

체적 전파 네트워크: 스테레오-LiDAR 장거리 깊이 추정을 위한 융합

최재성

, 주경돈

, 투바 임티아즈, 권인소

ID

초록 - 스테레오-LiDAR 융합은 고밀도 3D 정보(스테레오 카메라)와 고정밀 희소 포인트 클라우드(LiDAR)라는 두 가지 유형의 3D 인식을 실제 사용에 활용할 수 있다는 점에서 유망한 작업입니다. 그러나 다른 방식과 구조로 인해 센서 데이터를 정렬하는 방법은 성공적인 센서 융합의 핵심입니다. 이를 위해 우리는 체적 전파 네트워크라고 하는 장거리 깊이 추정을 위한 기하학 인식 스테레오-LiDAR 융합 네트워크를 제안합니다. 우리 네트워크의 핵심 아이디어는 통일된 3D 볼륨 공간에서 스테레오 이미지의 대응을 안내하기 위한 큐로 희소하고 정확한 포인트 클라우드를 활용하는 것입니다. 기존의 융합 전략과 달리 포인트 클라우드를 볼륨에 직접 삽입하여 유효한 정보를 볼륨의 가까운 복셀에 전파하고 대응의 불확실성을 줄일 수 있습니다. 따라서 두 가지 다른 입력 양식을 매끄럽게 융합하고 장거리 깊이 맵을 회귀할 수 있습니다. 우리의 융합은 이미지에 의해 안내되는 포인트 클라우드에 대해 새로 제안된 특징 추출 레이어인 FusionConv에 의해 더욱 향상되었습니다.

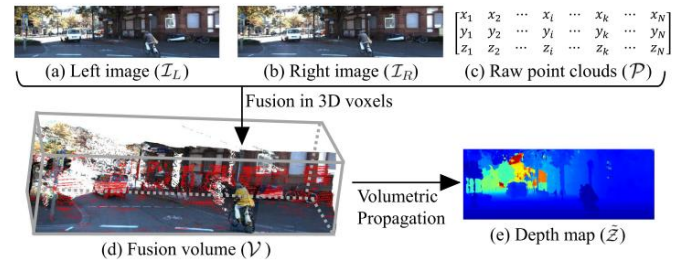


그림 1. 제안된 스테레오-LiDAR 융합의 파이프라인. (a), (b) 스테레오 이미지 및 (c) 포인트 클라우드가 주어지면 통합 3D 체적 공간인 (d) 융합 체적의 3D 복셀에서 두 가지 다른 입력 양식을 융합하여 체적에 의한 (e) 깊이 맵을 추론합니다. 번식.

FusionConv는 의미(2D 이미지 도메인)와 기하학적(3D 도메인) 관계를 모두 고려하고 볼륨에서 융합을 돕는 포인트 클라우드 기능을 추출합니다. 당사의 네트워크는 최근의 스테레오-LiDAR 융합 방법 중 KITTI 및 Virtual-KITTI 데이터 세트에서 최첨단 성능을 달성합니다.

색인 용어 - 자율 주행, 깊이 추정, 센서 융합, 스테레오-LiDAR 융합.

1. 서론

ENSOR 융합은 여러 센서의 데이터를 병합하는 프로세스로, 자율주행 또는 로봇 인식을 위한 3D 환경 (즉, 3D 인식)에 대한 이해를 심화할 수 있습니다. 각 센서는 고유한 특성을 가지며 융합을 통해 다른 센서의 한계를 보완할 수 있습니다. 특히 두 대의 카메라(스테레오 매칭) 또는 LiDAR와 한 대의 카메라(깊이 완성)와 같은 센서 융합을 통해

2020년 10월 15일 원고 접수; 2021년 2월 8일 승인. 발행일 2021년 3월 24일; 현재 버전 날짜 2021년 4월 13일. 이 서신은 부편집장 Lee S. Lee와 편집 Y.

심사위원들의 평가에 따라 최씨. 주경돈의 작업은 한국 정부(과기부)의 지원을 받는 정보통신기술기회평가원(IITP) 지원금으로 지원되었습니다.

(제2020-0-01336호, 인공지능대학원과정(UNIST)).

(교신저자 : 권인소)

최재성님이 KAIST 미래차과, Daejeon 34141, South Korea (e-mail: jaesung.choe@kaist.ac.kr)에 있습니다.

주경돈님은 울산 44919, 한국(이메일: kdjoo369@gmail.com) UNIST 컴퓨터공학과 인공지능 대학원에 있습니다.

투바 임티아즈와 권인소는 KAIST 전기공학부, Daejeon 34141, South Korea (e-mail: 14beetimtiaz@seecs.edu.pk; iskweon@kaist.ac.kr)에 있습니다.

이 편지에는 <https://doi.org/10.1109/LRA.2021.3068712>에서 다운로드할 수 있는 보충 자료가 있습니다. org/10.1109/LRA.2021.3068712, 저자 제공.

디지털 개체 식별자 10.1109/LRA.2021.3068712

정확한 깊이 정보를 추정합니다. 여러 연구에서 융합 기반 깊이 추정 알고리즘을 탐구했습니다[1]–[3]. 이러한 전통적인 방법으로 추정된 깊이의 품질은 딥 러닝의 출현으로 더욱 향상되었습니다[4]–[9].

최근에는 스테레오-라이다(stereo-LiDAR) 융합이 자율주행 실용화를 위해 더욱 주목받고 있다[10–12].

기존의 융합 기반 깊이 추정과 비교하여 스테레오 LiDAR 융합은 두 가지 유형의 3D 인식을 활용할 수 있는 새로운 작업입니다. 스테레오 카메라의 고밀도 3D 정보와 LiDAR의 희소 3D 포인트 클라우드로 깊이 품질을 더욱 향상시킬 수 있습니다. 이 융합 프레임워크 내에서 서로 다른 센서 데이터를 통합된 공간에 정렬하는 것은 깊이 추정 작업을 완전히 작동하는 데 필수적인 단계입니다. 이전 연구 [10]–[12]는 센서 간의 기하학적 관계 (즉, 내부 매개변수와 외부 매개변수)를 사용하여 2D 이미지 영역에서 융합을 수행합니다. 예를 들어, Wang et al. [10] 포인트 클라우드를 2D 이미지 도메인으로 투영하여 투영된 포인트 클라우드의 희소 깊이 맵과 이미지를 정렬합니다. 그러나, 2차원 영상 공간에서 이웃하는 픽셀은 3차원 미터법 공간에서 반드시 인접하지 않기 때문에, 영상 영역에서 이러한 2차원 융합은 깊이 단위의 공간적 연결성을 상실할 수 있고, 따라서 원거리 영역에서 정확한 깊이를 추정하는데 어려움이 있다. 스테레오 이미지와 포인트 클라우드의 기하학 인식 융합을 위해서는 통합된 3D 공간에서 공간적 연결성을 유지해야 합니다.

이 편지에서 우리는 체적 전파 네트워크라고 하는 장거리 깊이 추정을 위한 기하학 인식 스테레오-LiDAR 융합 네트워크를 제안합니다. 이를 위해 우리는 융합 체적이라고 하는 3D 체적 특성을 융합을 위한 통합 3D 체적 공간으로 정의합니다. 여기서 제안된 네트워크는 그림 1에서와 같이 스테레오 이미지와 포인트 클라우드 모두에서 대응을 계산합니다. 일치의 불확실성을 줄이기 위해 올바른 대응의 씨앗이 됩니다.

스테레오 이미지. 그런 다음 네트워크는 이 유효한 일치들 전체 볼륨에 전파하고 스테레오 이미지를 통해 나머지 볼륨의 일치 비용을 계산합니다. 또한 융합 볼륨에 포인트 기능을 포함하여 체적 전파를 촉진합니다. 포인트 피쳐는 FusionConv라고 하는 원시 포인트 클라우드에서 이미지 유도 피쳐 추출 레이어에 의해 추출됩니다. 우리의 FusionConv는 이미지 특징과 포인트 특징의 긴밀한 융합을 위해 의미론적(2D 이미지 도메인)과 기하학적(3D 도메인) 관계를 모두 고려합니다. 마지막으로 우리의 접근 방식은 KITTI 데이터 세트[13]와 Virtual-KITTI 데이터 세트[14]에 대한 스테레오-LiDAR 융합 방법 중 최첨단 성능을 달성합니다.

