ADAPTIVE AND NON-ADAPTIVE FUSION ALGORITHMS ANALYSIS FOR DIGITAL SURFACE MODEL GENERATED USING CENSUS AND CONVOLUTIONAL NEURAL NETWORKS

Hessah Albanwan [1], *Rongjun Qin [1,2,3]

[1] Geospatial Data Analytics Laboratory, Department of Civil, Environmental and Geodetic Engineering, The Ohio State University, Columbus, OH, USA - albanwan.1@osu.edu
[2] Department of Electrical and Computer Engineering, The Ohio State University - qin.324@osu.edu
[3] Translational Data Analytics Institute, The Ohio State University.

Commission VI, WG VI/4

KEY WORDS: Census, Matching Cost Convolutional Neural networks (MC-CNN), Adaptive Fusion, Non-Adaptive Fusion, Digital Surface Model (DSM)

ABSTRACT:

The digital surface models (DSM) fusion algorithms are one of the ongoing challenging problems to enhance the quality of 3D models, especially for complex regions with variable radiometric and geometric distortions like satellite datasets. DSM generation using Multiview stereo analysis (MVS) is the most common cost-efficient approach to recover elevations. Algorithms like Census-semi global matching (SGM) and Convolutional Neural Networks (MC-CNN) have been successfully implemented to generate the disparity and recover DSMs; however, their performances are limited when matching stereo pair images with ill-posed regions, low texture, dense texture, occluded, or noisy, which can yield missing or incorrect elevation values, in additions to fuzzy boundaries. DSM fusion algorithms have proven to tackle such problems, but their performance may vary based on the quality of the input and the type of fusion which can be classified into adaptive and non-adaptive. In this paper, we evaluate the performance of the adaptive and nonadaptive fusion methods using median filter, adaptive median filter, K-median clustering fusion, weighted average fusion, and adaptive spatiotemporal fusion for DSM generated using Census and MC-CNN. We perform our evaluation on 9 testing regions using stereo pair images from Worldview-3 satellite to generate DSMs using Census and MC-CNN. Our results show that adaptive fusion algorithms are more accurate than non-adaptive algorithms in predicting elevations due to their ability to learn from temporal and contextual information. Our results also show that MC-CNN produces better fusion results with a lower overall average RMSE than Census.

1. INTRODUCTION

The quality of the digital surface model (DSM) generated from satellite images has always been a crucial element to most remote sensing and photogrammetry applications. DSM generated using Multiview stereo (MVS) algorithms is very common due to its high efficiency and low cost, but it is limited performance due to its sensitivity to radiometric and geometric distortions in the stereo images, which lead to noise, occlusions, missing elevation values, etc. in the DSM. One of the most promising techniques that raised significant attention to improving the quality of DSM is fusion (Albanwan and Qin, 2020; Cigla et al., 2017; Papasaika et al., 2008). Fusion is the process of combining multi-temporal DSMs into a single high-quality DSM; it takes advantage of the redundant temporal information to compensate for the incorrect representations or missing elevation points (Albanwan and Qin, 2020; Cigla et al., 2017; Papasaika et al., 2008). DSM fusion algorithms can be categorized into 1) adaptive and 2) nonadaptive; the prior approach learns from the context, shape, and type of objects in the scene, in addition to the temporal information between DSMs, whereas non-adaptive approaches simply learn and predict elevation from the temporal information (Cigla et al., 2017; Wang and Gong, 2019). One of the oldest non-adaptive algorithms to perform fusion is median filtering, it is known to be robust to outliers and preserve boundaries of the objects

