ATTENTION-GUIDED COST VOLUME REFINEMENT NETWORK FOR SATELLITE STEREO IMAGE MATCHING

W. J. Jeong[1], S. Y. Park [1]*

[1] School of Electronic and Electrical Engineering, Kyungpook National University, Daegu 41566, South Korea,
- autowmacma@gmail.com, sypark@knu.ac.kr

Commission II, WG II/1

KEY WORDS: Satellite stereo images, Disparity estimation, Guided Cost Volume, Attention module, Residual network

ABSTRACT:

In remote sensing, disparity calculation using stereo images is a very necessary task and provides information for estimating the terrain elevation. The fields using disparity of stereo satellite images are used in various fields such as terrain models, autonomous driving using 3D maps, and content development. However, extracting disparity from stereo satellite images is a very difficult task, and inaccurate disparity may be extracted due to complex environments, façade areas of buildings, and texture-less areas. Our proposed method improves feature extraction and 3D aggregation steps based on Gwc-Net using stereo images rectified through RPC (Rational Polynomial Coefficients). To this achieve, we first improve the accuracy of the initial cost volume by extracting important features using the attention module 2D CBAM. In addition, in the aggregation step, we use 3D CBAM to extract important features from the cost volume and use GCE (Correlate-and-Excite) to guide image features to the cost volume to improve disparity. To evaluate the proposed method, the accuracy of disparity is evaluated using RPC-corrected stereo satellite images of DFC2019 data track2 of the US3D dataset. As a result of the experiment, the proposed method exhibited improvement compared to the baseline Gwc-Net.

1. INTRODUCTION

In remote sensing, disparity estimation using satellite images is a necessary task. A disparity map of a satellite image provides information needed to estimate the height or elevation of a terrain. The fields using disparity of satellite images are used in various fields such as terrain models, autonomous driving using 3D maps, and content development. The goal of disparity estimation is to calculate the horizontal displacement value by matching the corresponding pixels in the left and right images given a pair of modified stereo images. Among the disparity estimation methods, the classical method selects a four-step pipeline to estimate stereo matching (Scharstein and Szeliski, 2002). Disparity estimation is performed through several steps, including matching cost calculation, cost aggregation, disparity calculation, and disparity improvement. Recently, a learning-based stereo method using convolutional neural networks (CNN) has been proposed. Examples of these methods include GC-Net (Kendall et al., 2017), PSMNet (Chang and Chen, 2018), and GwcNet (Guo et al., 2019). GC-Net extracts feature from each image and concatenate the features of the left and right images to create a 4D cost volume. Next, the cost aggregation step is performed on the 4D cost volume to improve the noised initial cost volume. PSMNet extracts valuable features using the SPP module (Spatial Pyramid Pooling module) in the feature extraction of each image and improves the cost volume through the stacked hourglass module in the aggregation step. GwcNet proposes to construct a cost volume by fusion of group correlation and concatenation volume. The generated cost volume is aggregated similarly to PSM-Net to improve the cost volume to estimate disparity. However, remote-sensing image is more complex than natural images, making stereo-matching challenging. This is due to texture-less areas, repetitive patterns, complex structures, disparity discontinuity, and occlusion in certain areas caused by tall buildings. Therefore, the proposed method improves the accuracy of disparity by using 3D-CBAM (Huang et al., 2020) and GCE(Correlate-and-Excite) modules (Bangunharcana et al., 2021) in the aggregation step. The features of the reference image is calculated as weights through GCE module. Excite the geometric features of each region by multiplying the calculated weight by the cost volume. In addition, channel and spatial features are aggregated using 3D-CBAM. This reduces inconsistency of brightness and prevents the occurrence of high-frequency noise in fine structures. The proposed method is based on GwcNet and improves the accuracy of disparity by adding GCE and 3D-CBAM before and after each stacked hourglass module. In addition, in the feature extraction step, more valuable features are extracted using CBAM (Woo et al., 2018). The proposed method is tested on DFC2019 track 2 data of the US3D dataset (Bosch et al., 2019, Le Saux et al., 2019) rectified with RPC information. The results of the evaluation confirmed that the proposed method outperformed the GwcNet. The structure of this paper is as follows. First, Chapter 2 introduces related works to our proposed method. Section 3 describes the step-by-step configuration of the proposed method. Chapter 4 shows the experiment. Finally, in Section 5, we describe the conclusions.

