

# 딥러닝기반 입체 영상의 획득 및 처리 기술 동향

## Recent Technologies for the Acquisition and Processing of 3D Images Based on Deep Learning

윤민성 (M.S. Yoon, msyoon@etri.re.kr)

VR/AR콘텐츠연구실 책임연구원

### ABSTRACT

In 3D computer graphics, a depth map is an image that provides information related to the distance from the viewpoint to the subject's surface. Stereo sensors, depth cameras, and imaging systems using an active illumination system and a time-resolved detector can perform accurate depth measurements with their own light sources. The 3D image information obtained through the depth map is useful in 3D modeling, autonomous vehicle navigation, object recognition and remote gesture detection, resolution-enhanced medical images, aviation and defense technology, and robotics. In addition, the depth map information is important data used for extracting and restoring multi-view images, and extracting phase information required for digital hologram synthesis. This study is oriented toward a recent research trend in deep learning-based 3D data analysis methods and depth map information extraction technology using a convolutional neural network. Further, the study focuses on 3D image processing technology related to digital hologram and multi-view image extraction/reconstruction, which are becoming more popular as the computing power of hardware rapidly increases.

**KEYWORDS** 딥러닝, 깊이맵 추정, 3차원 영상 처리, 다시점 영상, 홀로그래픽 콘텐츠, 3D 복원

## 1. 서론

3차원 컴퓨터 그래픽스에서 깊이맵은 시점에서부터 피사체의 표면까지의 거리와 관련된 정보를 제공하는 이미지이다. 스테레오 센서, 깊이 카메라, 능동 조명계 및 시간 분해 검출기를 사용한 이미징 시스템 등은 자체 광원에 의해 정확한 깊이

측정을 수행할 수 있다. 깊이맵을 통하여 획득된 3차원 영상 정보는 3D 모델링, 자율 차량 내비게이션, 물체 인식·원격 몸짓감지, 해상도 증대된 의료영상, 항공·국방기술, 로봇공학 등에서 유용하게 쓰이고 있다. 또한 깊이맵 정보는 다시점 영상 추출·복원의 활용 및 디지털 홀로그램 합성에 필요한 위상(Phase) 정보 추출에 사용되는 중요한 데

\* DOI: <https://doi.org/10.22648/ETRI.2020.J.350510>

\* This work was supported by Institute by the Institute of Information & Communication Technology Planning & Evaluation(IITP) grant funded by the Korea government(MIST)[No.2020-0-00954, Development of high efficiency high transparency HOE screen].

이터이다.

본고는 GPU를 포함한 하드웨어의 컴퓨팅 성능이 빠르게 증가함에 따라 재조명되고 있는 딥러닝 데이터 분석 방법과 합성곱 신경망을 이용한 3D 깊이 정보 추출 기술, 그리고 디지털 홀로그램과 다시점 영상 추출·복원 관련 3D 영상 처리 기술에 대한 최근 연구동향을 보고한다.

## II. 기술 동향

### 1. 딥러닝기반 3D 이미지 획득 기술

#### 가. 깊이맵의 추정 기술

최근에 깊이 추정(Depth Estimation)의 연구 개발은 2D 이미지로부터 3D 복원을 수행하기 위해 합성곱 신경망(CNN: Convolutional Neural Network)을 활용하는 것에 중점을 두고 있는 실정이다. 이 절에서는 CNN을 통하여 다시점 컬러 영상 세트로부터 깊이맵(Depth Map)—이미지 내에 존재하는 각 픽셀들의 상대적인 거리를 grayscale 형식으로 구분하여 나타낸 이미지—을 추출하는 방법을 간략히 소개한다. 3차원 정보 중에서 깊이맵 이미지는 주어진 시점을 관찰하는 위치로부터 피사체의 표면까지의 거리와 관련된 정보를 제공하는 영상을 의미한다. 2D 컬러 이미지로부터 정확한 깊이 추정은 주어진 장면의 이해, 내비게이션, 이미지 분할(Segmentation) 맵, 초점면 변경(Focus/Defocus), 증강 현실 복원 등을 포함한 다양한 산업 부문들에서 핵심적임 요소이다.

기본적으로 다시점 이미지는 임의의 움직임에 대하여 카메라로 획득한 이미지에서 장면 깊이를 재구성하는 것이 주요 목적이다. 다음에서는, 이 문제를 해결하고자 시도되고 있는 최근에 소개된 심층분석법(DPSNet)과 인코더-디코더 구조를 갖는 DenseDepth를 소개한다.

