https://doi.org/10.2352/ISSN.2470-1173.2020.14.COIMG-342

ProPaCoL-Net: A Novel Recursive Stereo Image SR Network with Progressive Parallax Coherency Learning

Jeonghun Kim, Munchurl Kim; Korea Advanced Institute of Science and Technology; Daejeon; South Korea

Abstract

Recently, stereo cameras have been widely packed in smart phones and autonomous vehicles thanks to low cost and small-sized packages. Nevertheless, acquiring high resolution (HR) stereo images is still a challenging problem. While the traditional stereo image processing tasks have mainly focused on stereo matching, stereo super-resolution (SR) has drawn less attention which is necessitated for HR images. Some deep learning based stereo image SR works have recently shown promising results. However, they have not fully exploited binocular parallax in SR, which may lead to unrealistic visual perception. In this paper, we present a novel and computationally efficient convolutional neural network (CNN) based deep SR network for stereo images by learning parallax coherency between the left and right SR images, which is called ProPaCoL-Net. The proposed ProPaCoL-Net progressively learns parallax coherency via a novel recursive parallax coherency (RPC) module with shared parameters. The RPC module is effectively designed to extract parallax information in prior for the left image SR from its right view input images and vice versa. Furthermore, we propose a parallax coherency loss to reliably train the ProPaCoL-Net. From extensive experiments, the ProPaCoL-Net shows to outperform the very recent state-of-the-art method with average 1.15 dB higher in PSNR.

1. Introduction

Stereo cameras have the ability to simulate a pair of human eyes with laterally separated two cameras. Therefore, emerging devices adopt stereo cameras more and more for unmanned autonomous and aerial vehicles to replace human actions, and for smartphones to give three-dimensional features that allow for augmented reality and virtual reality applications. Recently, low cost and small-size packages of commercial cameras are rapidly accelerating the entry of stereo cameras into the markets.

Jeon *et al.* [11] presented a stereo image SR network, called StereoSR, that finds the correspondence information by stacking the shifted luma (Y-channel) of the right-view image with different intervals on top of a left-view Y-channel input image. The StereoSR is limited by using a predefined fixed maximum disparity that cannot cover stereo images with a variety of disparity. Wang *et al.* [12] proposed a parallax attention stereo SR network, called PASSRnet that integrates stereo correspondence into the stereo SR task using a parallax attention mechanism. The PASSRnet showed a limited capability of generating stereo SR images due to its simple visibility information handing and simultaneous learning of low-to-high frequency components in parallax information domain.

In this paper, we would focus on this stereo image SR problem via a systematic approach based on binocular vision. In order to overcome the limitation of the previous methods, we propose a novel and computationally efficient convolutional neural network (CNN) based deep SR network for stereo images by learning parallax coherency between the left and right SR images, which is called ProPaCoL-Net. The proposed ProPaCoL-Net progressively learns parallax coherency via a novel recursive parallax coherency (RPC) block with shared parameters. The RPC block can effectively extract parallax information in a prior pair for the left image SR from its right view input images and vice versa. Also, we propose a parallax coherency (PC) loss to reliably train the ProPaCoL-Net. Our main contributions are three-fold:

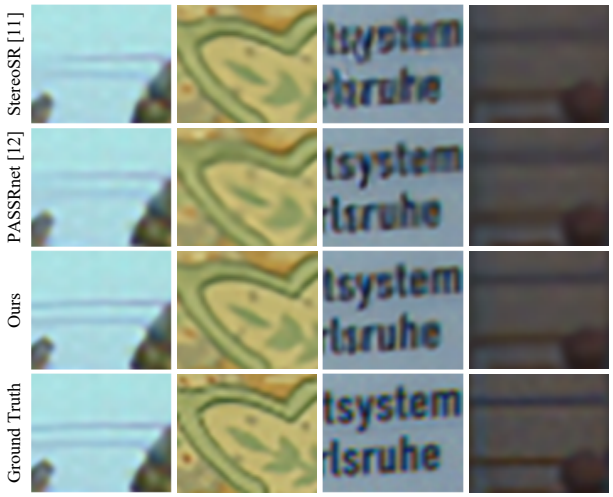


Figure 1: *Quantitative comparison with previous stereo image SR methods on some test images. Our method shows better stereo SR images.*

- Our proposed ProPaCoL-Net learns the parallax coherency between the two parallaxes from each stereo LR input pair (prior pair): one between a stereo SR pair and the other for a stereo HR ground truth pair. So, it can reconstruct stereo SR images with faithful binocular parallax in HR domain;
- Parallax information *in a prior pair* is effectively *extracted* by our novel Recursive Parallax Coherency (RPC) block, and is successively *transferred* into deeper layers to reconstruct the stereo SR images via a progressive learning method;
- Parallax information *in an HR pair* can be effectively learned by our novel Parallax Coherency (PC) loss.