2. RELATED WORK

The CNN-based approach in the stereo-matching task has shown great potential compared to classical methods. In particular, CNN can be applied in complex scenes, texture-less areas and occluded areas to obtain more accurate disparity. The first method to apply CNN to stereo matching is MC-CNN

* Corresponding author

(Zbontar and Yann, 2015), which estimates disparity by applying CNN to a 4-step pipeline, which is a classical method. MC-CNN improves disparity by calculating the similarity between features extracted using CNN during the matching computation. Recently, end-to-end methods have dominated the field of stereo matching, including DispNet (Mayer et al., 2016), GC-Net (Kendall et al., 2017), and PSM-Net(Chang and Chen, 2018). DispNet(Mayer et al., 2016) is the first model to apply end-to-end stereo matching. DispNet is designed as an encoder-decoder and is a method of estimating disparity by extracting features from left and right images and then calculating the similarity between left and right correlation. However, this method is less accurate than the 3D convolution method. GC-Net is a representative method using 3D convolution, which extracts the deep features of the left and right images, then concatenates the left and right features to create a 4D cost volume for the entire disparity range. The next step is cost aggregation using 3D convolution. PSM-Net utilizes the SPP module to extract feature maps with varying receptive fields. The following cost aggregation step is performed using a stacked hourglass module to improve accuracy. In CAR-Net(Huang et al., 2020), CBAM-ResNeXt, which combines attention, and 3D-CBAM, which is connected to 3D CNN for cost aggregation, is proposed. This method extracts valuable features by utilizing CBAM at each layer of ResNeXt(Xie et al., 2017) during the feature extraction step. In addition, this model used three hourglass models in the cost aggregation step and applied 3D- CBAM to each bottleneck model to improve accuracy by reducing high-frequency noise and inconsistency luminosity. CoEx(Bangunharcana et al., 2021) is a method that reduces the computational cost due to 3D convolution by using image features and improves the aggregation step. This method uses a reference image to improve the cost aggregation step. This module computes image feature weights and excites the cost volume. These methods improve cost volume by the interaction between context features and geometric features. GwcNet proposed a group-wise correlation method, which generates cost volume and improves cost volume by fusion of the concatenation method used in GC-Net and group-wise correlation. This method does not lose as much information as the single-channel correlation method (Mayer et al., 2016) and does not require more parameters in the cost aggregation step to learn correlation, like the concatenation method. This paper improves cost aggregation using CoEx's GCE module and 3D-CBAM introduced in CAR-Net as a GwcNet-based network. In addition, we enhance the features of each image using CBAM(Woo et al., 2018). Our method is similar to CAR-Net. However, we further improve disparity estimation by applying context information to cost volume using image features. The proposed method exhibits robustness against weather changes and luminance inconsistencies, which are common characteristics of satellite images, resulting in more accurate disparity estimation than GwcNet.

3. APPROCHS

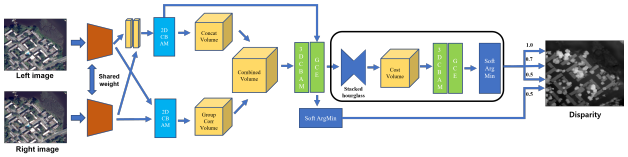

Figure 1. The overall framework of the proposed disparity estimation network.

The structure we propose is a stereo-matching network that improves disparity using attention and GCE modules (Bangunharcana et al., 2021). The proposed model consists of feature extraction, cost volume construction, cost aggregation, and disparity prediction based on Gwc-Net (Guo et al., 2019). We provide a detailed description of the proposed network's architecture and introduce the loss function in the network. The proposed network is composed of four main components: feature extraction, cost volume construction, cost aggregation, and disparity computation.