#### 1) DPSNet

먼저 2019년에 발표된, 상용 카메라로 촬영된 작은 모션 클립(Motion Clip)—시점 변화가 작은 이미지들로 이루어진 한 세트—으로부터, 별도의 카메라 보정(Calibration)을 필요로 하지 않으면서, 고품질의 깊이맵을 얻을 수 있는 알고리즘인 DPSNet(Deep Plane Sweep Network)을 살펴본다[1,2]. 이 방법은 보정되지 않은 단안용 카메라로 촬영된 작은 기준선들(Baselines)을 갖는 이웃한 연속 이미지들로부터 정확한 3D 점들의 특징 추출(Feature Extraction), 다시점 이미지용 비용 부피(Cost Volume) 생성 네트워크, 문맥 인식(Context-aware)과 관련된 합성곱 층 구조를 갖는 비용 집계(Cost Aggregation), 그리고 CNN을 통하여 조밀한 깊이맵을 유추하는 깊이맵 회귀(Regression) 부분으로 구성되어 있다. DPSNet은 종래의 각종 딥러닝 방법들에서 주로 사용되어 왔던 한 쌍의 이미지들로부터 깊이 또는 광학 흐름(Optical Flow)—연속된 두 개의 비디오 프레임들 사이에서 물체 이미지의 움직임 패턴—정보를 직접적으로 추정하는 방식 대신에, 평면 스위핑(Plane Sweep) 알고리즘—다각형과 같은 주어진 선분의 집합에서 선분 사이의 교점(Intersection)을 효과적으로 구하는 알고리즘—을 사용하여 심층 특징들로부터 비용 부피의 설정 및 이를 정규화함으로써 깊이맵을 효과적으로 유추할 수 있는 평면 스위핑 방식을 취하고 있다. 이 모델의 학습 과정은 Nvidia의 GPU 1,080 환경에서 PyTorch 사용자 버전을 통하여 4일 정도 소요되었다. 여기서 640×480 해상도 이미지의 각 세트를 처리하는 과정에서 특징 추출, 추적 및 번들 조정 등을 수행하는 데에 약 30초 정도 걸렸다.

또한, 제안된 방법은 단안 카메라로부터 얻은 작은 모션 클립 정보를 입력으로 하며, 카메라의 내장 및 렌즈 매개 변수도 이 작은 모션 클립으로부터

터 정밀도 높은 깊이맵을 재구성할 수 있는 것을 특징으로 한다.

제안된 이 DPSNet 모델을 토대로 한 실증 기술들을 더욱 고도화할 경우, 복잡한 표면 질감을 갖는 객체들로 구성된 이미지, 객체 주변에 상세한 구조의 전경을 담고 있는 장면 등 같이 다양한 도전적인 데이터 세트에서도 카메라 자체 교정 및 고품질 깊이 측정을 수행하도록 하는 완전 자동화형 파이프라인의 구현 가능성을 시사해 주고 있다.

## 2) LiDAR

다음은 LiDAR—레이저 빔을 목표물에 비추으로써 물성, 사물까지 거리, 또는 3D 영상 정보를 수집할 수 있는 기술—를 활용한 적응형 깊이맵 최적화 방법이 있다[3]. 기존 LiDAR 시스템이 갖는 고밀도의 3D 포인트 클라우드를 캡처하는 기능의 한계점을 극복하기 위해서 학습기반 깊이맵 완성 알고리즘이 개발되었다. RGB 이미지의 도움을 받아 누락된 깊이 정보를 보완하는 기법들이 적용되어 왔지만, 낮은 샘플링 속도로 인해 성공적이지 못하였다. 2020년에 미국의 스탠포드대학 연구팀에 의해 보고된 LiDAR 시스템에 대한 적응형 샘플링 방식이 차별화된 업스트림 작업을 통해 드문 깊이 샘플링 및 파손된 깊이 요소들을 중단 간 훈련함으로써 낮은 깊이 샘플링 조건에서도 깊이맵을 완성시킬 수 있음을 입증하였다.

적응형 샘플링 딥 네트워크는 RGB 이미지를 입력으로 하여 최적의 샘플링 패턴을 예측하고 고밀도의 깊이를 재구성한다. 사전 훈련된 단안 깊이 추정 네트워크는 주어진 장면의 초기 깊이 추정을 위하여 사용된다. 먼저 U-Net을 사용하여 샘플링 중요도 벡터 필드를 추출하면, 이 벡터 필드는 장면의 샘플링을 위하여 사용된다. 다른 하나의 U-Net을 통해 깊이 샘플을 단안 깊이 추정치와

함께 융합함으로써 조밀한 깊이를 예측한다. 깊이맵을 완성하는 작업에 사용된 이 알고리즘은 광학 위상 배열 하드웨어를 사용하여 깊이맵 이미지를 획득할 때, 처리 속도와 해상도 간 대립적 균형을 극복할 수 있음을 보여주었다. RGB 이미지로부터 정밀한 깊이맵 생성 작업에 활용된 이 적응형 샘플링 방법을 더욱 개선하면, 자율 주행 차량과 같이 주변 환경에 적응되도록 샘플링하고, 능동적으로 추론하는 이미징 시스템을 구축하는 데 활용될 것으로 예상된다.

## 3) DenseDepth

다음으로 소개할 깊이맵 추정 기술은 인코더—디코더 구조를 갖는 DenseDepth 모델이다[4]. 이것은 기존에 저해상도 및 낮은 선명도 근사 방법의 개선된 솔루션으로서, 전송학습을 통한 단일 RGB 이미지로부터 고해상도의 깊이맵 계산용 합성곱 신경망을 제시한다. 인코더—디코더 구조는 더 정확하고 상세한 경계면(Boundary)을 나타내는 고품질의 깊이맵 결과를 도출하도록 마련된 증강 학습 전략과 더불어 사전에 훈련된 고성능 네트워크를 사용하여 특징 추출을 수행한다.

깊이맵 추정을 위한 인코더는 입력 RGB 이미지를 ImageNet에 미리 훈련된 DenseNet-169을 사용하여 특징 벡터들로 인코딩된다. 이 벡터는 최종 깊이맵을 구성하기 위해 연속적인 업—샘플링 과정에 공급된다. 이 업 샘플링 층 및 스킵—연결은 디코더를 형성한다. 이 디코더는 Batch Normalization과 고급 레이어들을 포함하지 않는다. 인코더—디코더 구조기반 양질의 깊이맵 추정 방법은 일반적으로 두 가지로 나눌 수 있다. 첫 번째로는 원본 이미지에 대한 한 쌍의 좌/우 스테레오 2D 이미지들로부터 생성하는 방법이다. 두 번째로는 다 시점 영상들로 구성된 이미지 세트로부터 생성하



출처 M. Yoon, H. Kim, and H.S. Yeo, "Deep Depth Extraction of Multi-view Images for Holographic 3D Content," 2020 International Meeting on Information Display(IMID), 2020.