2. Related Work

Single Image and Video SR

Some traditional image SR techniques [4-8] have been still involved in deep learning based single SR problems. Cui *et al.* [13] utilized the internal example-based method [6] to produce patches having self-similarity property. Sparse-coding techniques have also been adapted for various deep learning SR methods. Especially Dong *et al.* [9] established relationship between each layer of [9] and each step in the sparse-coding in an end-to-end manner between LR and HR images. Kim *et al.* [10] proposed the

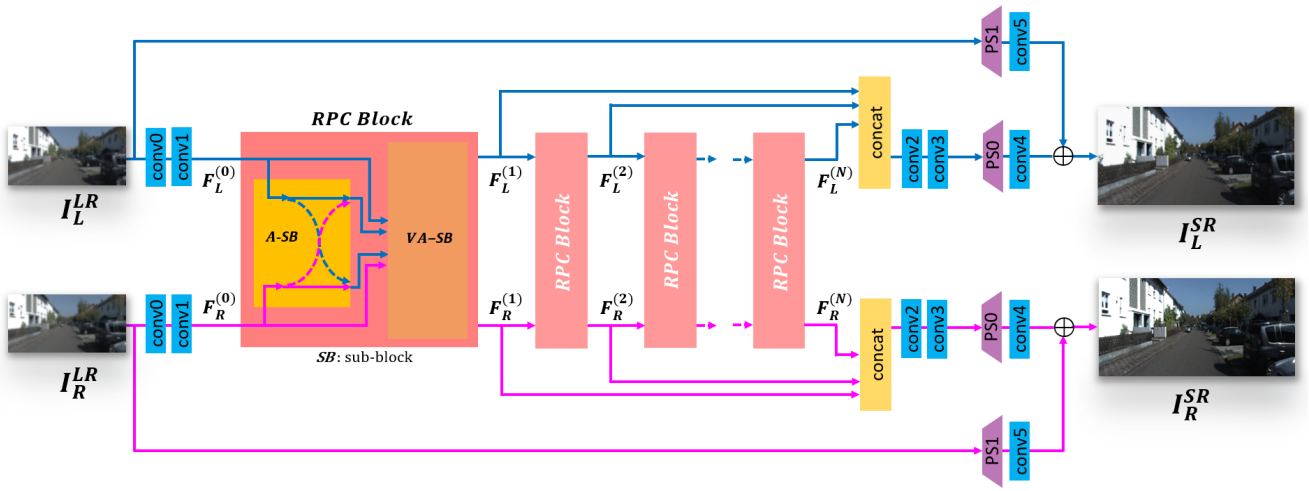


Figure 2: *The overall architecture of ProPaCoL-Net with Unfolded Recursive Parallax Coherency (RPC) block. The RPC Block is unfolded 6 times with all parameters shared, and consists of two sub-blocks: A-SB (alignment sub-block) and VA-SB (visibility attention sub-block). The configuration details for the ProPaCoL-Net are shown in Table 1.*

notable CNN-based SR method, very deep super-resolution, called VDSR. Residual learning and gradient clipping strategies improve the performance of single image SR. Zhang *et al.* [14] presented very deep residual channel attention networks, called RCAN, using a residual in residual structure for very deep network. Video SR [1-3] utilize the adjacent frame images for a given input frame to reconstruct an HR image where 2D motion estimation (ME) is often incorporated. Due to the epipolar constraint in stereo imaging, the correspondence matching is often limited to one dimension search, so 2D ME is not appropriate if it is directly applied for stereo image SR problems.

Stereo Image SR

The success of deep learning shown in image SR task has only recently stimulated stereo image SR. StereoSR [11] proposed by Jeon *et al.* enhanced the quality of stereo image by using a parallax prior generated by stacking the shifted luminance information of the right-view image with different intervals on top of the left luminance information and vice versa. The fixed maximum disparity value, however, induces lack of flexibility. Very recently, Wang *et al.* [12] presented a flexible solution with StereoSR by presenting PASSRnet, which captures a global correspondence, making it applicable for various stereo image pairs. While handing the areas such as occluded, textureless or reflected regions is the most common problem in stereo image task, PASSRnet takes a simple approach to handing visibility information by using morphology operation with empirically fixed parameters, which could not represent a variety of stereo images. In addition, the simply stacking method with estimated visibility information on aligned features during a visibility fusion operation counters the effectiveness of the already estimated visibility information. Note that [11-12] do not simultaneously generate a stereo SR image pair but a single SR image.

3. Proposed Method

Our proposed network aims to simultaneously reconstruct left and right images, I_L^{SR} and I_R^{SR}, from LR left and right input images, I_L^{LR} and I_L^{LR}. Fig. 2 shows the overall architecture of our proposed ProPaCoL-Net for stereo image SR.

Parallax Information

The stereo image SR methods must delicately handle the parallax information between the left and right images to effectively take advantage of the additional spatial information in the other view image, in reconstructing a high quality HR image pair. In the training, there exist two types of the crucial parallax information for a given stereo image SR problem: *a prior pair* $(I_L^{LR} \leftrightarrow I_R^{LR})$ and *an HR pair* $(I_L^{HR} \leftrightarrow I_R^{HR})$. We, therefore, assume that to reconstruct the high quality SR image pair, the parallax coherency should be consistent between a stereo SR pair $(I_L^{SR} \leftrightarrow I_R^{SR})$ and its ground truth stereo HR pair. This parallax coherency is obtained by learning the parallax information resided in prior pairs with the help of our PC loss during training. More specifically, our ProPaCoL-Net progressively captures the parallax information in the ground truth stereo HR pairs in a recursive structure with learned visual attention for both left and right view images.

Parallax Information in a Prior pair $(I_L^{LR} \leftrightarrow I_R^{LR})$

Firstly, for effectively maintaining the parallax information *in a prior pair* $(I_L^{LR} \leftrightarrow I_R^{LR})$ to the estimated SR outputs, we set our strategy to effectively *extract* and *transfer* the parallax information *in a prior pair*. Our Recursive Parallax Coherency (RPC) block, which is designed to be optimized in *extracting* and *transferring* the parallax information, consists of two sub-blocks: A-SB (alignment sub-block) for feature alignment and VA-SB (visibility attention sub-block) for handing visibility information on the left (right)-view aligned images from the right (left) images. Fig. 3 depicts the structure of the proposed RPC block.