3.1 Feature extraction

Feature extraction is a step for extracting meaningful features from images. Our structure consists of a CBAM module (Woo et al., 2018) and a network similar to ResNet used in GwcNet (Guo et al., 2019). We determined that extracting meaningful features from remote sensing images with unary feature maps is difficult and improved them using CBAM. In our network, we first extract the features with a network similar to ResNet to extract the features of the left and right images. During this process, the resolution size of each feature map is reduced to 1/4 of its original size. Next, we concatenate the three layers to extract unary feature maps. To enhance the quality of the extracted unary feature maps, we apply the CBAM module. The enhanced unary feature maps are employed to generate the concatenation volume. Additionally, we apply CBAM to compress and enhance the extracted unary feature maps, extracting features to generate the group-wise correlation volume. Finally, during the cost aggregation step, the two features used for concatenation volume generation and group-wise correlation volume generation are connected to create a feature map that guides the GCE module (Bangunharcana et al., 2021).

3.2 Cost volume construction

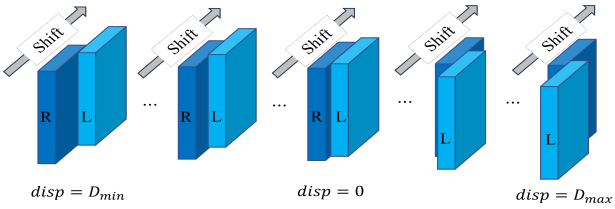

Figure 2. Cost volume construction work, the right feature shifts from -Dmin to Dmax, resulting in a disparity range up to [Dmin, Dmax], where Dmin is the negative disparity region and Dmax is the positive disparity region.

Using the two feature maps extracted in the previous step, as shown in Figure 1, first create two concatenation volumes and an initial cost volume of group-wise correlation volumes. These cost volumes have different characteristics. The concatenation volume contains abundant information, while the group-wise correlation volume provides correlation information between the left and right features. These cost volumes are created by shifting the left and right features. Typically, the cost volume creation method employs a network that ensures the production of only non-negative values. However, in the remote-sensed stereo images used in our proposed method, negative and positive disparities can be present due to the significantly different viewing angles. Therefore, as shown in Figure 2, we perform the shift operation to construct a cost volume that contains both positive and negative disparity (Tao et al., 2020). This is because the input image epipolar rectification, resulting in the disparity only existing in the horizontal direction. As shown in Figure 2, the reference feature is awaiting, and the target feature slides horizontally on the reference feature. Following the same procedure as shown in Figure 2, the 4D cost volume is obtained, proceeding with all disparity ranges. The

proposed method follows the same process as described above for both the group-wise correlation volume and the concatenation volume. These two cost volumes are concatenated to create a single cost volume.

3.3 Cost aggregation

In the next step, the cost aggregation process is conducted utilizing the initial cost volume. The cost aggregation stage involves three hourglass networks, four 3D-CBAM modules (Huang et al., 2020), and four GCE modules (Bangunharcana et al., 2021). As shown in Figure 1, our proposed method enhances the cost volume by incorporating 3D-CBAM and the GCE module before and after the output of the stacked hourglass module. First, an improved cost volume with noise removed is obtained by 3D-CBAM. 3D-CBAM is an extension of CBAM (Woo et al., 2018) that incorporates attention mechanisms at both the channel and spatial levels. It consists of spatial and positional information to a higher dimension than CBAM, comprising a 3D-channel attention sub-module and a 3D-spatial attention sub-module. As a result, this approach enhances the representation of the region of interest and suppresses irrelevant features, leading to an improved cost volume. Next, the GCE module is employed to guide the characteristics of the reference image to the initial cost volume. This is achieved using the features obtained during the reference image feature extraction step. The GCE module uses the reference image features as weights to enhance the cost volume. By guiding the cost volume using image feature weights, contextual information is effectively conveyed to the geometric features, resulting in improved disparity estimation. The GCE module is implemented using 2D point-wise convolution and sigmoid activation functions. It calculates weights based on the input image features. These weights then guide the cost volume, extracting relevant geometry features. This is achieved by multiplying the cost volume with the calculated weights. Next, the improved cost volume is input into Stack Hourglass again and aggregated to create a cost volume with improved noise issue and object boundary. Finally, after generating a volume with 1/4 resolution and conducting cost aggregation, we perform trilinear interpolation to resize the cost volume to match the size of the original image. This interpolated cost volume is utilized to calculate the disparity.

