Generative-Network Based Multimedia Super-Resolution For UAV Remote Sensing

Yash Turkar†, Christo Aluckal‡, Shaunak De◇, Varsha Turkar° and Yogesh Agarwadkar ‡

†University at Buffalo, SUNY,
‡InfiCorridor Solutions Pvt. Ltd.
°Don Bosco College of Engineering, Goa, India
◇Capella Space

***Abstract*—Unmanned Aerial Vehicle (UAV) based aerial mapping has taken over the surveying industry thanks to low costs and ease of use. Although these UAVs have relatively high- resolution imaging systems, there exists a near exponential relationship between the ground sampling distance (GSD) and the number of images required - which is a function of flight altitude. To tackle this, we use a generative network based super-resolution approach to increase the GSD of images which effectively reduces flight time. In this paper we test the efficiency and efficacy of this approach using two multimedia super-resolution implementations. We also provide quantitative results comparing the two using various image processing metrics.**

***Index Terms*—component, formatting, style, styling, insert**

I. Introduction

A. Super Resolution

Super Resolution techniques attempt to improve the spatial resolution an image incorporating additional information either based on multiple acquisitions or historic corpus data. Traditionally these techniques have involved capturing multiple images, usually from slightly different viewing geometries, and incorporating this information to improve level of detail. These techniques are well reviewed in [1]. In recent years, progress in the multimedia Super-Resolution domain has seen progress with Deep Neural Networks, especially in generating super-resolved images from a single input image, relying on additional information to be introduced from a data-corpus introduced during training of the network. A review of recent single image super resolution techniques is in [2].

The hypothesis proposed in this paper is that the learning manifold constructed from multimedia images can also help UAV Cameras since the modalities are similar enough in their image acquisition chains [3].

B. DEM Upscaling

Digital Elevation Models (DEM) are data structures that represent topographical elevation. Traditional methods of generating DEMs include SAR interferometry, ASTER GDEM, LiDAR which are readily available but may have lower spatial resolution. Although high-resolution DEMs can be generated using UAV photogrammetry their scope and availability is limited. Super-resolving DEMs is a convenient way of generating High Resolution DEM (HRDEM) from Low Resolution DEM (LRDEM). Super- resolving has attracted numerous research. Traditionally this is achieved by using interpolation techniques such as Fractal based methods [4], inverse distance weighting, natural neighbor, kriging, radial basis function, spline, etc. Zhang et. al (2021) [5] were able to create a Recursive Sub-Pixel Convolution Network (RSPCN) to estimate DEMs from augmented training samples. Argudo et. al (2018) [6] used a Fully Convolution Network to generate HRDEMs. By using a LRDEM along with an orthophoto of the same location, features could be extracted to aid in the super sampling method. DEM can also be super sampled by utilizing overlapping patches on a LRDEM and finding its corresponding HRDEM features based on non-local similarity conditions as shown by Chen et. al (2016) [7]. Instead of inputting a DEM directly into a CNN, Xu et. at (2019) [8] proposed a system where gradient maps are inputted into a pretrained CNN. The CNN can learn gradient features and is able to utilize transfer learning to generate HRDEMs. The mentioned works have all tackled super sampling DEMs using the DEM in its Low-Resolution state along with some external parameters in some cases

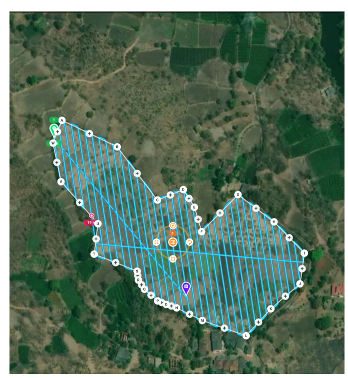

Fig. 1. Flight Path for Data Acquisition

II. Methodology

A. Data Acquisition and Study Area

The data was acquired over a mix rural-agriculture area in western India as shown in Figure 1 using commercial off- the-shelf UAVs in a grid like flight path with forward

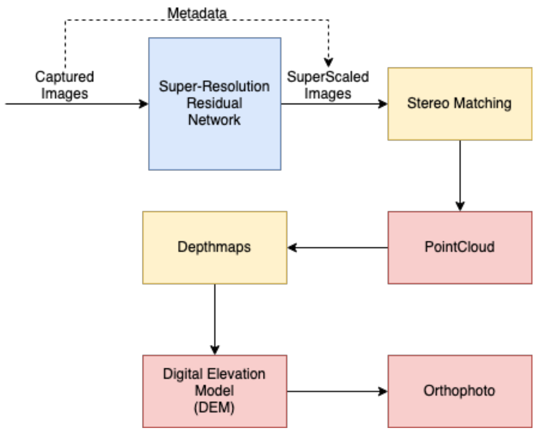

Fig. 2. Flight Path for Data Acquisition

and side overlap of 90% and 70% respectively. Traditionally such data is used to generate digital elevation models and orthomosaics. It is observed that the there exists a near exponential increase in the number of images as well as flight time (seen in fig d) with an increase in ground sampling distance or GSD (resolution) which can be expensive in terms of time and resources. In some cases, increasing GSD is not possible due to obstacles in flight path at low altitude. To overcome this limitation, this paper proposes adding a new step to the photogrammetry pipeline as seen in Figure 2 where we use popular multimedia super- resolution implementations to upscale raw images before feeding them to the above-mentioned pipeline.

