Improving Multi-View Stereo via Super-Resolution

Eugenio Lomurno(✉), Andrea Romanoni, and Matteo Matteucci

Politecnico di Milano, Milano, Italy
{eugenio.lomurno,andrea.romanoni,matteo.matteucci}@polimi.it

Abstract. Today, Multi-View Stereo techniques can reconstruct robust and detailed 3D models, especially when starting from high-resolution images. However, there are cases in which the resolution of input images is relatively low, for instance, when dealing with old photos or when hardware constrains the amount of data acquired. This paper shows how increasing the resolution of such input images through Super-Resolution techniques reflects in quality improvements of the reconstructed 3D models. We show that applying a Super-Resolution step before recovering the depth maps leads to a better 3D model both in the case of patchmatch and deep learning Multi-View Stereo algorithms. In detail, the use of Super-Resolution improves the average f1 score of reconstructed models. It turns out to be particularly effective in the case of scenes rich in texture, such as outdoor landscapes.

Keywords: Multi-View Stereo · Super-Resolution · Single-Image Super-Resolution · 3D reconstruction

1 Introduction

Recovering the 3D model of a scene captured by images is a relevant problem in a wide variety of scenarios, e.g., city mapping, archaeological heritage preservation, autonomous driving, and robot localization. In the Computer Vision community, this task goes under the name of Multi-View Stereo (MVS), and it aims to reconstruct 3D models as accurately and completely as possible.

Currently, the most successful workflow to perform such reconstructions starts from a Structure from Motion algorithm that estimates camera parameters such as their positions and orientations [18]. Then, it follows the depth maps estimation step, for which the most common approaches rely on patchmatch [1] or deep learning [24] techniques. The former approaches lead to very accurate results, while the latter produce more complete models, even if they still suffer scalability issues. As the last step, depth maps are projected on 3D space and fused together, obtaining a dense point cloud.

Under controlled scenarios, in which the hardware adopted to collect the images is not subject to particular constraints, it is relatively easy to acquire high-resolution images and obtain a high-quality reconstruction of the scene.

S. Sclaroff et al. (Eds.): ICIAP 2022, LNCS 13232, pp. 102–113, 2022.
https://doi.org/10.1007/978-3-031-06430-2_9

However, in several cases, the input of an MVS method consists of low-resolution images. For instance, when power consumption constrains the hardware, e.g., with drones or telescopes, or when processing images taken in low-resolution such as with old photos.

In these cases, the recovered 3D model most likely lacks details or is incomplete, regardless of the adopted MVS algorithm. We claim that algorithmically increasing input images resolution can overcome this issue by enhancing their information content and quality. This is possible via Super-Resolution techniques that have recently reached impressive performance in many application fields despite the possibility of generating some artefacts.

This paper shows the benefits for MVS pipelines of upscaling low-resolution images through Single-Image Super-Resolution (SISR) techniques. In particular, we test SISR contribution over COLMAP [19] and CasMVSNet [6] MVS pipelines, and validate it over ETH3D Low-resolution many-view [20] and Tanks and Temples [10] benchmarks. Perceptive and numerical results demonstrate that SISR improves the quality of the dense point clouds produced by MVS algorithms by effectively balancing the increased amount of generated points and their position in the space.

2 Related Work

In the literature, there are some attempts to exploit Super-Resolution (SR) with the goal of improving the quality of 3D reconstructions. For instance, Goldlücke *et al.* [5] proposed a variational method to improve 3D models appearance by estimating textures with SR techniques. More recently, Li *et al.* [13] proposed a novel model-based SR method that better exploits geometric features to enrich the texture applied to a 3D model.

Other approaches exploiting SR in the 3D reconstruction realm aims to increase depth maps resolution. Lei *et al.* [12] relied on bilinear interpolation of multiple depth maps to increase the resolution of a single depth map. The authors in [27] and [21] used high-resolution RGB images to guide a deep learning model to increase depth maps resolution.