3.4 Disparity computation

We use the soft-argmin operation proposed in GC-Net (Kendall et al., 2017) to convert the trilinear interpolated aggregated cost volume into a continuous disparity map that maintains subpixel accuracy. First, the softmax function transforms the inverse values of the cost volume into probability volumes. Then, the final disparity is calculated by summing the transformed probability volumes weighted by the corresponding probabilities for each disparity. The equation for calculating the disparity is

$$\hat{d} = \sum_{k=D_{min}}^{D_{max}} k \cdot \sigma(-c) \quad (1)$$

Here, k denotes a disparity candidate, and $\hat{d}$ indicates a predicted disparity. Also, σ means softmax, and -c means the opposite cost volume. The proposed method has both negative and positive disparity and calculates the final disparity by summing the probabilities from the minimum negative disparity candidate to the maximum positive disparity candidate.

3.5 Loss function

$$L = \sum_{i=0}^{3} \lambda_i \cdot \text{smooth}_{L_1}(\hat{d}_i - d_{gt}) \quad (2)$$

We compute the loss between the predicted disparity and the ground truth in the four output modules. The predicted disparity of the four output modules is expressed as d_i, and d_{gt} means the ground truth. Also, λ_i represents a coefficient for the i-th prediction disparity. The formula for Smooth L1 loss is as follows (Girshick, 2015).

$$\text{Smooth}_{L_1} = \begin{cases} 0.5x^2, & \text{if } |x| < 1 \\ |x| - 0.5, & \text{otherwise} \end{cases} \quad (3)$$

4. EXPERIMENTS

In this section, we first describe the metrics for quantitative evaluation with a description of the dataset and then describe the implementation details and quantitative and qualitative results for the proposed network. Finally, in the ablation study, we conduct evaluations by ablating each component of CBAM (Woo et al., 2018) used for feature extraction, as well as 3D-CBAM (Huang et al., 2020) and GCE (Bangunharcana et al., 2021) used for cost aggregation in the proposed method. This allows us to assess the individual effects of each module on the performance of the proposed method.

4.1 Dataset

We evaluated our proposed method's performance on the US3D track2 dataset, a challenge for the 2019 Data Fusion Contest (Bosch et al., 2019, Le Saux et al., 2019). This dataset contains high-resolution multi-view images collected by WorldView-3 between 2014 and 2016 in Jacksonville, Florida, and Omaha, Nebraska, USA. The stereo pairs in this data set are rectified to size 1024 × 1024 and geographically do not overlap. Jacksonville and Obama have 2139 and 2154 RGB image pairs, respectively, with ground truth disparity labels. We conducted training, validation, and testing for the experiment using the Jacksonville data. The Omaha data is only used for testing purposes. Table 1 provides an overview of the composition of each dataset utilized in the experiment.

Stereo Pair	Mode	Training/ Validation/Testing	Usage
Jacksonville	RGB	1500/139/500	Training, validation and testing
Omaha	RGB	-/-/2153	Testing

Table 1. Dataset configuration used in the experiment.

4.2 Implementation detail

Our model is trained with the Adam (Kingma and Ba, 2014) optimizer (b1:0.9, b2: 0.999). Normalize the input image to a pixel intensity level between -1 and 1 for data preprocessing. The input image size for model learning is 1024 x 1024 for object integrity, and learning proceeds without data cropping, data resizing, and data augmentation. We trained the model from scratch on the US3D dataset, set to 100 epochs in total. The initial learning rate is set to 0.001, and it drops by half every 10 epochs. The disparity range was set to [-64, 64], and the loss weights λ_1, λ_2, λ_3 and λ_4 were set to 0.5, 0.5, 0.7, and

1.0, respectively. We used an A6000 GPU for training, and it was implemented on Window 10 in the Pytorch environment. In addition, learning is performed with the batch size set to 2, and the test for the proposed model is also conducted in the same environment.