II. 관련 작업

스테레오 카메라, 모노 LiDAR 융합 및 스테레오-LiDAR 융합과 같은 센서 시스템의 유형에 따른 센서 융합 기반 깊이 추정 방법을 검토합니다.

스테레오 카메라: 스테레오 매칭은 한 쌍의 카메라에서 캡처한 3D 환경을 재구성하는 작업입니다[1]. 수정된 이미지 쌍 사이의 조밀한 픽셀 대응을 계산함으로써 스테레오 매칭은 역 깊이, 즉 디스패리티 맵을 추론합니다. 최근 딥 러닝 기술은 볼륨 기반 딥 아키텍처를 사용하여 보다 정확하고 강력한 매칭을 위한 기반을 마련했습니다[5], [15], [16]. 비용 볼륨은 참조 카메라 보기에서 3D 공간을 포함하는 널리 사용되는 볼륨 표현 중 하나입니다. 격자 축을 따라. 이 속성은 스테레오 이미지 간의 매칭 비용을 계산하는 데 유용합니다. 큰 개선포도 불구하고 스테레오 매칭은 여전히 먼 지역에서 정확한 깊이 추정이 부족합니다. 이 문제를 해결하기 위해 LiDAR는 장거리 깊이를 추정하는 탁월한 센서입니다.

Mono-LiDAR 융합: 깊이 완성은 단안 카메라와 LiDAR 센서를 사용하여 깊이 맵을 추정하는 작업입니다. LiDAR 센서에서 매우 정확하지만 희소한 점을 전파함으로써 이 작업은 이미지 정보의 도움으로 깊이 맵을 조밀하게 완성하는 것을 목표로 합니다. 최근의 딥 러닝 기반 방법[6], [7], [17], [18]은 깊이의 품질을 크게 향상시킵니다.

이러한 방법[6], [7], [18]은 포인트 클라우드를 이미지 도메인으로 투영하는 전처리 단계를 사용하고 이 희소 깊이 맵을 네트워크에 대한 입력으로 제공합니다 (즉, 초기 융합). 한편, Chen et al. [17]은 피쳐 공간에 중간 융합 방식을 도입했습니다. 그들은 초기에 각 양식에서 특징을 추출하고 포인트 특징을 투영하여 이미지 특징 공간에서 융합을 구현합니다. 이러한 개선에도 불구하고 깊이 완성 작업은 포인트 클라우드에 포함되지 않은 미지의 영역에서 깊이를 추정하는 데 어려움이 있습니다.

스테레오-LiDAR 융합: 스테레오 카메라-LiDAR 융합(간단히 스테레오-LiDAR 융합으로 표시됨)은 최근 추가 감각 정보를 활용하여 깊이 추정의 정확도를 더욱 높이는 것으로 입증되었습니다. 포인트 클라우드의 매우 정확하지만 희소한 깊이 맵에 의존하여 Park et al. [11] 희소 깊이 맵을 사용하여 스테레오 카메라의 디스패리티 맵을 개선합니다. 추정된 깊이의 품질이 향상되었음에도 불구하고 이 방법의 융합은 2D 이미지 영역에 있으며

따라서 장거리 깊이 추정을 위해 포인트 클라우드의 미터법 정확도를 유지하는데 충분하지 않습니다. 또 다른 최근 작업인 Wang et al. [10], 입력 융합 (즉, 스테레오 RGB-D 입력) 및 희소 디스패리티 (즉, 투영된 포인트 클라우드)에 의해 조절된 체적 정규화 의 아이디어를 소개합니다. 일반적으로 이 조건부 비용 볼륨 정규화[10]는 주로 3D 볼륨 집계에 영향을 미치며 우리의 볼륨 전파 아이디어에 더 가깝습니다. 그러나 이 방법은 오지에서 정확한 깊이를 추정하는데 어려움이 있다.

이 문제를 해결하기 위해 통합된 3D 볼륨 공간인 융합 볼륨에서 스테레오 이미지와 포인트 클라우드는 두 가지 입력 양식을 융합하는 것을 목표로 하는 체적 전파 네트워크를 도입합니다. 융합 볼륨 내에서 희소 지점을 스테레오 이미지 간의 유효한 일치들의 시드로 간주하고 전체 볼륨 공간에 유효한 일치들 전파합니다. 이를 위해 우리 네트워크는 (1) 융합 동안 포인트 클라우드의 메트릭 정확도를 유지하고 (2) 포인트 클라우드가 존재하지 않는 스테레오 매칭의 불확실성을 줄입니다. 포인트 클라우드용 특징 추출 레이어인 FusionConv를 통해 융합 볼륨에서 융합을 더욱 촉진합니다.

최근의 스테레오-LiDAR 융합 연구[10]-[12] 중에서 이 두 가지 기여로 우리는 놀라운 최첨단 깊이 추정 성능을 달성했습니다.

III. 개요

우리는 보정된 스테레오 이미지 IL 및 IR, LiDAR 포인트 클라우드 P 및 해당 보정 매개변수 (예: 카메라 고유 매트릭스 K)에서 참조 카메라 관점에서 장거리, 고정밀, 조밀한 깊이 맵 Z를 추정하는 것을 목표로 합니다. 포인트 클라우드는 N개의 3D 포인트 $P = \{p_i\}_N$ 의 집합으로 표시됩니다. 여기서 각 포인트 $p = [x, y, z]$ 는 참조 카메라 좌표에 있습니다.2 알려진 K에서 각각을 투영할 수 있습니다. 투영된 이미지 점 $x(p_i, K)$ 의 픽셀 위치에 위치할 수 있는 이미지 영역 $X \times Kp$ 로 가

보간과 함께 이 기하학적 관계를 사용하여 융합 체적(섹션 IV) 및 FusionConv(섹션 V)에서 두 가지 양식을 융합합니다. 제안된 접근 방식의 개요는 그림 2에 나와 있습니다.

IV. 체적 전파 네트워크

이 섹션에서는 체적 전파 방식으로 두 가지 다른 입력 양식인 스테레오 이미지와 원시 포인트 클라우드를 융합하기 위해 융합 볼륨 V를 구성하는 방법을 자세히 설명합니다. 제안된 융합 체적 V는 참조 뷰에서 장거리 깊이 맵 Z를 추정하기 위한 스테레오-LiDAR 융합을 위한 통합 3D 체적 공간입니다. 디스패리티 축을 따라 3D 데이터를 양자화하는 전통적인 복셀 표현 - 비용 볼륨 [5], [15], [19]와 달리, 융합 볼륨 V는 3D 환경을 깊이 범위를 따라 고르게 분포된 그리드 복셀 공간으로 설명합니다. 즉, 미터법 척도). 이것은 희소 점 P를 융합 볼륨에 포함할 때 양자화 손실을 줄이기 위한 것입니다. 비용 볼륨에 포인트를 임베딩하면 먼 지역에서 양자화 손실이 크게 증가하지만 미터법 규모의 융합 볼륨은 그렇지 않습니다. 구체적으로 정의하자면

2단순화를 위해 외부 매개변수를 사용하여 LiDAR 좌표의 포인트 클라우드를 참조 카메라 좌표로 변환합니다.

