

스테레오 영상에서 폐색에 강인하고 축소된 파라미터를 갖는 신경망

Neural network with occlusion-resistant and reduced parameters in stereo images

이 광엽*, 전영민*, 정준모**★

Kwang-Yeob Lee*, Young-Min Jeon*, Jun-Mo Jeong**★

Abstract

This paper proposes a neural network that can reduce the number of parameters while reducing matching errors in occluded regions to increase the accuracy of depth maps in stereo matching. Stereo matching-based object recognition is utilized in many fields to more accurately recognize situations using images. When there are many objects in a complex image, an occluded area is generated due to overlap between objects and occlusion by background, thereby lowering the accuracy of the depth map. To solve this problem, existing research methods that create context information and combine it with the cost volume or RoIselect in the occluded area increase the complexity of neural networks, making it difficult to learn and expensive to implement. In this paper, we create a depthwise separable neural network that enhances regional feature extraction before cost volume generation, reducing the number of parameters and proposing a neural network that is robust to occlusion errors. Compared to PSMNet, the proposed neural network reduced the number of parameters by 30%, improving 5.3% in color error and 3.6% in test loss.

요약

본 논문은 스테레오 매칭에서 깊이 맵의 정확도를 높이기 위해 폐색 영역의 매칭 오류를 줄이면서 파라미터의 수를 줄일 수 있는 신경망을 제안한다. 이미지를 이용한 상황인식을 보다 정확하게 하기 위해 많은 분야에서 스테레오 매칭기반 객체인식이 활용된다. 복잡한 이미지에 많은 객체가 있을 때 객체간의 겹침과 배경에 의한 가림으로 폐색영역이 발생하여 깊이 맵의 정확도를 낮추게 된다. 이를 해결하기 위해 context 정보를 만들어 cost volume에 결합하거나 폐색영역에 RoI를 만들어 선택하는 기존 연구 방법은 신경망의 복잡도를 높여서 학습의 어려움과 구현에 비용이 많이 들게 된다. 본 논문에서는 cost volume 생성전에 지역적인 특징추출을 보다 강화하는 depthwise separable 신경망을 만들어 파라미터의 수를 줄이고 폐색 오류에 강인한 신경망을 제안한다. 제안한 신경망은 PSMNet에 비하여 파라미터 수를 30% 줄이면서 폐색오류에서 5.3%, 테스트 손실에서 3.6% 개선하였다.

Key words : stereo matching, occlusion, cost volume, depth map, neural network

* Dept. of Computer Eng., Seokyeong University

** Dept. of Electronics Eng., Seokyeong University

★ Corresponding author

E-mail : kylee@skuniv.ac.kr, Tel : +82-2-940-7745

※ Acknowledgment

This Research was supported by Seokyeong University in 2024.

Manuscript received Mar. 18, 2024; revised Mar. 21, 2024; accepted Mar. 25, 2024.

This is an Open-Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License(<http://creativecommons.org/licenses/by-nc/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. 서론

이차원 이미지에서 단순한 객체검출 또는 인식을 벗어나 영상정보를 이용한 상황인식은 다방면에서 활용도가 높아지고 있다. 많은 사람이 모여있는 군중들 가운데에서 쓰러져있거나 앉아 있는 사람을 위급 상황에 있다고 인식하려면 사람의 자세 추정이 필요하며 상황에 따라 다른 사람 또는 사물, 배경에 신체의 일부가 가려지는 폐색(occlusion) 문제점을 해결해야 한다[1].

항공영상 기반의 공간정보 구축에서도 건물의 지붕, 그림자, 수목으로 발생한 폐색영역 보정을 위하여 다양한 각도에서 촬영된 영상을 사용하는 어려움을 갖고 있다[2].

폐색영역 문제점을 해결 방법으로 이미지 분할(segmentation)과 깊이맵(depth map)을 이용하는 것이 대표적이다. 깊이맵은 두 개의 카메라로 촬영된 스테레오 영상에서 추출되기 때문에 정확도에서는 깊이맵 방법이 더 높게 평가된다. 왼쪽과 오른쪽 두 개의 영상에서 계산된 시차 비용(disparity cost)을 모아(aggregation) 만들어진 cost volume으로부터 깊이맵이 생성되기 때문에 근소한 차이의 깊이도 인식하기 때문에 폐색된 부분을 다른 방법에 비하여 정확하게 찾게 된다.