4.3 Experimental results

We included DenseMapNet (Atienza, 2018), Stereonet (Khamis et al., 2018), PSMNet (Chang and Chen, 2018), and GwcNet (Guo et al., 2019) for comparison purposes. The DenseMapNet used in our comparison is the official baseline provided by the US3D data (Bosch et al., 2019). The proposed method demonstrates superior accuracy compared to the official baseline, DenseMapNet. Stereonet and PSMNet are commonly used models for stereo-based disparity extraction, while GwcNet is the foundation for our proposed method. We compared the proposed method with traditional methods using US3D Omaha data. As shown in Table 2, the performance of the proposed method is improved compared to the traditional methods. Good performance on the Omaha dataset means good generalization performance. For quantitative evaluation in Table 2, we used two metrics proposed by (Bosch et al., 2019), the average endpoint error (EPE) and the fraction of erroneous pixels (D1). Table 3 compares GwcNet and the proposed method for Jacksonville test data. For qualitative comparison, we present a selection of output disparity maps in Figure 3. Our network demonstrates enhanced robustness in texture-less areas like flat buildings and the ground. Figure 4 illustrates the error map between the ground truth and the estimated disparity to provide a clear representation. Our network achieves superior results compared to GwcNet.

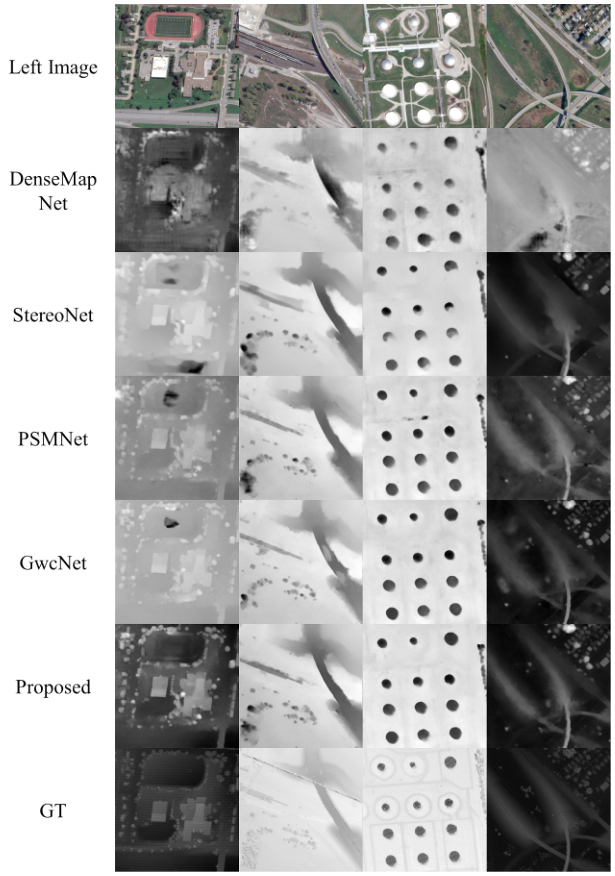

Figure 3. Disparity maps by different methods using the DFC 2019 Omaha dataset.

Method	Omaha	
	EPE(pixel)	D1(%)
DenseMapNet	2.0490	17.48
Stereonet	1.6496	11.41
PSMNet	1.5073	9.39
GwcNet	1.5015	9.07
Proposed	**1.4162**	**8.05**

Table 2. Quantitative evaluation on the whole Omaha data set. Best results are shown in bold.

Method	Jacksonville	
	EPE(pixel)	D1(%)
GwcNet	1.4125	9.50
Proposed	**1.3341**	**8.71**

Table 3. Quantitative evaluation on the Jacksonville data set. Best results are shown in bold.

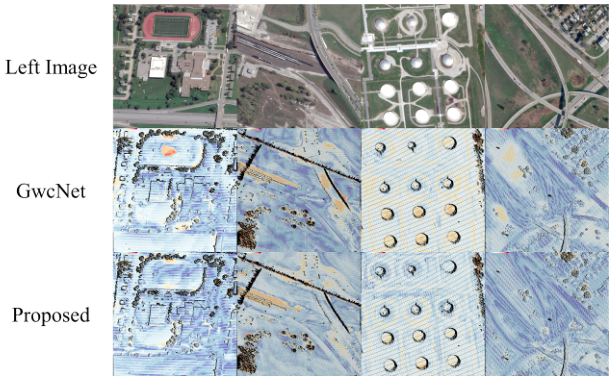

Figure 4. Error map visualization for the disparity in Figure 2. The blue color represents a correct estimate (<3px % error), and the red color represents a wrong estimate.

