Improving Stereo Matching Accuracy with Blind Super-Resolution Networks

Dinghao Zheng[1], Jiangtao Peng[2], Qiwei Xie[1], and Qian Long[1(✉)]

[1] Beijing Smarter Eye Technology Co., Ltd, Beijing 100085, China
longqian@tust.edu.cn
[2] Hubei Key Laboratory of Applied Mathematics, Faculty of Mathematics and Statistics, Hubei University,Wuhan 430062, China

Abstract. Due to the limitation of the camera's photoreceptor, the stereo images obtained from stereo cameras can only be presented in the form of pixel blocks, which cannot show the more minute object information in sub-pixel units between the pixel blocks, which may restricts the breakthrough of stereo matching in sub-pixel accuracy. For this reason, we innovatively use a blind super-resolution network to simulate the sub-pixel effect and improve the accuracy of stereo matching. We combine the super-resolution network with the stereo matching and design a new stereo matching process: the stereo image is first zoomed in by the blind super-resolution network to supplement the local information, and then the high-resolution image is inputted into the stereo matching algorithm to generate a dense and fine disparity map. Through experiments, we find that the blind super-resolution network BSRGAN can effectively improve the stereo matching accuracy in most scenarios. However, when facing repetitive and dense texture regions, the limitation of blind super-resolution network leads to a degradation of the matching accuracy. Nevertheless, our study provides a new idea and method for improving stereo matching accuracy.

Keywords: Stereo Matching, Blind Super-Resolution, Sub-Pixel

1   Introduction

Stereo matching is an important research direction in the field of computer vision, which aims to recover the depth information of a scene from two or more images with different viewpoints. This technique has a wide range of applications and importance in many fields such as 3D reconstruction, autonomous driving, and robot navigation. Given a pair of rectified stereo images, the goal of stereo matching is to compute the disparity d for each pixel in the reference image. Disparity refers to the horizontal displacement between a pair of corresponding pixels on the left and right images. For the pixel (x, y) in the left image, if its corresponding point is found at $(x - d, y)$ in the right image, then the depth of this pixel is calculated by $\frac{fB}{d}$, where f is the camera's focal length and B is the distance between two camera centers.

A typical stereo matching algorithm consists of four steps: matching cost calculation, cost aggregation, disparity calculation, and disparity optimization. In order to compute a more accurate disparity, many algorithms are designed and improved on these steps.

The SGM [1] improves the accuracy of the disparity by using sub-pixel optimization because the disparity maps obtained from disparity computation are integer pixels, which cannot meet the accuracy requirements in many applications, SGM uses quadratic curve interpolation to obtain sub-pixel accuracy, which fits a quadratic curve to the surrogate value of the optimal disparity and the possible values of the two disparities before and after it, and the disparity value corresponding to the extreme value point of the curve is the new sub-pixel disparity value. With the rapid development of deep learning, deep learning has been applied to stereo matching, deep learning stereo matching algorithm will build a parametric number of huge cost volume to help the network to find the possible disparity. Several aggregation-based methods [2-5] have shown significant progress in the domain of stereo matching in recent years. Gc-Net [6] firstly proposes an end-to-end supervised learning method to learn disparity maps from stereoscopic images directly. PSMNet [7] proposes a pyramidal stereo matching network consisting of spatial pyramid pooling and 3D CNN. The spatial pyramid pooling module uses global contextual information to create a cost volume through contextual aggregation at different scales and locations. IGEV-Stereo [8] uses an extremely lightweight 3D regularisation network to aggregate and regularise the cost volume, to obtain a Geometric Encoder Voxel (GEV), and ultimately combining the GEV with the all-pairs correlations in RAFT-Stereo [9]. However, these methods are all based on integer pixels for disparity calculation, which undoubtedly increases the difficulty of calculating more accurate disparities and makes the sub-pixel accuracy of stereo matching challenging. In view of this, we have to further explore other methods to assist stereo matching to compute disparity more accurately, thus breaking through the existing accuracy bottleneck.

