스테레오 정합을 위한 심층 자기유도 정합비용

윌리엄, 박인규
인하대학교 정보통신공학과
{22112195@inha.edu, pik@inha.ac.kr}

Abstract

In this paper, we present a deep self-guided cost aggregation method for obtaining accurate disparity maps. Conventional cost aggregation methods typically perform joint image filtering on each cost volume slice. Thus, it is necessary to have a guided image for the conventional methods to work well. Based on our empirical observation, each cost volume slice has its own features that can be used to guide itself. However, it is difficult to find the direct mapping function from the original and filtered cost volume slice without any guided image. To resolve this problem, we propose to utilize advanced deep learning technique for performing self-guided cost aggregation. We utilize the deep convolutional neural network and conscientiously decide the number of layers and filter size for a certain support region size. Experimental results show that the proposed method achieves comparable results even though it does not employ any guided image.

1. Introduction

Stereo matching has become a popular research topic in the last couple of decades to find the correspondence between two rectified images. Generally, stereo matching method consists of four steps: cost computation, cost aggregation, disparity computation, and disparity refinement.

Cost aggregation plays important role for local or global stereo matching. Most of the cost aggregation methods share similar concept with joint image filtering that utilizes a guided image [1, 2, 3]. Thus, the existence of a guided image is necessary. They assume that similar color in the guided image should have similar disparity value. However, this assumption often fails when there is highly textured region with similar disparity value.

In this paper, we propose a novel cost aggregation method that reduces the dependency of the guided image on the fly. We observe that the original cost volume has its own features to become guidance for itself. However, it is very complex to find the correlation between non-filtered and filtered cost volume without any guided image. Therefore, a deep convolutional neural network (CNN) architectures is utilized so that a cost volume slice can do self-guided cost aggregation.

2. Proposed Method

In this paper, we focus on local stereo matching to evaluate the performance of the proposed method. We choose truncated absolute difference of color and gradient as the matching cost model. After matching cost computation, we perform weighted cost aggregation. A guided image is required to measure the weight value for each pixel in the window. Instead of utilizing a guided color image, we use the input cost volume slice as a guidance. We notice that it is difficult to measure the weight directly from the input. Therefore, deep learning technique is utilized to measure the weight and perform weighted cost aggregation simultaneously.

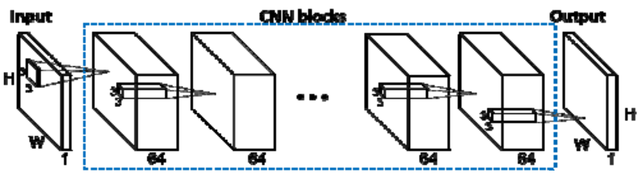

Fig. 1. Pipeline of the deep self-guided cost aggregation

Finally, we perform WTA (Winner-Take-All) strategy to obtain the disparity value. For the post processing, we compute both left and right disparity maps and then employ left-right consistency checking to detect pixels with occlusion and inaccurate disparity. To fill the hole, the lowest disparity value of the nearest pixels with correct disparity in the same scanline is assigned. Then, we refine the disparity by an adaptive median filter to also preserve the depth boundaries.

2.1 Network Architecture

The overview of the network architecture is shown in Fig. 1. We use D convolutional layers where rectified linear unit (ReLU) layer follows the first D – 1 layers. There are 64 kernels of size 3 × 3 × 64 for all layers except the first and last. It denotes that a small filter is performed across 64 channels for each layer. The first convolutional layer consists of 64 kernels of 3 × 3 × 1 and the last convolutional layer consists of a filter of size 3 × 3 × 64. The number of convolutional layers is set depending on the size of support region that we want to deal with. For 19 × 19 support region, we set D = 9 so that the center pixel inside the support region can influence all other pixels.

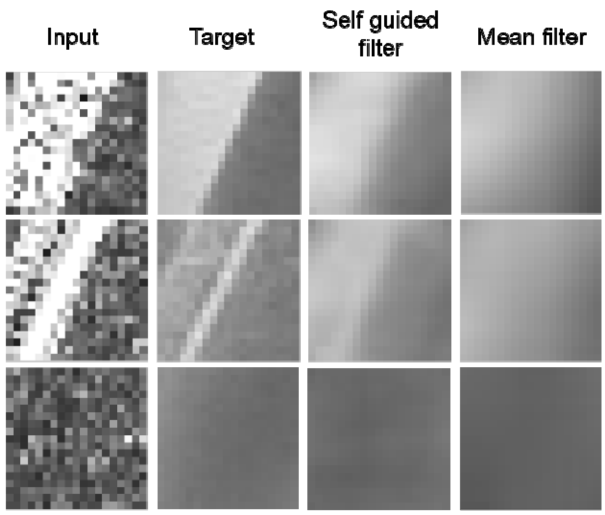

Fig. 2. Training dataset examples. The first and second rows are the patches near depth discontinuities. The third row is a patch on uniform depth.

