
Depth Estimation 기술의 원리 및 동향

김혜진^{1,*}, 이수영²

Principles and Trends of Depth Estimation

Hye-Jin S. Kim^{1*}, and Suyoung Chi²

5

요약

딥러닝의 발전에 따라 거리측정 기술 또한 비약적인 발전이 있었다. 본 논문에서는 양안 영상뿐만 아니라 단안영상에서의 거리측정 기술 동향에 알아본다. 또한, 단안 영상에서의 비지도학습 방법으로부터 카메라의 포즈를 측정을 통해 거리 측정 성능을 개선시키는 방법, 객체 모션 모델링을 통해 거리측정 성능을 개선시키는 방법에 대해 알아본다. 거리측정 기술은 다른 영상 기법의 기반이 되는 기술이며 GPU가 장착될 수 있는 자율주행, 로봇뿐만 아니라 스마트폰에서의 AR/VR, 드론 등에 접목하기 위해 경량화된 딥러닝에 기반한 거리측정 기술을 다룬다. 한 편, 2D 영상뿐만 아니라 360 영상과 같이 3차원의 영상에서의 거리 측정 기술도 함께 알아보려고 한다.

Abstract

With the advent of deep learning, the depth estimation based on images has made dramatic process for recent a few years. In this paper, we take a close look at not only stereo approach methods but also monocular approaches. In particular, unsupervised monocular methods consider the pose of the camera and by estimating the camera position and adopt to the current depth estimation. Besides, light-weight depth estimation network shows promising results that the network inference is able to execute on CPU. Moreover, we introduce the depth estimation method using 360 image inputs.

Keywords : Depth, Deep Learning, Stereo, Monocular, 360 Images

¹한국전자통신연구원 지능형로보틱스연구본부 (대전시 가정로 218), 선임연구원

²한국전자통신연구원 지능형로보틱스연구본부 (대전시 가정로 218), 책임연구원

*Corresponding Author : marisan@etri.re.kr

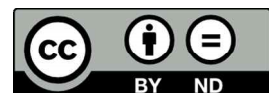
접수일자 : 2019. 05. 07.

1차 심사 : 2019. 05. 13.

2차 심사 : 2019. 05. 30.

게재확정 : 2019. 05. 30.

DOI: <http://data.doi.or.kr/10.22733/JITAE.2019.09.01.005>



1. 서론

영상기반 거리 측정기술은 Robot, AR/VR, Drone, 자율주행자동차 등 다양한 분야에서 활용되고 있다.

영상기반 거리 측정 기술은 영상 정보를 통해 거리를 측정하는 기술을 일컫는다. 이 기술은 컴퓨터 비전 분야에서 매우 오래된 기술 중 하나이나 최근 딥러닝 기술의 발전으로 괄목할 만한 성장을 이루었다.

기존에는 hand crafted feature를 추출에 의존한 stereo matching 위주의 방식이었다. 특징을 추출한 이후에는 calibration, matching을 통해 거리를 측정해 왔다 [1]. Hand-crafted feature에 의존하는 방식들은 이 feature에 의존성이 매우 높아 feature가 잘 뽑히는 환경에서는 성능이 높으나 조명 변화, texture 등에 따라 성능의 가변성이 높다. 특징 추출을 통해 단일 영상으로부터 거리를 추정하고자 하는 논문 [19]이 제안되었으며 매우 획기적인 접근으로 여겨졌다.

2014년 12월 NIPS에서는 딥러닝을 통해 단일 영상으로 거리 측정이 가능함을 보여줌으로써 딥러닝으로의 거리 측정 기술의 시작을 알렸다 [2]. 이후, 많은 연구들이 있어왔고, 양안 혹은 다차원 영상뿐만 아니라 단일 영상, 360영상 등에서의 높은 수준의 거리 측정이 가능함을 가능하게 해 주었다.

본 논문에서는 오늘날까지 짧은 기간에 괄목할 만한 성장을 보여준 딥러닝에서의 거리 측정 기술 동향에 대해 기술하고자 한다.