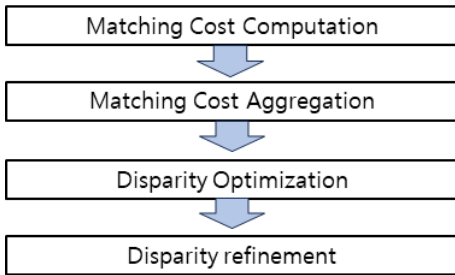


Fig. 1. Sequence of Stereo Matching.
그림 1. 스테레오 매칭 순서

깊이맵이 생성되는 과정은 그림 1과 같이 여러단계의 계산과정을 거치는데 이 과정을 기존의 영상처리 알고리즘보다 딥러닝을 이용하여 정확도를 높이는 방법이 최근 많은 연구를 통하여 사용되고 있다.

딥러닝을 이용한 스테레오 비전은 그림 2와 같이 좌, 우측 영상에서 충분한 특징값을 추출한 후 그림 1의 과정을 통하여 거리정보가 생성되는데 특징값을 추출하는 신경망 구조에 따라 시차비용과 cost volume이 만들어지며 최종 깊이맵의 정확도가 결정되기 때문에 특징 추출 신경망의 레이어 수가 증가하고 좌, 우측 대칭형 레이

어를 갖기 때문에 전체적인 파라미터의 수가 크게 증가할 수 밖에 없다. 이 문제는 파라미터 저장 메모리 증가로 학습 뿐만 아니라 추론에서도 처리 속도 저하 원인이 되고 있다. 따라서, 본 논문에서는 특징 추출 신경망의 파라미터 수를 축소하면서 스테레오 영상의 깊이맵 정확성에 제한을 주는 폐색 문제점을 개선하는 신경망 구조를 제안한다.

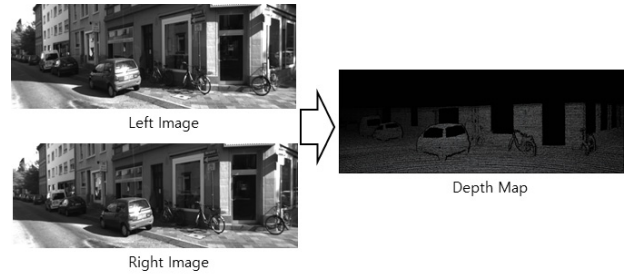


Fig. 2. Depth Map generated by Left, Right Image.
그림 2. 좌, 우 이미지로 생성된 깊이 맵 II. 관련 연구

스테레오 영상은 단안 카메라 영상보다 카메라 시점으로부터 객체들의 깊이 정보를 정확하게 제공하기 때문에 영상기반 상황인식 분야에서 오랫동안 연구되어 왔다. 그러나 스테레오 영상은 단안 영상과 달리 스테레오 매칭과정을 거쳐야 하며 이 과정의 정확도가 깊이 정보의 정밀도를 좌우하는 어려움이 있다.

스테레오 매칭을 위하여 좌, 우 영상에서 특징점을 추출하는데 높은 성능을 보이는 SIFT[3], SURF[4], ORB[5], CENSUS[6] 등의 알고리즘이 오랫동안 사용되어왔으나 딥러닝을 이용한 특징 추출이 더 우수한 성능을 보이면서 딥러닝 기반의 스테레오 매칭 연구로 급속히 전환되고 있다.

좌, 우 두 개의 이미지 패치를 매칭하고 그 결과로 스테레오 매칭 비용(cost)을 계산하는 과정을 CNN(Convolutional Neural Network)기반 학습으로 기존 영상 알고리즘 보다 좋은 성능을 보이기 시작했다[7]. 이 연구에서는 cross-based cost aggregation과 semiglobal matching을 CNN으로 구현하였다.