Image super-resolution networks has very important application value in the field of computer vision, the essence of image super-resolution is essentially a more detailed and dense representation of the original object. Dong et al. [10] propose SRCNN to learn the mapping from LR to HR images in an end-to-end manner, achieving superior performance against previous works. With the proposal of ResNet [11], RCAN [12] got good results in PSNR metrics by designing very deep residual structures. In order to improve visually, ESRGAN [13] borrowed the idea of GAN [14], and these networks do not use PSNR value as the main metric. Due to the many and complex degradations in the real world, it is not enough to just use bicubic to simulate high resolution to low resolution images, so in order to better simulate the degradations in the real world, so that the algorithm can work better in real world images, SRMD [15] proposes the idea of blind super-resolution, where SRMD splices the input image with degradation information and feeds it into a super-resolution model, which is the first blind image super-resolution method using deep learning. With the development of blind super-resolution networks, Real-ESRGAN [16] and BSRGAN [17] generate low-resolution blurred images from high-resolution images through multiple steps such as blurring, downsampling, and adding noise, and ultimately use these synthetic datasets for training. In addition, there are other blind super-resolution networks[18-19] that have contributed to blind super-resolution work.

In recent years, super-resolution networks have been widely used in other fields of computer vision. SSSR [20] combines image super-resolution algorithms with semantic

segmentation algorithms to improve segmentation performance. Noh et al. [21] proposes new feature-level super-resolution methods for improving small object detection performance. However, no scholars have yet directly combined super-resolution networks with stereo matching.

Due to the limitation of the camera's sensor, the stereo images we get can only be shown in the form of pixel blocks, which cannot show the more tiny object information in sub-pixel units between the pixel blocks, which limits the breakthrough of stereo matching in sub-pixel accuracy. In this paper, we improve disparity by combining super-resolution network and stereo matching algorithm, and enhance the sub-pixel accuracy and visual effect of disparity by simulating sub-pixel effect through blind super-resolution network.

Our main contributions are listed below:

- We are the first to combine blind super-resolution network with stereo matching, and theoretically analyze the influence and significance of blind super-resolution networks on stereo matching.
- We have designed a new process to assist stereo matching in calculating more accurate disparities, and provided practical application methods.
- Through experiments, we found that using blind super-resolution network BSRGAN can improve the accuracy of stereo matching in most scenarios. However, in some specific scenarios, BSRGAN can lead to worse stereo matching results. Through further research, we discovered that this is due to the poor performance of the super-resolution network in areas with highly repetitive and dense textures.

2   Theoretical Analysis

In this section we discuss the effect of super-resolution networks on stereo matching and analyse and conclude that effective super-resolution networks will have a positive impact on stereo matching to improve sub-pixel accuracy.

In the process of camera imaging, the image data obtained is a discretisation of the image, and due to the limitations of the capabilities of the sensory elements themselves, to the imaging surface each pixel represents only a nearby colour. For example, there is a 4.5 um spacing between the pixels on two sensory originals, macroscopically they are connected, microscopically there are countless tiny things that exist between them, these pixels that exist between the two actual physical pixels are known as sub-pixels, as shown in Fig. 1. Sub-pixels are actually supposed to exist, but there is just a lack of smaller sensors to detect them, so they can only be approximated in software. Maybe the super-resolution network can help the camera with additional information.
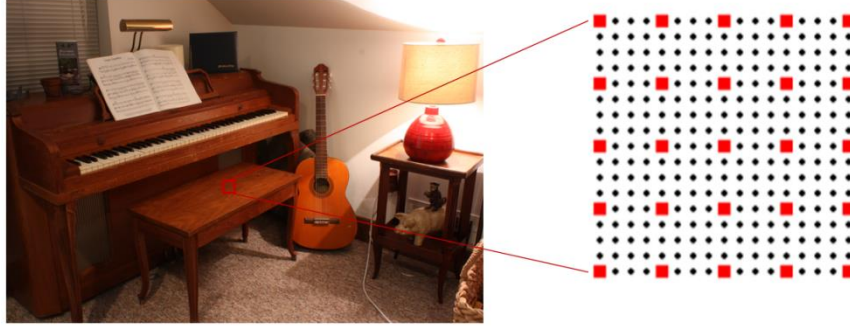
Fig. 1. The red block pixels represent the pixel blocks of the image, and the black circles represent tiny sub-pixel information that cannot be detected by the optical sensor.