Texture-less regions. A texture-less area generally occurs a lot in buildings or on the ground, where it is challenging to estimate disparity because the intensity of pixels between the reference image and the target image changes feebly. We selected and listed only texture-less regions from some of the disparity maps output in Figure 5. Roof areas of the same building usually have the same disparity; examples of such regions are roads and flat lawns. As shown in Figure 5, we can see that it predicts more consistent disparity compared to other networks. Figure 6 illustrates the error map between the ground truth and the estimated disparity. Table 4 shows the quantitative evaluation of the individual images in Figure 5.

Image number	EPE(pixel)		D1(%)	
	GwcNet	Ours	GwcNet	Ours
OMA-203-036-026	2.0850	**1.3618**	12.72	**1.38**
OMA-244-038-042	2.6241	**1.8545**	29.21	**7.58**
OMA-287-002-038	2.2660	**1.5913**	17.13	**6.44**
OMA-383-026-001	2.3910	**1.7580**	20.50	**5.57**

Table 4. Quantitative evaluation for each individual image in Figure 5. Best results are shown in bold.

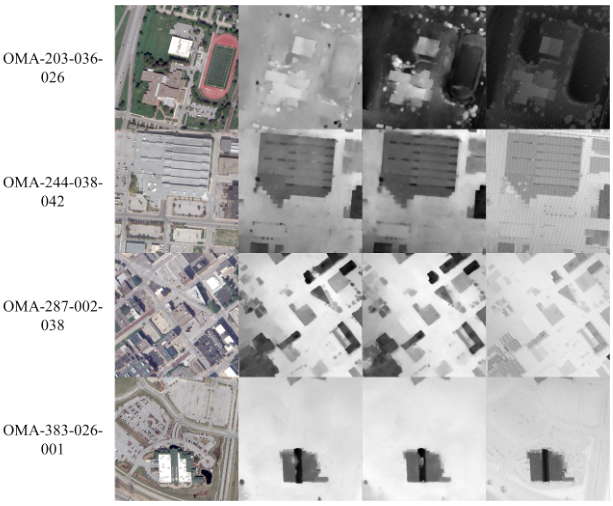

Figure 5. Results of disparity estimation of each individual image for various methods in the texture-less area. From left to right: Left image, GwcNet, Ours and Ground truth.

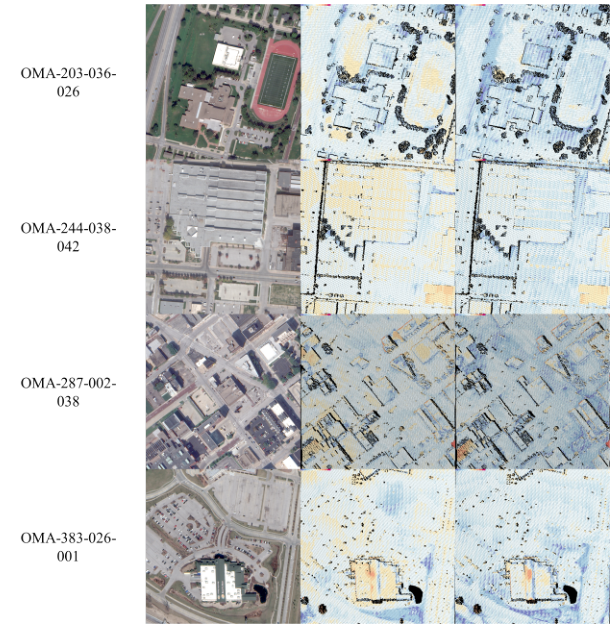

Figure 6. Error map visualization for the disparity in Figure 5. The blue color represents a correct estimate (<3px % error), and the red color represents an wrong estimate. From left to right: Left image, GwcNet and Ours.

Disparity discontinuities and occlusions. Disparity discontinuities have edge-fattening problems, and tall buildings can cause occlusion problems. In the case of an image with a high-altitude building, such as a satellite image, the occluded area is better observed. We show in Figure 7 the areas of occlusion and disparity discontinuities due to tall buildings. The proposed method improves the performance in the building edge area compared to the traditional network. Table 5 presents the quantitative evaluation of the individual images in Figure 7. Figure 8 illustrates the error map between the ground truth and the estimated disparity. Our results show improved performance in the evaluation of EPE and D1.