(Kuschk and D'Angelo, 2013; Ozcanli et al., 2015). Many fusion algorithms have upgraded simple median filter to a more robust approach by including the concept of adaptivity to scene objects, for instance, Qin, (2017), has proposed adaptive median filtering where he incorporates a flexible window that is formed on the shape of the object and applies median filtering on each object instead of using a fixed-sized window. Such adaptive methods are able to retain boundaries and shapes of objects in the scene (e.g., buildings, roads, etc.). Other studies have shown that the uncertainty of the DSM can be correlated with the class cover type (Albanwan and Qin, 2020), for instance, trees and grass changes based on the acquisition date and season, which can adversely influence the performance of MVS algorithm, and reduces the DMS uncertainty, whereas structures like buildings and roads are less changeable over time and most of the times have lower uncertainty. This led to the development of class-oriented fusion algorithms, as an example, (Albanwan and Qin, 2020) developed adaptive spatiotemporal fusion to impose different bandwidths based on the class of objects. On the other hand, other methods used the concept of k-median clustering to locate the cluster with the most consistent elevation points to reduce the number of outliers (Facciolo et al., 2017), where others used a pair ranking scheme based on a scoring technique to evaluate and sort stereo pairs by their quality and merge only the pairs with the best scores (Facciolo et al., 2017; Qin, 2019; Qin et al., 2020).

There are many factors influencing the fusion outcomes including the type of input and the fusion algorithm. Traditional

---

* Corresponding author

MVS algorithms generate DSMs using local window-based approaches, which extract and match similar feature correspondences to compute the disparity and then transforming it into a DSM. Census-Semi global matching (SGM) proposed by (Hirschmuller, 2005) is one of the simplest and most cost-efficient algorithms to generate DSMs; it is one of the widely used methods from 2005 until today. Although it is considered invariant to radiometric changes, it is still sensitive to illumination changes and window size, therefore many have proposed adaptive window to capture the size and shape of objects instead of rigid window (Han et al., 2020; Heiko Hirschmuller and Scharstein, 2007; Loghman and Kim, 2013). Nowadays, deep learning algorithms have captured a great interest in the area of dense image matching and elevation generation, where they are intended to enhance dense image matching by better understanding and learning from the scene components (Chang and Chen, 2018; Hamid et al., 2020; Zbontar and LeCun, 2015). (Zbontar and LeCun, 2015) are the first to introduce matching cost Convolutional neural networks (MC-CNN) for stereo analysis and disparity and DSM generation; their target was to enhance the matching cost and produce faster and better similarity matching results. Although MC-CNN is able to capture the shapes of objects well, it still requires a lot of post-processing and filtering. Many deep learning MVS algorithms are inspired by MC-CNN, which are developed to further enhance predictability, cost matching, and time efficiency. One of the drawbacks to deep learning algorithms is their limited performance due to the training process; they require a great amount of time and data along with the ground truth data for training. Additionally, the generalization of the training is a very critical matter, the trained model should be well reflected on the testing dataset or any other dataset regardless of the domain difference, otherwise, the network must be retrained.

In this work, we aim to evaluate fusion adaptive and non-adaptive algorithms for DSM generated using Census and MC-CNN, we use these algorithms as they have been widely used in the area of MVS algorithms. Understanding such work can help to close the holes in current algorithms of fusion and provide insights to improving their performance and output, in addition, to help the user understand how each data works under different fusion algorithms.

The paper is organized as follows: Section 2 includes the data description, pre-processing steps like DSM generation, and the fusion algorithms used, Section 3. Include the discussion and analysis of the results, and finally, we present our conclusion in the last section.

2. METHODOLOGY

2.1 Dataset description

In Our work, we use three different datasets from Omaha (OMA), Jacksonville (JAX), and Argentina (ARG) conducted from a very high-resolution satellite Worldview-3. Each dataset includes hundreds of multispectral image pairs with 0.3 meters spatial resolution. Every dataset includes three testing regions with varied spatial complexity and density, as some locations may be urban or suburban areas with either dense small houses or sparse large buildings or a mix of both as can be seen in Table 1. The total number of DSMs generated ranges between 91 – 482. The dataset images were captured within almost a one-year

time-span from September 2014- November 2015, October 2014- February 2016, and January 2015-January 2016 for OMA, JAX, and ARG respectively. For evaluation, we use a reference ground truth DSM generated from the LiDAR dataset.

