ISBN 978-981-14-1684-2
Proceedings of 2019 the 9th International Workshop on Computer Science and Engineering
(WCSE 2019 SUMMER)
Hong Kong, 15-17 June, 2019, pp. 219-226
doi: 10.18178/wcse.2019.06.033

Super-Resolution for Mixed-quality Stereo Images based on Patch Matching

Chengtao Cai[1], Bing Fan [1 +]and Haiyang Meng [1]

[1] College of Automation, Harbin Engineering University, Harbin, China

Abstract. We propose a novel method for mixed-quality stereo images super-resolution (SR) based on patch matching. Previous related methods always put effort on disparity estimation which cannot achieve the accuracy for SR. In this paper, we directly utilize the information from the high-resolution (HR) image to reconstruct the low-resolution (LR) image. More specifically, the vanishing point estimation algorithm is adopted to identify the planes correspondence in the pair of images. Then we search the best matching patch for LR from HR in the corresponding planar area. Furthermore, we define a curvature criterion that can keep the patches with high-frequency information during patch matching process. Compared with state-of-the-art methods, the proposed framework gains 1.39 PSNR improvement and 0.003 SSIM improvement.

Keywords: Stereo images, super-resolution, patch matching.

1. Introduction

Stereo images draw much attention in surveillance-oriented applicative fields since they are able to provide users a realistic 3D viewing experience of a scene [1]–[3]. For the past decades, most researchers focus on the disparity estimation from stereo images [4]–[6]. Stereo images are usually acquired by vertically-aligned cameras or horizontally -aligned cameras.

Stereo images can be classified into homogeneous stereo images[5][6] and heterogeneous stereo images [7] [8], [9]. The first one has the same image size which is obtained by two same vision sensors. The other one has different size and quality which can be generated by two situations:1) Stereo vision system with different vision sensors which aim at combining the advantage of a high-resolution and large field of view [8], [9]. 2) Asymmetric compression [7], i.e. two views are encoded with different spatial resolutions (resolution-asymmetry) in the communication or telepresence system to accelerate transmission rate or reduce storage. Although such heterogeneous images have the aforementioned advantages, they also pose challenges for image registration which is an essential step during the process of stereo images. To solve this issue, super-resolution (SR) for the low-resolution image reconstruction is demanded, which is able to normalize the heterogenous image pair. Zhi Jin [7] propose an end-to-end fully Convolutional Neural Network (CNN) for low-resolution images in mixed-quality 3D stereo images with depth maps by using the inter-view from 3D warping. However, accurate depth maps are hard to acquire during practical applications. Jain [10] reconstruct the low-quality video in mixed resolution videos by using the stereo correspondence, but pixel level accuracy cannot achieve the requirement of high-resolution image reconstruction where subpixel accuracy is needed. Hence, sophisticated approaches for mixed quality stereo images super-resolution are highly desired.

In this paper, we propose a novel method for mixed-quality stereo images super-resolution based on patch matching. Our goal is to develop an effective method which can accurately recover the high-frequency information of a low-resolution image in the mixed quality images pair by exploiting the prior knowledge

[+] Corresponding author. Tel.: + 86045182589656; fax: + 86045182589656.
E-mail address: fanbing@hrbeu.edu.cn.

from the corresponding high-resolution image without the calculation the disparity between stereo images. To this end, we adopt the vanishing point detection method to constrain the search area of patch matching in the HR images. Human being is more sensitive to the high-frequency content of the image [11]. Inspired by this knowledge, we adopt selective patch process (SPP) to select source patches with high-frequency information. By doing so, we are able to obtain the more plausible reconstructed image and improve computational effectiveness. Finally, we paste source patches from HR images at the location of the target patches by exploiting the transformation parameters. We validate our method in terms of qualitative evaluation and quantitative evaluation. The experimental results show that our method outperforms state-of-the-art SR methods.

2. Related Work

SR is able to recover a visual pleasing high-resolution image from a low-resolution image. Interpolation-based SR, which uses a convolution kernel to estimate the missing pixel, is the simplest algorithm. Some unfriendly visual artifacts will be introduced if the recovered scene is complicated. External learning-based SR methods are able to achieve a promising result. These methods mainly rely on machine learning techniques to learn a robust relationship between LR images and HR images by using external large datasets. For example, Yang [11] combine dictionary learning method with sparse representation to reconstruct LR image. Liang et.al [12] train two dictionaries to represent images. Although external learning-based SR methods obtain successful results, they are time-consuming, especially when the dataset is large.