적은 학습 데이터에서도 새로운 클래스의 이미지를 분류할 목적으로 두 개의 이미지를 입력으로 하고 동일한 두 개의 CNN을 거쳐 두 이미지의 임베딩 벡터 거리를 유사도 값으로 출력하는 siamese 신경망[8]은 좌, 우 이미지를 사용하는 스테레오 비전과 구조적으로 일치하기 때문에 스테레오 신경망의 기본 구조로 널리 사용되어 왔다. 임베딩 벡터 거리는 L1, L2 norm을 사용하고 두

이미지의 클래스의 거리에 따라 contrastive loss로 학습하는 방법이 적용되었다.

Siamese 신경망을 바탕으로 이미지 매칭 성능을 높이기 위해 입력 이미지를 저해상도와 고해상도로 나누어 멀티 해상도를 만들고 저해상도를 두 번 반복하여 매칭을 계산하는 central-surround two-stream network이 연구되었으며 그림 3에 나타내었다[9]. 그림 3과 같이 central은 저해상도이며 고해상도인 surround 계산이 또한 번 더 사용되어 두 개의 입력 패치 이미지의 매칭 계산의 정확도가 더욱 높아진다.

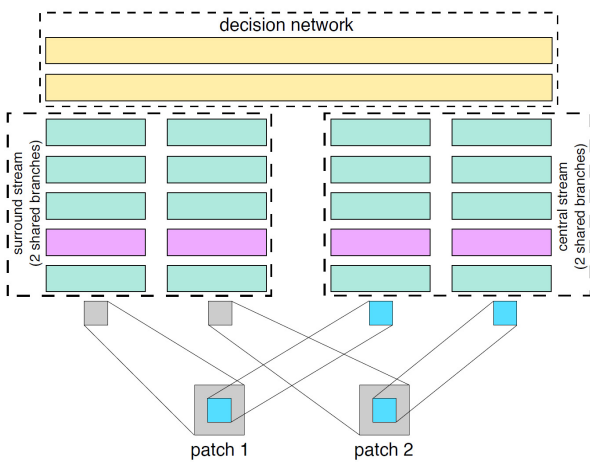


Fig. 3. Structure of Central surround network.
그림 3. Central surround 네트워크 구조

최근에는 스테레오 end-to-end 신경망을 이용하여 후처리 없이 disparity map을 생성하는 연구가 주를 이루고 있다. Pyramid stereo matching [10]은 context 정보를 강화하여 복잡한 사물에 대해서 정확한 대응점을 찾는다. 공간 정보를 담고 있는 영역을 8×8 , 16×16 , 32×32 , 64×64 크기로 단계적으로 pyramid처럼 쌓아 context 정보의 정확도를 최대화 하였다. Upsampling 과정에서는 stacked hourglass 구조를 사용하여 일반적인 encoder-decoder 구조에 비하여 더 높은 정확도의 신경망으로 depth map 생성이 가능하다.

스테레오 매칭과정에서 대상 이미지에 사물이 복잡하게 놓이게 되면 사물간에 또는 배경으로 인해 폐색(occlusion)이 발생하여 정확한 매칭점을 찾지 못하는 영역이 있게 된다. 이러한 문제를 해결하기 위해 딥러닝 신경망이 연구되어 왔다. 일반적인 스테레오 매칭은 cost volume 내에서 두 이미지의 픽셀기반으로 매칭을 계산하였으나 [11]연구에서는 그림 4와 같이 cost volume과 더불어 RoISelect와 2D-3D fusion을 결합하여 폐색영

역에 대한 cost volume의 의존도를 낮추었다. RoISelect는 높은 confidence의 volume pixel을 선택하고 보간 기법을 이용하여 연속적인 데이터 영역을 만들어 폐색 영역의 불연속성을 제거하여 문제를 해결한다.