Extracting. The RPC block in Fig. 3 takes two inputs into the A-SB: one is a left feature and the other is a right feature. The same left and right inputs are also fed into the VA-SB. The A-SB first aligns the auxiliary view feature to the main view feature and vice versa according to the same way [12], yielding $\boldsymbol{F}_{Ra}^{(n-1)}$ and $\boldsymbol{F}_{La}^{(n-1)}$ as shown in Fig. 3. Note that the right-view input feature

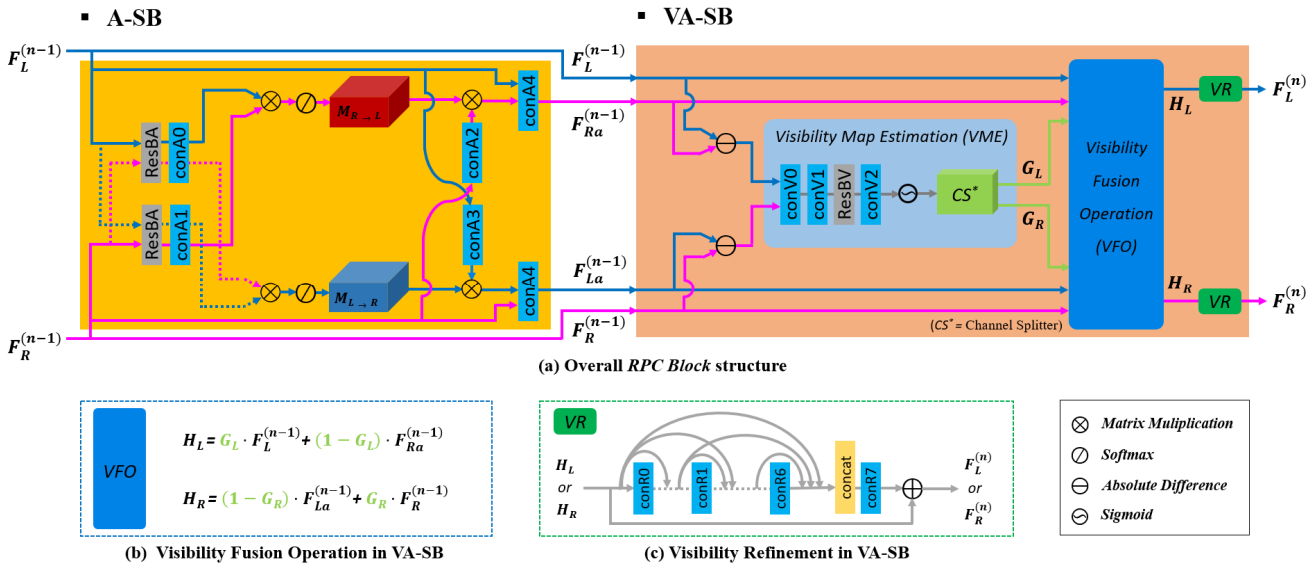


Figure 3: The *RPC Block and its component modules in ProPaCoL-Net. (a) Structure of RPC Block with alignment sub-block (A-SB) and visibility attention sub-block (VA-SB), (b) Visibility Fusion Operation (VFO) in VA-SB, (c) Visibility refinement (VR) in VA-SB. The configuration details for the RPC Block are shown in Table 1.*

$F_R^{(n-1)}$ becomes the auxiliary view one when the left-view input feature $F_L^{(n-1)}$ is the main view one, and vice versa. $M_{R\to L}$ and $M_{L\to R}$ indicate the parallax attention maps from a perspective of the auxiliary views to the main views [12]. The VA-SB inferences the left and right visibility maps. Those visibility attention maps G_L and G_R are generated from the visibility map estimation (VME) module as shown in Fig. 3.

With the separated structure of A-SB and VA-SB in the RPC block, the RPC blocks can better extract parallax features towards the left- and right-view SR reconstructions with the parallax coherence well maintained between them. Especially, the VA-SB estimates visibility attention maps for the main-view feature input, which can help selectively extract parallax information from the auxiliary view. It should be noted in Fig. 2 that the RPC block in Fig. 3 is unfolded N times (N = 6 in our experiments) to progressively learn the low to high frequency components of parallax information and to consistently transfer them to the final stage of the ProPaCoL-Net.

Especially, we drill down a more elaborate visibility handing function to improve a simple visibility handling approach of the very recent state-of-the-art method PASSRnet [12]. In the PASSRnet, the visibility map for the main view feature input is estimated from its own view information, which does not make a full use of available parallax information *in a prior pair*. Moreover, the estimated visibility map is refined simply by applying the morphology operation with empirically fixed parameters to remove small holes and objects. Then, during a fusion operation, the PASSRnet simply stacks the refined visibility map on top of the aligned features from the auxiliary view image, which may cause to diminish the effectiveness of the refined visibility map since its information may be ignored against the other information during the fusion operation.