To demonstrate the impact of super-resolution on stereo matching, let's consider an example as shown in Fig. 2. Before super-resolution, theoretically, when the red pixel is correctly matched, the disparity corresponding to that point is 1 pixel. However, after super-resolution, since the image is enlarged by a factor of 2, the step length required for correct pixel matching also needs to be doubled. Therefore, in theory, when the red pixel is correctly matched after super-resolution, the disparity corresponding to these four red pixels should be 2 pixels each. In other words, the disparity doubles after super-resolution. Consequently, after matching, we need to divide the obtained disparity by 2 to accurately represent the original scene.

However, in actual matching, the effect of the blind super-resolution network can result in closely similar pixel values among the newly generated four pixels. This can lead to the possibility of the stereo matching algorithm incorrectly matching a pixel point to another pixel point on the same row. For instance, point a might match to point c, or it might match to point d. In such cases, the disparity calculated for point a could be 2 pixels or 3 pixels. Similarly, point b might also match to point c or d, resulting in a disparity for point b that could be 1 pixel or 2 pixels. To accurately represent the original scene, we ultimately need to divide the calculated disparity by 2. Therefore, the disparity of the red pixel, which was originally 1 pixel, could potentially become 0.5 pixels, 1 pixel, or 1.5 pixels after super-resolution. By this reasoning, we can deduce a pattern: when the disparity of a certain pixel in the original image is i pixels, the disparity of these four pixels after super-resolution could potentially be i-0.5 pixels, i pixels, or i+0.5 pixels.
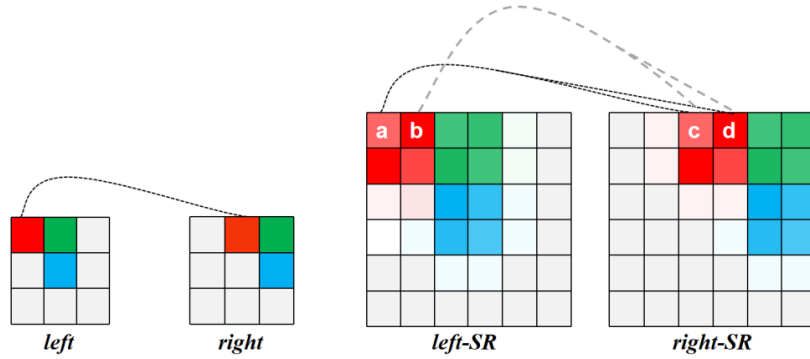
Fig. 2. The left and right show local regions located at the same coordinate positions in the left and right images, respectively, while the left-SR and right-SR represent the right image as well as the right image enlarged in twice by the blind super-resolution network, respectively.

Due to the design of deep learning stereo matching algorithms in terms of disparity cost calculation and disparity estimation, these algorithms will ultimately weight some possible disparities for summing or fitting, but the principle is to choose the best possible disparities to represent the disparity of the pixel as much as possible, if we apply the super-resolution mechanism to deep learning algorithms, the possible disparities the algorithms can choose from will be more fine-tuned in the range of the original possibilities. For example, as shown in Fig. 3, in the original disparity options may be 17, 18, or 19 pixels, but with the addition of the super-resolution algorithm, the disparity of the first pixel diffused by the super-resolution algorithm may be 15.5, 16, 16.5, 17, 17.5, or 18 pixels, and the disparity of the second pixel diffused may be 16, 16.5, 17, 17.5, 18, or 18.5 pixels. Then overall the diffused new pixels their disparity options might be 16.5, 17, 17.5, 18, 18.5, 19, or 19.5 pixels, which would calculate a finer disparity, and also open up the possibility of a breakthrough in stereo matching sub-precision.
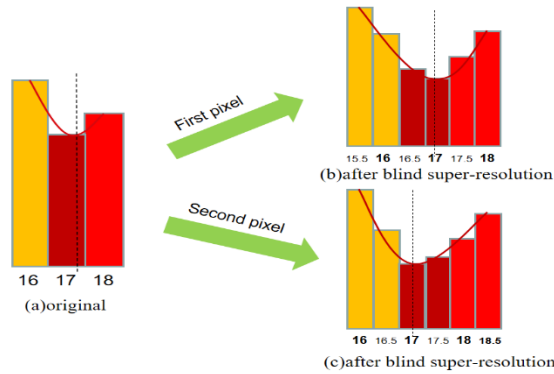