2. 거리측정 기술의 분류 및 동향

딥러닝 기반 거리측정 기술은 encoder-decoder 구조에 기반한다. Encoder 부분은 classification과 유사하게 특징을 추출하는 부분으로 input의 가로, 세로가 줄어들고 feature map 크기가 늘어난다. 반면 decoder 부분은 upsampling, deconvolution 등을 통해 input의 가로 세로가 늘어나 종단에는 최종 영상의 input 크기와

동일한 영상 크기가 되며 마지막 단에는 sigmoid 함수로 disparity, depth 값을 출력, 이 값으로부터 loss 함수를 계산하여 학습을 하게 되는 방식이다.

거리측정 기술을 이 논문에서는 크게 양안 방식 (stereo approach)와 단안 방식 (monocular approach)로 나누도록 하겠다. 고전적 방식인 hand-crafted feature에 기반한 접근 방법으로는 단안 방식의 거리측정 방식은 매우 어려운 기술이었으나, 딥러닝에 기반하여 양안 방식에 근접한 성능을 보여주고 있다.

또한, 본 논문에서는 기존에 딥러닝적 접근이 GPU의 한계점으로 인해 다양한 응용 분야에 활용이 어려웠던 점을 탈피하여 경량화된 거리 측정 기술을 개발하고 있어 이를 소개하고자 한다.

마지막으로, 2차원 영상 입력뿐만 아니라 360 영상 등 고차원화, Light-Field 영상으로 다차원화 되고 있는 다양한 영상 입력에 있어서 고전적 방식으로 해결이 어려웠던 거리 측정을 어떠한 식으로 접근하고 있는지에 대해서도 소개하고자 한다.

2.1 학습 Dataset

학습에 기반한 거리측정에서 널리 활용되는 공개 학습데이터로는 여러 가지가 있으나 그 중 널리 활용되는 것은 KITTI [5]에서 제공되는 데이터이다. KITTI에서는 KITTI 2012, KITTI 2015의 명칭으로 데이터들을 배포하고 있다. 이 데이터들은 차량에서 촬영된 영상과 그 영상에 해당하는 LiDAR로 구성되며 LiDAR에서 얻은 raw 파일로부터 각 영상에 해당하는 값을 계산하여 얻은 Ground Truth로 구성된다. LiDAR는 매우 sparse한 데이터이기 때문에 그림 1에서와 같이 완벽히 채워진 거리값이 아니라 성긴 데이터값이라 선이 그려져있는 것처럼 보이게 된다. 또한, LiDAR의 특성상 수직 FoV가 수평 FoV에 비해 좁기 때문에 수평으로 긴 GT를 얻게 되는 경향을 보인다.

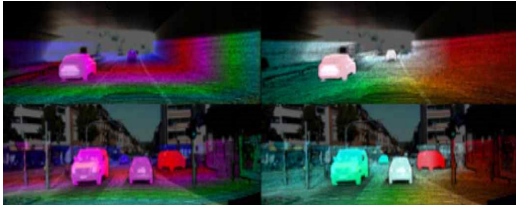


그림 1. KITTI 데이터셋 예시 [5].
Fig. 1. Examples in KITTI dataset [5].

또 다른 학습 데이터셋으로는 Cityscapes [6] 으로 이 데이터는 그림 2에서 보듯이 KITTI 와 유사하게 도시 거리의 영상을 주로 대상으로 한다. 사람과 다른 정보들의 semantic segmentation 정보도 함께 제공한다. 이들과 달리 실내를 주 대상으로 하는 NYU 의 데이터셋 [7]도 있다. 이 데이터는 Kinect로부터 거리를 측정 한 후 보정을 통해 Ground Truth를 획득했다.



그림 2. Cityscapes 데이터셋 예시 [6].
Fig. 2. Examples in Cityscapes dataset [6].

2.2 Stereo Approach

양안 영상을 이용하여 거리를 측정하는 방법은 인간의 거리 측정 메카니즘과 가장 유사한 방법이다. Calibration을 이용해 rectified된 영상으로부터 양쪽 영상의 disparity 차이를 구하고 이로부터 거리를 측정하는 방식이다.