OMA		
Dataset ID		
	I	II	III
Dimensions [W x H]	1168 x 1165	1206 x 1202	1187 x 1190
Number of DSMs	360	293	239
Type of region: Urban (U)-Sub-urban (S) / Dense (D)-Sparse (S)	U/S	S/S	U/D
JAX		
Dataset ID		
	I	II	III
Dimensions [W x H]	1163 x 1164	1122 x 1152	1133 x1164
Number of DSMs	138	91	247
Type of region: Urban (U)-Sub-urban (S) / Dense (D)-Sparse (S)	U/S	S/S	U/D
ARG		
Dataset ID		
	I	II	III
Dimensions [W x H]	2565 x 2561	4199 x 4256	4062 x 4208
Number of DSMs	479	472	482
Type of region: Urban (U)-Sub-urban (S) / Dense (D)-Sparse (S)	U/D	U/D	U/D

Table 1. The dataset information and testing regions.

2.2 Data pre-processing

Our pre-processing steps can be summarized as the following: 1) Image pair selection based on specific criteria, 2) geo-registration to ensure alignment, and finally 3) Disparity and DSM generation using census and MC-CNN cost metrics followed by semi-global matching for optimization (Hirschmuller, 2005).

2.2.1 Image pair selection
Since we have about 20 images for every dataset, we can obtain hundreds of DSMs for the stereo pairs, however, in practice, only a few numbers of DSMs can be available. Therefore, we choose the best 20 pairs to generate elevation and perform the fusion. The selection of the stereo pairs is performed using a scoring or ranking scheme that sorts the images based on the metadata information from the pair of images, we include geometrical information such as the intersection angles, sun angle difference,

and the number of days in which the images were acquired. We rank the stereo pairs based on their scores that are computed from the optimal values of sun angle difference and intersection angles as mentioned in (Qin, 2019).

2.2.2 Ortho-ready image, geo-registration, and image rectification
Each image is transformed into an ortho-ready image and registered to assure alignment of images using a reference image and using RPC Stereo Processor RSP software (Qin, 2016). We also rectify the images to produce epipolar images in which we use in the following steps for disparity and DSM generation using RSP software.

2.2.3 Disparity and digital surface models (DSM) generation
Initially, we generate the disparity from the rectified epipolar images, where we Census (Hirschmuller, 2005) and MC-CNN (Zbontar and LeCun, 2015) are used as the cost matching metrics to determine the horizontal displacement between feature correspondences, which leads to the disparity images. Census requires a predefined window to perform string bit/binary transformation, where any pixel lARGer than the central pixel takes a value of 1 and 0 otherwise. This transformation process is followed by a hamming distance to measure the score and compute the disparity based on the minimum score. On the other hand, we follow the CNN architect as in (Zbontar and LeCun, 2015), where we first train the MC-CNN using a satellite dataset, then we extract image patches with a size of 9x9. The image patches are then fed into a single layer of MC-CNNs each with size 5x5 and 32 kernels, followed by a couple of series of fully connected layers each with 200 neurons, which are then concatenated into a single layer of 400 neurons and passed to several layers with 300 neurons until the last layer which classified each pixel into a match or no match. The DSM is finally generated using Semi-global matching as proposed by (Hirschmuller, 2005) and implemented in RSP software.

2.3 DSM fusion and evaluation

For DSM integration, we use different fusion algorithms that vary between adaptive and nonadaptive approaches including:

1) Non-adaptive fusion algorithms:
   - Median filter
   - K-median clustering fusion (Facciolo et al., 2017)
   - Weighted average fusion (Papasaika et al., 2008)
2) Adaptive fusion algorithms:
   - Adaptive median fusion (Qin, 2017)
   - Adaptive spatiotemporal fusion (Albanwan and Qin, 2020)

