A Developed CNN for 3D Image Reconstruction of Concrete Surface

Haitao Gong, Jueqiang Tao
Texas State University
601 University Drive, San Marcos, Texas 78666

1. Problem Description

Characterization of the air-void system in hardened concrete has been a challenge for many years. Currently, the widely used method uses original data of 2 dimensional (2D) image of lapped concrete surface to approximate the values of the air-void system parameters. Given that the air-void system is a 3 dimensional (3D) model, it's intuitive that the application of 3D image can inherently improve the measurement accuracy compared with existing 2D image methods.

Computer stereo vision is the extraction of 3D information from 2D digital images. This method aims to estimate 3D geometry by computing disparities between matching pixels in a stereo image pair. Recently many approaches were introduced to address the reconstruction from 2D image to 3D image, with concentration on memory cost, computation cost, and precision loss [15]. 3D image reconstruction for air-void system analysis has a particularly high requirement for precision. Entrained voids in concrete have diameters from 0.01 mm to 1 mm, and those voids are uniformly distributed throughout the concrete. In order to recognize a majority of the air-voids, a minimum measurement resolution of 0.01 mm is strongly required. For example, for a concrete specimen surface of 40*40 mm, an 3D image with at least 16 megapixel (MP) is needed. This asks for two requirements from the adopted method: 1) be able to process high resolution 2D images within a reasonable time; and 2) be able to produce high precision 3D images.

2. Background

2.1. Literature Review

Accurately estimating 3D geometry from stereo imagery is a core problem for many computer vision applications [7]. In order to calculate the disparity of each pixel between a rectified stereo pair of images, it is important to compute the correspondence of each pixel between two images. A great number of researches have been conducted in this field. However, due to the difficulties in ill posed region such as textureless areas, reflective surfaces, thin structures and repetitive patterns, this is still a very challenge work to achieve robust results in real world.

Considering the air-void detection system requires a high-resolution image to capture 0.01mm diameter air-voids, it is necessary to find a high accuracy stereo matching method with acceptable computation expense. According to the taxonomy of typical stereo algorithms proposed by Scharstein and Szeliski, a typical stereo algorithm consists of 4 steps: (1) matching cost computation, (2) cost aggregation, (3) optimization, and (4) disparity refinement [11]. Matching cost measures the pixel similarity for potential corresponding stereo regions, and absolute difference, squared differences and truncated differences are commonly used measurements [5]. Sun etc. proposed a novel propagation-based stereo matching algorithm based on pixelwise line segments and 1D propagation [12]. The proposed method overcomes the inherent problem of error propagation for traditional propagation-based stereo matching algorithm and shows good performance in both efficiency and accuracy. Mei etc. proposed a near real-time stereo system based on several key techniques like AD-Census cost measure, cross-based support regions, scanline optimization and a systematic refinement process [9]. The proposed method presents good performance for the Middlebury data sets, while the sensitivity for noise information is not sure. Markov Random Field models were used by Bleyer etc. to solve object level image segmentation [1].

In the past few years, convolutional neural networks (CNN) have been shown to perform extremely well for stereo estimation. Zbontar etc. approached the matching cost computation problem by learning a similarity measure on small image patches using a convolutional neural network which shows great performance in challenging benchmarks, while the computation is expensive [14]. CNN yields significant improvements compared to conventional approaches, while it is still difficult to find accurate corresponding points in ill-posed regions. Luo etc. propose a matching network which is able to produce very accurate results in a reasonable GPU computation time [8]. Zhang etc. developed a much more efficient and effective guided matching cost aggregation strategies [15]. Chang etc. utilized spatial pyramid pooling module and dilated convolution to incorporate global context information into stereo matching for homologous points searching [6]. This method shows a robust result in ill-posed regions under acceptable runtimes.

Most of global stereo matching methods are computationally expensive, while local stereo methods are much easier to implicate and more effective. Support windows are always utilized in local stereo methods for reducing image ambiguity and it is desired to adapt its shape and size for disparity accuracy. Zhang etc. proposed a cross-based local stereo matching method which uses a cross-based local support window for matching cost aggregation [16]. The image content can be compressed when neighboring pixels possessing similar appearance [13]. Therefore, it is possible to optimized local matching costs within a global framework. Hirschuller proposed Semi-Global Matching method which uses a pixelwise, Mutual Information based matching cost for compensating radiometric differences of input images [4].

However, the matching costs between unaries can never be perfect, even when using a deep feature representation. Kendall etc. find regions like this can cause multi modal matching cost curves across the disparity dimension. Consequently, 3D CNN with deep encoder-decoder architecture is introduced in this research to learn the context information in disparity cost volume [7]. Chang etc. compared the performance of basic and stacked hourglass 3D CNN for cost volume regularization and the results shown that the stacked hourglass architecture shows a much better performance [6]. Gidaris et al. introduced a deep tructured model to decompose the task into three sub-blocks. It took advantage of deep CNNs to identify and correct outliers with the information of left color image and left initial disparity map [3].