딤러닝을 이용한 양안 방식으로 널리 알려진 논문은 Yann LeCun 랩에서 나온 논문으로 MC-CNN으로 널리 알려져 있다 [3]. 이 논문은 그 이전에 단안영상으로 딤러닝으로부터 거리를 학습하는 논문 [2] 이후 기존의 stereo matching 방식과 같은 식으로 딤러닝을 적용하는 방식을 제시했다는 점에 그 의의가 있다. MC-CNN은 양안영상을 활용하기 위해 같이 siamese network 구조를 가지고 있다. 왼쪽 영상과 오른쪽 영상 각각에 대해 convolution

과 relu의 여러 층을 쌓아 feature를 학습하고 이 두 영상을 concatenation 또는 dot product 를 하여 두 영상을 묶어 여러 층의 fully connected layer들과 sigmoid로 similarity score를 구하여 학습하도록 하였다. 이 논문은 딤러닝 기반으로 하는 양안 영상에 대한 초기 논문으로 학습을 통해 similarity score를 구하고 고전적인 방법에서 사용한 post processing 방법을 추가하여 거리 측정을 하였다.

GC-Net [4]은 거리 측정방법에 있어 최초 end-end로 학습시킨 논문이다. MC-CNN이 post processing과 분리되어 있었는데 GC-Net 가 가장 크게 기여한 부분을 depth를 cost volumn으로 측정한다는 것을 최초로 도입하여, end-to-end로 거리를 학습 가능하도록 바꾸어 주었다는 점이다. 즉, GC-Net은 양안 영상으로부터 disparity의 sub-pixel을 regression 을 통해 직접 구하도록 되어 있다. Cost volumn 이전까지는 feature를 구하는 부분이며 cost volumn은 disparity에 대한 cost 값을 width*height*disparity 형태를 갖도록 구해진 값으로 학습을 통해 3D conv-deconv로부터 disparity를 구하도록 되어 있다. GC-Net은 supervised learning의 방법으로 아래의 식을 통해 loss를 구하고 이 값의 backpropagation 으로부터 disparity를 구하게 된다.

$$Loss = \frac{1}{N} \sum_{n=1}^N \|d_n - \hat{d}_n\|_1 \quad (1)$$

이 식에서는 학습 데이터셋에서 제공한 GT 값과 예측된 값의 차이에 대한 loss 값으로부터 목적 함수를 구하게 된다. Supervised learning 에 기반한 접근 방법은 이 식과 유사한 loss 값을 사용해왔다. PSMNet 논문 [8]에서는 종래의 차이값 대신 아래의 smooth L1 loss 식을 도입하였다. 이 식은 객체식별에 있어 bounding box regression 에 많이 사용되는데 이 값이 앞선 loss 혹은 L2 loss 보다 outlier 에 강인한 경향이 있기 때문이다.

$$L(d, \hat{d}) = \frac{1}{N} \sum_{i=1}^N \begin{cases} 0.5(d_i - \hat{d}_i)^2, & \text{if } |d_i - \hat{d}_i| < 1 \\ |d_i - \hat{d}_i| - 0.5, & \text{otherwise} \end{cases} \quad (2)$$

PSMNet 논문은 spatial pyramid pooling 방법을 거리 측정에 도입하고 stacked hourglass 구조와 3D cost volumndf 활용하여 성능을 높일 수 있었다. 특히 이 논문은 spatial pyramid pooling을 적용하여 거리추출 성능을 높인 것으로 주목 받았는데, 이 방법은 여러 크기의 pooling layer들의 값 (64x64, 32x32, 16x16, 8x8)을 convolution 이후 upsampling 하는 layer 로 SPP module 로 명명하였다. 이 모듈은 기존에 여러 convolution layer 의 순차적으로 이어져왔던 layer에 spatial pyramid 형태로 변형한 것이다. 이 논문은 GC-Net에서 소개한 cost volumn 형태를 차용하고, Inception 네트워크처럼 여러 곳에서 loss 값을 주어 학습이 진행되도록 하여 성능 향상에 성공하였다.