Differently from previous works, we aim at improving models geometry instead of their texture appearance, by applying SR directly on input images. We argue that SR can improve the reconstruction from low-resolution images, and different stages of a 3D reconstruction pipeline could benefit from the availability of SR images, e.g., camera calibration and mesh refinement. Surprisingly, to the best of our knowledge, no paper has ever analyzed if and to what extent MVS 3D reconstruction pipelines can benefit from input images enhanced through SR.

2.1 Single-Image Super-Resolution

Single-Image Super-Resolution (SISR) aims at recovering a high-resolution image from a single low-resolution image [23]. In the last few years, we have seen how modern deep learning pipelines overtook non-learning-based algorithms, such as nearest-neighbours and bicubic interpolation.

As first attempt, Dong *et al.* [4] proposed a convolutional neural network to map low- to high-resolution images. This network architecture has been extended with a combination of new layers, and skip-connections by Kim *et al.* [9]. Subsequently, other methods have exploited different combinations of residual and dense connections [15,26].

Recent works show that networks with novel feedback mechanisms further improve the quality of the SR images. For instance, Li *et al.* [14] combine a feedback block with curriculum learning. In their most recent work, Haris *et al.* released the Deep Back-Projection Network architecture [7]. The idea is to generate numerous degraded and high-resolution hypothesis images that the network uses to improve the output result. In the last revision, the authors have implemented dense connections, adversarial loss and recurrent layers, making the entire architecture more scalable and performing. Chen *et al.* [3] proposed a self-supervised encoding network based on their implicit neural representation technique to learn continuous mappings for super-resolution.

2.2 Multi-View Stereo

MVS aims at recovering a dense 3D representation of a scene perceived by a set of calibrated cameras. The main step adopted by the most successful MVS methods is depth maps estimation, i.e., the process of computing the depth of each pixel belonging to each image. Once computed, these maps are fused into a dense point cloud or a volumetric representation.

The most performing depth estimation approaches are based on the patchmatch algorithm [1], which relies on the idea of choosing for each pixel a random guess of the depth and then propagating the most likely estimates to its neighbourhood. The work proposed by Schönberger *et al.* [19], named COLMAP, can be considered the cornerstone of modern patchmatch-based algorithms. It is a robust framework able to process high-quality images and jointly estimate pixel-wise camera visibility, as well depth and normal maps for each view. Since this method heavily relies on the Bilateral NCC Photometric-Consistency, it often fails in recovering areas with low texture. Recently, to compensate for this, TAPA-MVS [17] proposed to explicitly handle textureless regions by propagating in a planar-wise fashion the valid depth estimates to neighbouring textureless areas. Kuhn *et al.* [11] extended this method with a hierarchical approach improving the robustness of the estimation process.

Another family of MVS algorithms relies on deep learning. DeepMVS [8] and MVSNet [24] were the first approaches proposing an effective MVS pipeline based on DNNs. For each camera, both the approaches build a cost volume by projecting nearby images on planes at different depths, then they classify [8] or regress [24] the best depth for each pixel. Yao *et al.* [25] introduced an RNN to regularize the cost volume, while Luo *et al.* [16] built a model to learn how to aggregate the cost to compute a more robust depth estimate. MVS-CRF [22], finally, refines the MVSNet estimate through Markov Random Field, and Point-MVSNet [2] through a graph-based neural architecture. The huge limitation of learning-based approaches relies on their computational complexity. Usually, it is not feasible to handle high-resolution images as both memory and time costs grow cubically as the volume resolution increases, causing a limitation on the accuracy and completeness of the reconstructed models. The best attempt to handle this problem is the work of Xiaodong *et al.* [6], named CasMVSNet, in which they applied a coarse-to-fine approach that considerably improves the scalability of MVSNet-based methods.