1원쪽 카메라를 기준으로 간주하는 전통을 따릅니다. [1]

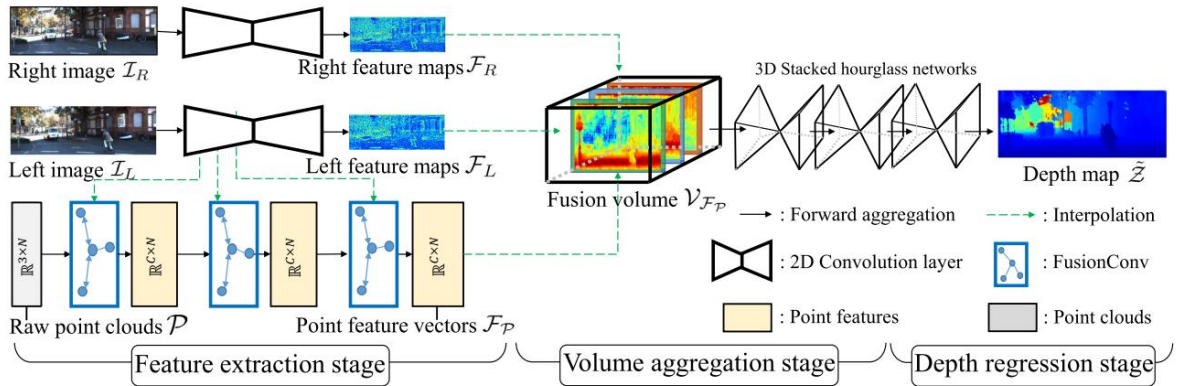


그림 2. 전체 아키텍처. 우리의 네트워크는 특징 추출 단계, 볼륨 집계 단계 및 깊이 회귀 단계의 세 단계로 구성됩니다. 우리는 처음에 입력 양식에서 특징을 추출합니다. 특히 우리는 FusionConv 레이어를 사용하여 포인트 피쳐 벡터 FP를 추론합니다. 이러한 추출된 기능은 퓨전 볼륨 VFP에 포함되어 스테레오 기능 맵 FL, FR과의 대응 관계를 계산합니다. 3D 적층 모래시계 네트워크를 통한 볼륨 집계 후, 최종적으로 참조 카메라 뷰에서 깊이 맵 Z를 얻습니다.

3D 볼륨 표현으로서의 융합 볼륨 V 및 각 복셀 $v \in V$ 는 특정 차원의 특징 벡터를 포함합니다.

즉, $V \in RW \times H \times D \times (2C+1)$, 여기서 W, H, D 는 각각 너비, 높이, 깊이 축에 따른 복셀의 수를 나타내고 C 는 이미지 특징 벡터의 차원을 나타냅니다. 볼륨 V 의 피쳐 치수는 $(2C+1)$ 로 설정되어 스테레오 이미지 피쳐와 포인트 클라우드 피쳐를 다른 채널을 따라 통합 볼륨 V 에 포함합니다.

처음에는 비어 있는 구성된 퓨전 볼륨에서 먼저 스테레오 정보를 채웁니다. 차등 비용 볼륨[5], [15]에서 영감을 받아 볼륨을 각각 데이터의 기능으로 채우면 중단 간 학습 가능한 복셀 표현이 깊이를 추정할 수 있습니다. 스테레오 이미지 $I_L, I_R \in \mathbb{R}^4 W \times 4H \times 3$ 에서 우리는 베이스라인 방법[15]의 피쳐 추출 레이어를 사용하여 스테레오 피쳐 맵 $FL, FR \in RW \times H \times C$ 를 추출합니다. 각 복셀 v 의 미리 정의된 위치를 이미지 도메인에 투영하여 융합 볼륨 V 를 스테레오 기능으로 채웁니다(그림 2 참조). 융합체적 조성에 대한 정확한 설명은 보충자료에 포함되어 있다.

우리의 융합 볼륨은 깊이 볼륨이 하는 것처럼 메트릭 스케일에 선형으로 3D 환경을 캡슐화합니다[20], [21]. 그러나 전통적인 깊이 볼륨[20], [21]이 초기 구축 비용 볼륨(즉, 불일치에서 깊이로)에서 변환된 결과인 반면, 우리의 융합 볼륨은 각각 기능에서 직접 구축되었습니다. 양자화 손실은 포인트가 초기 구축 볼륨에 포함될 때 발생하기 때문에 미터법 규모 볼륨을 직접 구성하면 특히 더 먼 영역에서 포인트의 양자화 손실을 줄일 수 있습니다. 이를 위해 알려진 카메라 행렬 K 와 볼륨 V 에서 각 복셀의 미리 정의된 위치를 사용하여 더 먼 영역에서 양자화 손실이 적은 이진 표현으로 포인트 클라우드 P 를 볼륨 V 에 포함할 수 있습니다. 점으로 포함된 복셀은 1(점유됨)로 채워지고 나머지 복셀은 0(비점유 또는 비어 있음)으로 채워집니다. 이를 위해 우리의 융합 볼륨은 통합된 볼륨 공간 내에서 스테레오 이미지와 포인트 클라우드의 기하학적 공간적 관계를 유지합니다.

지금까지 스테레오 기능 맵 FL, FR 및 원시 포인트 클라우드 P 를 융합 볼륨 V 에 통합했습니다. 이 볼륨 V 를 사용하여 임베딩 포인트를

스택된 모래시계 네트워크에서 다음 3D 컨볼루션 레이어를 통해 스테레오 기능(FL, FR) 또는 스테레오 기능과 포인트 클라우드(FL, FR, P) 간의 매칭 비용을 계산하기 위한 전체 볼륨입니다(그림 2 참조). 한편, 3D 컨볼루션 레이어는 공간적 대응과 볼륨 V 내의 채널별 특징을 모두 계산하므로 채널별 정보도 미터법 인식 융합에서 중요한 요소 중 하나입니다. 융합을 더욱 용이하게 하기 위해 다음 섹션에서 원시 포인트 클라우드 P 에서 특징 추출에 대해 논의합니다.

V. 융합 컨볼루션 레이어

최근에 많은 연구 [22]–[26]에서 포인트 클라우드에 대한 피쳐 추출 레이어를 제안했으며 포인트 피쳐 벡터를 집계할 때 로컬 이웃을 활용할 수 있는 가능성을 보여주었습니다. 포인트 클라우드에 대한 딥 아키텍처의 발전에도 불구하고 이전 방법[22]–[26]은 다른 요소와 융합하는 대신 원시 포인트 클라우드(분류 [27] 또는 세분화 작업[28], [29])를 활용하는 데 순전히 초점을 맞춥니다. 센서 정보 세트.

이 섹션에서는 센서 융합 기반 깊이 추정 작업에 특화된 FusionConv 라고 하는 포인트 클라우드용 이미지 유도 특징 추출 레이어를 소개합니다. 이미지와 포인트 클라우드 간의 알려진 기하학적 매핑 관계에 따라 여러 가지 방법으로 이미지 안내를 활용하는 FusionConv 레이어를 설계합니다. (1) 기하학적 관계(3D metric domain)와 의미적 관계(2D image domain)를 고려하여 각 점의 이웃을 적응적으로 클러스터링하여 관련 이웃을 결정할 수 있습니다. (2) 보간을 통해 입력 포인트 기능을 해당 이미지 기능과 직접 융합하여 암묵적으로 고유한 포인트 기능을 추출하는 데 도움이 됩니다.

제안된 FusionConv는 왼쪽 이미지 FL 의 입력 특징 맵과 입력 포인트 특징 벡터 $F_{in} \in \mathbb{R}^{C \times N}$ (원 포인트 클라우드 $P \in \mathbb{R}^{3 \times N}$ 에서 추출)을 융합하여 출력 포인트 특징 벡터 $F_P \in \mathbb{R}^{C \times N}$ 를 추정합니다. 관련 이웃을 고려하면서 두 가지 다른 기능을 제공합니다(그림 3 참조). 단순화를 위해 $p \in P$ 와 x 를 각각 3D 점이라고 하고 투영된 이미지 점을 K 로 지정합니다. 우리는 다음을 나타냅니다

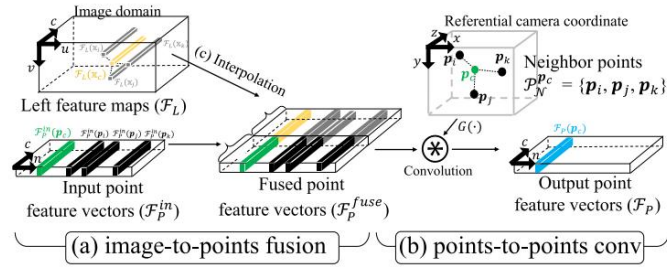


그림 3. FusionConv. 우리는 FusionConv 레이어의 프로세스를 보여줍니다. 인접 포인트 클라우드 P_{pc}에서 출력 포인트 특징 벡터 FP(pc)를 추출합니다. N 포인트 PC의. (a) 이미지 대 점 융합에서 보간을 사용하여 FL과

자세히 융합점 특징 벡터 Ffuse 생성

피. 그런 다음 (b)

융합점 P점 대 점 컨볼루션 Ffuse 피 및 기하학적 거리 인접점 사이의 G(·) P_{pc} N 출력 포인트 특징 벡터를 추론하기 위해 포인트 pc, FP(pc)로 표시. 전체 흐름은 클러스터링 단계 (P_{pc} = {파이, pj, pk}).