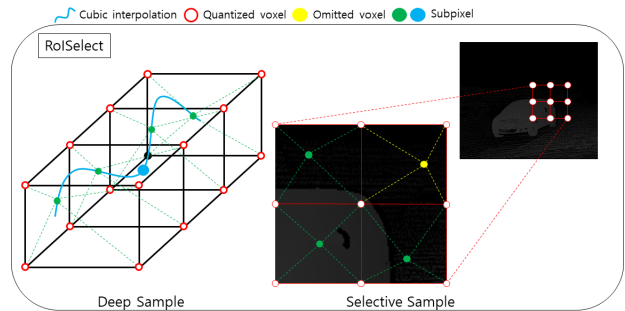


Fig. 4. RoISelect and 2D-3D fusion.
그림 4. RoISelect 와 2D-3D 결합

폐색영역 데이터를 비지도학습으로 증강하는 연구도 있다[12]. 좌, 우 이미지를 occlusion 신경망에 입력하여 occlusion map을 먼저 생성한 후 occlusion map과 입력 이미지를 concat하여 만들어진 이미지맵을 이용해 특징점 추출과 대칭점 생성을 하게된다. 특히, loss 계산에서는 occlusion 영역에는 loss를 주지않고 occlusion 이외부분에 대해서만 적용하여 학습하는 방식이다.

폐색영역에 집중하여 특징점을 추출하되 좌, 우 이미지의 매칭점을 넓은 영역에서 탐색하는 long-range dependency와 global context를 바탕으로 트랜스포머 구조를 적용하는 연구도 이루어지고 있다[13]. 폐색영역과 전역 정보와의 상관도를 계산하는 global aggregation 모듈을 이용하여 보다 향상된 depth map을 생성하는 방법이다.

III. 스테레오 매칭 신경망 설계

스테레오 매칭을 통하여 정밀도가 높은 깊이맵을 생성하면서 폐색영역과 같이 매칭점을 계산하기 어려운 문제에 강인한 end-to-end 스테레오 신경망의 설계가 본 논문에서 제안하는 내용이다. 설계된 신경망은 최근 스테레오 신경망의 근간이 되는 피라미드 풀링과 stacked hourglass구조를 기반으로 하며 관련연구들과 차별화된 폐색영역 매칭 문제 해결을 제안하였다.

1. Cost Volume 생성

객체 인식에 널리 활용되고 있는 YOLO 신경망의 근간이된 RCNN[14]은 region proposal network에서

바운딩 박스 이미지 2,000개를 대상으로 분류를 할 때 고정된 크기의 이미지를 사용하였다. 이 문제를 해결하기 위해 임의의 크기 입력 이미지에 대하여 영상 왜곡없이 특징맵을 압축하면 분류를 위한 FC 레이어에 맞도록 풀링하는 그림 5와 같은 Spatial Pyramid Pooling (SPP)-Net[10] 구조가 객체 인식에 좋은 결과를 보이기 때문에 본 논문에서도 스테레오 매칭 cost volume을 계산하기 위해 좌, 우 입력 이미지의 대칭점을 찾는 데 사용하였다. SPP-Net은 좌, 우 입력 이미지에서 추출된 특징에 context 정보를 결합하여 cost volume을 만든다.

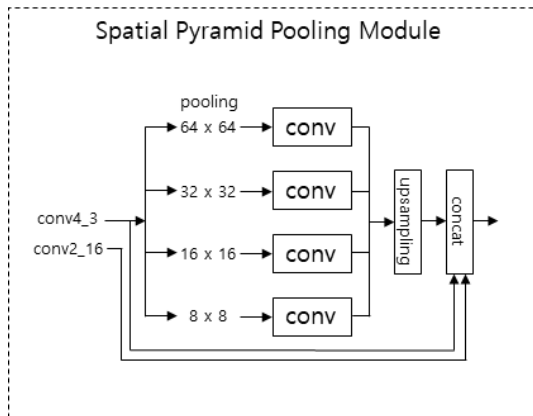


Fig. 5. Structure of Spatial Pyramid Pooling.
그림 5. 공간 피라미드 풀링 구조

2. Cost Volume 정규화

매칭 탐색에 따라 disparity 거리가 달라지고 spatial pyramid pooling을 통과하면서 다른 크기로 만들어지는 공간 정보로 인해 생성되는 cost volume이 다양하다. 따라서, cost volume aggregation을 통해 cost volume을 정규화해야 한다. 이를 위해서 일반적인 구조 보다는 그림 6의 stacked hourglass 구조를 사용하는 것이