2.2 Dataset Generation

Due to the absence of ground truth of the cost volume slice, it is difficult to utilize deep learning for the cost aggregation step. Thus, we utilize a guided filtered cost volume as the target. We only select the filtered cost volume that results in correct disparity. Note that the proposed deep network is an end to end learning, a set of patches are generated as the dataset instead of a set of images. There are two set of patches: patches around depth discontinuities and patches on uniform depth. The ratio of both sets is 50:50 so that the trained parameters can works well in both regions. Fig. 2 shows the example of training dataset in both regions.

3. Experimental Results

To evaluate the performance of deep self-guided cost aggregation method (DEEP), we compare with the state-of-the-art methods, such as GF [1] and NL [2]. In addition, we also compare the results with simple mean filter (BOX). Note that BOX is similar to the proposed method in terms of guided image free method.

Fig. 3 shows the qualitative comparison for Books dataset. It confirms that the proposed method performs well in highly textured regions with similar disparity. The red box shows the highly textured region and its zoom version for better understanding. Note that the guided image based weight is not equal to ground truth disparity based weight on highly textured region. While conventional methods fail obtaining accurate and smooth disparity on that region, our deep self-guided method achieves better results.

In Table I, we show the bad pixel percentage for each dataset. The bad pixel percentage is calculated only using the non-occluded region. It is shown that the proposed method achieves comparable results with other state-of-the-art cost aggregation methods. Note that our method is guided image free which is a big handicap.

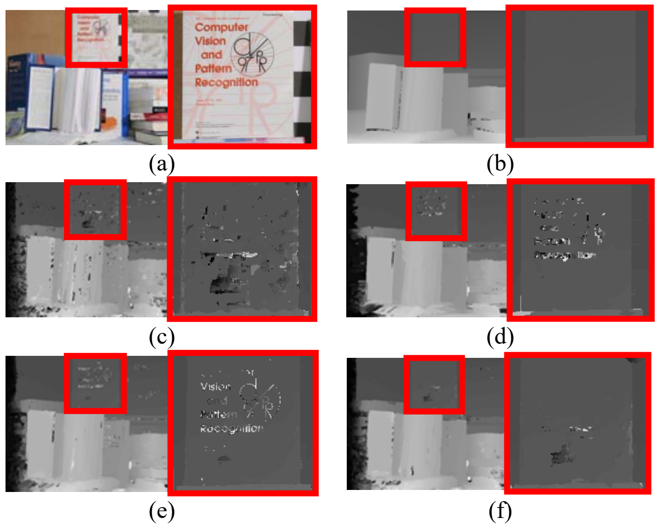

Fig. 3. Zoomed disparity maps comparison of Book data. (a) Input left image; (b) Ground truth; (c) BOX; (d) GF [1]; (e) NL [2]; (f) DEEP (Proposed method).

Table. I. Comparison of bad pixel percentage

Dataset	BOX	GF [1]	NL [2]	DEEP
Aloe	7.70	7.56	7.82	**6.40**
Books	17.56	11.59	13.46	**11.05**
Cloth2	5.63	5.49	6.63	**4.45**
Dolls	7.92	6.81	7.78	**6.78**

4. Conclusion

We proposed a deep self-guided cost aggregation method to obtain accurate stereo matching results. First, we randomly generated training patches from the training dataset. Then, we run deep convolutional neural network to learn the weight parameter. Instead of using a guided image, we performed the cost aggregation by self-guiding the cost volume slice. To our best knowledge, the proposed method is the first cost aggregation method that utilizes deep learning technique. It was shown through the experiments that the proposed method achieves comparable results to cost aggregation methods that use guided image.

Acknowledgement

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) (No. NRF-2016R1A2B4014731).

References

[1] A. Hosni, *et al.*, “Fast cost-volume filtering for visual correspondence and beyond,” IEEE TPAMI, vol. 35, no. 2, pp. 504-511, February 2013

[2] Q. Yang, “A non-local cost aggregation method for stereo matching,” Proc. of IEEE CVPR, pp. 989-998, June 2006

[3] K.-J. Yoon and I. S. Kweon, “Adaptive support-weight approach for correspondence search,” IEEE TPAMI, vol. 28, no. 4, pp. 650-656, April 2006