

딥러닝 기반 스테레오 영상의 시차지도 추정을 위한 개선된 스케일 불변 손실함수 구현

*박지원, 최태현, 설상훈
고려대학교 전기전자공학과

e-mail : jwpark@mpeg.korea.ac.kr, thchoi@mpeg.korea.ac.kr, sull@korea.ac.kr

Implementation of Improved Scale-invariant Loss Function for Predicting Disparity-Maps of Stereo Images Using Deep-learning

*Ji-Won Park, Tae-Hyeon Choi, Sang-Hoon Sull
School of Electrical Engineering
Korea University

Abstract

A improved scale-invariant loss function and a deep learning model is proposed in this paper. The proposed loss function induce a comparison of disparity difference between neighboring pixels, and pixel-wise comparison. The proposed deep learning model consists of three parts: disparity map prediction, a depth confidence prediction, and refinement. The learning model was implemented through Tensorflow.

I. 서론

스테레오 영상의 깊이(depth) 추정은 자율 주행 차량 및 드론 등을 포함하여 많은 컴퓨터 비전 응용 과제에서 중요한 문제이다. 영상의 깊이를 알기 위해서는 각각의 스테레오 카메라에 입력되는 영상사이의 시차(disparity)를 알아야 한다. 이 시차를 구하는 방법으로 CA, SGM등의 매칭 점을 계산하는 방법이 이용되고 있으며, 최근에는 인공지능을 이용한 딥러닝(deep-learning) 방법이 연구되고 있다. 본 논문에서는

화소 고유의 시차를 유지하며 시차간의 차이를 비교하여 영상 내 스케일에 불변(scale-invariant)한 특성을 가지도록 학습하는 새로운 손실함수와 이를 이용하여 스테레오 영상에 대한 시차지도를 생성하는 학습모델을 제안한다.

II. 본론

2.1 제안하는 구조

본 논문에서 제안하는 학습모델은 다음과 같다. 입력되는 스테레오 영상에 대해 예측시차지도를 생성하는 모델과 비교 기준이 되는 단안 영상에 대해 전경(foreground)과 배경을(background) 대비시키는 깊이 신뢰지도(depth-confidence-map)를 생성하는 모델, 그리고 이 두 지도와 영상크기를 조절한 스테레오 영상을 함께 입력받아 개선된 시차지도를 생성하는 개선모델로 이루어져 있다.

예측시차지도를 생성하는 모델은 residual 연결을 사용한 unary feature 모델을 사용하였다[1]. 스테레오 영상의 각 영상을 입력받는 인코더(encoder)부분은 서로 가중치를 공유하며, 각 영상들로 부터 생성된 특징지도(feature-map)들을 연결하여 인코더-디코더(encoder-decoder)를 통해 예측시차지도를 생성한다. 깊이신뢰지도를 생성하는 모델은 VGG16의 컨볼루션

(convolution)층을 사용하였으며 전경과 배경의 대비를 위해 ground-truth에 대해 지수함수를 취하여 학습하였다. 마지막으로 개선모델은 디코더를 통해 입력받은 영상의 해상도를 개선시켜주며 디코더의 마지막 층에서 영상의 크기를 키워준다. 상기의 과정에서 예측시 차지도의 오류도 함께 개선한다.

2.2 기존의 방법과 한계점

기존의 화소단위(pixel-wise) 손실함수[1]는 결과 시차지도를 ground-truth와 화소단위로 차이를 비교하여 학습을 진행한다. 이 방법은 생성된 결과지도가 각 물체 사이의 위치 등을 고려하지 않는 경향을 보인다. 이를 보완한 스케일 불변 방법의 손실함수가 제안되었다[2]. 하지만 이 또한 영상의 스케일 특성만을 학습하기 때문에 시차지도의 검증 error율은 화소단위 손실함수에 비해 높은 경향을 보이며, 이를 낮추기 위해 추가적인 입력 데이터와 학습이 요구된다.

2.3 제안하는 방법

$$L = \sum_i^n (|P_{i_cost} - P_{i_cost}^{gt}| \times |p_i - p_i^{gt}|) \quad (1)$$

제안하는 손실함수는 수식 (1). 과 같다. 기준 화소p의 시차 값과 임의의 화소만큼 떨어진 화소q의 시차 값을 양옆 화소의 시차 값을 이용해 수식 (2). 와 같이 평균화 한 뒤 이 두 값의 차이를 수식 (3). 과 같이 기준 화소p의 cost로써 정의한다. 이 cost를 ground-truth에서의 같은 위치의 화소p^{gt}의 cost와 비교한다.



그림 2. 기준이 화소p와 임의의 값만큼 떨어진 화소q

$$p_{i_aver} = \frac{(p_{i-1} + p_i + p_{i+1})}{3} \quad q_{i_aver} = \frac{(q_{i-1} + q_i + q_{i+1})}{3} \quad (2)$$

$$P_{i_cost} = p_{i_aver} - q_{i_aver} \quad P_{i_cost}^{gt} = p_{i_aver}^{gt} - q_{i_aver}^{gt} \quad (3)$$

구성된 cost를 비교하여 생성된 차이 값에 각 화소단위로 시차 값을 ground-truth와 비교한 화소단위 차이 값의 절댓값을 곱하는 것으로 최종 손실함수를 정의한다.