Fig. 3. (a) represents the situation of stereo matching fitting the disparity of a certain pixel. (b) and (c) depict the stereo matching fitting of disparities for two newly diffused pixels on the same row after applying the blind super-resolution network.

3   Method

We design a new process to improve the accuracy of stereo matching by enlarging the left and right maps through the blind super-resolution network BSRGAN, and calculating more accurate and denser disparity maps with the help of larger and clearer left and right images, so as to better represent the real world, as shown in Fig. 4.
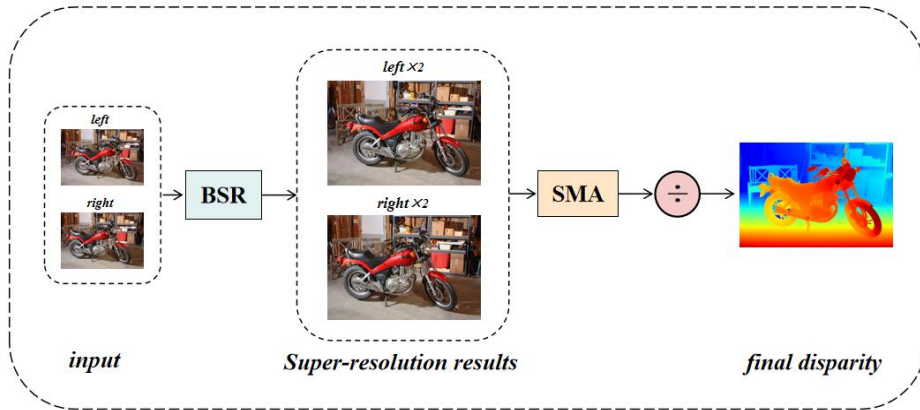

Fig. 4. Architecture overview of proposed method. Low-resolution left and right images are inputted into the Blind Super-Resolution Generative Adversarial Network (BSRGAN), obtaining high-resolution left and right images that are enlarged twice in size. At this point, the obtained high-resolution left and right images are inputted into the Stereo Matching Algorithm (SMA) to generate a disparity map for the scene. Finally, the disparity map is divided by 2 to obtain a dense disparity map that accurately represents the original scene.

In practical applications, using the aforementioned method, we obtain a new, denser, and more accurate disparity map that better represents the real-world scene.

4   Experiment

4.1   Experiment Details

Experimental Data. The experiment adopts the Middlebury 2014, which contains left and right images of different resolutions for the same scenes, along with corresponding disparity maps. In this experiment, the low-resolution left and right images and their corresponding ground truth of disparity are sourced from scenes with a resolution of 750×500, while the ground truth of disparity for the super-resolution images are derived from scenes with a resolution of 1500×1000.Due to the fact that the maximum disparity trained by the current stereo matching algorithm model is set to 192 by default, when the maximum disparity of low-resolution exceeds 96, the disparity of high-resolution will be greater than 192. Therefore, in the experimental process, the scenes from jadeplant (original

image maximum disparity of 160) and Vintage (original image maximum disparity of 190) will be excluded, while the remaining 13 scenes will be retained as experimental data.