3 Methods

Our work aims to provide an overview of the effects of Single-Image Super-Resolution (SISR) when applied in the head of Multi-View Stereo (MVS) algorithms. In order to generalize this phenomenon, we chose two different SISR algorithms to conduct our ablation studies. In particular, we identified the bicubic interpolation algorithm as a candidate algorithm for a more traditional SISR approach, while Deep Back-Projection Network (DBPN) by Haris *et al.* [7] to investigate the effects of a more recent pipeline based on deep learning. It is known in the literature that SR techniques are often prone to artifact generation, especially with the growth of the upscaling factor. In order to exploit both the SISR algorithms at the best of their performance, we fixed it to 2. Concerning DBPN we exploited the best set of weights provided by the authors for this upscaling factor, i.e., “DBPN-RES-MR64-3”.

Regarding the MVS pipelines, we chose two different approaches based on two different technologies. The first one is COLMAP [19], in which depth maps estimation is heavily based on patchmatch. Moreover, it turns out to be one of the most efficient algorithms of its family due to its parallel maps computation. The second one is CasMVSNet, the deep learning architecture based on MVSNet developed by Xiaodong *et al.* [6]. With the addition of a new cost volume technique, built upon a feature pyramid encoding geometry, it learns to estimate the depth space of the scene at gradually finer scales. In detail, it narrows the disparity range for every stage thanks to an iterative prediction made from the previous stages. It then gradually increases the cost volume resolution to obtain accurate outputs. This technique allowed us to exploit a deep learning approach, even with datasets composed of numerous images and limited hardware capabilities. For this algorithm, like for DBPN, we used the pre-trained model provided by the authors in order to facilitate the experiments reproducibility.

Table 1. F1 scores on the ETH3D low-resolution multi-view train set with COLMAP. We compare scores starting from low-resolution images against the ones obtained from bicubic interpolation and Deep Back-Projection Network Super-Resolution

τ (cm)	Overall			Indoor			Outdoor		
	Low-res	Bicubic	DBPN	Low-res	Bicubic	DBPN	Low-res	Bicubic	DBPN
1	35.80	39.85	**40.00**	40.68	**43.68**	43.42	32.55	37.29	**37.71**
2	53.41	54.83	**54.98**	56.15	**57.37**	57.09	51.59	53.14	**53.58**
5	72.16	72.58	**72.64**	**74.35**	74.05	73.62	70.70	71.59	**71.99**
10	81.83	**82.17**	82.13	**83.97**	83.24	82.69	80.40	81.46	**81.76**
20	88.98	**89.20**	89.14	**91.50**	90.55	90.11	87.30	88.30	**88.50**
50	95.29	**95.70**	**95.70**	**97.33**	97.19	97.15	93.93	94.71	**94.75**

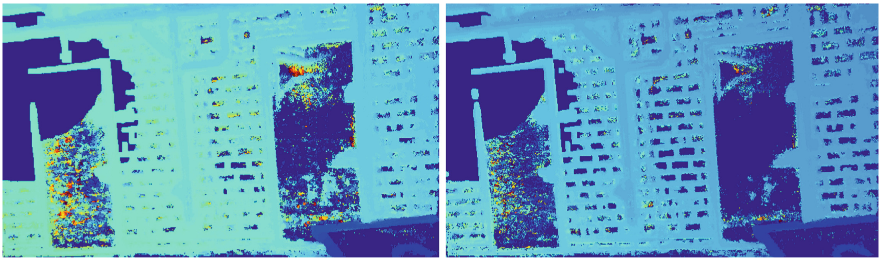

Fig. 1. COLMAP disparity maps of a sample from storage_room_2 dataset (Indoor). On the left side, the low-resolution estimation, on the right side, the one enhanced via Deep Back-Projection Network.

4 Ablation Study

In order to calibrate the chosen algorithms, we have conducted an ablation study over the training set of the ETH3D Low-resolution many-view benchmark [20], which is composed of five gray-scale datasets split in indoor (two) and outdoor (three). For this benchmark, the accuracy and completeness metrics are available to compute the goodness of the reconstruction with respect to the ground truth. In order to take both into account, we relied our analysis on their harmonic mean, i.e., the f1 score. This section's experiments have been executed on an Intel(R) Xeon(R) CPU E5-2630 v4 @ 2.20 GHz with an Nvidia GTX 1080Ti GPU.