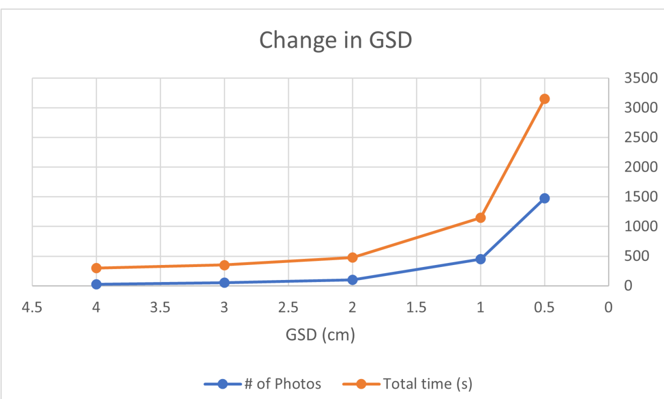

Fig. 3. Flight Path for Data Acquisition

B. Super-Resolution and Photogrammetry

Images captured are pre-processed and up-scaled using two different multimedia super-resolution techniques based on generative networks, Image Super Resolution (ISR) [9] and Super-resolution Generative-Adversarial-network (SRGAN) [10]. As seen in Figure 2, we upscale raw images captured by the UAV using ISR and SRGAN, the up-scaled images' metadata is modified to match that of the original image to maintain geo-location and camera orientation information provided by the UAV platform at the time of capture. Up-scaled images are then processed using a mix of commercial proprietary and open-source software to generate digital elevation models (DEM) and orthophotos / orthomosaics using stereo-matching and depth maps. DEMs generated by the two methods are compared to the native- high-resolution DEM to evaluate the performance of the proposed technique.

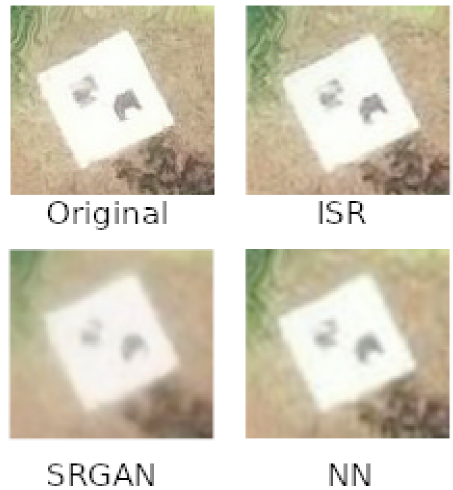

Fig. 4. Flight Path for Data Acquisition

TABLE I
COMPARISON METRICS

Metric	Original	SRGAN	ISR	NN
RMSE	-	0.010	0.002	0.002
PSNR	-	39.594	52.189	51.215
SSIM	-	0.926	0.993	0.991
SIFT Density (100 x 100 px)	-	18.54	205.75	157.07

Figure 4 shows the outputs of both the algorithms along with the original image and a nearest neighbor interpolated image. Comparative metrics have also been shown in Table 1 The next section focuses on the results and findings.

III. RESULTS AND DISCUSSION

Digital Elevation Models (DEM) generated using each method are evaluated against the DEM generated using native high-resolution imagery to assess the efficacy of using multimedia based super-resolution techniques for UAV remote sensing. Comparison metrics include Mean Error (ME), Root Mean Square Error (RMSE), Peak Signal to Noise Ratio (PSNR) and Structural Similarity Index (SSIM) where ME, RMSE and PSNR are calculated on the DEM data structure whereas SSIM is calculated by converting the DEMs into grayscale images.

TABLE II
COMPARISON METRICS

Metric	SRGAN	ISR	AVG
ME	-0.266	0.210	-0.024
MAE	0.390	0.312	0.199
RMSE	0.901	0.733	0.671
PSNR	29.95	31.74	32.51
SSIM	0.969	0.977	0.978

As seen in Table 2 it is observed that the DEM generated by ISR shows a lower RMSE which suggests that the ISR method produces DEMs with higher accuracy compared to SRGAN. Likewise, PSNR, being a function of RMSE shows comparable results. However, results become ambiguous when it comes to ME. It is observed that the DEM generated using SRGAN images has a negative ME while the DEM generated by ISR images has a positive ME. This suggests that there exists an inherent bias between the two upscaling methods when it comes to generating DEMs using multimedia super-resolution techniques.