Stereo Matching Algorithm. The stereo matching algorithm used for the experiments is the IGEV-Stereo which has excellent performance on public datasets, due to the powerful memory capability of deep learning, it is not appropriate to make comparisons on trained datasets, in order to make comparisons in a fairer way, The weights of the IGEV-Stereo trained on Scene Flow were chosen to be tested on the Middlebury 2014.

Blind Super-resolution Networks. We selected four blind super-resolution networks to participate in this experiment. The model used in the experiment is the blind super-resolution network BSRGAN trained on the produced multiple types of degraded data with ESRGAN as the baseline model, it can adapt to more stereo scenes with good generalisation ability and visual effect, since the maximum search range of the stereo matching algorithm is 192, we only use the weights of BSRGAN to enlarge the image by 2 times. We use other visually effective blind super-resolution network Real-ESRGAN, BSRMD and SRResCGAN as comparison networks.

4.2 Evaluation Metrics

Relative Error Rate(RER). we generally choose the mean absolute error to measure disparity:

$$\frac{1}{N}\sum_{(x,y)\in N}\left|d_{est}(x,y)-d_{gt}(x,y)\right| \tag{1}$$

Where N represents that the image has a total of N pixel points, (x,y) denotes a certain pixel point, $d_{est}(x,y)$ is the predicted disparity of the pixel point (x,y), and the true disparity of the pixel point (x,y) is $d_{gt}(x,y)$.

However, when the left and right images are processed with blind super-resolution network, the computed disparity is twice the original disparity, and the corresponding disparity error also increases to twice the original disparity. In this situation, if we still adopt the mean absolute error as the evaluation criterion, it is obviously not appropriate enough. In order to reflect more intuitively the changes in disparity before and after super-resolution and to facilitate comparison, we decided to use the relative error rate to judge the disparity advantages and disadvantages before and after blind super-resolution.

$$\frac{1}{N}\sum_{(x,y)\in N}\left|\frac{d_{est}(x,y)-d_{gt}(x,y)}{d_{gt}(x,y)}\right| \tag{2}$$

Match Error Rate(MER). For the original image we specify that the pixel error is within 1pixel for correct matching. When the image is processed by the blind super-resolution network, the step size of stereo matching will be enlarged by a factor of 2,

and the corresponding error will be enlarged by a factor of 2. At the same time, we will divide the final computed disparity by 2 as a practical application, and therefore we stipulate that the pixel error after super-resolution is within 2pixels as a correct match.

$$\frac{1}{N}\sum_{(x,y)\in N}\{|d_{est}(x,y) - d_{gt}(x,y)| > \delta_D\} \tag{3}$$

Where δ_D is taken as 1.0 and 2.0 respectively.

Distance Relative Error Rate(DER). To measure the accuracy of ranging, similar to the relative error rate, we calculate the distance relative error rate：

$$\frac{1}{N}\sum_{(x,y)\in N}\left|\frac{D_{est}(x,y)-D_{gt}(x,y)}{D_{gt}(x,y)}\right| \tag{4}$$

Where $D_{est}(x,y)$ is the predicted depth of the pixel point (x,y), and the ground truth of depth of the pixel point (x,y) is $D_{gt}(x,y)$.

The conversion between disparity(d) and depth(D) is expressed as:

$$\frac{Bf}{d} = D \tag{5}$$

Where f is the camera's focal length and B is the distance between two camera center.

To facilitate the calculation of the relative error of distance, we can simplify (4) according to (5)：

$$\frac{1}{N}\sum_{(x,y)\in N}\left|\frac{D_{est}(x,y)-D_{gt}(x,y)}{D_{gt}(x,y)}\right| = \frac{1}{N}\sum_{(x,y)\in N}\left|\frac{d_{est}(x,y)-d_{gt}(x,y)}{d_{est}(x,y)}\right| \tag{6}$$

4.3   Results

Based on the results on Middlebury2014, we found that combining real-ESRGAN and BSRGAN with the stereo matching algorithm IGEV-Stereo can reduce both the relative error rate and the match error rate. Notably, the best performance was achieved when BSRGAN was combined with IGEV-Stereo, However, we found that when BSRGAN is combined with IGEV-Stereo the distance relative error rate does not get smaller, but rather it becomes larger, as shown in Table 1. RER and MER are used as the main indicators, DER is susceptible to extreme values, so it is an auxiliary indicator.