그림 1 DenseDepth를 활용한 CGH 합성 개념도

는 방법이다. DenseDepth 아키텍처를 토대로 추출된 고품질의 깊이맵 이미지는 최근 디지털 홀로그램 콘텐츠의 고속 획득을 목적으로 FFT(Fast Fourier Transform) 알고리즘에 의한 CGH(Computer-Generated Hologram) 합성용 위상 정보 계산에 이용되고 있다(그림 1 참조)[5]. 특히 홀로그래피/AR 스마트 글라스 시스템에서 사용자와 상호작용이 가능하도록 디자인된 디지털 홀로그래피 콘텐츠는 주어진 장면에 관하여 3차원 공간 위치에 정확히 입체 영상으로 재생되도록 하려면, 실제 목표 깊이 정보에 대한 충실한 깊이맵 이미지를 획득하는 것은 매우 중요한 요소들 중의 하나이다.

인코더-디코더 구조를 갖는 CNN기반 Dense Depth 모델은 입력 데이터에 대한 특징 추출, 샘플링 축소(Down-Sampling), 결합, 샘플링 늘임(Up-Sampling) 등의 일련 과정을 통하여 최종적으로 깊이맵을 재구성하는 작업을 수행한다. 입력된 2D 컬러 이미지는 인코더 파트에 의해서 특징 추출과 다운-샘플링 과정을 거친다. 그리고 디코더 파트에서는 추출된 특징들에 대한 연쇄(Concatenation) 연산과 컬러 이미지의 사이즈를 참조하여 업-샘플링 과정이 수행된다. 그래서 정답 레이블인 깊이 이미지 정보와 손실 함수를 통하여 얻은 손실값을 최소화하는 방향으로 가중치가 업데이트된다. 이 모델을 통하여 생성된 깊이맵 추정 데이터와 정답 레이블에 해당되는 깊이맵 데이터

간에 대한 품질 비교는 SSIM(Structural Similarity), PSNR(Peak Signal-to-Noise Ratio), 배경 잡음(Background Noise), 물체 경계면에서의 프로파일 특성 등의 지표들의 분석을 통하여 평가될 수 있다.

#### 나. 깊이맵의 해상도 증대 기술

고품질의 심도 정보 획득은 지능형 차량 운영, 제스처 인식, 3D 모델 렌더링 등과 같은 비전 기술 관련 응용 분야에서 근본적으로 해결되어야 할 도전적 과제이다. 컬러 이미지의 초해상도(SR: Super-Resolution)에서 심층 신경망의 성공적인 적용에도 불구하고, 깊이맵 SR에서 심층 신경망 기술 발전은 상대적으로 뒤쳐져 있었다. 최근 저해상도 이미지와 고해상도 이미지 및 컬러이미지와 깊이맵 이미지 사이에 종단 간(end-to-end) 비선형 매핑학습을 함께 고려하여 심층 신경망 아키텍처를 고도화하므로 고품질의 깊이맵 추정이 가능하게 되었다.

#### 다. AGN기반 깊이맵 예측

고품질 깊이맵은 정확하고 상세하게 표현되어야 한다. 심층 신경망기반 방법들은 정확도뿐만 아니라 중요한 세부사항도 고려해 주어야 한다. 전형적인 딥러닝 방식들에 의한 깊이맵의 세부적인 손상은 CNN에서 집중된 다수의 스트라이드 합성곱들(Strided Convolutions) 및 공간 풀링(Spatial Pooling)을 사용한다는 사실에 기인된다. 중국 Beihang대학 연구그룹은 주의 블록 가이드 네트워크(AGN: Attention Guided Network) 아키텍처를 활용한 깊이맵 추출 모델을 통해 깊이맵에서 구조적 세부사항을 보존하면서 3D 재구성에서 합리적인 결과를 도출하였다[6]. ResNet과 확장된 합성곱(Dilated Convolution)을 결합한 DFE(Dense Feature Extractor) 모듈은 확장된 합성곱을 사용하여 특징맵들을 조밀하게

유지하면서 이미지에서 다중 스케일 특징들을 추출한다. 깊이맵 생성기(DMG: Depth Map Generator) 모듈은 주의(Attention) 메커니즘을 사용하여 특징맵들을 깊이맵으로 회귀시킨다. 고품질의 깊이맵을 얻기 위해서 CNN에 특징 추출기(DFE: Dense Feature Extractor)와 깊이맵 추출기(DMG: Depth Map Generator)를 추가한 아키텍처를 이용하여 일련의 후처리가 필요 없는 중단 간 학습 네트워크를 제안하였다.