To reduce the calculation burden, internal patch matching based SR methods [13][14] have been proposed by researchers recently. Shi et.al [15] develop an effective non-local self-similarity dictionary learning method with low complexity. Huang et.al [14] propose a search strategy which allows patches perspective deformation. This method obtains excellent result if the scene contains a large number of self-similar patterns.

Recently, SR methods have developed rapidly due to the evolution of the deep learning technique [16]–[18]. The first work of deep learning based SR method is super resolution convolutional neural network (SRCNN)[16]. They found the traditional sparse-coding-based SR method could be simulated as an end-to-end mapping. The state-of-the-art CNN based SR framework is enhanced deep super-resolution network (EDSR) [19]. They optimize the training modules by removing the unnecessary parts. Their network is able to address SR with different upscale factors. While deep learning-based methods get excellent performance and more real-time computation, they rely on GPU during the training phase, which is limited in common devices.

3. Proposed Super-resolution Framework

The block diagram of mixed quality stereo images super-resolution is shown in Figure 1. The input of our framework is the high-resolution right view image and the low-resolution left view image. The output is the reconstructed high-resolution left view image. The proposed framework includes three parts: plane detection, informative patch selection, nearest neighbor field (NNF) estimation. Each part will be described in detail in the following sections.

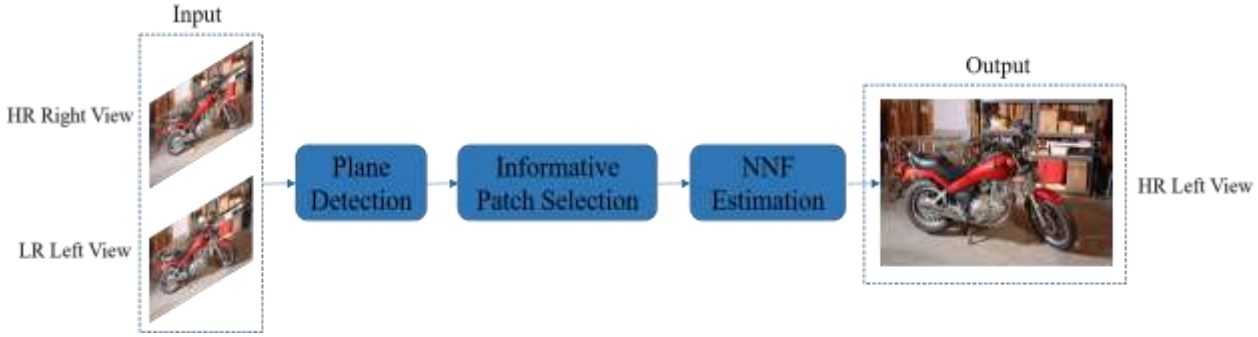


Fig. 1: Block diagram of the proposed framework.

3.1. Plane Detection

Many algorithms [20]–[22] for plane detection have been proposed. We choose a relative standard and simple one [23]. The detected planes are used to guide the search of corresponding patches.

To determine the orientation of the plane in the stereo image pair, we first detect the vanishing points (VPs) of two images respectively. The Cascaded Hough Transform is adopted since only one Hough space is accumulated in this transformation space. To reduce the computational complexity of the vanishing points process, it is operated on the extracted edges of the image. We detect three groups of vanishing points since a normal image usually contains three plane orientation.