Image number	EPE(pixel)		D1(%)	
	GwcNet	Ours	GwcNet	Ours
JAX-416-007-002	1.7428	**1.4274**	10.73	**10.04**
JAX-467-001-016	2.1423	**1.8250**	12.90	**11.75**
OMA-212-006-037	1.4487	**1.3583**	8.71	**5.10**
OMA-281-036-032	1.9013	**1.3030**	9.29	**2.93**

Table 5. Quantitative evaluation for each individual image in Figure 7. Best results are shown in bold.

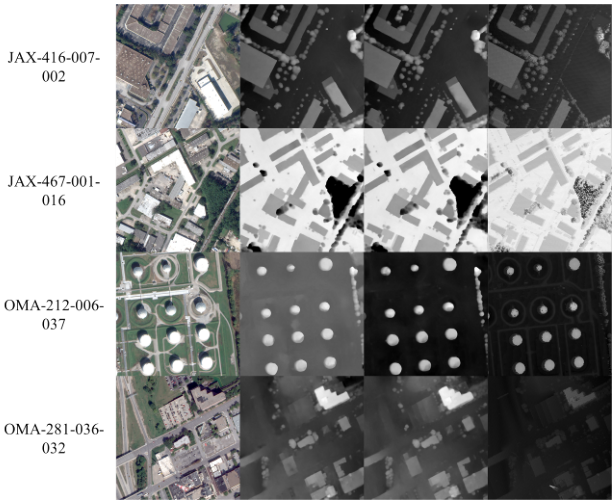

Figure 7. Estimation results of various model disparity in disparity discontinuities and occlusion region. From left to right: Left image, GwcNet, Ours and Ground truth.

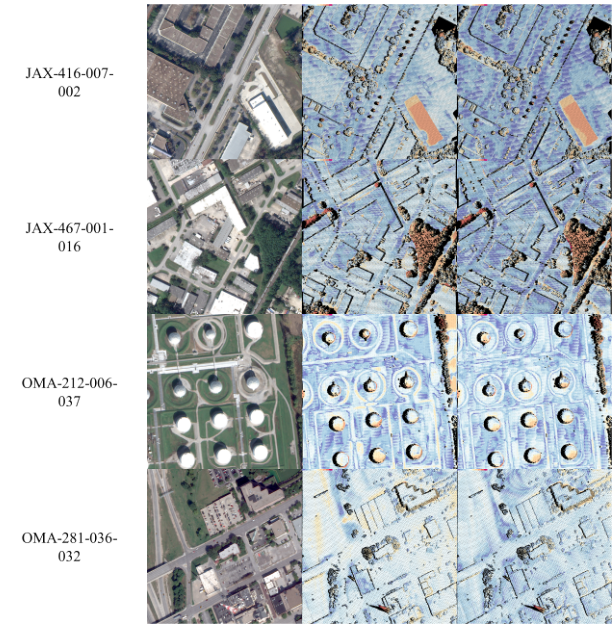

Figure 8. Error map visualization for the disparity in Figure 7. The blue color represents a correct estimate (<3px % error), and the red color represents a wrong estimate. From left to right: Left image, GwcNet and Ours.

4.4 Ablation study

In this section, an ablation study on the US3D dataset is performed and evaluated to validate the module used in the proposed network. To evaluate the performance of the proposed method, we conducted experiments using different configurations, including the use of only 2DCBAM in the feature extraction step, the addition of 3DCBAM in the cost aggregation step, and the incorporation of the GCE module. In Table 6, the "CBAM" results indicate the performance when CBAM is added to the feature extraction step in GwcNet, while the "CBAM + 3D CBAM" results indicate the performance when 3D CBAM is added to the cost aggregation step. The last row of Table 6 corresponds to the proposed method. As presented in Table 6, the performance of the proposed method is further improved.

Model	Jacksonville		Omaha	
	EPE(pixel)	D1(%)	EPE(pixel)	D1(%)
CBAM	1.3919	9.16	1.4599	8.51
CBAM+ 3D-CBAM	1.3344	8.86	1.4237	**7.80**
Ours	**1.3341**	**8.71**	**1.4162**	8.05

Table 6. Ablation study on US3D dataset. Best results are shown in bold.