이 그룹은 제안한 모델을 평가하기 위해 NYU-depth v2 데이터 세트 중에서 640×480 크기의 이미지 데이터를 사용하여 15fps의 속도로 처리하였다. 밀집된 특징 추출기(DEF: Dense Feature Extractor)는 CNN을 이용한 선행 연구들과는 달리 일반적인 이미지 분류를 위해 디자인된 아키텍처를 그대로 사용하지 않고, 특징들의 해상도를 최대한 유지하면서 높은 수준의 컨텍스트 정보를 얻기 위해 확장된 컨볼루션을 접목한 필터를 사용하여 특징들을 추출하였다. 그리고 DMG는 DFE에 의해 추출된 다양한 스케일의 컨텍스트 정보를 혼합하기 위해 제시된 구조이며, 주의 융합 블록(afb: Attention Fuse Block)과 채널 축소 블록(CRB: Channel Reduce Block)으로 구성되어 있다. AFB는 주의 메커니즘을 이용하여 입력된 입력 신호에서의 컨텍스트 중에 가장 중요한 파트들을 계산하고, CRB는 ResNet의 나머지 블록(Residual Block)에서 제시한 방식대로 입력된 고차원 컨텍스트 정보들을 재정의하는 역할을 담당한다.

afb의 내부는 다음과 같다. 먼저 두 개의 입력 단과 연결된 후, 글로벌 평균 풀링 레이어와 2개의 1×1 컨볼루션 레이어가 각각 이어진다. 여기서 입력 채널 수(C)에 의해 1×1×C 텐서가 생성되는데, 이 1×1×C 텐서의 값은 특징맵의 해당 채널에서 가중치를 나타낸다. Res-Block에 의해 생성된

특징맵은 1×1×C 텐서와 곱셈(Multiplication)하여 가중치를 적용하고, 그 결과는 AFB에 의해서 생성된 특징맵에 최종 더해진다.

앞에서 기술한 DFE 및 DMG 모듈을 결합함으로써 한 장의 컬러 영상으로부터 깊이맵을 예측하는 이 새로운 중단 간 학습 네트워크 접근법은 경계면의 스무딩 및 저해상도 출력 문제 개선, 그리고 자세한 모서리 유지 등과 같이 구조적 세부 사항을 유지하고 선명하게 출력되도록 할 수 있다. 상세한 네트워크 구조는 참고문헌에서 확인할 수 있다.

**라. 깊이맵 초해상도용 딥 네트워크**

이 절에서 비선형 매핑을 직접 학습하는 방법으로, 즉 고해상도 컬러 이미지와 저해상도 깊이맵 간의 구조적 상관관계를 활용한 중단 간 CNN 학습기반 초해상도 깊이맵 출력 알고리즘을 살펴보고자 한다. 즉, 2017년에 West Virginia대학에서 개발된 컬러 이미지기반 깊이맵 초해상도 깊이맵 예측 모델(deep FCN: deep Fully Convolutional Network)은 그림 2와 같이 크게 네 개의 파트들로 구성된다[7]. 먼저 이 FCN 모델의 입력부로서 파트 ①, 즉 컬러기반 예측 네트워크(CBPN: Color-Based Prediction Network)는 HR 컬러 이미지의 특징맵

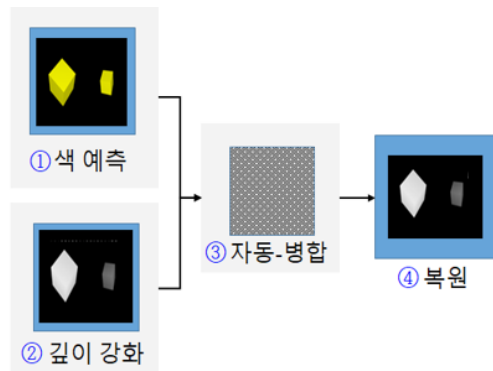


그림 2 깊이맵 출력을 위한 FCN 개념도[7]

(Feature Map)을 추출하고, 10개의 완전한 합성곱 층들(Fully Convolutional Layers)로 구성되어 있다. 모든 층들은 주어진 입력에 대하여  $3 \times 3$ 의 필터를 이용한 각 합성곱(Convolution) 연산을 통해 64개의 특징맵들을 생성하며, 이 특징맵들은 ReLU 활성화 함수를 거친 후에 다음 층으로 전달된다. 여기서 특징맵의 사이즈를 입력 이미지 사이즈와 동일하게 유지시키기 위해서 각 층마다 영값-채우기 기법(Zero-padding)을 적용한다. 이 CBPN의 주요 기능은 HR 깊이맵을 위한 고주파 성분들을 예측하는 것이다. 참고로, 더 많은 합성곱 층들을 사용하게 되면, 성능을 향상시킬 수 있으나, 계산의 복잡성은 더욱 증가된다.

한편, 이 모델의 다른 입력부로서 파트 ②는 깊이 증대 네트워크(DEN: Depth Enhancement Network)로서 LR 깊이 맵 이미지의 특징맵을 추출하며, 입력 이미지가 LR 깊이맵 이미지라는 차이 이외는 파트 ①의 구조와 동일하다. 이 DEN의 핵심 기능은 LR 깊이맵으로부터 누락된 고주파수 성분들 및 잡음 성분을 동시에 추정하는 것이다. 그리고 두 세트의 특징맵들(CBPN과 DEN)로부터 선택된 스위칭 네트워크를 설계하는 접근법 대신에, 이 깊이맵 초해상도용 FCN 모델은 그림 2의 파트 ③에서 보여주는 것과 같이 자동-병합부(Auto-merging)를 사용하는 것을 특징으로 한다.