After obtaining three groups VPs, we can calculate the plane orientation from every group. The parameters of the plane m are represented by vanishing lines which connects VPs from two distinct areas. The vanishing lines can be formulated to:

$$I_{\infty}^{m} = [l_1^m, l_2^m, l_3^m]^T \tag{1}$$

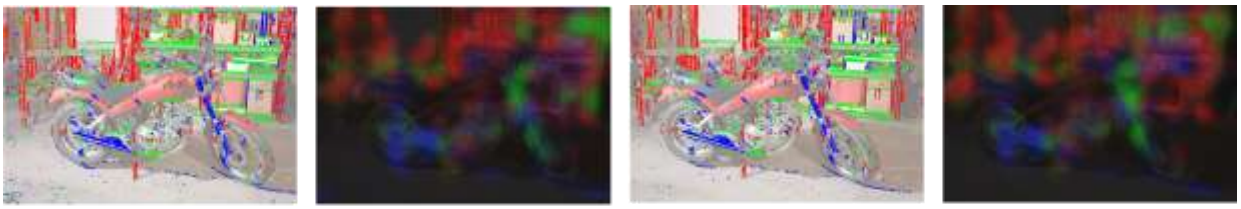
To ensure the corresponding area of each vanishing line, we use the Gaussian kernel to diffuse the spatial support. The Gaussian kernel can be expressed as:

$$h_{\sigma} = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \tag{2}$$

where σ is the bandwidth which controls the range of Gaussian kernel. Therefore, spatial support S_m around vanishing lines can be obtained with:

$$S_m = h_{\sigma} \otimes I_{\infty}^{m} \tag{3}$$

The value distribution after kernel convolution is regarded as the plane location density. Visualization of vanishing point detection and spatial support is shown in Figure 2. Red, green and blue represent lines from three different orientation planes respectively.



(a) (b) (c) (d)

Fig. 2: (a)(c) Vanishing point detection for LR left to view and HR right view. (b)(d) The spatial support for LR left to view and HR right view.

3.2. Informative Patch Selection

We divide the input stereo image pairs X^l and X^h into N patches $\{x_i^l\}_{i=1}^N$, $\{x_i^h\}_{i=1}^N$ for better correspondence. To acquire image patches with high-frequency components, we represent features based on a high-pass filter. In the literature, a high-pass filter is used for sharpening images or extract the edges, texture, and noise of the image. The simplest filter is ideal high pass filter. In our method, the first- and second-order derivatives are used as the gradient operators.

The gradient features are extracted by using five gradient operators:

$$\begin{aligned} f_1 &= [-1, 0, 1] \\ f_2 &= f_1^T \\ f_3 &= [1, 0, -2, 0, 1] \\ f_4 &= f_3^T \\ f_5 &= [1, 0, -1; 0; -1, 0, 1] \end{aligned} \tag{4}$$

The extracted features are denoted as f_x, f_y, f_{xx}, f_{yy} and f_{xy}. Hence, the curvature difference can be expressed as:

$$D_i = \| \mathrm{f}_i^{\varepsilon} | - | \mathrm{f}_i^{\eta} \| \tag{5}$$

where,

$$f^{\varepsilon} = \frac{f_x^2 f_{xx} + 2f_x f_y f_{xy} + f_y^2 f_{yy}}{f_x^2 + f_y^2} \tag{6}$$

$$f^{\eta} = \frac{f_y^2 f_{xx} + 2f_x f_y f_{xy} + f_x^2 f_{yy}}{f_x^2 + f_y^2} \tag{7}$$

By using these five gradient operators, we represent each patch by using five feature vectors, which are converted into one representative vector. According to the analysis of [24], different feature values indicate different information from images:

1) For edges, $|f_i^{\varepsilon}|$ is large but $|f_i^{\eta}|$ is small.
2) For smooths, $|f_i^{\varepsilon}|$ and $|f_i^{\eta}|$ are small.
3) For noises, $|f_i^{\varepsilon}|$ and $|f_i^{\eta}|$ are large.

Our objective is to select the patches with high-frequency information, so we define D_i as a median value. Based on this analysis, our method not only minimizes the distances of low-frequency base structures but also preserves the high-frequency detailed structures accurately.