대응하는 3D 특징 벡터와 2D 이미지 특징 벡터

$$FP(p) \in RC \times 1 \text{ 및 } FL(x) \in RC \times 1.$$

매핑 관계를 통해 FusionConv 계층 클러스터

로컬을 집계하기 위한 인접 포인트 특징 벡터

점 특징 벡터의 응답. 많은 접근 [23] 동안 -

[26] 클러스터 포인트 클라우드 P는 미터법 공간에서만 필요합니다.

복셀을 따라가는 인접 포인트 클라우드를 클러스터링하기 위해

참조 카메라 보기에서도 정렬되는 볼륨 V의 정렬. 구체적으로, P의 각 점 p에 대해, 우리는

상주하는 인접 포인트 클라우드 PN을 동적으로 클러스터링

복셀 창 크기의 미리 정의된 크기 내에서. 예를 들어

그림 3에는 세 개의 인접한 점 pi, pj 및 pk가 있습니다.

중심이 포인트 pc를 나타내는 복셀 창 내.

이 세 개의 이웃 점과 중심점 pc 자체에서,

FusionConv 레이어는 로컬 응답을 계산할 수 있습니다.

융합과 정렬된 점 pc의 점 특징 벡터

볼륨 V.

FusionConv 레이어의 다음 프로세스는 다음으로 구성됩니다.

이미지 대 점 융합 및 점 대 점 집계. 를 위해

융합 프로세스, 우리는 먼저 FL을 FP로 보간합니다.

융합점 특징을 얻기 위한 투영 매핑 관계

벡터 퓨즈 피. 이러한 융합된 기능 Ffuse 피 같은 번호를 가지고

포인트 수는 N이지만 채널 길이가 확장됨

FL과 Fin을 모두 포함 피. 이미지-포인트 융합 후,

융합점 특징 벡터 Ffuse를 컨볼루션합니다. 피

및 집계 할 인접

포인트 PN 사이의 기하학적 거리

점(PN) 대 점 (Ffuse)이라고 하는 로컬 응답

피) 컨볼루

션. 예를 들어 그림 3에서 가중된 기하학적 거리

중심점 pc = [xc, yc, zc]에서 인접 지점까지의 G(·)

포인트 pi=[xi, yi, zi]는 다음과 같이 계산됩니다.

$$G(pc - pi) = G(\Delta p) = A_0 + A_1 \cdot \Delta x + A_2 \cdot \Delta y + A_3 \cdot \Delta z, \quad (1)$$

여기서 A₀, A₁, A₂ 및 A₃는 학습 가능한 가중치이며 Δp = [Δx, Δy, Δz]를 Δx=xc - xi, Δy=yc - yi로 정의합니다.

Δz=zc - zi. 이 가중 기하학적 거리 G(·)는

융합된 점 특징 벡터가 있는 컨볼루션의 가중치

퓨즈 점 pc에서 로컬 응답을 다음과 같이 계산합니다.

$$FP(PC) = \frac{1}{|P_{PC}|} \sum_{\text{파이} \in P_{PC}} (\text{파이}) \cdot G(pc - \text{파이}), \quad (2)$$

어디서 |P_{PC}|는 점 근처에 있는 인접 점의 수입니다.

PC 따라서 이것은 서로 다른 가중치 G(·)를 부과하는 방식입니다.

인접 포인트 기능 Fin

피 및 왼쪽 피쳐 맵 FL

중심점 특징 벡터 FP(pc)를 추출할 때. 완전히

모든 포인트 클라우드에서 출력 특징 벡터를 계산하고,

FusionConv 레이어는 출력 포인트 기능을 반복적으로 계산합니다.

벡터 FP :

$$FP = FP(p_1), FP(p_2), \dots, FP(p_N). \quad (3)$$

마지막으로 출력 포인트 특징 벡터 FP ∈ 를 추출합니다.

원시 포인트 클라우드 P 및 왼쪽 피쳐 맵 FL의 RC × N.

이 점 특징 벡터는 수정된 파일에 포함됩니다.

융합체적 VFP ∈ R³ C × D × H × W 그림 2. 내장형

FP의 위치는 해당 공간 위치와 동일합니다.

확장된 채널별 복셀 내 원시 포인트 클라우드 P

2C + 1(V) → 3C(VFP)에서 미터법 정확도를 유지합니다.

원시 포인트 클라우드 P. 이 융합 볼륨 VFP를 사용하여 융합할 수 있습니다.

서브픽셀 매칭을 계산하기 위한 두 가지 다른 방식

그림 2와 같이 다음 3D 컨볼루션 레이어에 의한 비용.

VI. 깊이 맵 회귀

융합에서 두 가지 양식의 기능을 융합한 후

볼륨 VFP, 우리는 전체에 포인트 기능을 전파

누적된 볼륨을 통해 매칭 비용을 계산합니다.

모래 시계 네트워크 [15], [31] 깊이 맵을 회귀합니다. [15]에 따라 다음을 집계하여 비

용 집계 프로세스를 수행합니다.

깊이 차원과 공간 차원에 따른 융합 체적. 쌓인 모래시계 네트워크에서 네트워크는

3개의 인코더-디코더 네트워크로 구성됩니다.

중간 손실을 통해 비용 집계를 개선 [15], [19], [31].

집계 후 비용 집계는 채널을 줄입니다.

VFP를 3D 구조로 변환 A ∈ R^D × H × W A에서 다음을 수행할 수 있습니다.

픽셀(u, v)에서 깊이 값 z~u, v를 다음과 같이 추정합니다.

$$z \sim u, v = \frac{D-1}{d=0} \frac{d}{D-1} \cdot z_{\max} \cdot \sigma(\text{ad } u, v), \quad (4)$$

여기서 z_{max}는 최대 범위를 정의하는 하이퍼 매개변수입니다.

깊이 추정의 σ(·)는 softmax 연산을 나타내며,

그리고 평균 피. 비용 집계 벡터의 d번째 값입니다.

(u, v)에서 au, v ∈ R^D × 1. 실험을 위해 하이퍼 매개변수를 D = 48 및 z_{max} = 100으로 설정했습니다. 구체적으로 Z~ 및

다음과 같이 추정된 깊이 맵에서 실제 깊이 맵 Z에서 깊이 손실 L_{depth}를 입력합니다.

$$\text{깊이} = \frac{1}{\text{중}} \text{SmoothL1}(z_u, v - z \sim u, v), \quad (5)$$

여기서 M은 정규화를 위한 Z의 유효한 픽셀 수입니다.

인자 z~u, v는 픽셀에서 예측된 깊이 맵 Z의 값입니다.

위치(u, v), 그리고 smoothL1(·)은 부드러운 L1 손실 함수입니다.