DeMon [9]과 DPSNet [10]은 ego-motion에 따른 거리 측정방법을 제시할 뿐만 아니라 카메라의 포즈 정보까지를 얻고자 하는 학습기반 structure from motion 대표적인 논문이라 할 수 있다. 이 논문들은 camera motion을 통해 거리 측정 정확도를 올릴 수 있는 점에 착안하였다. 단안영상 기반의 거리 측정은 정보량이 부족하기 때문에 절대 크기를 알 수 없다는 단점을 갖고 있으나 카메라는 하나이나 움직임에 따라 두 장 혹은 그 이상의 영상을 획득하여 거리를 추정하는 방법을 제시하였다. DeMon [9]에서는 양안 영상으로부터 거리를 추정할 뿐만 아니라 normal 데이터와 optical flow 추정을 통해 depth를 추정하고, 이 autoencoder network를 iterative 하게 여러 차례 반복하여 학습시킨다. 이후 refinement network를 통하여 보다 정확한 depth 와 카메라 포즈 (R, t)를 구하였다.

DPSNet [10]은 여러 장의 영상이 있을 때 camera의 자기 위치 추정을 활용해 거리를 측정하는 방법을 제시하였다. 기존 방법들이 calibrated 된 영상 학습 데이터 (예: KITTI)를 학습하는 것에만 의지하는 반면 여러 영상으로부터 plance (sweep) 가정을 통해 camera

pose 값을 활용해 보다 정확한 거리 측정에 대한 방법을 제시하였다.

2.3 Monocular Approach

단안 영상으로 거리를 측정하는 방법은 고전적인 접근 방법으로는 시도는 되었으나 매우 어려운 분야로 간주되어 왔었다. 왜냐하면 단일 영상에서는 실제 크기에 것을 측정하기 어렵기 때문이다. 이 덕분에 NIPS에서 딥러닝으로 최초로 거리측정 알고리즘으로 소개된 단안영상 거리 측정 방법 [2]은 많은 관심을 끌었고 딥러닝 접근 방법의 거리 측정 가능성을 보여 주었다. 초기의 접근은 input 영상 크기의 1/4 크기로 축소된 depth 결과를 예측하였다. Eigen은 coarse network와 fine network로 나누어 학습하였다. Course network는 여러 층의 convolution layer와 마지막 단에서 fully connected layer를 통해 분류 방법과 유사하게 값을 측정하도록 하였으며 fine network에서 좀 더 정밀하게 보장하는 거리 측정 방법을 제안하였다.

Godard [11]는 고전에서 사용되어 왔던 기법을 loss 함수로 활용하여 단일 영상, unsupervised learning (비지도학습)으로 높은 성능의 거리 추정 기술을 개발하였다. C_{ap} , C_{ds} , C_r 세 가지 loss 값을 설정하여 정답데이터 없이 거리측정 모델을 만들었다. 첫 번째로 C_{ap} 는 appearance의 유사도에 대한 loss를 정의하였다. 이 네트워크는 단일영상의 성능을 향상시키기 위하여 양안의 반대쪽 영상으로부터 픽셀을 샘플링하여 영상을 만들게 되는 방식을 취하였다. SSIM은 3x3 블록 filter를 의미하며 광학적 영상복원의 비용함수로 입력영상 I_{ij}^l 과 N개의 픽셀을 갖는 복원된 영상 \tilde{I}_{ij}^l 를 비교하는 비용함수로 L1 regularizer가 함께 적용되었다.

$$C_{ap} = \frac{1}{N} \sum_{ij} \alpha \frac{1 - SSIM(I_{ij}^l, \tilde{I}_{ij}^l)}{2} + (1 - \alpha) \|I_{ij}^l - \tilde{I}_{ij}^l\| \quad (3)$$

C_{ds} 는 disparity의 smoothness에 대한 loss를 측정하였다. 거리측정에 있어 거리의 비연속성은 영상의 그라디언트에서 발생한다고 알려

져 있다 [16]. Disparity의 그라디언트에 영상 그라디언트를 이용해 가중치를 부여하는 방식으로 smoothness에 대한 cost를 정의하였다.