3.3. NNF Estimation

To find the source patch in the HR image for target patch in the LR image, we convert it to the nearest neighbor field estimation (NNF) problem. The key idea of our cost function is the square of the Euclidean distance [25] which is also adopted by the classic feature detection algorithm SIFT [26]. The objective function is defined as:

$$(\mathbf{t}_i, \mathbf{s}_i) = \arg\min_{\mathbf{t}_i, \mathbf{s}_i} \sum_{i \in \mathfrak{V}} E_u(\mathbf{t}_i, \mathbf{s}_i, T_i) \tag{8}$$

where T_i is the deformable parameters of source patches; $\mathfrak{V}$ is the set of pixel indices of the stereo image pair. $E_u(\mathbf{t}_i, \mathbf{s}_i, H_i)$ can be expressed as:

$$E_u(\mathbf{t}_i, \mathbf{s}_i, T_i) = \| P(\mathbf{t}_i) - S(\mathbf{s}_i, T_i) \|_2^2 \tag{9}$$

where $P(\mathbf{t}_i)$ is the intensity value of the pixel $\mathbf{t}_i$ of the LR image and $S(\mathbf{s}_i, T_i)$ is the intensity value of the pixel $\mathbf{s}_i$ of the HR image. T_i is the transformation parameters from the source patch to target patch. To guarantee the accuracy of transformation parameters, we calculate it by combining the coordinates $\mathbf{t}_i, \mathbf{s}_i$ with the corresponding plane index m_i, which also be used in [14].

Suppose that $\mathbf{H}_{m_i}$ is the homograph which rectifies the planes in the scene. $\mathbf{h}_1, \mathbf{h}_2, \mathbf{h}_3$ are three vectors of $\mathbf{H}_{m_i}$. The source patches and target patches in the rectified space can be expressed as :

$$\mathbf{t}_i^{'} = [\mathbf{h}_1 \mathbf{t}_i, \mathbf{h}_2 \mathbf{t}_i, \mathbf{h}_3 \mathbf{t}_i]^T \tag{10}$$

$$\mathbf{s}_i^{'} = [\mathbf{h}_1 \mathbf{s}_i, \mathbf{h}_2 \mathbf{s}_i, \mathbf{h}_3 \mathbf{s}_i]^T \tag{11}$$

Let (d^x, d^y) be the displacement vector from target to source patch positions in the rectified space. Hence, the position of the source patch $\mathbf{s}_i^{'}$ can be formulated as:

$$\mathbf{s}_i^{'} = \begin{bmatrix} \mathbf{h}_1 + \mathbf{h}_3 d^x \\ \mathbf{h}_2 + \mathbf{h}_3 d^x \\ \mathbf{h}_3 \end{bmatrix} \mathbf{t}_i \tag{12}$$

$$\mathbf{s}_i = \mathbf{H}_{m_i}^{-1}\mathbf{s}_i^{'} = \mathbf{H}_{m_i}^{-1}\begin{bmatrix}\mathbf{h}_1 + \mathbf{h}_3 d^x \\ \mathbf{h}_2 + \mathbf{h}_3 d^x \\ \mathbf{h}_3\end{bmatrix}\mathbf{t}_i \tag{13}$$

Therefore:

$$\mathbf{s}_i = \mathbf{H}_{m_i}^{-1}\begin{bmatrix}\mathbf{h}_1 + \mathbf{h}_3 d^x \\ \mathbf{h}_2 + \mathbf{h}_3 d^x \\ \mathbf{h}_3\end{bmatrix}\begin{bmatrix}1 & 0 & t_i^x \\ 0 & 1 & t_i^y \\ 0 & 0 & 1\end{bmatrix}\begin{bmatrix}0 \\ 0 \\ 1\end{bmatrix} = \mathbf{T}_i\begin{bmatrix}0 \\ 0 \\ 1\end{bmatrix} \tag{14}$$

Finally, the source patches are pasted on the target patches by using the transformation parameter $\mathbf{T}_i$.

4. Experiment

To acquire the mixed quality stereo images datasets for the experiment, we downsample left images from stereo images datasets and keep the HR right images. In order to evaluate our method, we compare its performance with several state-of-the-art algorithms: SRCNN [16]**,** SelfExSR [14]**,** MMPM [27]**.** Among these methods, [14][27] belong to traditional example based algorithms and [16] is deep learning-based algorithms. We use 5x5 patches in the experiment. The hardware configuration of this experiment was a computer equipped with a dual-core Intel Pentium G2020 29 GHz, and 4 GB of RAM, running Windows 10. All codes are implemented by C++.