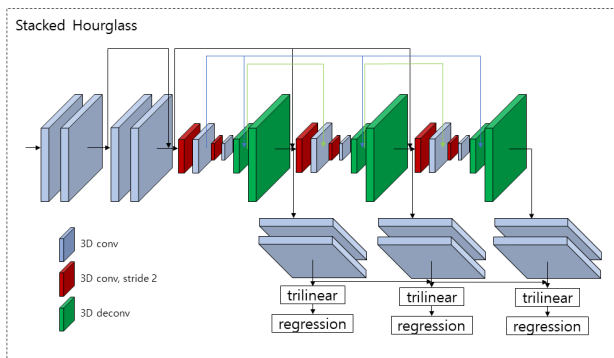


Fig. 6. Cost Aggregation using Stacked hourglass.
그림 6. Stacked hourglass를 이용한 비용 결합

효과적이다. Stacked hourglass는 단계마다 supervision 과정을 거치기 때문에 context 정보를 잘 학습한다. Stacked hourglass 모델은 3개의 hourglass 모델이 반복되며 각 모델은 각각 loss 값을 만들기 때문에 기존의 encoder-decoder 구조에 비하여 학습할 때 3개의 loss 출력이 반영되어 높은 학습 정확도를 만든다.

3. 폐색 영역에 강인한 스테레오 매칭

스테레오 매칭점을 계산할 때 폐색 영역에서는 해당 객체의 정보가 상실되어 있기 때문에 이 문제에 강인하기 위해 기존 방법으로는 cost volume과 disparity map에서 3D와 2D RoI로 폐색영역을 선택하여 보간(interpolation)함으로써 연속적인 깊이 정보를 만들어 cost volume의 폐색된 부분을 보완할 수 있다. 그러나 이 방법은 RoIselect 모듈과 보간으로 cost volume을 다시 생성하여 복잡도가 매우 높다. 또 다른 방법은 셀프 어텐션을 갖는 트랜스포머를 좌, 우 이미지의 특징맵에 적용하여 폐색영역 마스크를 만든다. 그리고 이를 이용하여 context 정보를 수정하여 disparity map을 수정 보완한다. 여기에는 셀프 어텐션 트랜스포머와 context 갱신을 포함하는 복잡도를 갖는다.

본 논문에서는 기존과는 달리 스테레오 매칭 신경망의 전체 복잡도를 줄이면서 폐색 영역 오류를 줄이는 방법으로 전역 정보를 결합하는 방식이 아닌 지역 정보를 세분화하여 강화함으로써 폐색 영역의 크기를 최소화하여 오류를 줄이는 방법을 적용하였다.

4. 제안하는 스테레오 매칭 신경망 설계

Depthwise separable convolution은 커널과의 연산을 채널 대신 depth로 대체하여 관심영역에서 지역적인 특징을 더 많이 추출할 수 있으며 가중치 파라미터가 크게 감소하여 신경망을 경량화 하는데 큰 장점이 있다.

특히 좌, 우 이미지에서 특징을 추출하는데 사용되는 CNN은 두 개의 병렬 CNN으로 설계되기 때문에 신경망의 크기에 큰 비중을 차지한다. 따라서, 특징 추출 CNN의 가중치 파라미터를 줄이면 전체 신경망 경량화에 큰 효과를 나타낸다.

스테레오 매칭 cost volume은 좌, 우 이미지에서 추출된 특징값으로 계산되기 때문에 지역적 특징을 많이 포함한 특징맵으로부터 계산된 cost volume은 폐색 영역의 불연속구간을 최소화하는 비용값으로 구성되어 복잡한 cost volume aggregation이나 refine과정 대신 복잡도가 낮은 구조로 폐색에 강인한 깊이맵을 생성한다.

Table 1. Proposed Neural Network Architecture.