According to Lillesand et. al (2004) [11], for a stereoscopic DEM generation, the elevation of a point is determined using

$$h_{x,y} = H - \frac{B * f}{p_{x,y}} \tag{1}$$

h - elevation of a point
H - elevation of the point above a datum (usually MSL)
B - baseline or the distance between two stereo images
f - focal length of the camera
p - parallax / distance of the same point in two stereo images

To address this discrepancy, we take the DEMs generated by both methods and generate a DEM named AVG by shifting and averaging values from the former. It is observed that the new DEM is more accurate, thus implying that combining results from both methods is feasible and effective. The efficiency of this process is assessed below.

The three DEMs image shown before represents the differences between the DEMs generated using ISR and SRGAN upscaled images as well as the combined result AVG and the original DEM. Green pixels indicate an undershoot in altitude estimation while blue pixels indicate an overshoot compared to native-high-resolution DEM. The combined DEM, AVG, gives a much better estimation and compensates for the overshoots and undershoots. As is evident, the values from the SRGAN method show a negative bias while the values belonging to ISR show a positive bias. The combination of the two (Averaging) produces results that close to the original. Figure 5 is the absolute elevation difference between the averaged DEM compared to original. An overall increase in accuracy is observed with the large blue patches indicating small errors.

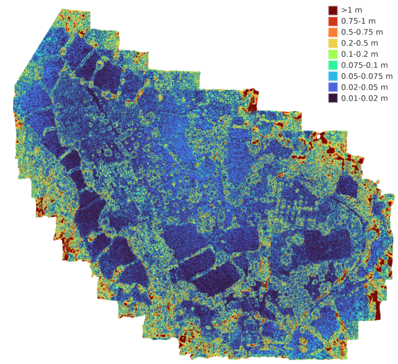

Fig. 5. Flight Path for Data Acquisition

IV. CONCUSION

Generating High-Resolution Digital Elevation Models (HR-DEM) from Low-Resolution Digital Elevation Models (LR-DEM) has been previously studied using traditional interpolation techniques as well as modern approaches like neural networks which tend to have an added benefit of learning certain patterns and approximating them with high fidelity. However, in this paper we generate HRDEMs from super resolved images in contrast to upscaling the elevation models. From the results it is evident that using multimedia super-resolution techniques to generate DEMs from low- resolution images can result in HRDEMs that have a small bias. Additionally, combining the results from the two methods can result in a HRDEM that is more accurate. Further research can include tuning the generative network for specific use to improve DEM accuracy and decrease processing time.

REFERENCES

[1] L. Yue, H. Shen, J. Li, Q. Yuan, H. Zhang, and L. Zhang, "Image super-resolution: The techniques, applications, and future," *Signal Processing*, vol. 128, pp. 389–408, 2016.
[2] S. M. A. Bashir, Y. Wang, M. Khan, and Y. Niu, "A comprehensive review of deep learning-based single image super-resolution," *PeerJ Computer Science*, vol. 7, p. e621, 2021.
[3] M. Cramer, H. Przybilla, and A. Zurhorst, "Uav cameras: Overview and geometric calibration benchmark," *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 42, p. 85, 2017.
[4] K. Arai, Y. Terayama, and K. Nakamura, "A comparative study on interpolation methods for digital elevation model," in *Proceedings of IGARSS '93 - IEEE International Geoscience and Remote Sensing Symposium*, 1993, pp. 1975–1977 vol.4.
[5] R. Zhang, S. Bian, and H. Li, "Rspcn: Super-resolution of digital elevation model based on recursive sub-pixel convolutional neural networks," *ISPRS International Journal of Geo-Information*, vol. 10, no. 8, 2021. [Online]. Available: https://www.mdpi.com/2220-9964/10/8/501
[6] O. Argudo, A. Chica, and C. Andujar, "Terrain super-resolution through aerial imagery and fully convolutional networks," *Computer Graphics Forum*, vol. 37, no. 2, pp. 101–110, 2018. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/cgf.13345

[7] Z. Chen, X. Wang, Z. Xu, and H. Wenguang, "Convolutional neural network based dem super resolution," *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XLI-B3, pp. 247–250, 06 2016.

[8] Z. Xu, Z. Chen, W. Yi, Q. Gui, W. Hou, and M. Ding, "Deep gradient prior network for dem super-resolution: Transfer learning from image to dem," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 150, pp. 80–90, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0924271619300413

[9] F. C. et al., "Isr," https://github.com/idealo/image-super-resolution, 2018.

[10] R. Hasnain, "Fast sr-gan," https://github.com/HasnainRaz/Fast-SRGAN, 2019.

[11] T. Lillesand, R. Kiefer, and J. Chipman, *Remote Sensing and Image Interpretation (Fifth Edition)*, 01 2004, vol. 146.