4.1. Qualitative Evaluation

To illustrate the applicability of our method, we select different datasets from [28] and [29] for qualitative evaluation. In [28], they provide accurate indoor static stereo images. In [29], the stereo images are synthesized by Microsoft. Figure 3 shows the subjective comparison for 2x SR of three images. Partial regions are zoomed in red rectangles. Blurry is introduced by SRCNN[16]**.** Some unfriendly artifacts appear in the result of SelfExSR [14] and MMPM [27]**.** Our method has a more human-pleasure nature and sharper edges. That is because we utilize the HR right views as prior knowledge during the patch searching process.

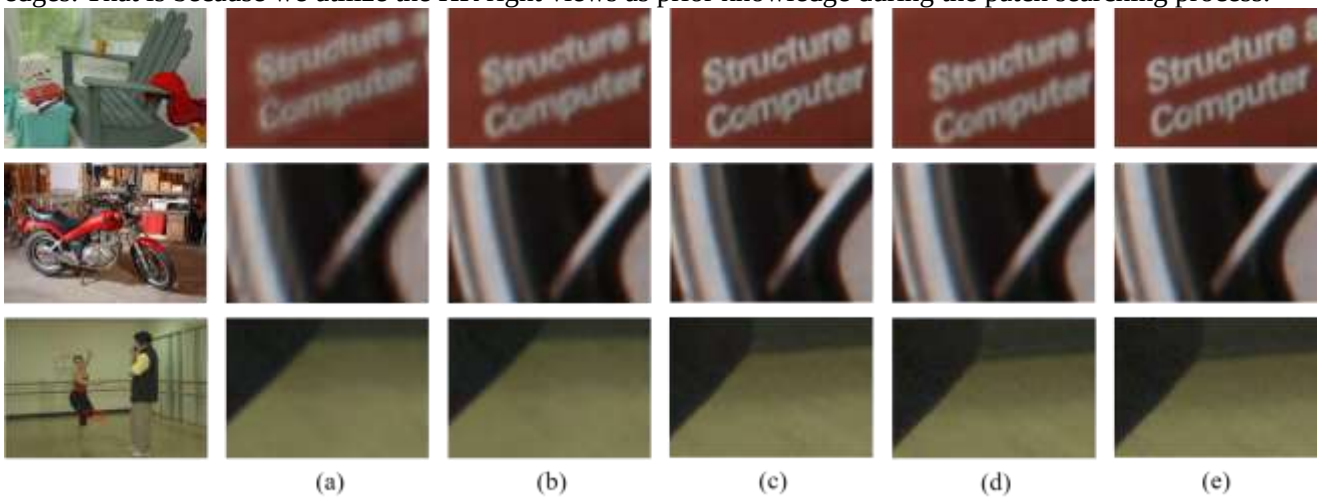


Fig. 3: Visual comparison for 2x super-resolution. (a) Bicubic (b) SRCNN [16] (c) SelfExSR [14] (d) MMPM [27] (e) Our method.

4.2. Quantitative Evaluation

We use three datasets for testing, including KITTI [30] and Middlebury [5]. Objective comparison with 2x, 3x and 4x upscale factors is shown in Tabel I. The best result is marked by bold font. PSNR is used to measure the squared intensity differences of the reconstructed and ground truth image pixels. SSIM is adopted to measure the structural similarity between the reconstructed and ground truth image pixels. Compared with other methods, the proposed method improves PSNR by 1.39 on average and SSIM by 0.003 on average. With the increase of upscale factors, the quality of all results decreases linearly. From these results, it is clear that for all cases, the quality of the LR left views has been improved significantly by the proposed method.

4.3. Time consumption

We also evaluate the computation time of our method to verify if it meets the requirement of real-time application. Table II shows time consumption for 240x240 image super-resolution of SRCNN [16], SelfExSR [14], MMPM [27] and our method. Time consumption of our method is relatively longer than the other two. Thus, there is a trade-off between performance and time-consuming.