The median filter simply takes the median of the temporal DSMs at any pixel position and produces median DSMs. The adaptive median filter as suggested by (Qin, 2017) applies the same concept except using an adaptive window that captures the shape of the object and applies median filtering on individual objects. K-median clustering in fusion is proposed by (Facciolo et al., 2017) to merge several pre-ranked stereo pairs into a single DSM by clustering the input temporal elevations and picking the cluster with minimum cost to fuse its elevation points. Weighted average fusion on the other hand is a broader algorithm (Papasaika et al., 2008); first, it computes the residual map

between two DSM images, then use these residuals to get weights and multiply by the elevations as follows:

$$F = \frac{\sum_{i=1}^{t} w_i * DSMi}{\sum_{i=1}^{t} wi} \quad (1)$$

Where F= the fused DSM
  w_i= is the weight computed from the residual maps
  t= the number of temporal DSMs

Finally, we perform adaptive spatiotemporal fusion by applying different bandwidths for different classes, where highly complex and seasonally variable classes like trees, grass, and water take higher bandwidths, while static rigid objects like buildings and roads take smaller bandwidths. In Adaptive spatiotemporal fusion, we first compute the median and generate masks as suggested by (Albanwan and Qin, 2020), then difference the spatial and temporal DSMs from the median to calculate the weight and impose different bandwidths for each type of class, the algorithm is as follows:

$$F(x,y) = \frac{1}{WT} \sum_{x=1}^{width} \sum_{y=1}^{height} Wr * Ws * Wh * hmed(x,y,t)$$
(2)

Where F = fused DSM
  x, y = position of pixels in the DSM
  h_{med} = median height from the temporal DSMs
  W_r= spectral weight
  W_s= spatial weight
  W_h= temporal height weight
  W_T= total weight

Wr and Ws computes spatial and range weights from the orthophoto as the bilateral filter, as follows:

$$Wr(x,y) = \exp\left(-\frac{||I(x,y)-I(k,j)||^2}{2\sigma_r^2}\right) \quad (3)$$

$$Ws(x,y) = \exp\left(-\frac{||(x-k)^2+(y-l)^2||}{2\sigma_s^2}\right) \quad (4)$$

Where I = the orthophoto image x, y, l, k = the position of current and neighboring in the window
  σ_r = range bandwidth
  σ_s = spatial bandwidth

The adaptivity for classes is computed using the following equation:

$$W_h(x,y) = exp\left(-\frac{||hmed-h(i,j,t)||^2}{2\sigma 2h}\right) \quad (5)$$

Where σ_h= the height bandwidth relative to each class

We perform the fusion on the 20 best pairs, and evaluate the fusion output using root mean squared error (RMSE) and the LiDAR data as the ground truth (GT) as follows:

$$RMSE(t) = \sqrt{\frac{\sum_{i=1}^{N}(F-GT)^2}{N}} \quad (6)$$

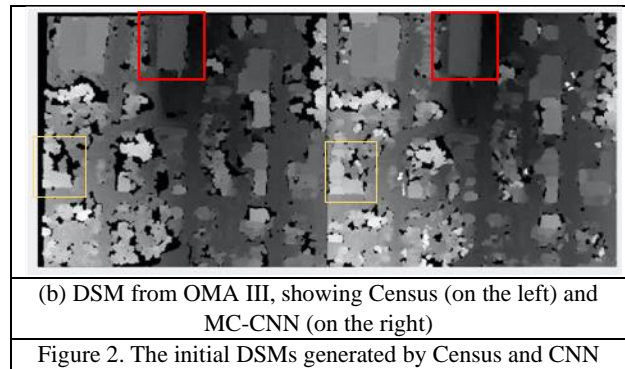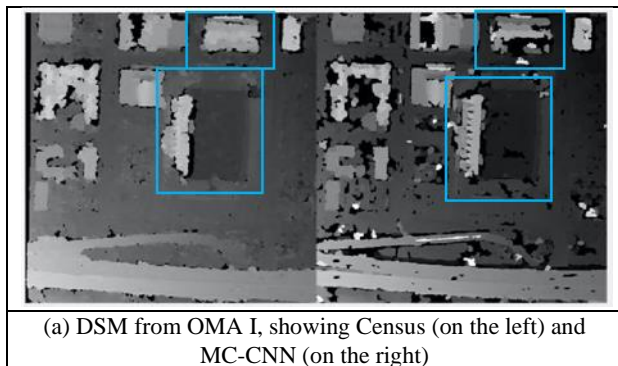Where F is the fused DSM and N is the number of valid pixels.