손실을 계산하는 데 사용 [5], [15]. 마지막으로 총 손실 L_{total}

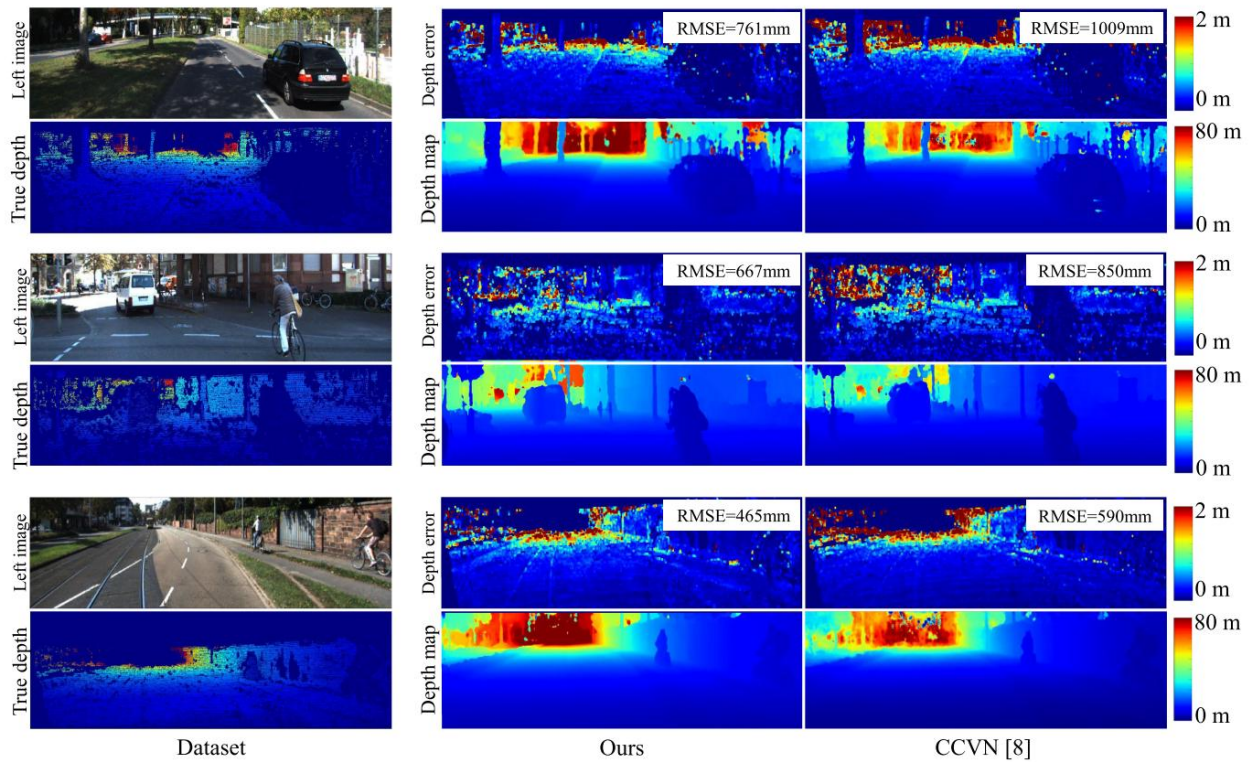


그림 4. KITTI 데이터셋에 대한 정성적 결과. 우리는 우리의 최근의 스테레오-라이다 방법(CCVN [10])의 깊이 맵과 깊이 오류를 세 가지 다른 경우에 시각화합니다. 깊이 메트릭 RMSE도 포함합니다(낮을수록 좋음).

우리 네트워크의 3개 적응 모래시계 네트워크 [15], [31]에서 깊이 맵의 3가지 다른 중간 결과로부터 다음과 같이 계산됩니다.

$$\text{총계} = \sum_{i=1}^3 w_i \cdot \text{리 깊이}, \quad (6)$$

여기서 w_i 는 i 번째 깊이 손실 L_i 의 가중치입니다. (우리는 가중치를 $w_1=0.5$, $w_2=0.7$ 및 $w_3=1.0$ 으로 설정했습니다). 평가하는 동안 그림 2와 같이 마지막 네트워크에서 예측된 깊이 맵만 고려합니다.

VII. 실험적 평가

이 섹션에서는 네트워크의 구현 세부 정보를 설명합니다. 우리의 네트워크는 두 개의 독립적인 데이터 세트인 KITTI 데이터 세트[13]와 Virtual-KITTI 데이터 세트[14]에 대해 훈련되었습니다. 기존 기술과 비교하여 각 데이터 세트의 깊이 정확도를 별도로 평가합니다. 또한, 우리는 방법의 각 주요 구성 요소를 검증하고 융합 부피 VFP에서 초기 융합 [10]과 중간 융합을 대조하기 위해 질적 연구를 수행합니다.

가. 건축

우리의 네트워크는 그림 2와 같이 특징 추출 단계, 블록 집계 단계 및 깊이 회귀 단계의 세 단계로 구성됩니다. 특징 추출 단계에서는 Chang 및 Chen [15]에서와 같이 이미지의 특징 추출 계층의 아키텍처를 따릅니다.], 하지만 중간 왼쪽 피쳐 맵을 추출합니다.

FusionConv 레이어에서 이미지 대 점 융합을 작동합니다. 3개의 FusionConv 레이어를 사용하여 점 특징 벡터를 추론합니다.

그런 다음 입력 양식에서 추출된 기능은 섹션 IV 및 V에서 설명한 대로 융합 볼륨에 포함됩니다. 3개의 적응 모래시계 네트워크를 통해 볼륨을 집계하고 마지막으로 섹션 VI에서 설명한 대로 깊이 맵을 회귀합니다.

B. 데이터 세트 및 교육 계획.

KITTI 데이터 세트. KITTI Raw 벤치마크[32]는 다양한 도로 환경에서 순차적인 스테레오 이미지와 LiDAR 포인트 클라우드를 제공합니다. 벤치마크 내에서 KITTI 완료 벤치마크[13]는 42,949개의 훈련 샘플과 1,000개의 검증 샘플로 구성된 실제 깊이 맵과 해당 센서 데이터를 제공합니다. 입력된 감각 데이터가 주어지면 학습률을 5 Epoch에 대해 0.001로 설정하고 30 Epoch에 대해 0.0001로 설정하여 네트워크를 훈련합니다. 우리는 훈련을 위해 3개의 NVIDIA 1080-Ti GPU를 사용하고 배치 크기는 9입니다. 전체 훈련 계획은 3일이 걸리고 테스트 중 네트워크의 추론 속도는 0.71FPS(프레임당 1.40초)입니다. 훈련 단계에서 무작위 자르기 증강을 사용합니다. 훈련을 위해 잘린 이미지의 크기는 256×512 입니다. 또한 잘린 왼쪽 이미지 내에 있는 포인트 클라우드 P를 자릅니다.

일반적으로 학습 단계에는 ~5K 포인트 클라우드가 있지만 이미지 내에서 잘린 영역의 위치에 따라 다를 수 있습니다. 테스트를 위해 추가 또는 필터링 없이 이미지의 원래 모양과 원시 포인트 클라우드(이미지당 ~25K 포인트)를 완전히 활용합니다.