For each dataset, we computed its enhanced twins via bicubic interpolation and DBPN. Then, we applied COLMAP and CasMVSNet pipelines on all these datasets to obtain the dense point clouds. Finally, we evaluated the results with respect to the ground truths. We compare, with different distance tolerances, the performance of each dataset.

According to the literature, and after some preliminary trials, we noticed that COLMAP achieves poor performance in indoor datasets despite their resolution. This issue is related to its patchmatch algorithm heavily based on the Bilateral Normalized Cross-Correlation Photometric-Consistency, which is translated into

Table 2. F1 scores on the ETH3D low-resolution multi-view train set with CasMVS-Net. We compare scores starting from low-resolution images against the ones obtained from bicubic interpolation and Deep Back-Projection Network Super-Resolution

τ (cm)	Overall			Indoor			Outdoor		
	Low-res	Bicubic	DBPN	Low-res	Bicubic	DBPN	Low-res	Bicubic	DBPN
1	38.28	39.58	**39.59**	37.24	38.43	**38.46**	38.97	40.34	**40.35**
2	49.00	49.65	**49.66**	48.08	48.07	**48.09**	49.61	50.70	**50.71**
5	60.58	**61.17**	61.15	**61.10**	60.49	60.40	60.25	61.63	**61.65**
10	67.59	**68.43**	68.37	**69.39**	69.25	69.10	66.39	**67.89**	**67.89**
20	74.00	**74.93**	74.84	77.33	**77.44**	77.25	71.79	**73.26**	73.23
50	82.60	83.59	**83.61**	87.14	**87.31**	**87.31**	79.56	81.10	**81.14**

artifacts and poor estimates in textureless regions, e.g., monochromatic and reflective surfaces. In order to cope with this unwanted behavior, we modified COLMAP parameters when dealing with Indoor scenes by increasing the robustness of the depth estimates and thus trading-off with a higher computational cost. Specifically, we reduce the minimum NCC threshold and increase the window radius by 2. Then we filter the resulting depth maps with a speckle filter algorithm before fusing them. For this widely used filter, we have chosen a max depth range equal to 5 and set the maximum speckle size to 1% of the depth map dimension. These parameters have been tuned to maximize the average performance over both low-resolution datasets and their versions enhanced via SR.

Table 1 shows the results of this first set of experiments computed via COLMAP. On average, both SR techniques lead to noticeable improvements in the quality of the reconstructed models. This improvement tends to be more evident when tolerances are computed concerning little points neighborhoods. It is also observable a different behavior between indoor and outdoor scenes: in the first case, we observe good synergies between SR and the MVS algorithms concerning small tolerances, while the effects turn out to degrade the performance by considering more flexible evaluation criteria. As can be seen in the disparity map example displayed in Fig. 1, on the one hand, the effect of DBPN is translated into less noisy depth estimations, on the other, it erases the small number of points belonging to the textureless regions such as the inner part of the bricks or doors.

In the second case, the advantages of applying SISR are spread with evidence among the tolerances, reaching remarkable f1 score improvements (+5.16% for $\tau = 1\,\mathrm{cm}$). An example of the benefits can be visually appreciated in the disparity maps comparison in Fig. 2. In this case, it is evident how COLMAP is able to exploit the increased amount of input information, on the one hand, to identify pieces of bushes in the foreground, on the other, to better perceive the image depth and produce a more detailed disparity map.