3. RESULTS AND ANALYSIS

3.1 Parameter selection for fusion methods

For the adaptive median filtering, we use the default setting as (Qin, 2017), and for the k-median clustering, we use elevation point from the temporal and spatial domains, where the window size is set to 5 and the threshold to determine the cost and when to stop clustering is set to less than 10. For the weighted average filter, we used the median fused DSM as a reference to compute the residuals map and compute the corresponding weights. Finally, for the adaptive spatiotemporal fusion we used a window size of 5, we obtain 4 to 5 classes for each dataset including buildings, roads and ground, trees, grass, and water. The spectral and spatial bandwidths are chosen empirically and set to $\sigma_r = 30$ and $\sigma_s = 5$; the elevation bandwidths for the classes are set to $\sigma_{Buildings} = 3$, $\sigma_{Road\_Ground} = 3$, $\sigma_{Trees} = 7$, $\sigma_{Grasss} = 7$, and $\sigma_{Water\_surface} = 7$.

3.2 DSM analysis

The first step in the analysis is DSM quality inspection, which is necessary for the evaluation and comparison between the before and after fusion results. Figure 2. Represents a sample of the initial DSM generated by Census and MC-CNN from two OMA testing regions (OMA I and OMA III). We can notice that regardless of the method, the generated DSM is always associated with problems such as noise, outliers, missing elevation points, and fuzzy representation of object shapes, which can raise due to mismatching in the dense image matching process. These issues can vary DSMs and may be addressed in some of the generation algorithms, for instance, Census produce DSMs that are smoother, fuller, and well distributed as can be seen in Figure 2. (a) and highlighted by the blue boxes. On the other hand, MC-CNN includes more missing elevation points which can be obvious from the black spots in the right image of Figure 2. Nevertheless, MC-CNN captures the edges and boundaries of buildings better than Census, which can be indicated in the red and yellow box around the buildings in Figure 2. (b).

(a) DSM from OMA I, showing Census (on the left) and MC-CNN (on the right)

(b) DSM from OMA III, showing Census (on the left) and MC-CNN (on the right)

Figure 2. The initial DSMs generated by Census and CNN

3.3 Statistical analysis

We provide a comprehensive analysis and comparison of the overall performance of the adaptive and nonadaptive fusion algorithms, in addition to the performance of fusion algorithms for the non-deep learning and deep learning (i.e., Census and MC-CNN) DSM generation algorithms. We present a visual and statistical evaluation for the results as presented in Figure 3. and Table 2. From the bar representation in Figure 3., we can notice that the adaptive methods such as the adaptive median fusion and adaptive spatiotemporal fusion always produce less uncertainties in comparison to the other fusion algorithms, which can be indicated in the orange and purple bars in Figure 3. They also had the lowest average RMSE of 5 (meters) as can be seen in Table 2. Their performance is consistent over different datasets regardless of the DSM generation algorithm, which is indicated by having less varying RMSE values (See Figure 3. and Table 2.). On the other hand, we can notice that fusion methods like weighted average fusion and k-median clustering fusion have the highest ranges of RMSE as indicated from green and yellow bars in Figure 3., their RMSE also ranges between 5 to 35 (meters).

