Global Contextual Complementary Network for Multi-View Stereo

Yongrong Cao[1], Suping Wu[1,*], Xing Zheng[1], Bin Wang[1], Pan Li[1], Zhixiang Yuan[1], Lei Lin[1], Yuxin Peng[1]

[1]School of Information Engineering, Ningxia University, Yinchuan, China

Motivation

- Existing reconstruction methods mostly rely on convolutional neural networks, which limits the ability of the network to capture the global context of images, resulting in a lack of a certain complete representation of the final depth map.
- Convolution operation has strong ability to extract local feature information, such as texture and color. However, for a whole input image, the correlation degree of the relevant information of the image itself seriously affects the learning of the global features of the object.

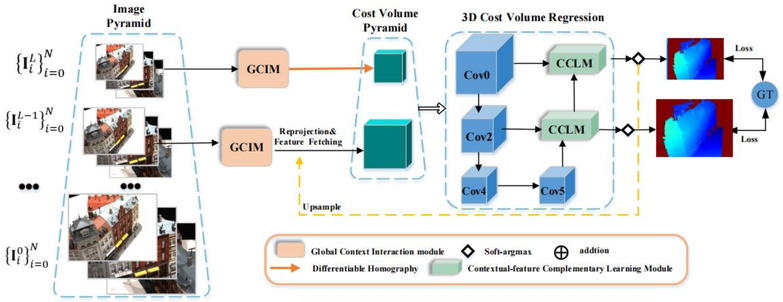
Objective

- In this paper, we propose a Global Context Complementary Network (GCCN), which aims to enhance the complete representation of depth maps with a global context complementary learning strategy.

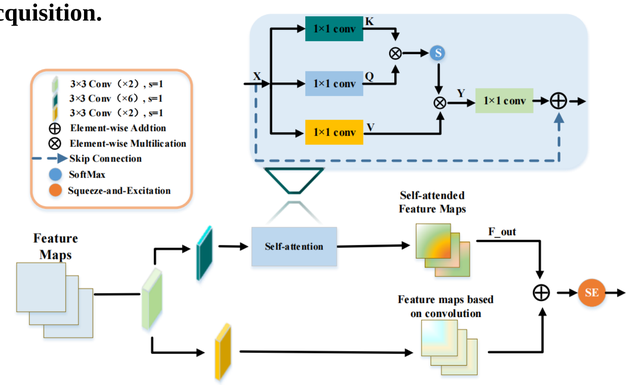
Main Contribution

- We propose a novel end-to-end deep neural framework, namely Global Context Complementary Network (GCCN), for robust long-range global context aggregation within images. Moreover, the combination of local and global information contributes to converge network.
- In addition, to better regress the depth map, we introduce a contextual-feature complement-ary learning module to restore the 3D structure information of the scene.
- Our method achieves state-of-the-art results on the DTU dataset and the Tanks & Temples benchmark.

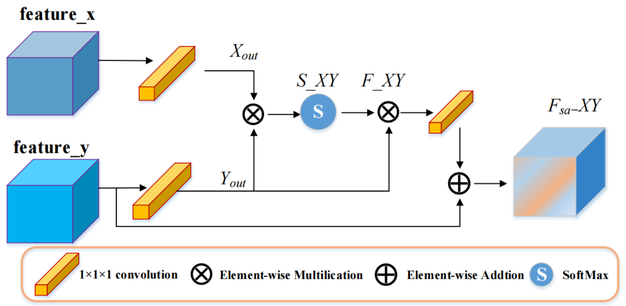
GCCN Architecture



A Global Context Interaction Module(GCIM) designed, which involves two key points: local detail extraction and global feature acquisition.

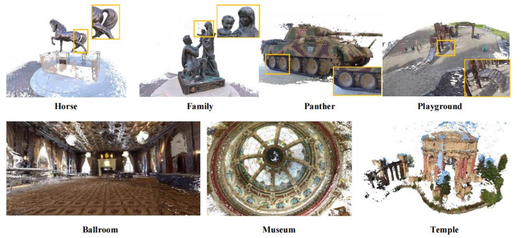


GCIM



CCLM

Qualitative Results



Quantitative Results

Table 1. Results of Different Methods on DTU

	Methods	DTU test set Acc.(mm)	Comp.(mm)	Overall(mm)
Traditional	Camp [2]	0.835	0.554	0.695
	Furu [6]	0.613	0.941	0.777
	Tola [20]	0.342	0.190	0.766
	Gipuma [7]	**0.283**	0.873	0.578
Learning-based	SurfaceNet [11]	0.450	1.040	0.745
	MVSNet [26]	0.456	0.646	0.551
	R-MVSNet [27]	0.383	0.452	0.417
	P-MVSNet [16]	0.406	0.434	0.420
	MVSCRF [24]	0.371	0.426	0.398
	EF-MVS [15]	0.402	0.375	0.388
	AA-RMVSNet[23]	0.376	0.339	0.357
	Cascade-MVS[8]	0.325	0.385	0.355
	CVP(Baseline) [25]	0.296	0.406	0.351
	Ours	0.371	**0.303**	**0.337**

Loss

$$Loss = \sum_{l=0}^{L} \sum_{p \in P_{valid}} \| d(p) - d^{p}(p) \|_1$$

Ablation Results

Table 2. Ablation study of our GCCN.

Model Architecture		Mean Distance(mm)		
GCIM	CCLM	Overall	Comp.	Acc.
✗	✗	0.351	0.406	**0.296**
✓	✗	0.353	0.307	0.400
✗	✓	0.345	0.310	0.380
✓	✓	**0.337**	**0.303**	0.371

Conclusion

In this paper, we have designed a Global Context Complementary Network (GCCN) for high completeness MVS reconstruction, which focuses on long-distance dependencies, thereby maximizing the preservation of on global information of the scene. Specifically, our global context interaction module comprises effective context-aware information within images, which focuses on global structure information of objects. In addition, we utilize the contextual feature complementary learning module to effectively fuse the cost volume features of different depths, and regress a higher completeness depth map.