5. CONCLUSIONS

This paper presents an enhanced network based on GwcNet for estimating disparity in high-resolution satellite stereo images. The proposed network enhances the accuracy of the feature extraction step by adding CBAM and improves the cost aggregation step by including 3D-CBAM and GCE module. In the experimental results, through a comparison with traditional methods, and performance evaluation, our method outperforms traditional methods in terms of performance. In the future, we will study methods to effectively fusion image features into the cost volume by enhancing the performance of the GCE module and improve the proposed method's performance further.

ACKNOWLEDGEMENTS

This work has supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT)(No. 2021R1A2C2009722) and this work was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education (No. 2021R1A6A1A03043144)

REFERENCES

Scharstein, D., & Szeliski, R., 2002. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International journal of computer vision*, 47, 7-42.

Kendall, A., Martirosyan, H., Dasgupta, S., Henry, P., Kennedy, R., Bachrach, A., & Bry, A., 2017. End-to-end learning of geometry and context for deep stereo regression. In Proceedings of the *IEEE international conference on computer vision* (ICCV).

Chang, J. R., & Chen, Y. S., 2018. Pyramid stereo matching network. *In Proceedings of the IEEE conference on computer vision and pattern recognition* (CVPR).

X. Guo, K. Yang, W. Yang, X. Wang and H. Li., 2019. Group-Wise Correlation Stereo Network. *IEEE Conference on Computer Vision and Pattern Recognition* (CVPR).

Zbontar, J., & LeCun, Y., 2015. Computing the stereo matching cost with a convolutional neural network. *In Proceedings of the IEEE conference on computer vision and pattern recognition* (CVPR).

Mayer, N., Ilg, E., Hausser, P., Fischer, P., Cremers, D., Dosovitskiy, A., & Brox, T., 2016. A large dataset to train convolutional networks for disparity, optical flow, and scene flow estimation. *In Proceedings of the IEEE conference on computer vision and pattern recognition* (CVPR).

Woo, S., Park, J., Lee, J. Y., & Kweon, I. S., 2018. Cbam: Convolutional block attention module. *In Proceedings of the European conference on computer vision* (ECCV).

Huang, G., Gong, Y., Xu, Q., Wattanachote, K., Zeng, K., & Luo, X., 2020. A convolutional attention residual network for stereo matching. *Ieee Access*, 8, 50828-50842.

A. Bangunharcana, J. W. Cho, S. Lee, I. S. Kweon, K. -S. Kim and S. Kim., 2021 Correlate-and-Excite: Real-Time Stereo Matching via Guided Cost Volume Excitation. *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems* (IROS), Prague, Czech Republic.

Xie, S., Girshick, R., Dollár, P., Tu, Z., & He, K., 2017. Aggregated residual transformations for deep neural networks. *In Proceedings of the IEEE conference on computer vision and pattern recognition* (CVPR).

Tao, R., Xiang, Y., & You, H., 2020. An edge-sense bidirectional pyramid network for stereo matching of vhr remote sensing images. *Remote Sensing*, 12(24), 4025.

Girshick, R., 2015. Fast r-cnn. *In Proceedings of the IEEE international conference on computer vision* (ICCV).

Bosch, M., Foster, K., Christie, G., Wang, S., Hager, G. D., & Brown, M., 2019. Semantic stereo for incidental satellite images. *In 2019 IEEE Winter Conference on Applications of Computer Vision* (WACV).

Le Saux, B., Yokoya, N., Hansch, R., Brown, M., & Hager, G., 2019. 2019 data fusion contest [technical committees]. *IEEE Geoscience and Remote Sensing Magazine,* 7(1), 103-105.

Kingma, D. P., & Ba, J., 2014. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

Atienza, R., 2018. Fast disparity estimation using dense networks. *IEEE International Conference on Robotics and Automation* (ICRA).

Khamis, S., Fanello, S., Rhemann, C., Kowdle, A., Valentin, J., & Izadi, S., 2018. Stereonet: Guided hierarchical refinement for real-time edge-aware depth prediction. *In Proceedings of the European Conference on Computer Vision* (ECCV).