III. 실험 환경 및 결과

실험 및 구현은 Tensorflow 라이브러리를 사용하였으며, 학습 환경은 NVIDIA TITAN X를 사용하여 진행하였다.

학습은 스테레오 영상에 대한 예측시차지도 학습, 단안 영상에 대한 깊이신뢰지도 학습과 두 추정 결과를 통한 최종 개선된 시차지도 학습을 동시에 진행하였다. 학습을 위한 영상은 가로, 세로 길이가 각각 640, 480인 NTSD 스테레오 영상[3]을 영상의 중심을 기준으로 512, 256 크기로 자른 영상을 사용하였으며, 총 1800쌍의 영상을 1440쌍의 학습영상과, 검증영상과 실험영상으로써 각 180쌍을 사용하였다. 입력영상과 추정영상은 최소값 0, 최대값 1이 되도록 표준화를 수행하였으며, 최종 결과영상은 최대 시차 값이 96이 되도록 표준화하여 학습을 수행하였다. 각 생성모델의 초기 학습율은 0.0001로 설정하고, 개선 모델을 제외한 나머지 모델의 학습율은 50 에폭 이후 10%씩 낮추고 개선 모델은 60 에폭 이후 10%씩 낮추는 방법으로 총 100 에폭의 학습을 수행하였다. 초기 가중치는 임의의 값을 갖도록 설정하였으며, Adam Optimizer를 통해 최적화를 수행하였다.

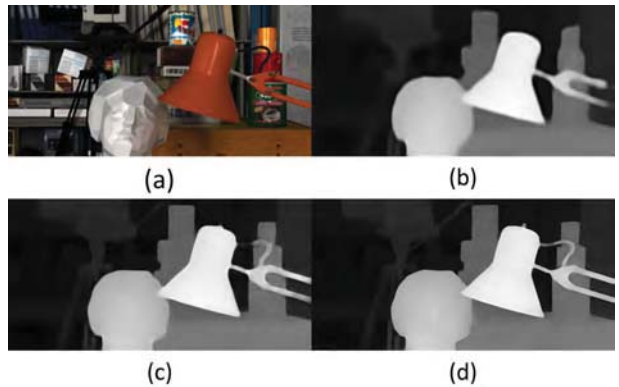


그림 3. 실험 결과

표 1. 실험 평가 결과

	(b)	(c)	(d)
EPE (<3px)	0.092 (9.2%)	0.113 (11.3%)	0.063 (6.3%)
SC-Inv error	0.895	0.682	0.576

성능평가는 화소간의 스케일 불변 특성을 판단하는 scale-invariant error와 시차의 일치 정도를 판단하는 end point error를 사용하였다. 그림 3. 의 (a)는 입력된 스테레오 영상의 좌측 영상이며, (b), (c), (d) 각각 화소단위의 손실함수, 스케일 불변 손실함수, 제안된 손실함수를 사용하여 학습한 결과영상이다. 결과 (b)의

경우 화소단위의 학습을 하기 때문에 화소 당 error율은 낮지만 영상 전체에 번짐 현상이 보이며, (c)의 경우 스케일 학습을 통해 불연속 부분이 비교적 날카롭게 구별 되지만 각 화소의 시차 값은 비율만을 학습할 뿐 error율은 제일 높은 결과를 보인다. 반면 제안된 손실함수로 학습한 결과 (d)의 경우 스케일 특성을 유지하며 화소단위 학습도 병행하기 때문에 error율도 낮은 수치를 보임을 표1.을 통해 확인할 수 있다.

IV. 결론 및 향후 연구 방향

스테레오 영상의 깊이를 추정하기 위해서는 영상 사이의 시차의 정확성 외에도 물체들의 위치 관계에 대한 이해가 중요하며, 이를 위해 각 화소의 고유 시차 추정 뿐 아니라 주변 화소와의 시차차이를 추정 할 수 있는 연구 또한 선행되어야 한다. 본 논문에서는 스테레오 영상의 시차지도를 화소의 시차 값을 유지하며 스케일 불변 특성을 지니도록 학습하는 손실함수와 딥러닝 모델을 제안하였다. 제안된 손실함수는 기존의 두 손실함수의 한계점을 보완하여 두 가지 특성을 모두 유지함으로써 시차지도의 정확성을 높였다.

Acknowledgement

“본 연구는 과학기술정보통신부 및 정보통신기술진흥센터의 대학ICT연구센터육성 지원사업의 연구결과로 수행되었음” (IITP-2017-2016-0-00464)

참고문헌

- [1] A Kendall 외, “End-to-End Learning of Geometry and Context for Deep Stereo Regression”, March, 2017.
- [2] Benjamin Ummenhofer 외, “DeMoN: Depth and Motion Network for Learning Monocular Stereo”, December, 2016.
- [3] <http://cvlab-home.blogspot.kr/>