표 1. 제안하는 신경망 구조

Layer Name	Proposed Neural Network
Input	$H \times W \times 3$
Feature Extraction	
conv0_1	$3 \times 3, 32$
conv0_2	$3 \times 3, 32$
conv0_3	$3 \times 3, 32$
conv1_x	$\begin{bmatrix} 3 \times 3, 32 \\ 3 \times 3, 32^* \end{bmatrix} \times 3$
conv2_x	$\begin{bmatrix} 3 \times 3, 32 \\ 3 \times 3, 32^* \end{bmatrix} \times 16$
conv3_x	$\begin{bmatrix} 3 \times 3, 32 \\ 3 \times 3, 32^* \end{bmatrix} \times 3$
conv4_x	$\begin{bmatrix} 3 \times 3, 32 \\ 3 \times 3, 32^* \end{bmatrix} \times 3, \text{dila} = 2$
* Depthwise Seperable Convolution	

구현한 stereo matching 네트워크의 구조는 Feature Extraction, SPP Module, Stacked Hourglass 3D CNN으로 3가지 부분으로 구성되어 있다. Cost Volume을 생성하기 위한 과정은 Siamese 구조로 이루어져 있으며 Feature Extraction, SPP Module은 같은 가중치를 공유한다.

Baseline Model의 Feature Extraction은 Residual Module을 수십개 쌓은 형태이며, 각 Residual Module은 2개의 2D Convolution Layer로 구성되어 있다. 본 논문에서 제안하는 네트워크는 표 1과 같이 Feature Extraction은 1개의 2D Convolution Layer와 1개의 Depthwise Seperable Convolution Layer로 구성된 Residual Module을 수십개 쌓은 형태이다. 이러한 구조가 가지는 장점은 2D Convolution Layer만을 사용한 구조에 비해 전체 학습 파라미터 수를 상당히 줄일 수 있으며, 전체 네트워크의 정확도에는 큰 손실이 없다.

Feature Extraction을 통해 얻어낸 Image Feature는 Global Context 정보를 추출해낼 수 있는 SPP Module에 입력된다. SPP Module은 서로 다른 Kernel Size를 가지는 4개의 Average Pooling Layer로 구성되어 있으며 다양한 Scale의 Context를 학습한다. SPP Module의 Kernel Size는 64×64 , 32×32 , 16×16 , 8×8 로 pyramid 방식으로 처리된다. SPP Module의 Output은 Convolution Later를 거친 뒤 Left Output과 Right Output에 대해 Concatenation을 수행하여 Cost Volume을 생성한다. 생성된 Cost Volume은 4차

원의 형태를 가진다(Height, Width, Disparity, Feature Size).

Cost Volume은 Stacked Hourglass 3D CNN을 거쳐 Cost Aggregation을 수행한다. Stacked Hourglass 3D CNN은 Encoder-Decoder 구조가 3번 반복되며 더 많은 Context 정보를 학습한다.

IV. 학습 및 측정결과

1. 학습 데이터셋

모델의 성능을 측정하기 위해 KITTI 2012[15], SceneFlow[16] 데이터셋을 사용하여 실험을 진행하였다. KITTI 2012 데이터셋은 자율주행을 위한 데이터셋으로, 실제 도로에서 주행하는 차량의 시점에서 수집한 스테레오 이미지 쌍을 제공한다. Scene Flow 데이터셋은 컴퓨터 그래픽으로 렌더링한 다양한 종류의 스테레오 이미지 쌍을 제공한다.

KITTI 2012 데이터셋은 194쌍의 트레이닝 이미지 쌍, 195쌍의 테스트 이미지 쌍으로 구성되어 있다. Scene Flow 데이터셋은 39,000쌍이 넘는 다양한 종류의 렌더링 된 그래픽 스테레오 이미지 쌍을 제공한다. 적용한 데이터 셋의 좌측 이미지와 ground-truth depth map은 그림 7과 같다.



Fig. 7. Training Data Sets.

그림 7. 학습 데이터 셋

2. 측정결과

본 실험에서는 Baseline이 되는 PSMNet과 제안하는 모델을 KITTI 2012 데이터셋을 사용하여 각각 300 에포크씩 학습을 진행하였으며 Scene Flow 데이터셋을 사용하여 3 에포크 학습을 진행하였다. 동등한 비교를 위해 Pretrained Model은 사용하지 않았다. 훈련 과정에서 Optimizer로 Adam Optimizer를 사용하였으며 Loss Function으로 Smooth L1 Loss를 사용하였다.

