matching algorithm. *Image Vis. Comput.* **2004**, *22*, 983–1005. [CrossRef]
16. Song, W.; Le, A.V.; Yun, S.; Jung, S.W.; Won, C.S. Depth completion for kinect v2 sensor. *Multimed. Tools Appl.* **2017**, *76*, 4357–4380. [CrossRef]
17. FHG-HHI. Stereo-Video Database. 2020. Available online: http://sp.cs.tut.fi/mobile3dtv/stereo-video/ (accessed on 22 October 2020).
18. Chen, M.J.; Su, C.C.; Kwon, D.K.; Cormack, L.K.; Bovik, A.C. Full-reference quality assessment of stereopairs accounting for rivalry. *Signal Process. Image Commun.* **2013**, *28*, 1143–1155. [CrossRef]
19. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [CrossRef] [PubMed]
20. Wang, Z.; Li, Q. Information content weighting for perceptual image quality assessment. *IEEE Trans. Image Process.* **2010**, *20*, 1185–1198. [CrossRef] [PubMed]
21. Ponomarenko, N.; Silvestri, F.; Egiazarian, K.; Carli, M.; Astola, J.; Lukin, V. On Between-Coefficient Contrast Masking of DCT Basis Functions. In Proceedings of the Third International Workshop on Video Processing and Quality Metrics, Scottsdale, AZ, USA, 25–26 January 2007; Volume 4.
22. Zhang, L.; Zhang, L.; Mou, X.; Zhang, D. FSIM: A feature similarity index for image quality assessment. *IEEE Trans. Image Process.* **2011**, *20*, 2378–2386. [CrossRef] [PubMed]
23. Wang, Z.; Bovik, A.C. *Modern Image Quality Assessment*; Synthesis Lectures on Image, Video, and Multimedia Processing; Morgan & Claypool: San Rafael, CA, USA, 2006; Volume 2, 156p.
24. BT.500. Methodologies for the Subjective Assessment of the Quality of Television Images. 2020. Available online: https://www.itu.int/rec/R-REC-BT.500 (accessed on 22 October 2020).