Table 1. Experimental Results on Middlebury 2014.

Method	RER(%)	MER(%)	DER(%)
IGEV-Stereo	3.17	12.82	4.62
BSRDM+IGEV-Stereo	3.98	18.18	**4.09**
SRResCGAN+IGEV-Stereo	3.40	12.56	5.44
Real-ESRGAN+IGEV-Stereo	3.11	12.59	6.88
BSRGAN+IGEV-Stereo	**2.85**	**12.17**	4.87

According to Fig. 5, BSRGAN improves the visual effect of disparity maps computed by stereo matching algorithm and improves the ability of stereo matching algorithm to capture details. Specifically, we can look at the error map of the disparity map, the bluer the error map indicates that the error in the region is smaller, and the disparity error obtained by processing IGEV-Stereo through BSRGAN is smaller.
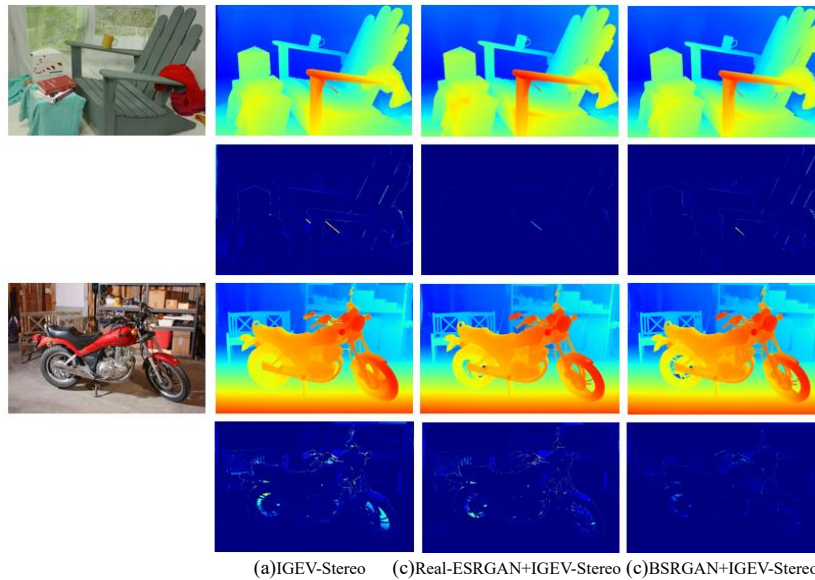

(a)IGEV-Stereo   (c)Real-ESRGAN+IGEV-Stereo (c)BSRGAN+IGEV-Stereo

Fig. 5. Results of disparity estimation for Middlebury 2014.The first and third rows represent the scene and the corresponding disparity map. The second and fourth rows correspond to the error maps of the disparity.

4.4 Cause Analysis

In order to figure out why DER has become larger on the premise that RER and MER have decreased, we list the experimental results for each scene. As shown in Table 2，take the best-performing combination BSRGAN+IGEV-Stereo as an example, compared with IGEV-Stereo alone, it is noteworthy that after being processed by BSRGAN, both the piano and playtable experienced a decrease in accuracy. For convenience, we will refer to IGEV-Stereo simply as IGEV.

Table 2. The experimental results for each scene in Middlebury 2014.