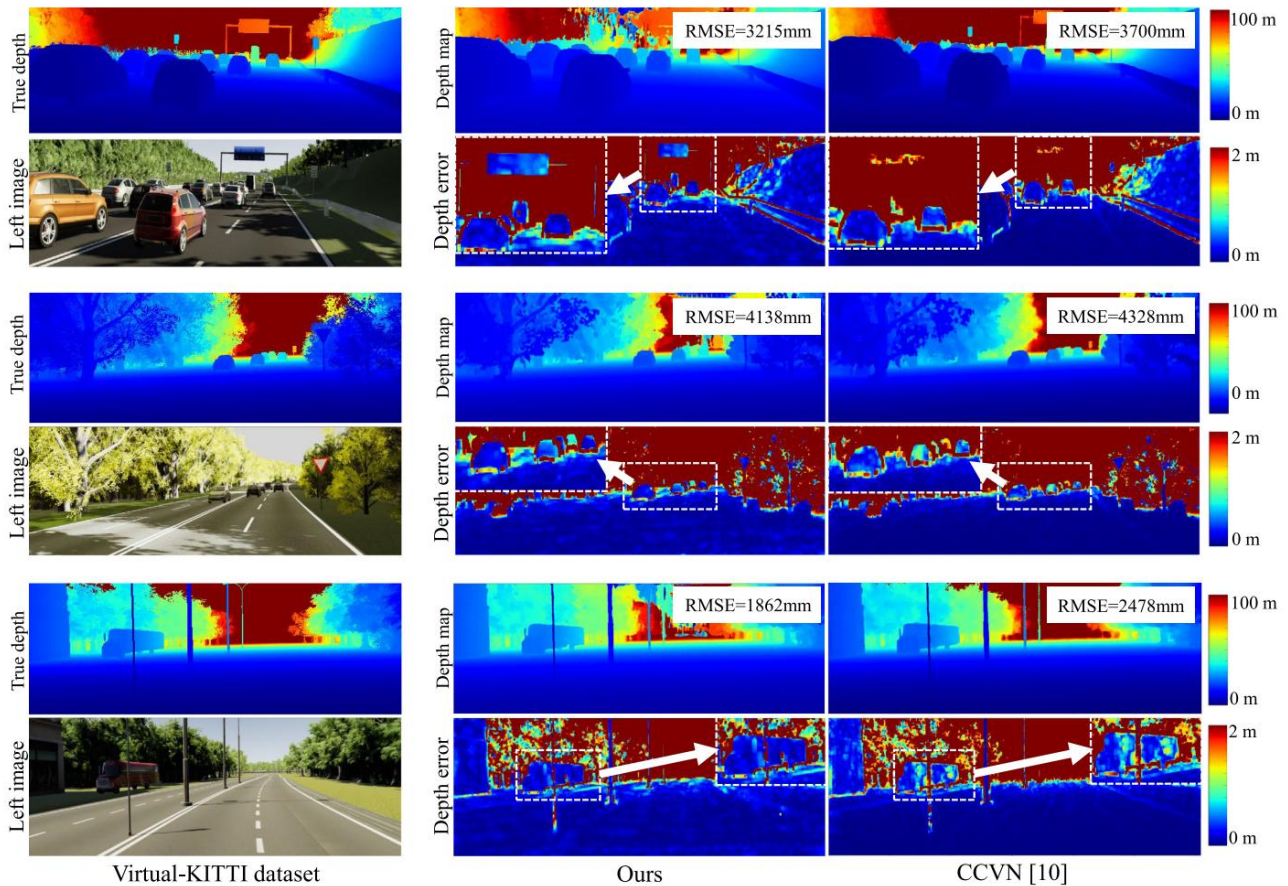


그림 5. Virtual-KITTI 2.0 데이터셋의 정성적 결과. 우리는 우리 네트워크와 Wang et al.의 최근 Stereo-LiDAR 융합 네트워크에서 추정된 깊이를 평가합니다. [10]. 이 합성 데이터는 최대 655m의 광범위한 깊이를 다루지만 표 II에서와 같이 평가 중에 실제 깊이 맵과 추정된 깊이 맵을 최대 100m까지 고정합니다. 자세한 시각화를 위해 각 프레임에서 깊이 오차 맵의 일부를 자르고 확대합니다. 주로 잘린 이미지는 장거리 깊이 추정을 검증하기 위해 더 먼 영역에 해당합니다.

가상 KITTI 2.0 데이터 세트. Virtual KITTI 2.0[14]은 이전 버전의 virtual-KITTI 1.3.1[33]보다 훨씬 사실적인 이미지를 제공하는 최근에 발표된 데이터 세트입니다.

합성 환경의 장점은 더 먼 영역에서 밀집된 ground truth에 대한 액세스를 포함하는 반면 KITTI 원료 벤치마크[13]는 상대적으로 희박한 ground truth를 제공합니다. 이 데이터 세트에는 5개의 장면이 있으며 각 장면에는 비, 일몰 등과 같은 10개의 서로 다른 시나리오가 있습니다. 우리는 네트워크 훈련을 위해 두 개의 장면(Scene01, Scene02)을 설정하고 다른 장면은 깊이의 정확도를 평가하기 위해 활용됩니다. 지도, 각 장면에 대해 훈련 및 평가를 위한 유일한 시나리오(왼쪽 15도)를 사용합니다. 훈련 세트에는 총 680개의 이미지가 있고 테스트 세트에는 1,446개의 이미지가 있습니다. 이 데이터 세트에는 원시 포인트 클라우드 데이터가 제공되지 않지만, 실제 깊이 픽셀을 무작위로 샘플링하고 픽셀을 포인트 클라우드로 간주합니다. KITTI 데이터 세트와 동일한 수의 선택된 포인트 클라우드를 선택합니다 (즉, 교육용 5K 포인트 및 테스트용 25K 포인트). KITTI 데이터 세트의 사전 훈련된 가중치를 감안할 때 KITTI 데이터 세트에 대해 채택한 것과 유사한 훈련 방식 및 증대 방법으로 네트워크를 미세 조정하지만 훈련 이미지가 더 적기 때문에 훈련 시간을 5K 반복으로 줄입니다. 가상 KITTI 데이터 세트에서.

측정 항목. 우리는 네트워크에 의해 추정된 깊이의 품질을 평가합니다. 우리는 다음에서 제안한 미터법을 따릅니다.

Eigen et al. [34] 공식 KITTI 깊이 완성 벤치마크[13], 즉 RMSE, MAE, iRMSE, iMAE와 동일하며, 여기서 RMSE 및 MAE는 대상 메트릭이거나 메트릭 거리를 평가합니다. 이러한 메트릭 공식은 Virtual-KITTI 데이터 세트[14]의 성능을 측정하기 위해 동등하게 적용됩니다. Virtual-KITTI 데이터셋은 원시 포인트 클라우드 데이터를 제공하지 않기 때문에 우리는 ground truth 깊이 픽셀을 5번 반복 샘플링하여 깊이 메트릭을 평가합니다. 표 II의 결과 메트릭은 샘플에 대한 평균입니다.

대상 메서드의 오픈 소스 코드에 액세스하여 네트워크를 재교육하고 표 I 및 II와 같이 재평가 결과를 '*' (예: PSMnet*)로 표시합니다.

비교. KITTI 원료 검증 벤치마크(표 I) 및 Virtual-KITTI 2.0(표 II)에 대한 기존 방법[10]–[12]에 대해 네트워크를 평가합니다. 또한 스테레오 매칭 방법 [5], [15] 및 깊이 완성 연구[6], [7], [18], [30]와 같은 다른 깊이 추정 네트워크의 성능도 포함합니다. 기존의 방법들 중 우리의 방법은 Table I, II와 같이 RMSE와 MAE에서 최첨단 깊이 성능을 달성하였다. 이러한 결과는 우리의 방법이 먼 지역의 깊이 추정에서 더 높은 정확도를 나타냄을 시사합니다.

이는 미터법-선형 체적 설계를 생성하려는 우리의 의도와 일치합니다. 우리 방법의 강점을 추가로 검증하기 위해 KITTI 데이터 세트에 대한 정성적 결과를 제공하고

표 I KITTI

원료 검증 벤치마크 의 깊이 추정 네트워크 의 정량적 결과 . * 는 재현 된 결과 를 나타냅니다 .

Method	Modality	Depth Evaluation (Lower the better)			
		RMSE (mm)	MAE (mm)	iRMSE (1/km)	iMAE (1/km)
GCnet [5]	Stereo	1031.4	405.4	1.6814	1.0356
PSMnet* [15]	Stereo	884	332	1.649	0.999
Sparse2Dense* [6]	Mono + LiDAR	840.0	-	-	-
Guidenet [7]	Mono + LiDAR	777.78	221.59	2.39	1.00
NLSPN [18]	Mono + LiDAR	771.8	197.3	2.0	0.8
CSPN++ [30]	Mono + LiDAR	725.43	207.88	-	-
Park <i>et al.</i> [11]	Stereo + LiDAR	2021.2	500.5	3.39	1.38
LiStereo [12]	Stereo + LiDAR	832.16	283.91	2.19	1.10
CCVN [10]	Stereo + LiDAR	749.3	252.5	1.3968	0.8069
Ours	Stereo + LiDAR	636.2	205.1	1.8721	0.9870

표 II 가상 키

티 2.0 데이터 세트 의 깊이 추정 네트워크 의 정량적 결과 . * 는 재현 된 결과 를 나타냅니다 . S, M+L 및 S+L 은 각각 라이다 가 장착된 스테레오 카메라, 라이다 가 장착된 모노 카메라 및 라이다 가 장착된 스테레오 카메라 를 나타냅니다 .