$$C_{ds}^l = \frac{1}{N} \sum_{i,j} |\partial_x d_{ij}^l| e^{-\|\partial_x d_{ij}^l\|} + |\partial_y d_{ij}^l| e^{-\|\partial_y d_{ij}^l\|} \quad (4)$$

이 논문은 학습 시 왼쪽영상과 오른쪽 영상 모두에 대해 각각의 disparity d^l , d^r 를 구하고 이들의 차이로부터 비용함수 C_{lr} 를 아래 식과 같이 계산하였다.

$$C_{lr}^l = \frac{1}{N} \sum_{i,j} |d_{ij}^l - d_{ij}^r| \quad (5)$$

최종 loss 값 C_s 는 아래와 같이 오른쪽 영상, 왼쪽 영상에서 얻은 값과 weight들로 구성했다.

$$C_s = \alpha_{ap}(C_{ap}^l + C_{ap}^r) + \alpha_{ds}(C_{ds}^l + C_{ds}^r) + \alpha_{lr}(C_{lr}^l + C_{lr}^r) \quad (6)$$

고전적인 방법으로도 광학적 특성의 유사도, disparity의 smoothness 등을 반영하기 위해 여러 가지 방법들이 개발되었다 [16]. 학습에 기반한 접근 방법에 있어서도 이러한 특징들을 반영하기 위해 loss 값을 활용하게 되며 Godard [11]는 그 좋은 예라 할 수 있다.

Ego-Motion [12]은 GT 데이터 없이 비지도학습으로 단일 2차원영상으로부터 거리 정보와 카메라의 포즈 정보를 학습을 통해 얻도록 하는 3차원 정보를 획득하는 방법을 제시했다. 단일 영상으로부터 거리를 예측하고, 인접한 영상 프레임들로부터 포즈를 예측한 후 프로젝션하여 픽셀간의 차이를 비용함수로 네트워크를 학습하여 지도학습에 근사한 결과를 얻었다. 이 논문에서는 카메라 위치 측정 성능을 ORB-SLAM과 비교하였다. ORB-SLAM은 hand-crafted feature인 ORB 특징에 기반하여 SLAM에 적용한 것으로 비지도학습의 단일 영상을 사용하였음에도 short-term 결과에서는 ORB-SLAM보다 나은 성능을 보여 주었다.

Struct2depth [13]는 구글에서 AAI'19에 발표한 논문으로 단일카메라 비디오로부터 비

지도학습으로 3차원 장면과 개별 객체의 움직임 모델링함으로서 거리와 ego-motion을 학습한 방법으로 뛰어난 성능을 얻었을 뿐만 아니라 영상 내에서 객체의 속도 측정까지도 가능한 모델을 제시하였다. 움직임에 대한 학습을 통해 Google은 움직이는 객체에 대한 거리를 정확히 측정할 수 있게 하였다.

2.4 Light-Weight Approach

거리측정 방법은 여러 영상처리 알고리즘의 센서 정보와 같은 형태로 제공되었었다. 그렇기 때문에 거리측정 알고리즘 자체가 가볍고 빠른 처리 속도를 갖는 방향으로 발전이 중요하다.

이런 측면에서 딥러닝 기반의 거리측정 기술의 경량화 [14][20]는 바람직한 일이라 하겠다. [20]은CNN 네트워크의 효율성은 reduction과 expansion의 반복이 될 때 커진다는 점에 착안하여 output 채널수를 줄이는 방식으로 걸 측정 알고리즘의 경량화를 시도하였다. 또한, 그림 3에서와 같이 메모리 및 연산량 뿐만 아니라 mobile GPU TX2에서 에너지 사용 측면에서도 효율적인 알고리즘을 개발하였다.

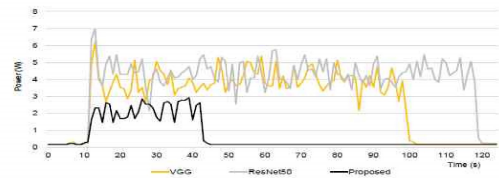


그림 3. Nvidia TX2에서의 경량화 네트워크 에너지 사용량.

Fig. 3. Energy consumption of NVIDIA Jetson TX2.