In Table 2 are instead summarized the results obtained by carrying out the same set of experiments with CasMVSNet as MSV algorithm. In general, the behavior of this algorithm when its input is enhanced with SR is coherent with the one demonstrated by COLMAP. In fact, also in this case, it is evident a

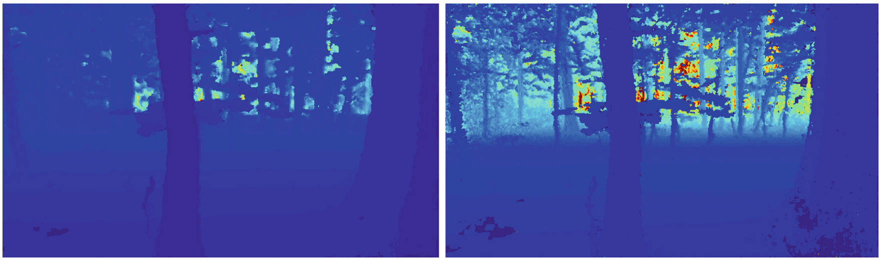

Fig. 2. COLMAP disparity maps of a sample from forest dataset (Outdoor). On the left side, the low-resolution estimation, on the right side, the one enhanced via Deep Back-Projection Network.

Table 3. F1, accuracy and completeness scores over ETH3D low-resolution multi-view benchmark. We compare the presented models grouped in many subsets with a tolerance $\tau = 1\,\text{cm}$

Model	Overall			Indoor			Outdoor		
	F1	Acc	Comp	F1	Acc	Comp	F1	Acc	Comp
COLMAP	36.6	**40.7**	33.8	34.4	**38.1**	31.7	38.0	**42.4**	35.1
COLMAP (DBPN)	**40.6**	37.2	**45.8**	**36.5**	35.7	**38.2**	**43.3**	38.2	**50.8**
CasMVSNet	36.8	**42.4**	34.3	27.8	36.0	23.7	42.7	**46.8**	41.3
CasMVSNet (DBPN)	**37.7**	41.4	**36.9**	**28.9**	**37.4**	**24.7**	**43.6**	44.0	**45.0**

generalized f1 score improvement, especially while considering the most restrictive tolerances.

We can thus argue that, in general, both patchmach and deep learning MSV algorithms demonstrate on average to benefit from the presence of richer input information. Moreover, from the results of this ablation study is possible to assert that this is true regardless of the SR algorithm chosen.

5 Experiments and Results

After having calibrated the algorithms, we evaluated their performance in a broader set of experiments. For this purpose, we executed two independent validation runs, both computed on an Intel(R) Xeon(R) CPU E5-2630 v4 @ 2.20 GHz with an Nvidia GTX 1080Ti.

5.1 Evaluation over ETH3D Benchmark

We evaluated the proposed approach over the whole ETH3D Low-resolution many-view benchmark [20], which is composed of ten gray-scale datasets split into indoor (four) and outdoor (six).

We used DBPN as SR algorithm for these experiments and COLMAP and CasMVSNet for the MVS pipeline. From Table 3 it is possible to summarize that,

Table 4. F1, precision and recall scores over Tanks and Temples train benchmark. We compare the downscaled version (low-res) with the SISR approaches (bicubic and DBPN) and with a theoretical high-resolution version (high-res). All the reconstructions are computed via COLMAP.

Model	Overall			Indoor			Outdoor		
	F1	Prec	Rec	F1	Prec	Rec	F1	Prec	Rec
Low-res (1x)	9.56	20.40	6.47	8.93	18.97	6.24	10.02	21.47	6.63
Bicubic (2x)	31.44	37.96	27.91	26.24	36.39	21.86	35.35	39.14	32.46
DBPN (2x)	**32.94**	**40.09**	**29.04**	**26.85**	**36.39**	**22.15**	**37.5**	**41.74**	**34.27**
High-res (2x)	37.65	45.28	33.19	32.03	42.63	26.57	41.87	47.27	38.15

concerning the most significant evaluation criterion, i.e., $\tau = 1\,\text{cm}$, the trade-off represented by the f1 score is strongly raised up from the completeness metric. Like in the ablation study, we can observe a general benefit brought about by the SR usage, both in the case of patchmatch and deep learning MVS pipelines.

From these results and the visual comparison of reconstruction details is shown in Fig. 3, we argue that the improvement is related to the increased amount of reconstructed points allowed by the increased amount of pixels in input. In fact, the completeness boost means that there are many more points close to ground truth points regarding the low-resolution case. On the other hand, there are also many points that are far away from the ground truth, and this is translated into an accuracy drop.