We can also notice during our analysis that MC-CNN performs better than Census in terms of robustness to outliers, like in the case of median filter, k-means filtering, adaptive spatiotemporal fusion, and weighted average fusion for dataset ARG I, MC-CNN was able to suppress these invalid elevation errors that were not resolved in Census. This can lead to concluding that fusing Census DSMs is not robust to outliers as in MC-CNN. Table 2., also show that the error range for MC-CNN was about 4-12, whereas for Census the range is between 6 to 31.
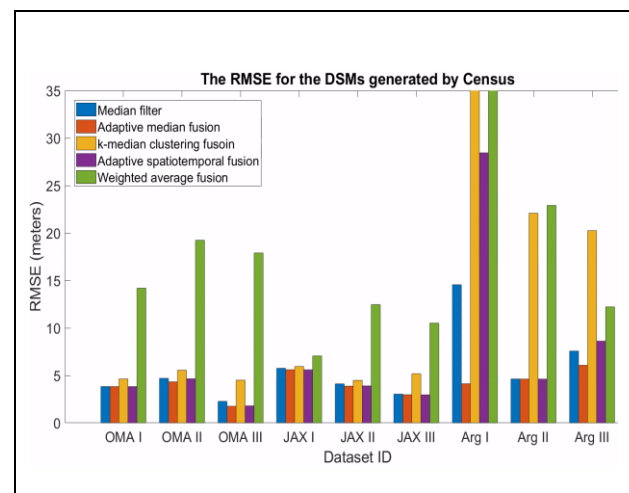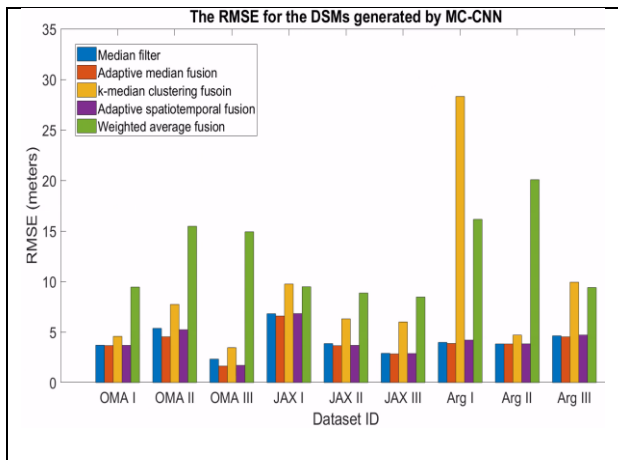
Figure 3. A comparison between fusion algorithms for DSMs generated from Census and MC-CNN.

OMA		MF	AMF	KMF	ASPF	WAF
I	Census	6.39	6.38	13.96	**6.36**	18.29
	MC-CNN	7.85	**7.68**	7.79	**7.68**	11.84
II	Census	6.47	**4.33**	7.82	4.65	36.33
	MC-CNN	7.21	**4.53**	8.53	5.30	33.57
III	Census	2.36	**1.76**	4.83	1.79	18.36
	MC-CNN	2.41	**1.64**	3.81	1.70	15.28
JAX		MF	AMF	KMF	ASPF	WAF
I	Census	5.76	**5.60**	5.95	**5.60**	7.06
	MC-CNN	6.81	**6.59**	9.86	6.82	9.92
II	Census	4.11	**3.87**	5.27	3.90	14.92
	MC-CNN	3.86	**3.64**	22.88	3.68	9.68
III	Census	3.03	2.97	5.57	**2.96**	17.19
	MC-CNN	2.89	**2.84**	7.53	2.87	8.51
ARG		MF	AMF	KMF	ASPF	WAF
I	Census	14.55	**4.12**	74.65	28.45	37.85
	MC-CNN	3.98	**3.87**	28.31	4.20	16.15
II	Census	4.63	**4.61**	22.08	**4.61**	22.91
	MC-CNN	3.82	**3.82**	4.70	3.83	20.07
III	Census	7.57	**6.07**	20.25	8.62	12.22
	MC-CNN	4.63	**4.52**	9.93	4.70	9.40

Table 2. The uncertainty of the fused DSMs. Note: The abbreviations stand for median filter (MF), Adaptive median filter (AMF), adaptive spatiotemporal fusion (ASPF), K-median clustering fusion (KMF), and weighted average fusion (WAF), and Bold numbers indicate the lowest RMSE for all fusion methods in each testing region and each DSM generation algorithm.