Scenes	RER (%)		MER (%)		DER (%)	
	IGEV	BSRGAN +IGEV	IGEV	BSRGAN +IGEV	IGEV	BSRGAN +IGEV
Adirondack	1.77	1.17	4.41	1.71	4.13	0.92
ArtL	3.10	2.81	16.32	13.12	10.32	15.85
Motorcycle	2.90	2.01	10.02	7.19	2.62	2.05
MotorcycleE	2.92	2.09	10.31	7.51	2.64	2.06
Piano★	2.32	**2.77**	9.13	**13.09**	2.09	**5.02**
PianoL	7.24	7.07	28.84	29.40	12.23	16.87
Pipes	6.23	5.45	16.81	13.99	6.01	4.64
Playroom	2.74	2.07	17.20	12.69	4.91	2.55
Playtable★	1.71	**2.09**	8.90	**10.68**	1.68	**2.35**
PlaytableP★	1.63	**2.08**	8.44	**10.77**	1.58	**2.51**
Recycle	1.87	1.83	7.96	9.09	2.48	1.57
Shelves	3.32	3.62	18.38	21.49	5.89	4.85
Teddy	3.43	1.95	9.92	7.48	3.49	2.09

After our careful comparison and analysis, we found that the super-resolution network has some limitations in some regions, which leads to poor super-resolution effect and indirectly to worse stereo matching. According to Fig. 6. BSRAGN and Real-ESRGAN are both obtained by using ESRGAN model as the baseline training, these models improve the image resolution and visual effect, but also accompanied by artifacts and excessive smoothing problems, these problems are also the current research of super-resolution network faces the problem. The most important thing in stereo matching is to find the points corresponding to the left and right images, after which the disparity value is calculated through the positional relationship, when the above situation occurs, it will probably lead to the inconsistency of the super-resolution effect of the corresponding pixels of the left and right images, which will lead to the inconsistency of the left and right images, and thus the matching error occurs in the process of stereo matching. Nevertheless, blind super-resolution can improve disparity calculation in most scenarios.
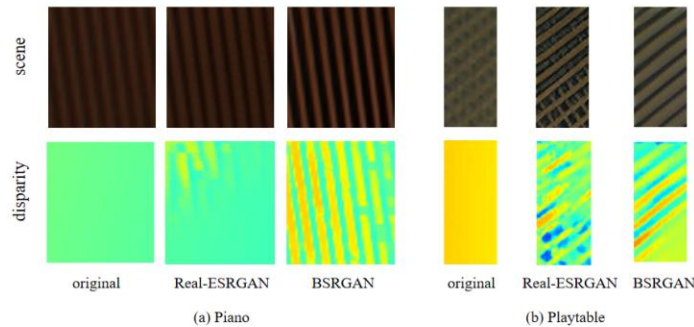
Fig. 6. Image texture and disparity map of certain regions of the image before and after super-resolution networks processing.

5   Conclusion

We propose an innovative approach to improving stereo matching accuracy using blind super-resolution networks. Due to limitations of camera photoreceptors, stereo images are presented in pixel blocks, lacking sub-pixel details which restrict stereo matching accuracy. We discuss and conclude that effective super-resolution networks positively impact stereo matching and enhancing sub-pixel accuracy.

Our method uses a blind super-resolution network to simulate sub-pixel effects and improve stereo matching accuracy. The process involves upscaling stereo images with a super-resolution network to enhance local details, then inputting the high-resolution images into a stereo matching algorithm to generate detailed disparity maps. Experiments show that the BSRGAN significantly improves stereo matching accuracy, particularly when combined with IGEV-Stereo, reducing both relative error and matching error rates. However, in regions with repetitive and dense textures, the limitations of blind super-resolution networks can degrade accuracy. Specifically, BSRGAN and Real-ESRGAN, based on ESRGAN models, improve image resolution and visual quality but can introduce artifacts and excessive smoothing. These issues can cause inconsistencies between the left and right images, leading to matching errors.

In summary, improving stereo matching accuracy with the help of blind super-resolution networks is a promising research direction. By continuously optimising the blind super-resolution networks, we are expected to achieve more accurate and efficient stereo matching.