On the other hand, our VA-SB in the RPC block is a dedicated module to estimate the two visibility attention maps: one for the left-view feature input and the other for right-view feature input. As shown in Fig. 3, each visibility attention maps (G_L or G_R) in unfolded RPC blocks are estimated via their shared VME module by fusing the two residues inputs of $|F_L^{(n-1)} - F_{Ra}^{(n-1)}|$ and $|F_R^{(n-1)} - F_{La}^{(n-1)}|$ crosswise from $F_L^{(n-1)}$ and $F_R^{(n-1)}$. By doing so, the two input sources from the main and auxiliary views are best fused to produce the feature outputs, $F_L^{(n)}$ and $F_R^{(n)}$, via G_L and G_R at each iteration. The feature outputs $F_L^{(n)}$ and $F_R^{(n)}$ are the visibility refinements (VR) for the outputs H_L and H_R of VFO via a shared densely connected VR module as shown in Fig. 3-(c). In the following visibility fusion operation, our VFO fuses the main view feature and the aligned auxiliary view feature by weighting them with the estimated visibility attention maps. This boosts the focus of attention for relatively important information on the main and aligned auxiliary view feature inputs.

Transferring. The extracted features that contain the parallax information *in a prior pair* ($I_L^{LR} \leftrightarrow I_R^{LR}$) should be *transferred* to the final stage to predict a stereo SR output pair by maintaining the obtained parallax information from the input *prior pair*. For this, we take a progressive *transferring* learning approach where our network could focus on learning low, middle and high frequency restorations through iterations at the unfolded stage of the RPC blocks in a recursive manner. Fig. 4 shows the interim features at evolved iterations in our effective *transferring* processing. Note that the previous methods [16-17] for stereo image SR simultaneously learn the information from the low frequency to the high frequency, which limits the network's performance due to the load of network's learning.

Parallax Information in an HR pair ($I_L^{HR} \leftrightarrow I_R^{HR}$)

In order to ensure the consistency between the parallax information *in an HR pair* ($I_L^{HR} \leftrightarrow I_R^{HR}$) and that of a predicted stereo SR output pair ($I_L^{SR} \leftrightarrow I_R^{SR}$), we propose a novel Parallax

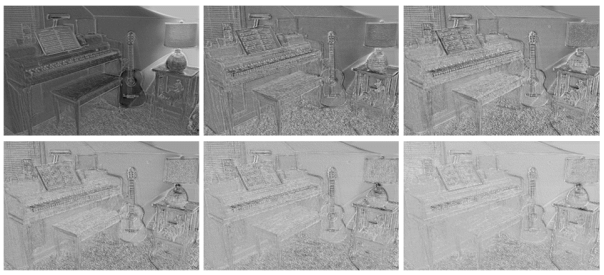


Figure 4: *Progressive learning method – Progressive six outputs of the unfolded Recursive Parallax Coherency (RPC) blocks from the top left to the bottom right. The output of the RPC block indicates a residual information where its amount is progressively reduced as the recursion evolves progressively. It is interesting to note that the RPC block progressively learns to restore higher frequency information.*

Coherency (PC) loss. Our proposed PC loss L_{PC} is defined as:

$$L_{PC\_diff} = \left\|(I_L^{SR} - I_R^{SR}) - (I_L^{HR} - I_R^{HR})\right\|_1 , \quad (1)$$

$$L_{PC\_mult} = \left\|(I_L^{SR} \cdot I_R^{SR}) - (I_L^{HR} \cdot I_R^{HR})\right\|_1 , \quad (2)$$

$$L_{PC} = L_{PC\_diff} + L_{PC\_mult} . \quad (3)$$

where (1) and (2) indicate the disparity and structural differences, respectively, due to the binocular parallax between SR and HR pairs. To the best of our knowledge, this is the first method to design the loss directly involving the parallax information *in an HR pairs* in stereo image SR problem.

4. Experiments

Dataset

For the datasets of LR and HR stereo image pairs, we used the stereo images from the Middlebury dataset [15], the KITTI2012 dataset [16], KITTI2015 dataset [17] and Flickr1024 dataset [12] which are commonly used for the stereo image tasks. Since the Middlebury dataset consists of the relatively higher resolution images than the others (KITTI2012, KITTI2015, Flickr1024), we down-scaled the images of the Middlebury dataset into half sizes in similar resolutions of [12], [16-17].

For the training images, we used 60 images from the Middlebury dataset [15] and 800 images from the Flickr1024 training dataset [12]. From the training samples, 164,040 and 19,952 patches of 40×120 size were extracted with stride 25 for scale factors 2 and 4, respectively. These patches were randomly flipped only horizontally for the data augmentation. Note that, the randomly vertical flipping and rotation were not performed due to epipolar constraints. For the testing images, we selected 5, 20 and 20 images from the Middlebury [15], KITTI2012 [16] and KITTI2015 [17] datasets respectively, which were not used in the training.

Implementation Details

All the convolution filters in the ProPaCoL-Net were initialized by the Xavier initialization [18] using both the number of input and output neurons of the layer. The parameters were updated using back propagation with a recently introduced optimizer *Lookahead* [20] with k =5, α = 0.5 combined with Adam optimizer [19] with β_1= 0.9, β_2= 0.999.

Table 1 summarizes the configuration of our ProPaCoL-Net used for the experiments. The ProPaCoL-Net consists of total 25

Table 1: The configuration of our ProPaCoL-Net.