기존 PSMNet의 학습 파라미터 수는 8,564,320개이며 본 논문에서 제안하는 모델의 학습 파라미터 수는 표 2와 같이 5,999,392개이다.

학습이 완료된 모델의 성능을 측정하였을 때, KITTI 2012 데이터셋을 사용한 실험에서는 폐색 영역에 대해서 기존 모델보다 5%가량의 정확도 향상을 보였으며 비 폐색 영역에 대해서는 13% 가량의 성능 향상을 보였다. 3-pixel error는 예측 depth map과 실제 depth map의 각 픽셀 간의 수치 차이가 3보다 큰 픽셀들의 비율로 측정하였다. Scene Flow 데이터셋을 사용한 실험에서는 본 논문에서 제안하는 모델의 Test Loss가 기존 모델의 Test Loss에 비해 3%가량 감소하는 결과를 얻었으며 표 3에 수록하였으며 입력 이미지 대비 출력된 depth map 결과 비교는 그림 8과 같다. Test loss는 예측 데이터와 실제 데이터 사이의 차이를 loss function으로 계산되었다.

Table 2. Comparison of Number of Parameters.

표 2. 파라미터 수 비교

	Baseline (PSMNet) [10]	Proposed Network
Number of Parameters	8,564,320	5,999,392

Table 3. Comparison of occlusion error and test loss with 2 Training Data Sets.

표 3. 2개 학습 데이터 셋을 이용한 폐색 오류와 검증오류 비교

	KITTI 2012(3-pixel error)		Scene Flow (test loss)
	occ(%)	noc(%)	
Baseline (PSMNet) [10]	1.955	1.496	1.0752
Proposed Network	1.850	1.318	1.0361

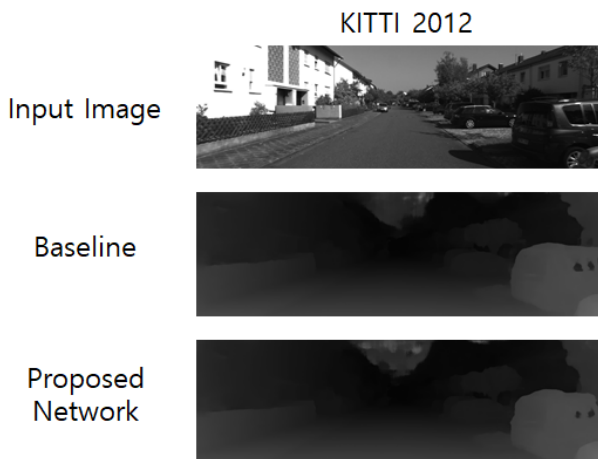


Fig. 8. Comparison of Depth Map Results.
그림 8. 깊이 맵 결과 비교

V. 결론

스테레오 매칭을 활용한 이미지의 깊이 값은 많은 분야에서 상황인식을 위해 널리 사용되고 있다. 최근에는 매칭 결과인 깊이 값의 정확도와 폐색 영역으로 인한 매칭 오류를 보상하기 위해 딥러닝 기반의 신경망 활용이 확대되고 있으나 많은 연구를 필요로 하고 있다. 본 논문에서 제안하는 Depthwise Seperable Convolution으로 Feature Extraction 단계의 Residual Module을 대체하는 구조는 기존 PSMNet에 비해 전체 학습 파라미터 수를 30%가량 감소시켰다. 실험 결과 KITTI 2012 데이터셋을 사용한 실험과 Scene Flow 데이터셋을 사용한 실험에서는 정확도가 개선되는 결과를 얻을 수 있었다.