3.4 Visual analysis

In general, we can notice that fusion has solved many problems related to missing or incorrect elevation points (See Figure 4). Moreover, the visual illustration in Figure 4. confirms our findings in Section 3.3. that adaptive methods produce the best fused results not only statistically but also visually, this can be

seen from the second and last rows in Figure 4, where they produced smooth and sharp DSMs, we can also see that the buildings are well captured and most of the noise in the original DSMs have been reduced. Other methods like k-median clustering fusion and weighted average fusion are less robust to noise and outliers in the DSM, which is evident from the third and fourth images in Figure 4.

Additionally, we can see that fusion using Census is better in terms of generating smoothed fused DSM for all fusion methods (Seen Figure 3), especially in the case of k-median clustering fusion and weighted average fusion, whereas fusion of MC-CNN DSM can produce noisy results in these methods. Nevertheless, unlike Census, MC-CNN fused DSMs seem better in edge and boundary preservation in all fusion methods, as can be noticed in Figure 3 where buildings are sharp and better outlined.
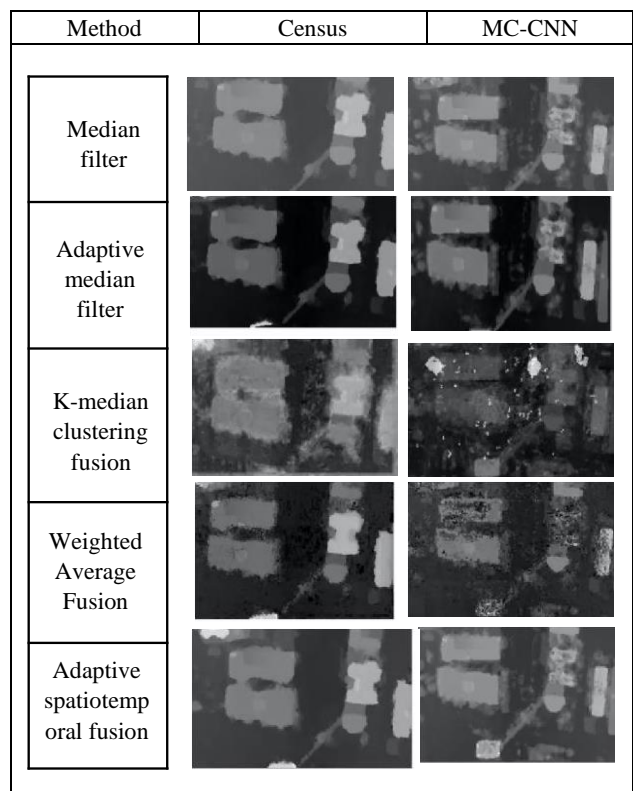

Figure 4. sample subsection of JAX II dataset to show the difference between Census and MC-CNN in the fusion algorithms.

4. CONCLUSION

To conclude, our work has shown that in general median filter produces fairly good results with low computational cost. However, adaptive fusion methods such as adaptive median fusion and adaptive spatiotemporal fusion always produce the best results due to their robustness towards outliers and flexibility to learn and predict elevation from homogeneous objects and a consistent set of neighboring pixels. The generalized non-adaptive fusion method did not perform as well due to the lack of contextual and temporal information and correlation in the elevation prediction and fusion process. We also evaluate the usage of DSMs generated using different dense

image matching algorithms in the fusion process. We found that in general, MC-CNN performs better than Census for most fusion algorithms, due to the architecture and mechanism of MC-CNN and since it can learn similarity from highly complex features and produces more detailed DSM. Therefore, combining its results can help achieve better fusion outcomes. Census on the other hand may generate less accurate DSMs than MC-CNN but leverages between quality and computational extensive algorithms. Overall, Census can generate satisfactory results and is appropriate for when there are limited resources and expertise to avoid training and deep learning algorithms. Although MCCNN has shown its ability to preserve the shape of objects, they still require post-processing and refinement to reduce the noise and outliers in the resultant fused DSMs. Such work can further be extended to improve fusion methods by combining DSMs from both MC-CNN and Census to take the advantage of both methods, however, the distribution of data must be taken into consideration, since different methods generate different DSMs.