이 파트 ③은 파트 ①에서 생성된 64개의 특징맵과 파트 ②에서 생성된 64개의 특징맵을 연쇄(Concatenation) 연산하여 128개의 특징맵을 생성한다. 여기서  $3 \times 3$ 의 필터를 이용한 1개의 합성곱 층(Convolutional layer)을 통해 특징맵을 병합한 후, 파트 ①에서 추출된 특징(HR 컬러)과 파트 ②에서 추출된 특징(LR 깊이)이 반영된 64개의 특징맵을 도출한다. 이 자동-병합 파트는 DEN 및 CBPN에 의해 생성된 특징맵들이 새로운 특징맵의 세트로 자

동적으로 결합시키는 역할을 한다. 그리고 마지막 구성 요소인 파트 ④는 복원부(Reconstruction)로서 HR 컬러 이미지에서 추출된 특징 및 LR 깊이맵 이미지에서 추출된 특징을 연쇄 연산하여 얻은 특징맵으로부터 HR 깊이맵을 재구성한다. 이때 1개의 합성곱 층을 통하여 최초 입력이었던 LR 깊이 이미지에 연쇄 특징맵(Concatenated Feature Map)인 HR 컬러 이미지 특징 및 LR 깊이 이미지 특징을 투영시켜 새로운 HR 깊이맵 이미지를 취득한다. 또한 취득된 이 HR 깊이맵 이미지에 목표 해상도(Target Resolution)를 더한 후 손실 함수를 최소화함으로써 누락된 고주파수(Missing High-frequency) 성분들을 학습하여 원치 않는 잡음 성분을 억제시킨다. 비행시간(TOF: Time-of-Flight) 방식의 상용 카메라 제품들에 의해 생성된 깊이맵 이미지의 낮은 품질 특성 때문에, 공간 분해능을 높이면서 원하지 않는 잡음들을 억제할 수 있도록 컬러 이미지의 도움을 받아서 깊이맵 이미지의 복구 및 품질 개선 방법론들은 지속적으로 발전할 것으로 예상된다.

## 2. 딥러닝기반 3D 이미지 처리 기술

컴퓨터에 의해 합성되는 홀로그램(CGH: Computer-Generated Hologram)은 임의의 광학 필드(Optical Field)를 재생하는 간섭 패턴을 계산하는 기술이다. CGH를 반복적으로 재기록할 수 있는 공간광변조기(SLM: Spatial Light Modulator)에 의해서 이 광학 필드가 복원된다. CGH를 합성하는 대표적인 방법으로 반복 최적화 알고리즘인 Gerchberg-Saxton(GS) 방법, 최적 각도 스펙트럼 계산 방법 등이 있다[8]. 이 CGH의 유망한 응용 분야는 광 세기가 출력되는 3차원 공간에서 CGH에 의해서 제어되는 동적 빔 성형(Shaping) 산업, 홀로그래픽

3D 디스플레이 산업 등이 있다. 그러나 스펙클 잡음, 다운-샘플링 효과, 0차 광, 공액(Conjugate) 광 등으로 인하여 공간 분해능 및 시야각 제한, 낮은 이미지 품질 등과 같은 해결해야 할 문제들이 남아 있다[8-10].

최근 들어 머신 러닝의 응용으로서 광학 감지 및 광 제어 등과 같은 광학 분야에서 기계학습 기술들이 도입되고 있다. 여러 딥러닝(또는 심층 신경망)기반 파동광학 연구 프로젝트들이 위상 정보 복원, 초해상도, 고스트 이미징 문제 등을 해결하기 위해 수행되고 있다. 이 절에서 향후 디지털 홀로그래피 서비스를 향한 최신 연구 결과들을 살펴 보고자 한다.

**가. 홀로그램 합성용 ResNet 모델**

2018년에 일본 Osaka대학의 Tanida 교수 연구팀은 딥러닝을 기반으로 컴퓨터 생성 홀로그램(CGH)을 합성하는 방법을 제시하였다[11]. 광파동의 역전파 과정이 다수의 스펙클 이미지 데이터 셋들을 활용하여 회귀되는 방법을 사용함으로써 비반복적인 CGH 계산이 가능함을 보였다.

제안된 이 방법은 GS 알고리즘 방법으로 획득된 위상 CGH 데이터와 비교하여 실험적으로 검증되었다. 이 연구팀이 제안한 CGH 계산용 비반복적 방법은 다중-스케일의 ResNet들(Convolutional Residual Networks)로 이루어진 네트워크를 활용하고 있다. 이 ResNet은 스킵 연결들(Skip Connections)과

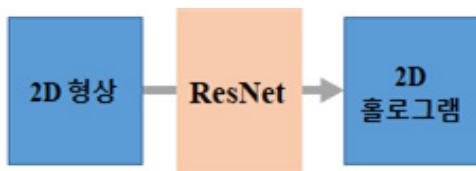


그림 3 ResNet기반 홀로그램 합성 개념도[11]

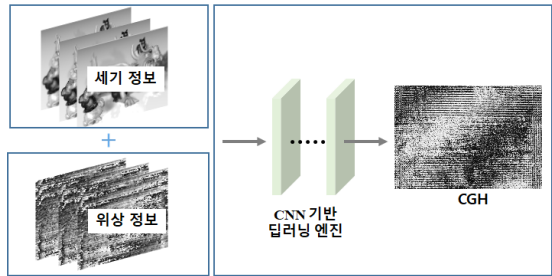


그림 4 홀로그램 합성 네트워크 개념도[9,10]

함께 하여 정체 문제(Stagnation)를 방지함으로써 딥러닝의 최적화를 가능케 하고, 특히 위상 정보 추출용으로 사용되고 있다(그림 3 참조). 기존의 방법 대비 제안된 이 방법에 의해서 얻어진 홀로그램으로부터 세기 패턴 복원의 바람직한 영상 품질 및 빠른 연산 처리 시간 등의 장점이 있음을 보였다. 그러나 이 모델은 2D 패턴에 대한 위상 전용 홀로그램에서만 검증된다는 점이 제한된 이 모델의 한계점이다.