Given the obtained performance, we can argue that this trade-off brings positive outcomes that are perceptively translated into denser reconstructions and characterized by a reduction of texture holes.

5.2 Evaluation over Tanks and Temples Benchmark

Finally, we have conducted a further experiment over the Tanks and Temples train benchmark [10], composed of 7 RGB high-resolution datasets split into indoor (three) and outdoor (four).

Given its images good quality, we performed an evaluation test to estimate the goodness of the SISR approaches with respect to theoretical high-resolution data. In detail, we downscaled every dataset by a factor of 4, considering it as the low-resolution benchmark. We did the same but with a factor of 2, considering this as the high-resolution benchmark. Finally, starting from the low-resolution datasets, we computed with bicubic interpolation and DBPN algorithms their twins enhanced via SR with upscaling factor 2, and we compared all the results. For these experiments, we have used COLMAP as MVS pipeline.

For this benchmark, the precision and recall metrics are available to compute the goodness of the reconstruction with respect to the ground truth. In order to compare these results with the ones of the previous experiments, these metrics can be intended from the reconstructed points perspective (if ETH3D benchmark

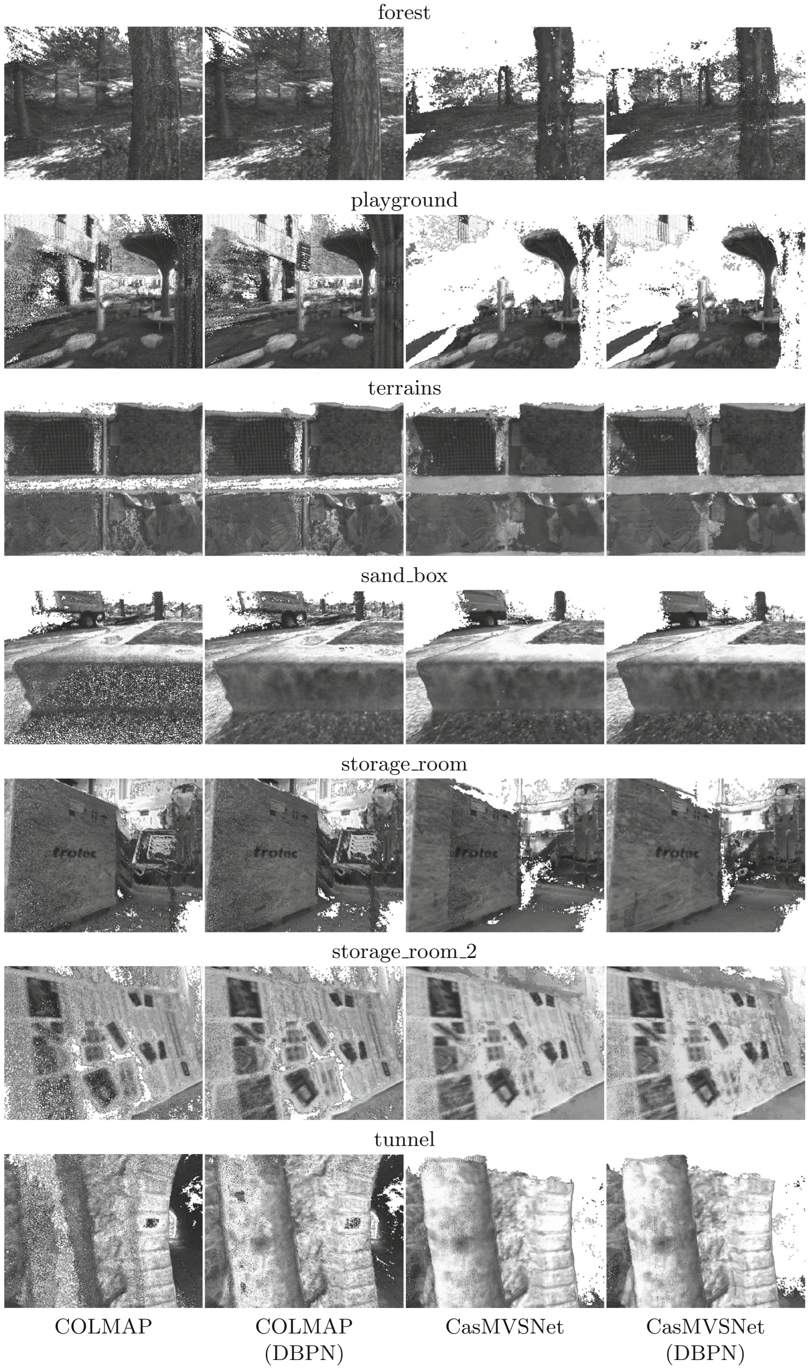

Fig. 3. Details of ETH3D low-res many-view benchmark 3D reconstructions. We compare the same view for each proposed pipeline in both low-resolution and enhanced via DBPN versions.

completeness can be assumed as the accuracy of the reconstructed points with respect to the nearest point of the ground truth, given a fixed tolerance, now we are dealing with its precision and its recall). Also in this case, in order to take both into account, we relied our analysis mainly on the f1 score. This time, the tolerance is fixed and different for each dataset according to the authors evaluation rules.