PyDNet [14]은 depth estimation을 경량화 네트워크로 만들면서도 좋은 성능을 보이는 네트워크를 설계하고자 하였다. PyDNet 구조는 GoogleNet [17]과 유사하게 여러 개의 sigmoid 함수를 주어 depth output이 여러 레벨에서 출력될 수 있도록 하였다. 또한, sigmoid 이전의 출력은 deconv에 연결된 후 하위 레이어에 연결되어 skip connection으로 형태를 만들어 성능이 향상되도록 하였다. 이 논문은 GPU 뿐만 아니라 CPU에서도 동작할

수 있음에 강점이 있다 하겠다.

2.5 Beyond 2D Images

360영상이 YouTube와 같은 곳에서 쉽게 접할 수 있는 요즘이다. 이러한 영상들에서도 depth를 뽑는 방법을 제시한 논문 (OmniDepth [15])이 발표되었다. 360 영상에서의 거리측정 기술은 다시점 영상을 만들거나 360 VR 콘텐츠를 제작함에 있어 유용하게 활용될 수 있다. 그러나 360 영상에서의 Ground truth 획득의 어려움이 있기 때문에 이 논문에서는 가상의 데이터를 사용하여 학습을 시켰다. 360 영상에서의 거리추정은 걸음마 단계이나 향후 많은 발전이 있을 것으로 예상된다.

이 뿐만 아니라 Light Field에서도 학습에 기반한 거리를 측정하는 기술이 활용되고 있다 [18].

3. 결론 및 고찰

본 논문에서는 딥러닝에 기반한 다양한 거리 추정 알고리즘에 대하여 알아보았다. 양안 방식에만 제한되어 있던 방식이 딥러닝과 접목되어 비지도학습, 단안인 상태에서도 우수한 성능을 확보하였으며 움직임을 모델링하여 움직이는 물체에 대한 거리 측정 정확도도 높일 수 있었다. 또한, 2차원 영상 데이터에서 벗어나 360 영상, Light Field 영상과 같이 다양한 차원의 영상입력으로부터도 거리를 측정하는 기술을 개발되고 있음을 확인하였다. 더불어, 딥러닝 기반이긴 하나 GPU가 아닌 CPU에서도 사용될 수 있는 알고리즘을 개발하기 위한 많은 노력이 이루어지고 있다.

표 1과 표 2는 각각 양안방식과 단안 방식에서의 KITTI2015 [5] 데이터셋을 이용하여 거리측정 기술의 성능을 비교한 표이다. 2014년부터 시작하여 짧은 시간 안에 기하학 정보의 부재에서 발생하는 문제를 해결하는 방법으로 카메라의 파라미터를 학습을 통해 알아내고 이로부터 거리측정을 보완하거나 의미론적 정보, 구조적 정보를 이용하여 거리 정보를 개선하는

등 기존 접근 방법에 비해 학습을 통해 풍부한 정보를 활용해 거리 측정의 정확도를 높이는 기술을 선보이며 꾸준히 성능을 높여 왔다.

그러나, ORB-SLAM과의 비교에서처럼 근거리에서는 기존보다 강인한 성능을 보이나 장거리에서는 그 성능이 저하되는 등 여전히 한계점이 있다. 그러나, 기존 방법이 정확한 calibration이나 두 장 이상의 카메라가 있어야 하는 한계점이 있는 반면, 딥러닝에 의한 거리 측정 방법은 카메라 포즈와 같은 다양한 정보를 활용하여 가까운 시일에 이러한 차이를 극복한 연구가 진행되어야 할 것이다. 또한, 학습데이터가 NYU를 제외하고는 야외 도시환경에 데이터가 편중되어 있다. 다양한 분야에서의 요구에 맞게 이러한 학습데이터의 다변화가 필요하다.

표 1. 양안영상기반 거리 측정 기술 비교.
Table 1. Performance comparison of stereo approach (KITTI 2015).

	D1-bg	D1-fg	D1-all
MC-CNN	-	-	3.25
GC-Net	2.21	6.16	2.87
PSMNet	1.86	4.62	2.32
DPSNet	4.21	7.58	4.77

표 2. 단안영상기반 거리 측정 기술 비교.
Table 2. Performance comparison of monocular approach (KITTI 2015).