Method	Modality	Depth Evaluation (Lower the better)			
		RMSE (mm)	MAE (mm)	iRMSE (1/km)	iMAE (1/km)
PSMnet [15]*	S	5728	2235	9.805	4.380
Sparse2Dense [6]*	M+L	3357±8	1336±2	12.136±0.045	6.243±0.013
CCVN [10]*	S+L	3726.83±10.83	915.6±0.4	8.814±0.019	2.456±0.004
Ours	S+L	3217.16±1.84	712±2	7.168±0.048	2.694±0.011

표 III 깊이 추

정 을 위한 볼륨 유형의 절제 연구 . 볼륨 유형 을 YOU ET AL 에 따라 비용 볼륨 [5] (IE, COTV), 깊이 볼륨 (IE, DEPTHV) 으로 구분합니다 . [35], 그리고 우리 의 퓨전 볼륨 V (IE, FUSIONV). 귀하 등의 DEPTH VOME (IE, DEPTHV) 에 유의 하십시오 . [35] 는 비용 볼륨 에서 변환 된 볼륨 이지만 우리 의 방법 은 센서 데이터 를 볼륨 V 에 직접 보간 합니다 .

우리는 KITTI 원료 검증 에 대한 각 방법 을 평가 합니다
벤치마크 [13]

	Preserve (✓)			Depth Evaluation (Lower the better)			
	Type of Volume			RMSE (mm)	MAE (mm)	iRMSE (1/km)	iMAE (1/km)
	CostV [5]	DepthV [35]	FusionV (ours)				
1	✓			884	332	1.649	0.999
2		✓		797	286	2.046	1.070
3			✓	636.2	205.1	1.8721	0.9870

그림의 가상 KITTI 데이터 세트. 3과 4. 우리는 이러한 개선이 두 입력 양식의 기능 간의 일치 비용을 동시에 계산하는 융합 볼륨에서의 융합에 기인합니다. 추론 시간에 대한 더 많은 정량적 결과와 정보는 보충 자료에서 확인할 수 있습니다.

VIII. 절제 연구

이 섹션에서는 신경 모듈인 융합 볼륨과 FusionConv 레이어를 광범위하게 조사합니다. KITTI 데이터셋[13]에 대해 다음과 같은 실험을 수행하였다.

퓨전 볼륨. 우리는 표 III에 나와 있는 것처럼 비용 볼륨[5] 및 깊이 볼륨[35]과 같은 다른 볼륨 표현과 융합 볼륨을 비교합니다. 깊이 볼륨[35]은 비용 볼륨을 통해 구축되지만 융합 볼륨은 포인트 특징 벡터 FP를 미터법 인식 복셀로 직접 인코딩합니다. 공정한 비교를 위해 동일한 볼륨에서 각 유형의 볼륨을 평가합니다.

표 IV 깊이 추

정 을 위한 융합 유형의 절제 연구 . 각 유형의 방법 은 QI 등 의 MLP (다층 퍼셉트론) 의 RAW POINT CLOUD P, POINT 기능 벡터 와 같은 퓨전 볼륨에 다른 메트릭 정보 를 포함 합니다 . [22] FUSIONCONV 레이어 의 포인트 퍼처 벡터

	Preserve (✓)			Depth Evaluation (Lower the better)			
	Type of point network			RMSE (mm)	MAE (mm)	iRMSE (1/km)	iMAE (1/km)
	Raw points	MLP [22]	FusionConv (ours)				
4	✓			669	226	2.169	1.120
5		✓		652	212	2.099	1.066
6			✓	636.2	205.1	1.8721	0.9870

딥 아키텍처 (예: FusionConv) 및 입력 감각 데이터 (예: 스테레오-라이다 융합)와 같은 조건. Table III에서 우리의 fusion volume은 다른 voxel 표현 중에서 가장 높은 정확도를 보인다. 우리는 우리의 volume이 metric point를 volume으로 직접 인코딩할 수 있기 때문이라고 추론한다. 대조적으로, 다른 복셀 표현[35]은 초기 비용 볼륨 구성 동안 메트릭 정보를 잃을 수 있는 비용 볼륨을 통해 구성됩니다.

FusionConv. Table IV에서 우리는 FusionConv를 convolution, cluster, fusion의 3가지 요소로 분해하여 제안된 계층의 각 구성요소를 분석한다. 기준선(표 IV의 방법 4)으로 섹션 IV에 설명된 대로 원시 포인트 클라우드 P를 볼륨 VP에 포함합니다. 또한 클러스터링을 통해 점 특징 벡터 FP를 유추하지 않는 점 특징 추출[22](방법 5)을 위한 다층 퍼셉트론 계층 (즉, 완전 연결 계층)으로 또 다른 기준 네트워크를 설정했습니다. 두 가지 기본 방법의 깊이 품질이 이전 방법[10]을 능가하지만 FusionConv에서와 같이 인접 지점 PN 간의 컨볼루션 연산은 깊이 추정의 정확도를 더욱 높입니다(방법 6). 이것은 우리의 클러스터링 전략과 이미지 대 점 융합이 센서 융합 기반 깊이 추정 작업에 효과적임을 확인시켜줍니다. 마지막으로 FusionConv 레이어(방법 6)는

맨스.

초기 융합. 우리의 방법은 융합 볼륨을 통해 중간 융합이라고 하는 특징 공간에서 융합을 제안합니다. 그러나 이전 방법[10]–[12]은 내부적으로 사전 처리 단계를 사용하여 원시 포인트 클라우드 P를 이미지 도메인에 투영하여 이를 초기라고 하는 RGB 이미지와 연결합니다.

표 V 다양한 수준의 융합에 대한 절제 연구. . 얼리 퓨전(Early FUSION)은 퓨전을 위해 이미지 도메인에 포인트 클라우드를 투사하기 위한 사전 처리 단계 를 거치는 반면, 중급 퓨전(INTERMEDIATE FUSION)은 퓨전 볼륨(예: 퓨전 볼륨)과 같은 기능 공간에서 퓨전을 제안합니다.

	Preserve (✓)			Depth Evaluation (Lower the better)			
	Level of fusion		Volume	RMSE (mm)	MAE (mm)	iRMSE (1/km)	iMAE (1/km)
	Early fusion	Intermediate fusion (ours)					
7	✓		CostV	744	249	2.026	1.022
8	✓		FusionV	650	215	1.912	0.964
9		✓	FusionV	636.2	205.1	1.8721	0.9870

퓨전. 이 절제 연구를 통해 초기 융합과 중간 융합 사이의 성능 격차를 검증하여 3D 미터법 공간에서 융합의 효율성을 증명합니다. 서로 다른 융합 수준과의 공정한 비교를 위해 표 V에 나열된 볼륨 유형이 다른 유사한 아키텍처를 사용합니다. 초기 융합을 위해 포인트 클라우드를 볼륨에 포함하지 않고 이전 작업에서 제안한 대로 전처리 단계를 거칩니다. 10)–[12].

방법 9는 우리의 방법을 나타냅니다. Table V에서 3D 복셀 공간에서 중간 융합(방법 8, 9)이 초기 융합 접근(방법 7)보다 더 나은 결과를 나타냄을 알 수 있습니다. 우리는 우리의 융합 방식이 원활한 융합을 위해 포인트 클라우드와 스테레오 이미지의 공간적 연결성을 고려하고 장거리 깊이 맵을 획득한다고 추론합니다.

IX. 결론

이 편지에서 우리는 스테레오-LiDAR 융합을 위한 체적 전파 네트워크를 제안하고 장거리 깊이 추정을 수행합니다. 이를 위해 우리는 통합 체적 공간에서 융합을 촉진하기 위해 두 가지 주요 모듈인 fusion volume과 FusionConv를 설계합니다. 융합 체적 내에서 우리는 희소 점 특징과 조밀하게 정렬된 스테레오 이미지 특징의 공간적 연결성을 고려하여 체적 전파로 융합을 공식화합니다. 우리의 방법은 KITTI 및 Virtual-KITTI 데이터 세트에 대한 최첨단 성능을 보여주고 기하학적 인식 스테레오-LiDAR 융합에 대한 메시지를 전달합니다.