TABLE 1. OBJECTIVE EVALUATION OF OUR METHOD WITH STATE-OF-THE-ART SR ALGORITHMS

Dataset	**Scale**	**SRCNN** [16]		**SelfExSR** [14]		**MMPM** [27]		**Proposed Method**	
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Middlebury	x2	31.48	0.9505	31.62	0.9511	32.21	0.9528	**33.52**	**0.9561**
	x3	28.76	0.9136	28.89	0.9087	29.33	0.9094	**31.08**	**0.9128**
	x4	27.11	0.8814	27.54	0.8854	27.69	0.8862	**29.13**	**0.8924**
KITTI 2012	x2	29.27	0.9162	29.18	0.9136	29.74	0.9145	**31.05**	0.9137
	x3	26.98	0.8533	25.87	0.8548	26.31	0.8549	**27.13**	**0.8607**
	x4	25.34	0.8002	26.12	0.8015	27.44	0.8074	**28.64**	**0.8133**
KITTI 2015	x2	28.50	0.9193	29.02	0.9197	30.05	0.9209	**31.89**	**0.9218**
	x3	26.19	0.8530	26.51	0.8537	27.83	0.8578	**29.25**	**0.8594**
	x4	24.44	0.7951	24.97	0.7960	25.75	0.7985	**27.14**	**0.7994**

TABLE 2. TIME CONSUMPTION

Methods	**SRCNN**	**SelfExSR**	**MMPM**	**Our method**
Time(s)	7.06	4.02	3.88	5.41

5. Conclusion

We have presented a novel super-resolution approach for mixed quality stereo images based on patch matching, which avoids the estimation of parallax in the stereo images. To select the source patch for target patch accurately, we use the detected perspective planes in each plane as soft constraints to guide the patch searching. Selective patch process (SPP) is adopted for obtaining the high frequency information from source patch. Our best patch matching selection is actually a nearest neighbour field estimation (NNF) problem, which is represented by the solution of the defined cost function. Meaningful qualitative and quantitative measures verify the superior performance of our approach. In future work, we plan to focus on improving applicability of our approach in real system, such as hybrid vision system (stereo vision system with heterogeneous cameras).

6. Acknowledgements

This work has been supported in part by the National Natural Science Foundation of China via grants 61203255 and 61175089. The Fundamental Research Funds for the Central Universities. (HEUCF180405).

7. References

[1] L. Zhang and W. J. Tam, “Stereoscopic Image Generation Based on Depth Images for 3D TV,” *IEEE Trans. Broadcast.*, vol. 51, no. 2, pp. 191–199, Jun. 2005.

[2] T. Aykut, M. Karimi, C. Burgmair, A. Finkenzeller, C. Bachhuber, and E. Steinbach, “Delay Compensation for a Telepresence System With 3D 360 Degree Vision Based on Deep Head Motion Prediction and Dynamic FoV Adaptation,” *IEEE Robot. Autom. Lett.*, vol. 3, no. 4, pp. 4343–4350, Oct. 2018.

[3] L. Stelmach, Wa James Tam, D. Meegan, and A. Vincent, “Stereo image quality: effects of mixed spatio-temporal

resolution," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 10, no. 2, pp. 188–193, Mar. 2000.

[4] Woontack Woo and A. Ortega, "Overlapped block disparity compensation with adaptive windows for stereo image coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 10, no. 2, pp. 194–200, Mar. 2000.

[5] N. Mayer *et al.*, "A Large Dataset to Train Convolutional Networks for Disparity, Optical Flow, and Scene Flow Estimation," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 2016, pp. 4040–4048.

[6] A. Geiger, M. Roser, and R. Urtasun, "Efficient Large-Scale Stereo Matching," in *Asian Conference on Computer Vision*, 2010.

[7] Z. Jin, H. Luo, L. Luo, W. Zou, X. Lil, and E. Steinbach, "Information Fusion based Quality Enhancement for 3D Stereo Images Using CNN," in *2018 26th European Signal Processing Conference (EUSIPCO)*, Rome, 2018, pp. 1447–1451.

[8] Chung-Hao Chen, Yi Yao, D. Page, B. Abidi, A. Koschan, and M. Abidi, "Heterogeneous Fusion of Omnidirectional and PTZ Cameras for Multiple Object Tracking," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 8, pp. 1052–1063, Aug. 2008.

[9] Y. Bastanlar, "A simplified two-view geometry based external calibration method for omnidirectional and PTZ camera pairs," *Pattern Recognit. Lett.*, vol. 71, pp. 1–7, Feb. 2016.

[10] A. K. Jain and T. Q. Nguyen, "Video super-resolution for mixed resolution stereo," in *2013 IEEE International Conference on Image Processing*, Melbourne, Australia, 2013, pp. 962–966.