3D 공간 형상과 같은 더 높은 차원의 빔 패턴 형성, 다채널 스펙트럼 홀로그램, 복소수 홀로그램 및 회절광학 소자, 생물학용 광학 핀셋, 고속 병렬 포커싱, 광유전학용 멀티 스팟 자극, 그리고 동화상 볼륨 이미지 구현 등으로까지 적용 가능하도록 성능을 향상시키려면, 그림 4의 예시도와 같은 기존 홀로그램 계산 알고리즘들이 CNN기반 합성 방법으로 대체되면서 복잡도가 높은 홀로그램 합성으로 확장성이 보장되는 다양한 네트워크 아키텍처들에 대한 심화 연구가 필요하다.

**나. 홀로그래픽 영상 복원용 HRNet**

디지털 홀로그래피는 진폭과 위상을 포함하는 물체의 전체 파면 특성을 기록한 정보를 갖고 있다. 이 물체를 수치적으로 복원하려면, 각 스펙트럼 방법, 컨볼루션 방법 등과 같은 Fresnel - Kirch-

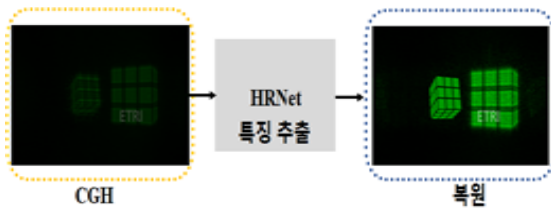


그림 5 홀로그래픽 영상 복원용 HRNet 구조의 모식도[12]

hoff 적분기반 알고리즘으로 이 홀로그램을 역전파 시키게 된다. 이러한 접근 방법은 물체 거리, 광원 파장, 조사되는 두 빔 사이의 입사각 등과 같은 사전 지식이 요구되며, 추가적인 컴퓨팅 하드웨어가 필요할 뿐만 아니라 계산 처리에서 복잡한 요구사항들이 많아진다. 예를 들어 홀로그램의 복원 시, 발생하는 원치 않은 0차 광 및 쌍둥이 이미지는 추가적인 필터링 작업을 통해 제거되어야 한다. 또한 위상 이미징의 경우, 위상 수차도 보정되어야 하고, 특히 주어진 물체의 실제 두께를 복구하려면, 잡음과 왜곡에 민감한 언래핑(Unlapping) 처리 단계가 필요하다. 이와 같은 홀로그래픽 복원 과정에서의 문제점들을 해결하기 위해서, 2019년 홍콩대 연구그룹은 홀로그래픽 복원 네트워크 HR-Net(Holographic Reconstruction Network)이라고 불리는 종단 간 딥러닝 프레임워크를 제안하였다[12].

그림 5에서 보여주는 것과 같이 이 방법은 입력 블록, 특징 추출 블록, 그리고 재구성 블록으로 구성된다. 첫 번째 블록에서 입력은 진폭 홀로그램, 위상 홀로그램, 또는 2단면 물체의 홀로그램이다. HRNet의 구조로 된 두 번째 블록은 잔여 단위(Residual Unit) 및 서브픽셀 컨볼루션 층을 포함한다. 세 번째 블록은 이 특정한 입력에 따라서 수치적으로 복원되는 이미지를 출력한다. 이 네트워크를 통해 사전 지식을 필요로 하지 않으면서 잡음이 없는 홀로그래픽 이미지의 복원이 가능함과 깊이맵 생

성 및 위상 이미징용으로 활용이 가능함을 보였다. 제안된 HRNet 모델은 시간 소모적인 계산 과정, 중간 알고리즘 디자인 등을 피할 수 있다. 특히 현미경기반 바이오 및 산업용 표본에서 수집된 데이터를 토대로 한 물체 감지, 입자 추적 및 고해상도 영상 서비스 분야 등에서 향후 적용될 잠재성을 가지고 있다.

다. 홀로그램 초해상도

최근 디지털 홀로그래픽 이미징은 생물학적 응용과 산업적 응용에서 3차원 물체의 파면 정보를 제공할 수 있는 강력한 기술로 부상되고 있다. 그러나 이미징 센서에서 픽셀 구조의 제한적 특성으로 인해 획득된 디지털 홀로그램 데이터는 홀로그래픽 복원 영상의 화질에 직접적인 영향을 미치고 있다. 이러한 제약을 극복하기 위해서 2019년에 노스웨스턴 폴리텍 대학/홍콩대학의 연구팀은 고품질의 디지털 홀로그램 콘텐츠를 획득할 수 있는 초해상도(SRH: Super-Resolution for Hologram) 딥러닝 모델을 제안하였다[13](그림 6 참조).