References

1. Hirschmuller, H.: Accurate and efficient stereo processing by semi-global matching and mutual information. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 807-814(2005)
2. Xu, G., Cheng, J., Guo, P., Yang, X.: Attention concatenation volume for accurate and efficient stereo matching. In: 2022 IEEE/CVF conference on computer vision and pattern recognition, pp. 12981-12990(2022)
3. Zhang, J., Wang, X., Bai, X., Wang, C., Huang, L., Chen, Y., Gu, L., Zhou, J., Harada, T., Hancock, E. R.: Revisiting domain generalized stereo matching networks from a feature consistency perspective. In: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 13001-13011(2022)
4. Liu, B., Yu, H., & Qi, G.: Towards domain generalized stereo matching with a broad-spectrum and task-oriented feature. In: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 13012-13021(2022)
5. Xu, G., Wang, Y., Cheng, J., Tang, J., & Yang, X.: Accurate and efficient stereo matching via attention concatenation volume. In: 2023 IEEE Transactions on Pattern Analysis and Machine Intelligence(2023)
6. Kendall, A., Martirosyan, H., Dasgupta, S., Henry, P., Kennedy, R., Bachrach, A., Bry, A.: End-to-end learning of geometry and context for deep stereo regression. In: 2017 IEEE international conference on computer vision, pp. 66-75(2017)
7. Chang, J. R., Chen, Y.: Pyramid stereo matching network. In: 2018 Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 5410-5418(2018)

8. Xu, G., Wang, X., Ding, X., Yang, X.: Iterative geometry encoding volume for stereo matching. In: 2023 Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 21919-21928(2023)
9. Lipson, L., Teed, Z., Deng, J.: Multilevel recurrent field transforms for stereo matching. In: 2021 International Conference on 3D Vision (3DV), pp. 218-227(2021)
10. Dong, C., Loy, C. C., He, K., Tang, X.: Learning a deep convolutional network for image super-resolution. In: Computer Vision–ECCV 2014, pp. 184-199(2014)
11. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In 2016 IEEE conference on computer vision and pattern recognition, pp. 770-778(2016)
12. Zhang, Y., Li, K., Li, K., Wang, L., Zhong, B., Fu, Y.: Image super-resolution using very deep residual channel attention networks. In: 2018 European conference on computer vision (ECCV), pp. 286-301(2018)
13. Wang, X., Yu, K., Wu, S., Gu, J., Liu, Y., Dong, C., Qiao,Y., Change Loy, C.: Enhanced super-resolution generative adversarial networks. In: 2018 European conference on computer vision (ECCV) workshops, pp. 0-0(2018)
14. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courvill, A., Bengio, Y.: Generative adversarial nets. In: neural information processing systems 27 (2014)
15. Zhang, K., Zuo, W., Zhang, L.: Learning a single convolutional super-resolution network for multiple degradations. In: 2018 IEEE conference on computer vision and pattern recognition, pp. 3262-3271(2018)
16. Wang, X., Xie, L., Dong, C., Shan, Y.: Real-esrgan: Training real-world blind super-resolution with pure synthetic data. In: 2021 IEEE/CVF international conference on computer vision, pp. 1905-1914(2021)
17. Zhang, K., Liang, J., Van Gool, L., Timofte, R.: Designing a practical degradation model for deep blind image super-resolution. In: 2021 IEEE/CVF International Conference on Computer Vision, pp. 4791-4800(2021)
18. Yue, Z., Zhao, Q., Xie, J., Zhang, L., Meng, D., Wong, K. Y. K.: Blind image super-resolution with elaborate degradation modeling on noise and kernel. In: 2022 IEEE/CVF conference on computer vision and pattern recognition, pp. 2128-2138(2022)
19. Umer, R. M., Foresti, G. L., Micheloni, C.: Deep generative adversarial residual convolutional networks for real-world super-resolution. In: 2020 IEEE/CVF conference on computer vision and pattern recognition workshops, pp. 438-439(2020)
20. Wang, L., Li, D., Zhu, Y., Tian, L., Shan, Y.: Dual super-resolution learning for semantic segmentation. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 3774-3783(2020)
21. Noh, J., Bae, W., Lee, W., Seo, J., Kim, G.: Better to follow, follow to be better: Towards precise supervision of feature super-resolution for small object detection. In: 2019 IEEE/CVF international conference on computer vision, pp. 9725-9734(2019)