The results of this experiment are displayed in Table 4. In this scenario, it is evident how the SISR turns out to be very effective for every metric taken into account. The reason behind such a remarkable improvement has to be addressed to the shallow resolution of the low-res benchmark, which made very difficult for COLMAP to estimate good depths and subsequently fuse them into accurate dense point clouds. The most interesting result relies on comparing reconstructions from the benchmark enhanced via DBPN and the high-resolution one. In fact, despite there is still a performance gap between them, this gap is much lower than the one with the scores obtained starting from the low-resolution benchmark.

Summarizing this result with the ones obtained from previous analysis, we can conclude that SISR algorithms can help MVS techniques to produce better and denser reconstructions when put on top of the pipeline. Despite the presence of artifacts, the results are qualitatively closer to high-resolution theoretical reconstructions, and this approach can be applied both on MVS algorithms based on patchmatch and deep learning.

6 Conclusions

In this paper, we presented a study on how to improve 3D reconstruction starting from low-resolution images through the use of Single-Image Super-Resolution techniques, demonstrating Super-Resolution effectiveness for Multi-View Stereo algorithms based on both patchmatch and deep learning. Moreover, we have demonstrated the existence of a strong correlation between starting images and 3D models qualities and that an increased amount of input information provided by Super-Resolution is effectively translated into more robust and dense representations in the 3D space by Multi-View Stereo pipelines. Despite the Super-Resolution algorithm chosen, we have shown how the 3D models obtained results to benefit from the Single-Image Super-Resolution improvement of the input images the more they do not have a starting high-resolution.

References

1. Bleyer, M., Rhemann, C., Rother, C.: PatchMatch stereo-stereo matching with slanted support windows. In: BMVC, vol. 11, pp. 1–11 (2011)
2. Chen, R., Han, S., Xu, J., Su, H.: Point-based multi-view stereo network. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1538–1547 (2019)