참조

[1] R. Hartley 및 A. Zisserman, Computer Vision의 다중 보기 기하학. 캠브리지 대학 출판부, 2003.

[2] H. Hirschmuller, "세미 글로벌 매칭 및 상호 정보를 통한 정확하고 효율적인 스테레오 처리", Proc. 회의 계산 비교 패턴 인식, 2005, pp. 807–814.

[3] F. Guney 및 A. Geiger, "Displets: 객체 지식을 사용하여 스테레오 모호성 해결", Proc. 회의 계산 비교 패턴 인식, 2015, pp. 4165–4175.

[4] N. Mayer et al., "디스플레티, 광학 흐름 및 장면 흐름 추정을 위해 컨볼루션 네트워크를 훈련하기 위한 대규모 데이터 세트", Proc. 회의 계산 비교 패턴 인식, 2016, pp. 4040–4048.

[5] A. Kendall, H. Martirosyan, S. Dasgupta 및 P. Henry, "심층 스테레오 회귀를 위한 기하학 및 컨텍스트의 종단 간 학습", Proc. IEEE 국제 회의 계산 Vis., 2017, pp. 66–75.

[6] F. Mal 및 S. Karaman, "Sparse-to-dense: 희소 깊이 샘플 및 단일 이미지에서 깊이 예측", Proc. IEEE 국제 회의 로봇. Automat., 2018, pp. 4796–4803.

[7] J. Tang, F.-P. Tian, W. Feng, J. Li 및 P. Tan, "깊이 완성을 위한 학습 안내 컨볼루션 네트워크", 2019, arXiv:1908.01238.

[8] 최재이, 주경, 임타자, 권이스, "스테레오 객체 매칭 네트워크", in Proc. IEEE 국제 회의 로봇. 오토메트, 2021.

[9] 최재이, 주경, 예프 라모, 심지, 권이스, "Segment2regress: 2단계 단안 3D 차량 위치 파악", Robotics: Sci. 시스템, 2019.

[10] T.-H. 왕, H.-N. Hu, CH Lin, Y.-H. Tsai, W.-C. Chiu 및 M. Sun, "조건부 비용 볼륨 정규화 와 함께 스테레오 매칭 네트워크를 사용하는 3D 라이다 및 스테레오 융합", Proc. IEEE/RSJ 국제 회의 인텔. 로봇 시스템, 2019.

[11] 박경, 김성수, 손경, "3D 라이다와 스테레오 융합을 통한 고정밀 깊이 추정," in Proc. IEEE 국제 회의 로봇. Automat., 2018, pp. 2156–2163.

[12] J. Zhang, MS Ramanagopal, R. Vasudevan 및 M. Johnson-Roberson, "Listereo: 라이다 및 스테레오 이미지에서 고정밀 깊이 맵 생성", 2019, arXiv:1905.02744.

[13] J. Uhrig, N. Schneider, L. Schneider, U. Franke, T. Brox 및 A. Geiger, "희소성 불변 cnns," in Proc. 국제 회의 3D Vis., 2017, pp. 11–20.

[14] Y. Cabon, N. Murray 및 M. Humenberger, "가상 키티 2", 2020, arXiv:2001.10773.

[15] J.-R. 장과 Y.-S. Chen, Proc. 의 "피라미드 스테레오 매칭 네트워크". 회의 계산 비교 패턴 인식, 2018, pp. 5410–5418.

[16] S. Im, H.-G. Jeon, S. Lin, IS Kweon, "Dpsnet: 종단 간 딥 플레인 스위프 스테레오", Proc. 국제 회의 배우다. 대표, 2019.

[17] Y. Chen, B. Yang, M. Liang 및 R. Urtasun, "깊이 완성을 위한 공동 2D-3D 표현 학습", IEEE Int. 회의 계산 비스., 2019.

[18] 박재이, 주경, 후지, C.-K. Liu, IS Kweon, "깊이 완성을 위한 비-로컬 공간 전파 네트워크", Proc. 유로 회의 계산 비스., 2020.

[19] F. Zhang, V. Prisacariu, R. Yang 및 PH Torr, "GA-net: 종단 간 스테레오 매칭을 위한 유도 잡계 네트워크", Proc. 회의 계산 비교 패턴 인식, 2019, pp. 185–194.

[20] Y. Wang, W.-L. Chao, D. Garg, B. Hariharan, M. Campbell 및 KQ Weinberger, "시각적 깊이 추정의 Pseudo-LiDAR: 자율 주행을 위한 3D 물체 감지의 격차 해소", in Proc. 회의 계산 비교 패턴 인식, 2019, pp. 8445–8453.

[21] Y. Chen, S. Liu, X. Shen, J. Jia, "DSGN: 3D 물체 감지를 위한 심층 스테레오 기하학 네트워크", 2020, arXiv:2001.03398.

[22] CR Qi, H. Su, K. Mo 및 LJ Guibas, "PointTNet: 3D 분류 및 세분화를 위한 포인트 세트에 대한 딥 러닝", Proc. 회의 계산 비교 패턴 인식, 2017, pp. 652–660.

[23] S. Wang, S. Suo, W.-C. Ma, A. Pokrovsky 및 R. Urtasun, "깊은 매개변수 연속 컨볼루션 신경망", Proc. 회의 계산 비교 패턴 인식, 2018, pp. 2589–2597.

[24] Y. Li, R. Bu, M. Sun, W. Wu, X. Di 및 B. Chen, "PointTCNN: x-변환된 점에 대한 컨볼루션", Proc. 고급 신경 정보 프로세스. 시스템, 2018, pp. 820–830.

[25] Y. Wang, Y. Sun, Z. Liu, SE Sarma, MM Bronstein 및 JM 슬로몬, "포인트 클라우드 학습을 위한 동적 그래프 cnn", ACM Trans. 그래프., vol. 38, 아니. 5, pp. 1–12, 2019.

[26] J. Mao, X. Wang 및 H. Li, "3D 포인트 클라우드 이해를 위한 보간된 컨볼루션 네트워크", Proc. IEEE 국제 회의 계산 Vis., 2019, pp. 1578–1587.

[27] AX Chang et al. "ShapENet: 정보가 풍부한 3D 모델 저장소", 2015, arXiv:1512.03012.

[28] I. Armeni et al., "대규모 실내 공간의 3D 의미 분석", Proc. 회의 계산 비교 패턴 인식, 2016, pp. 1534–1543.

[29] L. Yi et al., "3D 모양 컬렉션에서 영역 주석을 위한 확장 가능한 활성 프레임워크", ACM Trans. on Graph., vol. 35, 아니. 6, 2016년 1~12페이지.

[30] X. Cheng, P. Wang, C. Guan 및 R. Yang, "CSPN++: 학습 컨텍스트 및 깊이 완성을 위한 리소스 인식 컨볼루션 공간 전파 네트워크", Proc. AAAI 회의 아티프. Intel., 2020, pp. 10615–10622.

[31] A. Newell, K. Yang 및 J. Deng, "인간 포즈 추정을 위한 누적된 모래시계 네트워크", Proc. 유로 회의 계산 Vis., 2016, pp. 483–499.

[32] A. Geiger, P. Lenz, C. Stiller, R. Urtasun, "비전과 로봇 공학: 키티 데이터 세트" Int. 제이 로봇. Res., vol. 32, 아니. 11, pp. 1231–1237, 2013.

[33] A. Gaidon, Q. Wang, Y. Cabon 및 E. Vig, "다중 객체 추적 분석을 위한 프록시로서의 가상 세계", in Proc. 회의 계산 비교 패턴 인식, 2016, pp. 4340–4349.

[34] D. Eigen, C. Puhrsch 및 R. Fergus, "다중 규모 심층 네트워크를 사용하는 단일 이미지의 깊이 맵 예측", Proc. 고급 신경 정보 프로세스. 시스템, 2014.

[35] Y. You et al., "자율주행에서 3D 물체 감지를 위한 Pseudo-LiDAR 정확한 깊이", Proc. 국제 회의 배우다. 표현, 2020.