CNN기반으로 한 이 SRH 이미징 알고리즘은 수집된 대규모 데이터를 컨볼루션 신경망으로 훈련시킴으로써 저해상도(USAF 1951 resolution target, 512x512)의 홀로그램 데이터를 고품질(1,024x1,024)로 출력시킬 수 있는 홀로그램 해상도 향상 기술이다. 여기서 실험적으로 렌즈 없는 비축

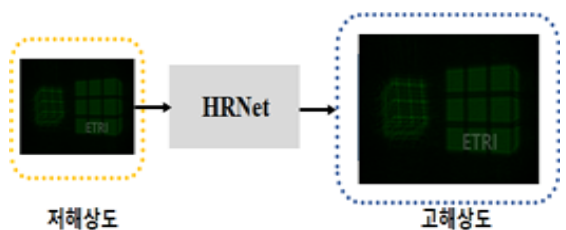


그림 6 SRH 네트워크의 모식도[13]



(Off-axis) 타입의 홀로그래프를 사용하여 검증하였다. 또한 이 모델은 추가적인 고급 하드웨어나 파라미터 최적화 프로세스 없이 홀로그래프 이미징 시스템의 공간 대역폭 품질을 향상시킬 수 있다. 이 기법은 단층 촬영·의료 진단, 3D 현미경 광학, X-선 영상 분야에서 저해상도 검출기 또는 낮은 신호 대 잡음비 등과 같은 극한 조건하에서 이미징 처리하는 작업에 특히 유용하다. 참고로 본 연구 조사에서 다루지 않지만, 저해상도 홀로그래프의 품질 개선 관련하여 초해상도 홀로그래픽 이미징용 딥러닝 알고리즘들이 UCLA의 Oscan 교수 연구팀에 의해서 활발히 개발되고 있다[14].

#### 라. 자동 초점 조절용 DeepFocus 모델

가상현실용 VR 체험 영상 서비스에서 눈이 편안하면서 실제와 같은 경험을 제공하는 기술의 핵심은 초점이 맞는 효과 및 초점이 맞지 않은 효과의 조합을 지원하는 영상 콘텐츠를 실시간으로 구현하는 것이다. 우리의 눈은 작은 카메라와 같아서 주어진 물체에 초점을 맞출 때, 장면 속의 다른 깊이에 있는 부분은 흐릿하게 보인다. 이러한 흐릿한 영역이 존재함으로써 우리의 시각 시스템은 실제 세계의 3D 구조를 자연스럽게 느끼도록 도와주고, 또한 우리가 다음 순간에 눈을 응시할 다른 곳을 결정하도록 도와준다.

이 절에서는 마지막으로 다시점 영상을 관찰하는 단안에서 입체감을 증강시킬 수 있는 초점 조절(Accommodation) 효과를 효율적으로 지원하도록 설계된 중단 간 컨볼루션 신경망인 DeepFocus 방법을 소개한다[15].

이 DeepFocus는 미국 Facebook사의 Reality Labs 연구원들이 2018년 12월에 발표한 렌더링 시스템이다. 이 시스템은 AI 툴을 사용하여 가변 초점(Varifocal) 기능을 제공하는 헤드셋 HW에서 초실감형

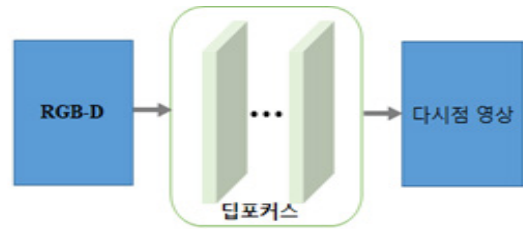


그림 7 DeepFocus 모델의 개념도[15]

영상을 실시간으로 생성할 수 있다.

이 모델은 실시간 렌더링을 위한 신경망 초-샘플링(Neural Super-sampling) 솔루션을 포함하며, 이 실시간 렌더링 기술은 저해상도 입력 이미지를 고해상도 출력으로 변환하는 기계학습 접근 방식을 기반으로 하고 있다(그림 7 참조). 여기서 업-샘플링 프로세스는 장면 통계에 대한 학습 신경망들을 사용하여 장면의 선명한 세부 정보를 복원함으로써 종래의 실시간 응용 프로그램에서 이러한 세부 사항을 직접 렌더링하는 계산상의 부담을 크게 줄일 수 있다.

초점 조절 지원이 가능한 HMD(Head-Mounted Display)를 위해 모든 범위의 계산 작업을 수행하는 이 DeepFocus 네트워크 기술은 비초점-흐림 효과 및 초점-선명 효과들을 정확하게 실시간에 가까운 속도로 합성해 낼 수 있는 특징을 갖는다. 이 결과는 제안된 이 모델을 통해 일반적으로 사용 가능한 RGB-Depth 이미지 데이터만 사용하여 자연스러운 초점 조절 표현 효과를 HMD 단말에서 제공할 수 있음을 의미한다. 또한 이 팀은 제안한 새로운 접근 방법을 통해 공간 및 시간 충실도가 높도록 렌더링된 콘텐츠를 16배 빠른 초-샘플링 처리가 가능함을 보고한 바 있다. 즉, 가변 초점형 VR 헤드셋들을 통해 시청자가 집중하는 곳은 어디든지 선명한 이미지를 제공할 수 있으며, DeepFocus 기술을 적용 시, 실제 장면처럼 나머지 장면을 자

연스러운 흐림 효과로 빠른 속도로 렌더링할 수 있다. 이 Facebook의 연구팀은 향후 깊이 자동 조절형 콘텐츠와 연동된 고화질 차세대 VR 시스템 및 실제와 구분이 불가능한 수준의 콘텐츠 체험 서비스를 선보일 준비를 하고 있다.

### III. 결론

이상에서 딥러닝 방법과 합성곱 신경망을 이용한 깊이 정보 추출과 디지털 홀로그램·다시점 영상 관련 3D 이미지의 획득·처리 기술에 대한 최신 연구동향을 살펴보았다. 딥러닝 연구 개발에 있어서 현재의 추세라면, 이동 중인 자율차량 내에서 단일 카메라로부터 촬영한 컬러 영상으로부터 홀로그램을 실시간 처리하게 되는 시대도 멀지 않을 것이다. 또한 장면 내 3D 물체들 간의 가려지는 현상이 발생 시, 가려진 영역의 추론 및 예측 형상을 디스플레이할 수 있을 것이다. 나아가 각 개인별 모바일 단말에서 촬영 중인 여러 장의 컬러 영상들로부터 360° 전방향에서 재생되는 초실감(Super-realistic) 영상을 원격 대화면으로 체험할 수 있는 시대도 수년 내에 도래할 것으로 전망된다.