[11] Jianchao Yang, J. Wright, T. S. Huang, and Yi Ma, "Image Super-Resolution Via Sparse Representation," *IEEE Trans. Image Process.*, vol. 19, no. 11, pp. 2861–2873, Nov. 2010.

[12] Shenlong Wang, Lei Zhang, Yan Liang, and Quan Pan, "Semi-coupled dictionary learning with applications to image super-resolution and photo-sketch synthesis," in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, Providence, RI, 2012, pp. 2216–2223.

[13] Jian Zhang, Debin Zhao, and Wen Gao, "Group-Based Sparse Representation for Image Restoration," *IEEE Trans. Image Process.*, vol. 23, no. 8, pp. 3336–3351, Aug. 2014.

[14] J.-B. Huang, A. Singh, and N. Ahuja, "Single image super-resolution from transformed self-exemplars," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, USA, 2015, pp. 5197–5206.

[15] W. Shi, C. Chen, F. Jiang, D. Zhao, and W. Shen, "Group-based sparse representation for low lighting image enhancement," in *2016 IEEE International Conference on Image Processing (ICIP)*, Phoenix, AZ, USA, 2016, pp. 4082–4086.

[16] C. Dong, C. C. Loy, K. He, and X. Tang, "Image Super-Resolution Using Deep Convolutional Networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 2, pp. 295–307, Feb. 2016.

[17] C. Ledig *et al.*, "Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, 2017, pp. 105–114.

[18] W. Shi *et al.*, "Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 2016, pp. 1874–1883.

[19] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, "Enhanced Deep Residual Networks for Single Image Super-Resolution," in *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Honolulu, HI, USA, 2017, pp. 1132–1140.

[20] Z. Zhang, A. Ganesh, X. Liang, and Y. Ma, "TILT: Transform Invariant Low-Rank Textures," *Int. J. Comput. Vis.*, vol. 99, no. 1, pp. 1–24, Aug. 2012.

[21] D. Aiger, D. Cohen-Or, and N. J. Mitra, "Repetition Maximization based Texture Rectification," *Comput. Graph. Forum*, vol. 31, no. 2pt2, pp. 439–448, May 2012.

[22] O. Chum and J. Matas, "Planar Affine Rectification from Change of Scale," in *Computer Vision – ACCV 2010*, vol. 6495, R. Kimmel, R. Klette, and A. Sugimoto, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 347–360.

[23] R. Hartley and A. Zisserman, “Multiple view geometry in computer vision,” *Kybernetes*, vol. 30, no. 9/10, pp. 1865–1872, 2008.

[24] Q. Chen, P. Montesinos, Q. S. Sun, P. A. Heng, and D. S. Xia, “Adaptive total variation denoising based on difference curvature,” *Image Vis. Comput.*, vol. 28, no. 3, pp. 298–306, Mar. 2010.

[25] P. E. Danielsson, “Euclidean distance mapping,” *Comput. Graph. Image Process.*, vol. 14, no. 3, pp. 227–248, 1980.

[26] D. G. Lowe, “Distinctive Image Features from Scale-Invariant Keypoints,” *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, Nov. 2004.

[27] Y. Huang, J. Li, X. Gao, L. He, and W. Lu, “Single Image Super-Resolution via Multiple Mixture Prior Models,” *IEEE Trans. Image Process.*, vol. 27, no. 12, pp. 5904–5917, Dec. 2018.

[28] D. Scharstein *et al.*, “High-Resolution Stereo Datasets with Subpixel-Accurate Ground Truth,” in *Pattern Recognition*, vol. 8753, X. Jiang, J. Hornegger, and R. Koch, Eds. Cham: Springer International Publishing, 2014, pp. 31–42.

[29] C. L. Zitnick, S. B. Kang, M. Uyttendaele, S. Winder, and R. Szeliski, “High-quality video view interpolation using a layered representation,” *ACM Trans. Graph.*, vol. 23, no. 3, p. 600, Aug. 2004.

[30] A. Geiger, P. Lenz, and R. Urtasun, “Are we ready for autonomous driving? The KITTI vision benchmark suite,” in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, 2012, pp. 3354–3361.