Method	RMSE	RMSE log	Abs Rel	Sq Rel	$\delta \leq 1.25$	$\delta \leq 1.25^2$
Eigen [2]	6.307	0.282	0.203	1.548	0.702	0.890
Godard [11]	5.370	0.208	0.133	1.158	0.841	0.949
Vid2Depth [12]	6.220	0.250	0.163	1.240	0.762	0.916
struct2depth [13]	4.7503	0.1866	0.1087	0.8250	0.8738	0.9577

감사의 글

본 논문은 산업통상자원부 i-Ceramic제조혁신폐플랫폼기술개발사업으로 지원된 연구결과입니다 (세라믹산업 제조혁신을 위한 클라우드

기반 빅데이터 플랫폼 개발, 20004367).

참고 문헌

- [1] D. Scharstein, R. Szeliski, and R. Zabih, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms", Proc. SMBV, 2001.
- [2] D. Eigen, C. Puhrsch, and R. Fergus, "Depth map prediction from a single image using a multi-scale deep network", Proc. NIPS, 2014.
- [3] J. Zbontar, Y. LeCun, "Stereo matching by training a convolutional neural network to compare image patches", J. of Machine Learning Research, Vol. 17, 2016.
- [4] A. Kendall, et al., "End-to-end learning of geometry and context for deep stereo regression", Proc. ICCV, 2017.
- [5] M. Menze, C. Heipke, and A. Geiger, "Object scene flow", Journal of Photogrammetry and Remote Sensing, Vol. 140, pp. 60-76, 2018.
- [6] M. Cordts, et al., "The cityscapes dataset for semantic urban scene understanding", Proc. CVPR, 2016.
- [7] N. Silberman, P. Kohli, D. Hoiem, and R. Fergus, "Indoor segmentation and support inference from RGBD images", Proc. ECCV, 2012.
- [8] J.-R. Chang, Y.-S. Chen, "Pyramid stereo matching network", Proc. CVPR, 2018.
- [9] B. Ummenhofer, et al., "DeMoN: depth and motion network for learning monocular stereo", Proc. CVPR, 2017.
- [10] S. Im, H.-G. Jeon, S. Lin, and I. S. Kweon, "DSPNet: end-to-end deep plane sweep stereo", Proc. ICLR, 2019.
- [11] C. Godard, O. M. Aodha, and G. J. Brostow, "Unsupervised monocular depth estimation with left-right consistency", Proc. CVPR, 2017.
- [12] R. Mahjourian, M. Wicke, and A. Angelova, "Unsupervised learning of depth and ego-motion from monocular video using 3D geometric constraints", Proc. CVPR, 2018.
- [13] V. Casser, S. Pirk, R. Mahjourian, and A. Angelova, "Depth prediction without sensors: leveraging structure for unsupervised learning from monocular video", Proc. AAAI, 2019.
- [14] M. Poggi, F. Aleotti, F. Tosi, and S. Mattoccia, "Towards real-time unsupervised monocular depth estimation on CPU", Proc. IROS, 2018.
- [15] N. Zioulis, A. Karakottas, D. Zarpalas, and P. Daras, "OmniDepth: dense depth estimation for indoors spherical panoramas", Proc. ECCV, 2018.
- [16] P. Heise, S. Klose, B. Jensen, and A. Knoll, "Patchmatch with huber regularization for stereo matching", Proc. ICCV, 2013.
- [17] C. Szegedy, et al., "Going deeper with convolutions", Proc. CVPR, 2015.
- [18] H. Schilling, M. Diebold, C. Rother, and B. Jahne, "Trust your model: light field depth estimation with inline occlusion handling", Proc. CVPR, 2018.
- [19] A. Saxena, S. H. Chung, and A. Y. Ng, "Learning depth from single monocular images", Proc. NIPS, 2006.
- [20] S. Y. Oh, J. E. Lee, and H. J. S. Kim, "Fast and light-weight unsupervised depth estimation for mobile GPU hardware", Proc. CVPR-W, 2018.