3. Chen, Y., Liu, S., Wang, X.: Learning continuous image representation with local implicit image function. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 8628–8638 (2021)
4. Dong, C., Loy, C.C., He, K., Tang, X.: Image super-resolution using deep convolutional networks. IEEE Trans. Pattern Anal. Mach. Intell. **38**(2), 295–307 (2015)
5. Goldlücke, B., Aubry, M., Kolev, K., Cremers, D.: A super-resolution framework for high-accuracy multiview reconstruction. Int. J. Comput. Vis. **106**(2), 172–191 (2014)
6. Gu, X., Fan, Z., Zhu, S., Dai, Z., Tan, F., Tan, P.: Cascade cost volume for high-resolution multi-view stereo and stereo matching. arXiv preprint arXiv:1912.06378 (2019)
7. Haris, M., Shakhnarovich, G., Ukita, N.: Deep back-projection networks for super-resolution. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1664–1673 (2018)
8. Huang, P.H., Matzen, K., Kopf, J., Ahuja, N., Huang, J.B.: DeepMVS: learning multi-view stereopsis. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2821–2830 (2018)
9. Kim, J., Kwon Lee, J., Mu Lee, K.: Accurate image super-resolution using very deep convolutional networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1646–1654 (2016)
10. Knapitsch, A., Park, J., Zhou, Q.Y., Koltun, V.: Tanks and temples: benchmarking large-scale scene reconstruction. ACM Trans. Graph. (ToG) **36**(4), 1–13 (2017)
11. Kuhn, A., Lin, S., Erdler, O.: Plane completion and filtering for multi-view stereo reconstruction. In: Fink, G.A., Frintrop, S., Jiang, X. (eds.) DAGM GCPR 2019. LNCS, vol. 11824, pp. 18–32. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-33676-9_2
12. Lei, J., Li, L., Yue, H., Wu, F., Ling, N., Hou, C.: Depth map super-resolution considering view synthesis quality. IEEE Trans. Image Process. **26**(4), 1732–1745 (2017)
13. Li, Y., Tsiminaki, V., Timofte, R., Pollefeys, M., Gool, L.V.: 3d appearance super-resolution with deep learning. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 9671–9680 (2019)
14. Li, Z., Yang, J., Liu, Z., Yang, X., Jeon, G., Wu, W.: Feedback network for image super-resolution. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3867–3876 (2019)
15. Lim, B., Son, S., Kim, H., Nah, S., Mu Lee, K.: Enhanced deep residual networks for single image super-resolution. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 136–144 (2017)
16. Luo, K., Guan, T., Ju, L., Huang, H., Luo, Y.: P-MVSNet: learning patch-wise matching confidence aggregation for multi-view stereo. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 10452–10461 (2019)
17. Romanoni, A., Matteucci, M.: TAPA-MVS: textureless-aware patchmatch multi-view stereo. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 10413–10422 (2019)
18. Schonberger, J.L., Frahm, J.M.: Structure-from-motion revisited. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4104–4113 (2016)
19. Schönberger, J.L., Zheng, E., Frahm, J.-M., Pollefeys, M.: Pixelwise view selection for unstructured multi-view stereo. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9907, pp. 501–518. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46487-9_31

20. Schops, T., et al.: A multi-view stereo benchmark with high-resolution images and multi-camera videos. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3260–3269 (2017)
21. Voynov, O., et al.: Perceptual deep depth super-resolution. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 5653–5663 (2019)
22. Xue, Y., et al.: MVSCRF: learning multi-view stereo with conditional random fields. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 4312–4321 (2019)
23. Yang, W., Zhang, X., Tian, Y., Wang, W., Xue, J.H., Liao, Q.: Deep learning for single image super-resolution: a brief review. IEEE Trans. Multimedia **21**(12), 3106–3121 (2019)
24. Yao, Y., Luo, Z., Li, S., Fang, T., Quan, L.: MVSNet: depth inference for unstructured multi-view stereo. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds.) ECCV 2018. LNCS, vol. 11212, pp. 785–801. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-01237-3_47
25. Yao, Y., Luo, Z., Li, S., Shen, T., Fang, T., Quan, L.: Recurrent MVSNet for high-resolution multi-view stereo depth inference. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 5525–5534 (2019)
26. Zhang, Y., Tian, Y., Kong, Y., Zhong, B., Fu, Y.: Residual dense network for image super-resolution. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2472–2481 (2018)
27. Zuo, Y., Wu, Q., Fang, Y., An, P., Huang, L., Chen, Z.: Multi-scale frequency reconstruction for guided depth map super-resolution via deep residual network. IEEE Trans. Circ. Syst. Video Technol. **30**(2), 297–306 (2019)