Name			Function	Input	Output
conv0			3×3	H×W×3	H×W×64
conv1			3×3	H×W×64	H×W×64
RPC(1) ⋮ RPC(N)	A-SB	ResBA	H×W×64	H×W×64	H×W×64
		conA0	1×1	H×W×64	H×W×64
		conA1	1×1 Reshape	H×W×64	H×W×64
		$M_{R\to L}$ or $M_{L\to R}$	⊗	H×W×64 H×64×W	H×W×W
		conA2	1×1	H×W×64	H×W×64
		conA3	1×1	H×W×64	H×W×64
		Mat.Mul	⊗	H×W×W H×W×64	H×W×64
		conA4	1×1 L.ReLU	H×W×128	H×W×64
	VA-SB	conV0	3×3 L.ReLU	H×W×128	H×W×64
		conV1	3×3 L.ReLU	H×W×64	H×W×32
		ResBV	3×3 L.ReLU 3×3	H×W×32	H×W×32
		conV2	1×1	H×W×32	H×W×2
		CS	Channel Splitter	H×W×2	H×W×1 H×W×1
	VR	conR0	3×3 L.ReLU	H×W×64	H×W×64
		.....			
		conR6	3×3 L.ReLU	H×W×6·64	H×W×64
		conR7	1×1	H×W×7·64	H×W×64
conv2			1×1	H×W×64·N	H×W×64
conv3			1×1	H×W×64	H×W×64
PS0			1×1 Pixel shuffle	H×W×64	s·H×s·W×64
conv4			3×3	s·H×s·W×64	s·H×s·W×3
PS1			1×1 Pixel shuffle	H×W×3	s·H×s·W×64
conv5			3×3	s·H×s·W×64	s·H×s·W×3

convolution layers while the PASSRnet [12] and Stereo SR [11] use 45 and 61 convolution layers in total, respectively. The initial learning rate is set to 2×10^{-4}, where the learning rate is reduced with 0.95 times every epoch. The mini-batch size is 4. The training takes around 4 days using an Nvidia TITAN^{TM} XP GPU for the scale factor 2. The inference time on the same GPU was taken about 0.035ms for the scale factor 2 in yielding each 380×1200 stereo SR pair. All models were implemented using Pytorch package. The total epochs are 27 (1,100k iterations). Quantitative performances were measured in terms of Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Metric (SSIM).

5. Results

In this section, we compare our ProPaCoL-Net with recent single image SR methods and stereo image SR methods. Also, we perform an ablation study for the ProPaCoL-Net to inspect the effectiveness of its individual components via intensive experiments.

Comparison to the state-of-the-arts methods

The methods for comparison include two single image SR methods with SRCNN [9] and VDSR [10], and the state-of-the-

art (SOTA) stereo image SR methods with StereoSR [11] and the very recent state-of-the-arts PASSRnet [12]. We use three benchmark datasets [15-17] commonly used in stereo image tasks and the codes provided by the authors [9-12]. Table 2 shows the quantitative comparisons for the six methods under comparison where the highest PSNR and SSIM values are in red color and the

Table 2: Quantitative evaluation of our ProPaCoL-Net with state-of-the-arts methods for x2, x4 scale up factors on the commonly used stereo datasets [15-17].

Scales	Dataset	Metric	Bicubic	Single Image SR		Stereo Image SR		
				SRCNN [9]	VDSR [10]	StereoSR [11]	PASSRnet [12]	Ours
x2	Middlebury (5images)	PSNR	29.60	32.44	33.15	33.17	**34.14**	**35.29**
		SSIM	0.8509	0.9225	0.9318	0.9609	**0.9670**	**0.9732**
	KITTI2012 (20images)	PSNR	27.10	29.23	29.62	28.75	**30.07**	**32.47**
		SSIM	0.8589	0.9028	0.9090	0.9376	**0.9487**	**0.9530**
	KITTI2015 (20images)	PSNR	28.04	30.01	**30.31**	29.90	31.16	**31.38**
		SSIM	0.8863	0.9187	0.9231	0.9553	**0.9632**	**0.9664**
x4	Middlebury (5images)	PSNR	25.70	27.55	28.01	27.70	**28.62**	**29.04**
		SSIM	0.7362	0.7887	0.8051	0.8777	**0.8933**	**0.9005**
	KITTI2012 (20images)	PSNR	23.45	24.79	25.16	23.84	**25.43**	**25.53**
		SSIM	0.7037	0.7595	0.7757	0.8281	**0.8615**	**0.8645**
	KITTI2015 (20images)	PSNR	24.15	25.73	26.14	24.84	**26.58**	**26.75**
		SSIM	0.7504	0.7974	0.8141	0.8671	**0.8984**	**0.9022**