ACKNOWLEDGEMENTS

REFERENCES

Albanwan, H., and Qin, R., 2020. Enhancement of DSM by Fusion Using Adaptive and Semantic-Guided Spatiotemporal Filtering. *Annals. Photogramm. Remote Sens. Spatial Inf. Sci.* 2020. ISPRS Congress (2020/2021). ISPRS, Nice, Fr.

Cigla, C., Brockers, R., and Matthies, L., 2017. Gaussian Mixture Models for Temporal Elevation Fusion. *2017 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 889–897. https://doi.org/10.1109/WACV.2017.104

Facciolo, G., De Franchis, C., and Meinhardt-Llopis, E., 2017. Automatic 3D Reconstruction from Multi-date Satellite Images. *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 1542–1551. https://doi.org/10.1109/CVPRW.2017.198

Han, Y., Liu, W., Huang, X., Wang, S., and Qin, R., 2020. Stereo Dense Image Matching by Adaptive Fusion of Multiple-Window Matching Results. *Remote Sensing*, *12*(19), 3138. https://doi.org/10.3390/rs12193138

Hirschmuller, H., 2005. Accurate and Efficient Stereo Processing by Semi-Global Matching and Mutual Information. *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, *2*, 807–814. https://doi.org/10.1109/CVPR.2005.56

Hirschmuller, H., and Scharstein, D., 2007. Evaluation of Cost Functions for Stereo Matching. *2007 IEEE Conference on Computer Vision and Pattern Recognition*, 1–8. https://doi.org/10.1109/CVPR.2007.383248

Kuschk, G., and d'Angelo, P., 2013. Fusion of Multi-Resolution Digital Surface Models. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XL-1/W3, pp. 247–251. https://doi.org/10.5194/isprsarchives-XL-1-W3-247-2013

Loghman, M., and Kim, J., 2013. SGM-based Dense Disparity Estimation Using Adaptive Census Transform. *2013 International Conference on Connected Vehicles and Expo (ICCVE)*, 592–597. https://doi.org/10.1109/ICCVE.2013.6799860

Ozcanli, O. C., Dong, Y., Mundy, J. L., Webb, H., Hammoud, R., and Tom, V., 2015 A Comparison of Stereo and Multiview 3-D Reconstruction Using Cross-Sensor Satellite Imagery. *2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 17–25. https://doi.org/10.1109/CVPRW.2015.7301292

Papasaika, H., Poli, D., and Baltsavias, E., 2008. A Framework for the Fusion of Digital Elevation Models. *Int Arch Photogramm Remote Sens Spatial Inf Sci*, 37(B2):811–817

Qin, R., 2016. RPC Stereo Processor (RSP) – A Software Package Digital Surface Model and Orthophoto Generation from Satellite Stereo Imagery. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, *III–1*, 77–82. https://doi.org/10.5194/isprs-annals-III-1-77-2016

Qin, R., 2019. A Critical Analysis of Satellite Stereo Pairs for Digital Surface Model Generation and a Matching Quality Prediction Model. *ISPRS Journal of Photogrammetry and Remote Sensing*, *154*, 139–150. https://doi.org/10.1016/j.isprsjprs.2019.06.005

Qin, R., 2017. Automated 3D Recovery from Very High-Resolution Multi-View Satellite Images, *Proceedings of the ASPRS (IGTF) Annual Conference,* Baltimore, Maryland, p. 10.

Qin, R., Song, S., Ling, X., and Elhashash, M., 2020. *3D Reconstruction Through Fusion of Cross-View Images. In Kwan, C. (Eds.)*, Recent Advances in Image Restoration with Applications to Real World Problems. IntechOpen. https://doi.org/10.5772/intechopen.93099

Wang, N., and Gong, X., 2019. Adaptive Fusion for RGB-D Salient Object Detection. *IEEE Access*, *7*, 55277– 55284. https://doi.org/10.1109/ACCESS.2019.2913107