Kendall etc. proposed a differentiable soft argmin operation for disparity regression, which is fully differentiable and able to train CNN model end-to-end to sub-pixel accuracy without any additional post-processing or regularization. [End-to-End Learning of Geometry and Context for Deep Stereo Regression] The soft argmin operation shows a high disparity estimation accuracy with significantly faster computation efficiency.

2.2. Research Object

This research aims to propose a developed CNN for 3D image reconstruction of concrete surface. A major effort will be focused on improving the precision of 3D image construction within acceptable computation expense.

3. Data Descritpion

This paper uses the KITTI 2012 and 2015 benchmarks [2] [10]. for model training and testing. The 2012 stereo / flow benchmark consists of 194 training image pairs and 195 test image pairs, saved in loss less png format. The stereo 2015 / flow 2015 / scene flow 2015 benchmark consists of 200 training scenes and 200 test scenes (4 color images per scene, saved in loss less png format). Compared to the stereo 2012 and flow 2012 benchmarks, it comprises dynamic scenes for which the ground truth has been established in a semi-automatic process. Pairs of concrete sample surface images will be used for testing. These images are taken by Sony Alpha a7R digital camera with 42.4-megapixel. Compared with KITTI benchmarks, the size of concrete surface images are much larger.

4. What have been done

(1)Comparison of existing stereo matching approaches;
(2)Built an initial structure for learning model;

5. What remains to be done

(1)Write program for learning model and implement it in the data selected.
(2)Prepare the concrete surface images, and use this data for model testing.
(3)Analyze the experiment result.
(4)Finish paper writing.

References

[1] M. Bleyer, C. Rother, P. Kohli, D. Scharstein, and S. Sinha. Object stereo—joint stereo matching and object segmentation. *In CVPR 2011*, pages 3081–3088, 2011.

[2] A. Geiger, P. Lenz, and R. Urtasun. Are we ready for autonomous driving? the kitti vision benchmark suite. *In 2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 3354–3361, 2012.

[3] S. Gidaris and N. Komodakis. Detect, replace, refine: Deep structured prediction for pixel wise labeling. *In Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5248–5257, 2017.

[4] H. Hirschmuller. Stereo processing by semi-global matching and mutual information. *IEEE Transactions on pattern analysis and machine intelligence*, pages 328–341, 2007.

[5] H. Hirschmuller and D. Scharstein. Evaluation of cost functions for stereo matching. *In 2007 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–18, 2007.

[6] C. Jiaren and C. Yongsheng. Pyramid stereo matching network. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5410–5418, 2018.

[7] A. Kendall, H. Martirosyan, S. Dasgupta, P. Henry, and R. Kennedy. End-to-end learning of geometry and context for deep stereo regression. *In Proceedings of the IEEE International Conference on Computer Vision*, pages 66–75, 2017.

[8] W. Luo, AG. Schwing, and R. Urtasun. Efficient deep learning for stereo matching. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2016*, pages 5695–5703, 2016.

[9] X. Mei, X. Sun, M. Zhou, S. Jiao, H. Wang, and X. Zhang. On building an accurate stereo matching system on graphics hardware. *International Conference on Computer Vision Workshops*, pages 467–474, 2011.

[10] M. Menze and A. Geiger. Object scene flow for autonomous vehicles. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3061–3070, 2015.

[11] D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*, 47:7–42, 2002.

[12] X. Sun, X. Mei, S. Jiao, M. Zhou, and H. Wang. Stereo matching with reliable disparity propagation. *International Conference on 3D Imaging, Modeling, Processing, Visualization and Transmission*, pages 132–139, 2011.

[13] F. Tombari, S. Mattoccia, L. Di Stefano, and E. Addimanda. Classification and evaluation of cost aggregation methods for stereo correspondence. *IEEE transactions on circuits and systems for video technology*, pages 1–8, 2008.

[14] J. Zbontar and Y. LeCun. Computing the stereo matching cost with a convolutional neural network. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1592–1599, 2015.

[15] F. Zhang, V. Prisacariu, R. Yang, and P. Torr. Ga-net: Guided aggregation net for end-to-end stereo matching. *Computer Vision and Pattern Recognition (CVPR)*, pages 185–194, 2018.

[16] K. Zhang, J. Lu, and G. Lafruit. Cross-based local stereo matching using orthogonal integral images. *IEEE transactions on circuits and systems for video technology*, pages 1073–1079, 2009.