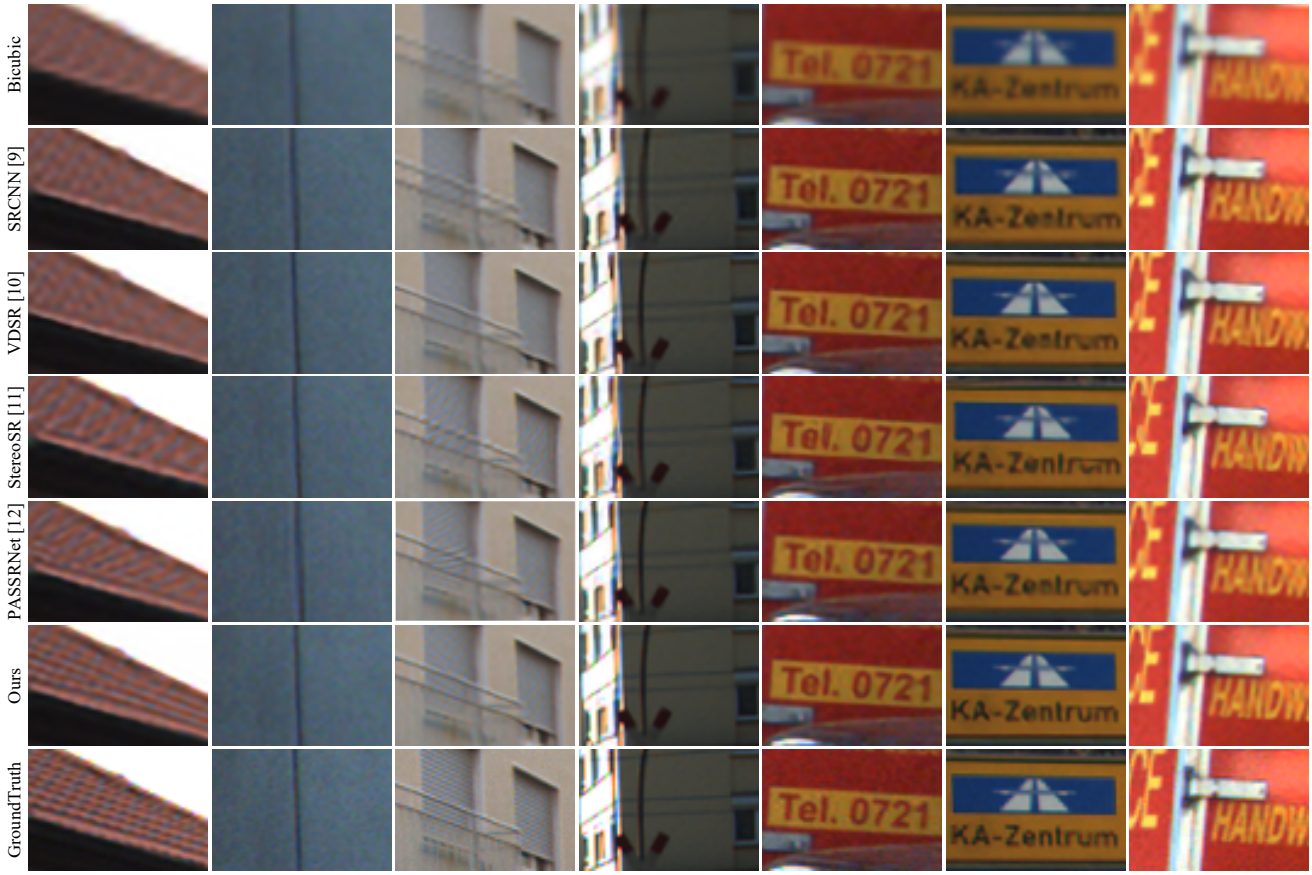


***Figure 5**: Visual comparison for x2 SR tested on the KITTI dataset. Ours shows the detail information such as the lines with higher contrast (1st~ 4th columns) and the letters with higher contrast (5th~7th columns). Especially, the two previous stereo image SR methods of StereoSR and PASSRnet show the aliasing artifacts in the highly repeated horizontal lines in the window (4th and 5th row of 3rd column) due to the wrong parallax inferencing. On the other hand, our ProPaCoL-Net does not suffer from such aliasing artifact (6th row of 3rd column).*

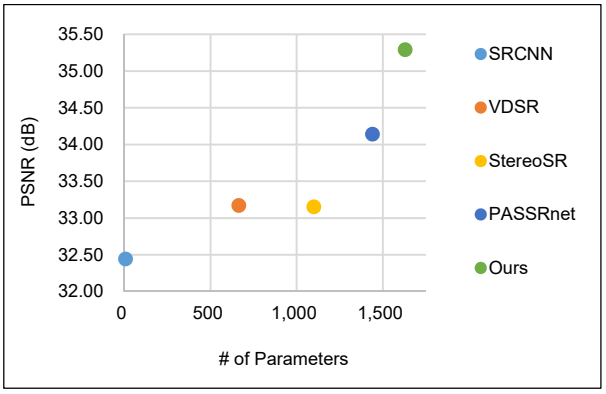


***Figure 6**: PSNR comparison in terms of the number of the network parameter on Middlebury dataset for x2 scale factor. Our method shows the much higher performance than PASSRnet with similar network size.*

Table 3: Ablation study tested on Middlebury dataset for x2 scale factor.

Models	PSNR	SSIM
ProPaCoL-Net-v1	35.22	0.9729
ProPaCoL-Net-v2	35.24	0.9731
ProPaCoL-Net-v3	35.16	0.9726
ProPaCoL-Net-v4	35.11	0.9723
ProPaCoL-Net (Baseline)	35.29	0.9732

second highest ones are in blue color. As shown in Table 2, our ProPaCoL-Net outperforms all the other methods for all experiments, yielding **1.15dB higher in PSNR** than the SOTA method, PASSRnet [12], for the Middlebury dataset at a scale factor 2. Fig. 5 shows the reconstructed left view SR images for visual comparison. Our method shows the clear lines and no aliasing artifacts, leading to perceptually plausible images. Finally, Fig. 6. shows the PSNR comparison results in terms of the number of network parameters, indicating the efficiency of network design. Our ProPaCoL-Net exhibits much higher performance than the PASSRnet in a similar network size.