앞에서 살펴 본 깊이맵·홀로그램·다시점 이미지의 생성과 후처리 방법들은 입력 영상과 준비된 정답 영상을 이용하여 다량의 데이터를 학습한 후, 테스트하는 일반적인 신경망기반 알고리즘들이다. 이러한 신경망들은 전체 데이터 중 일부를 학습 데이터로 활용한 다음, 나머지 데이터로 테스트를 진행한다. 그 결과, 전반적으로 우수한 성능을 확보할 수 있다. 반면에 특정 새 영상이 적용될 시, 과적합(Overfitting) 발생 외의 다른 원인들에 의해 성능 저하가 발생하게 된다.

무엇보다 딥러닝기반 실감 영상처리 SW는 현재 해외 의존도가 매우 높은 기술 분야이며, 우리나라

는 선진국에 비해 전문 인력 및 기술 경쟁력이 낮은 편이다. 그러나 이 분야의 관련 기관들이 원천 기술 확보가 가능한 공백 기술·핵심 요소 부문을 발견해 내고, 더불어 선택과 집중형 정책적인 지원도 반드시 필요하다고 판단된다. 나아가 학습 데이터의 신뢰도 및 일반화의 성능저하 극복이라는 과제를 해결해 나감으로써 국내의 딥러닝기반 3차원 영상 처리 분야는 그 응용 범위를 더욱 넓혀갈 것이다.

#### 용어해설

**Accommodation** 관찰자가 물체를 바라볼 때, 물체의 상이 관찰자의 망막에 정확히 맺히도록, 수정체의 형태학적 변화에 의한 초점 조절 작용을 의미. 단안에 의해 깊이 정보를 판단할 수 있는 생리적인 인지 단서(cue)들 중의 하나임

**CGH** 홀로그램 기록면에서 3차원 장면을 재생하도록 하는 광파동 장(field)을 컴퓨터에 의해 계산된 불연속적인 복소수 데이터

**CNN** 이미지 인식에 사용되는 신경망의 한 종류로서, 입력 이미지의 특징을 얻기 위한 합성곱 계층(convolutional layer), 다운-샘플링을 위한 풀링층(pooling layer), 분류를 위한 완전 결합(fully-connected) 신경망으로 이루어짐

#### 약어 정리

CGH	Computer-Generated Hologram
CNN	Convolutional Neural Network
FFT	Fast Fourier Transform
GPU	Graphics Processing Unit
LiDAR	Laser/Light Detection And Ranging

#### 참고문헌

- [1] S. Im et al., "Deep Depth from Uncalibrated Small Motion Clip," IEEE Trans. Pattern Anal. Mach. Intell., 2019, pp. 1-14.
- [2] S. Im et al., "DPSNet: End-to-end Deep Plane Sweep Stereo," ICLR2019, 2019.
- [3] A. W. Bergman, D. B. Lindell, and G. Wetzstein "Deep Adaptive LiDAR: End-to-end Optimization of Sampling and Depth Completion at Low Sampling Rates," IEEE Int. Conf. Comput.

- Photography, 2020, pp. 1-11.
- [4] I. Alhashim and P. Wonka, "High Quality Monocular Depth Estimation via Transfer Learning," arXiv preprint arXiv:1812.11941, 2018.
- [5] M. Yoon, H. Kim, and H.S. Yeo, "Deep Depth Extraction of Multi-view Images for Holographic 3D Content," 2020 International Meeting on Information Display (IMID), 2020.
- [6] Z. Hao et al., "Detail preserving depth estimation from a single image using attention guided networks," Int. Conf. 3D Vision (3DV), 2018, pp. 304-313.
- [7] W. Zhou, X. Li, and D. Reynolds, "Guided deep network for depth map super-resolution: How much can color help?" IEEE Int. Conf. on Acoustics, Speech Signal Process., 2017, pp. 1457-1461.
- [8] 윤민성, "디지털홀로그래피콘텐츠기술," 전자통신동향분석, 32권 5호, 2017, pp. 49-64.
- [9] 윤민성, 회절광학 및 홀로그래피, BOOKK, 2019.
- [10] 윤민성, 디지털 홀로그램과 디스플레이(개정판), BOOKK, 2020.
- [11] R. Horisaki, R. Takagi, and J. Tanida, "Deep-learning-generated holography," Appl. Opt., vol. 57, no. 14, 2018, pp. 3859-3863.
- [12] Z. Ren, Z. Xu, and E. Lam, "End-to-end deep learning framework for digital holographic reconstruction," Advanced Photonics, vol. 1, no. 1, 2018, p. 016004.
- [13] Z. Ren, H. So and E. Lam, "Fringe Pattern Improvement and Super-Resolution Using Deep Learning in Digital Holography," IEEE Trans. Ind. Inf., vol. 15, no. 11, 2019, pp. 6179-6186.
- [14] Y. Rivenson and A. Ozcan, "Deep learning microscopy," Optica, Vol. 4, no. 11, 2017, pp. 1437-1443.
- [15] L. Xiao et al., "DeepFocus: learned image synthesis for computational display," ACM Trans. Graph., vol. 37, no. 6, 2018, Article no. 200.