References

[1] Hyun-Jae Bae, Jin-Pyung Kim, Jee-Hyong Lee, "Semantic Occlusion Augmentation for Effective Human Pose Estimation," *KIPS Trans. Softw. and Data Eng.*, Vol.11, No.12 pp.517-524, 2022. DOI: 10.3745/KTSDE.2022.11.12.517

[2] Kyoung-Ho Bae, Hong-Gi Park, "A Study on the Applicability of Deep Learning Algorithm for Detection and Resolving of Occlusion Area," *Journal of the Korea Academia-Industrial cooperation Society*, vol.20, No.11, pp.305-313, 2019. DOI: 10.5762/KAIS.2019.20.11.305

[3] Patricio Loncomilla, Javier Ruiz del Solar, "Improving SIFT-based object recognition for robot applications," *ICIAP 2005*, pp.1084-1092, 2005. DOI: DOI:10.1007/11553595_133

[4] H. Bay, T. Tuytelaars, L. V. Gool, "SURF: Speeded Up Robust Features," *9th European Conf. Computer Vision*, pp.404-417, 2006. DOI: 10.1007/11744023_32

[5] E. Rublee, V. Rabaud, K. Konolige, "ORB: An efficient alternative to SIFT or SURF," *Computer Vision and 2011 IEEE International Conference*, pp.2564-2571, 2011. DOI: 10.1109/ICCV.2011.6126544

[6] Wade S. Fife, James K. Archibald, "Improved

Census Transforms for Resource-Optimized Stereo Vision," *IEEE Tr. on Circuits and Systems for Video Technology*, Vol.23, Issue1, pp.60-73, 2013. DOI: 10.1109/TCSVT.2012.2203197

[7] J. Zbontar and Y. LeCun, "Computing the stereo matching cost with a convolutional neural network," in *IEEE CVPR*, 2015, pp.1592-1599. DOI: 10.1109/CVPR.2015.7298767

[8] Gregory Koch, Richard Zemel, Ruslan Salakhutdinov, "Siamese Neural Networks for One-shot Image Recognition," *ICML deep learning workshop*, 2015.

[9] Sergey Zagoruyko, Nikos Komodakis, "Learning to compare Image Patches via Convolutional Neural Networks," *Proc. of the IEEE CVPR 2015*, pp.4353-4361, 2015. DOI: 10.48550/arXiv.1504.03641

[10] Jia-Ren Chang, Yong-Sheng Chen, "Pyramid Stereo Matching Network," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.5410-5418, 2018. DOI: 10.1109/CVPR.2018.00567

[11] Jaesung Choe, Kyungdon Joo, Francois Rameau, In So Kweon, "Stereo Object Matching Network," *2021 IEEE International Conference on Robotics and Automation (ICRA)*, 2021. DOI: 10.1109/ICRA48506.2021.9562027

[12] Ang Li, Zejian Yuan, "Occlusion Aware Stereo Matching via Cooperative Unsupervised Learning," *2018 Asian Conference on Computer Vision, Springer*, pp.197-213, 2018. DOI: 10.1007/978-3-030-20876-9_13

[13] Zihua Liu, Yizhou Li, Masatoshi Okutomi, "Global Occlusion-Aware Transformer for robust Stereo Matching," *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp.3535-3544, 2024. DOI: 10.48550/arXiv.2312.14650

[14] Ross Girshick, Jeff Donahue, Trevor Darrell, Jitendra Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.

580-587, 2014. DOI: 10.1109/CVPR.2014.81

[15] <https://www.cvlibs.net/datasets/kitti/>

[16] <https://lmb.informatik.uni-freiburg.de/resources/datasets/SceneFlowDatasets.en.html>

BIOGRAPHY

Kwang-Yeob Lee (Life Member)



1985 : BS degree in Electronics Engineering, Sogang University
1987 : MS degree in Electronics Engineering, Yonsei University.
1994 : PhD degree in Electronics Engineering, Yonsei University.

1989~1995.2 : Senior Researcher, Hyundai Electronics Inc.

1995.3~present : Professor, Dept. of Computer Engineering, Seokyeong University

Young-Min Jeon (Member)



2025 : BS degree candidate in Computer Engineering, Seokyeong University.

Jun-Mo Jeong (Member)



1985 : BS degree in Electronics Engineering, Hanyang University.
1987 : MS degree in Electronics Engineering, Hanyang University.
1992 : PhD degree in Electronics Engineering, Hanyang University.
1995~presently : Professor in Seokyeong University

Main interesting field : Micro Processor, Integrated Circuit, Circuit Design & Test