Ablation Study

In order to see the effectiveness of the individual components in our ProPaCoL-Net, we implement different versions of the ProPaCoL-Net which are shown in Fig. 7 and described as follows:

- ProPaCoL-Net-v1: The cascade connection of A-SB and VA-SB in the RPC block is reconfigured into a parallel connection. See Fig. 7-(a).
- ProPaCoL-Net-v2: The combined VME module in the RPC block for the generation of the left and right visibility attention maps is reconfigured into two separate VME modules: one for $\boldsymbol{G}_L$ and the other for $\boldsymbol{G}_R$. See Fig. 7-(b).
- ProPaCoL-Net-v3: The VFO module is excluded. Instead, the three inputs, $\boldsymbol{F}_L$, $\boldsymbol{F}_{Ra}$ and $\boldsymbol{G}_L$ (or $\boldsymbol{F}_R$, $\boldsymbol{F}_{La}$ and $\boldsymbol{G}_R$) are stacked and the following 1×1 convolution results are then fed into the left (or right)-view VR module. See Fig. 7-(c).
- ProPaCoL-Net-v4: This version is the ProPaCoL-Net trained without the PC loss.

Table 3 shows the PSNR and SSIM performance for the different

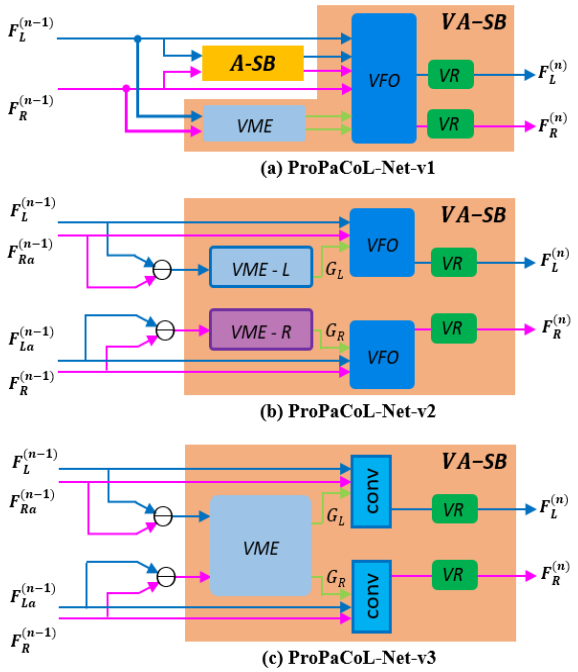


***Figure 7**: The different versions of ProPaCoL-Net: (a) ProPaCoL-Net-v1, (b) ProPaCoL-Net-v2, (c) ProPaCoL-Net-v3.*

versions of the ProPaCoL-Net. As shown for the ProPaCoL-Net-v4 in Table 3, the PC loss most significantly affects the stereo SR performance with 0.18 dB drop in PSNR when it is not used for training. As also shown for the ProPaCoL-Net-v3, the second most significant component is the VFO module, which drops 0.13 dB in PSNR without it. From the ProPaCoL-Net-v1 and ProPaCoL-Net-v2, it can be noted that our cascaded structure and the combined VME in the ProPaCoL-Net also effectively work. In conclusion, we ensure that our ProPaCoL-Net is well configured to appropriately generate stereo image SR images with parallax coherency well-maintained.

6. Conclusion

The parallax information in stereo image pairs is one of the important features for the stereo image SR problems. We, therefore, assume that to obtain the high quality stereo image SR outputs, the two parallaxes between the stereo SR output pairs and between HR pairs should be consistent. Our proposed ProPaCoL-Net learns the parallax coherency between the two parallaxes from the prior pairs. Especially, our novel RPC block effectively extracts the parallax information from the prior pairs and successively transfers it into deeper layers by a progressive learning method. The parallax learning is done in a mutually beneficial way between the left and right views by inputting the features crosswise from the left to right view input and vice versa. Also, we introduce an effective PC loss that helps the ProPaCoL-Net learn the parallax information in HR pairs. As a result, by learning the consistency of the intrinsic parallax in stereo images, our ProPaCoL-Net could effectively improve the performance of stereo image SR with faithful binocular parallax reconstructed in HR domain, outperforming the state-of-the-art PASSRnet with average 1.15dB higher PSNR.

References

[1] Z. Na, R. Liao, X. Tao, L. Xu, J. Jia, and E, Wu. Handling motion blur in multi-frame super-resolution, Proceedings of the Computer Vision and Pattern Recognition, pp. 5224-5232, 2015.

[2] C. Liu and D. Sun, On bayesian adaptive video super resolution, Proceeding of the IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 36, no 2, pp. 346-360, 2013

[3] R. Liao, X. Tao, R. Li, Z. Ma, and J. Jia, Video super resolution via deep draft-ensemble learning, Proceeding of the IEEE Inter-national Conference on Computer Vision, pp. 531–539, 2015.

[4] Radu Timofte, Vincent De Smet, Luc Van Gool, Anchored Neighborhood Regression for Fast Example-Based Super-Resolution, Proceeding of the IEEE International Conference on Computer Vision, pp. 1920-1927, 2013.

[5] Radu Timofte, Vincent De Smet, Luc Van Gool, A+: Adjusted Anchored Neighborhood Regression for Fast Super-Resolution, Proceedings of the Asian Conference on Computer Vision, pp. 111-126, 2014.

[6] Daniel Glasner, Shai Bagon, Michal Irani, Super-Resolution from a Single Image, Proceedings of the IEEE International Conference on Computer Vision, pp. 349-356, 2009.

[7] Jianchao Yang, John Wright, Thomas S. Huang, Yi Ma, Proceeding of the IEEE Transactions on Image Processing, 19(11), pp. 2861-2873, 2010.

[8] Samuel Schulter, Christian Leistner, Horst Bischof, Fast and Accurate Image Upscaling with Super-Resolution Forests, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3791-3799, 2015.

[9] Chao Dong, Chen Change Loy, Kaiming He, Xiaoou Tang, Image Super-Resolution Using Deep Convolutional Networks, Proceeding of the IEEE Transactions on Pattern Analysis and Machine Intelligence, 38(2), pp. 295-307, 201

[10] Jiwon Kim, Jung Kwon Lee, Kyoung Mu Lee, Accurate Image Super-Resolution Using Very Deep Convolutional Network, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1646-1654, 2016.

[11] Daniel S. Jeon, Seung-Hwan Baek, Inchang Choi, and Min H. Kim, Enhancing the spatial resolution of stereo images using a parallax prior, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp.1721-1730, 2018.

[12] Longguang Wang, Yingqian Wang, Zhengfa Liang, Zaiping Lin, Jungang Yang, Wei An, Yulan Guo, Learning Parallax Attention for Stereo Image Super-Resolution, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019.

[13] Z. Cui, H. Chang, S. Shan, B. Zhong, and X. Chen, Deep network cascade for image super-resolution. Proceedings of the European Conference on Computer Vision, pp. 49–64, 2014.

[14] Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu, Image Super-Resolution Using Very Deep Residual Channel Attention Networks, Proceedings of the European Conference on Computer Vision, 2018.

[15] Daniel Scharstein, Heiko Hirschm¨uller, York Kitajima, Greg Krathwohl, Nera Nesic, Xi Wang, Porter Westling, High-resolu-tion stereo datasets with subpixel-accurate ground truth, Proceed-ing of the German Conference on Pattern Recognition, volume 8753, pp 31–42, 2014.

[16] Andreas Geiger, Philip Lenz, Raquel Urtasun, Are we ready for autonomous driving? The KITTI vision benchmark suite, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3354–3361, 2012.

[17] Moritz Menze, Andreas Geiger, Object scene flow for autonomous vehicles, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3061–3070, 2015.

[18] Xavier Glorot, Yoshua Bengio, Understanding the difficulty of training deep feedforward neural networks, Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, 9:249-256, 2010.

[19] Diederik P. Kingma, Jimmy Ba, Adam: A Method for Stochastic Optimization, Proceeding of the International Conference for Learning Representations, 2015.

[20] Michael R. Zhang, James Lucas, Geoffrey Hinton, Jimmy Ba, Lookahead Optimizer: k steps forward, 1 step back, https://arxiv.org/-abs/1907.08610, 19 Jul 2019.

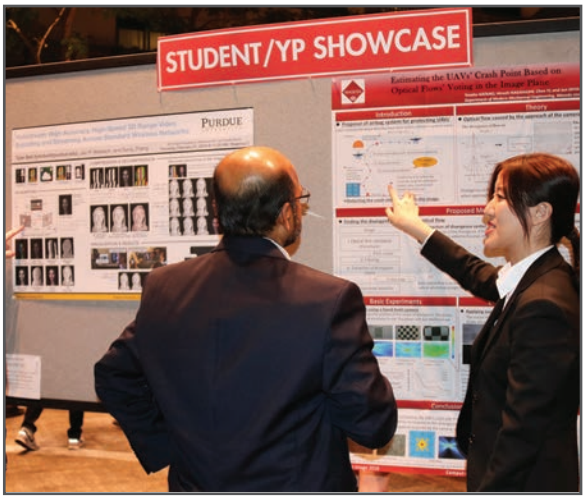
JOIN US AT THE NEXT EI!

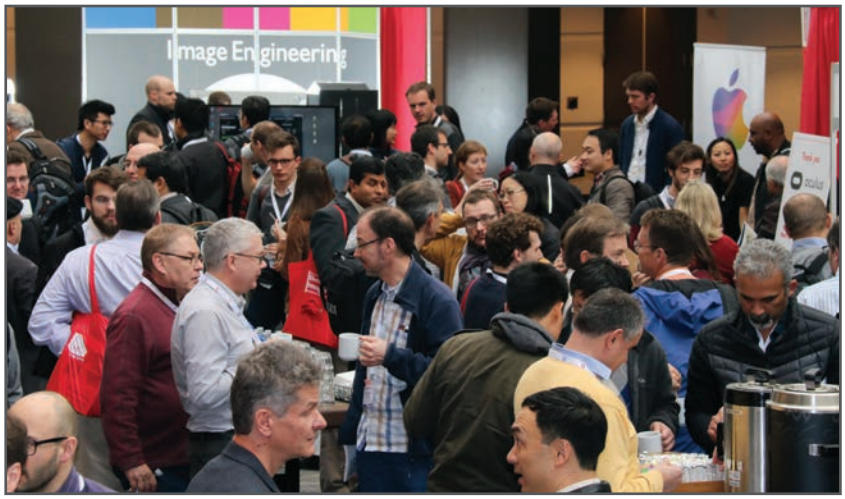
IS&T International Symposium on

Electronic Imaging

SCIENCE AND TECHNOLOGY

Imaging across applications . . . Where industry and academia meet!





• SHORT COURSES • EXHIBITS • DEMONSTRATION SESSION • PLENARY TALKS •
• INTERACTIVE PAPER SESSION • SPECIAL EVENTS • TECHNICAL SESSIONS •

www.electronicimaging.org

IS&T
imaging.org