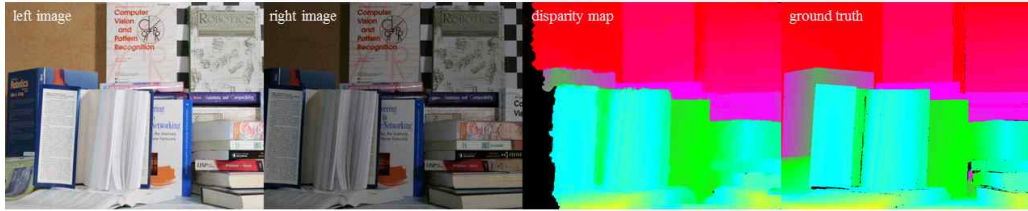
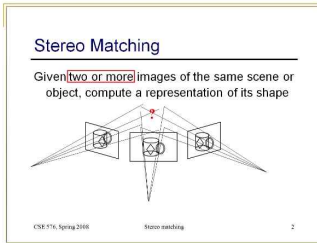


# GA-Net: Guided Aggregation Net for End-to-end Stereo Matching

## 1. Stereo Matching 이란? (as known as Disparity Estimation)

→ 각각 다른 시선에서 픽셀들을 매칭시키는 것.



## 2. 이 논문이 기여한 점

- 1) 기존의 GC-Net의 경우, 3D stereo matching 연구에서 3D Conv를 사용
- 2) 이 3D conv는 고비용/높은 복잡도를 요구 → 연산 시 소요되는 시간이 길다.
- 3) 기존의 SGA(Semi-Global Aggregation)을 다른 방향으로 개선
- 4) 여기에 추가적으로 LGA(Local Guided Aggregation) 개발
- 5) SGA+LGA는 3D Conv를 대체 가능하고, 1/100 수준의 복잡도를 가진다.
- 6) 15-20fps로 실시간 평가 및 대조 가능

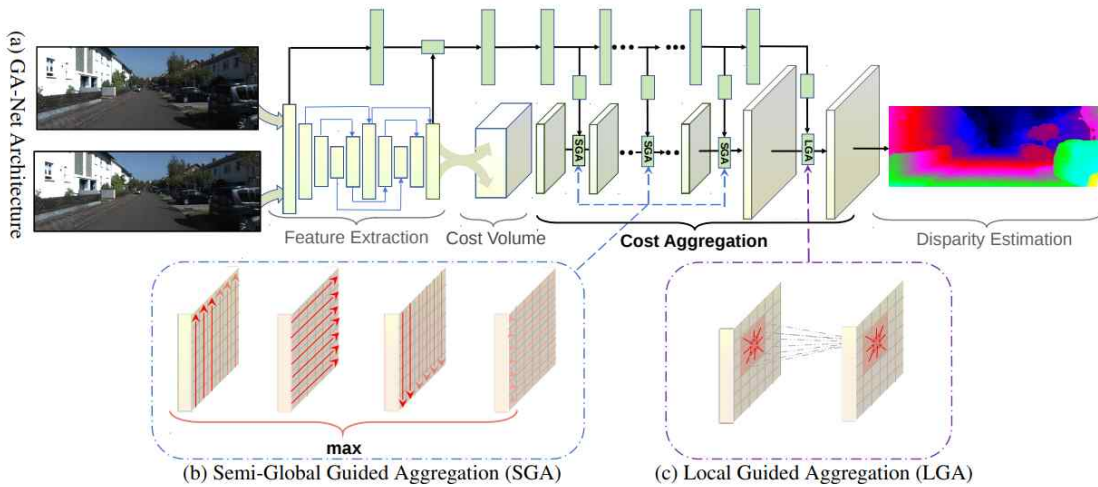
## 3. SGA 개선

- 1) 기존의 SGA는 직접 조정(tuning)하기 힘든 사용자 설정 파라미터가 있음, 이 파라미터는 네트워크 학습 도중에 가변성이 매우 높다 → unstable 함.
- 2) SGM으로 픽셀 매칭 시 너무 다른 환경에 영향을 많이 받음.
- 3) 그래서 SGA를 세 방향으로 개선함.
  - 3-1) 사용자 설정 파라미터를 학습 가능하게 변경(learnable)
  - 3-2) 첫 번째/외부 최소 선택을 가중치의 합으로 대체함, 기존 연구에서 strides는 max-pooling layer 대체해서 정확도 손실이 없음을 밝혔음.
  - 3-3) 내부/두 번째 최소 선택(selection)을 최대 선택으로 변경 → 본 모델의 학습하고자 하는 타겟은 매칭 코스트를 최소화하는 것 대신 ground truth의 예상 점수(confidence score와 비슷한 개념)를 최대화하는 것을 중점으로 보기 때문(다시 말해서, 최대한 잘 맞추는 것에 집중함)

#### 4. LGA

→ 영상 내의 얇은 구조물과 객체의 가장자리(엣지)를 처리하기 위함.

5. GA-Net 구조: 3개의 SGA 레이어와 1개의 LGA 레이어 확인 가능.



#### 6. 실험

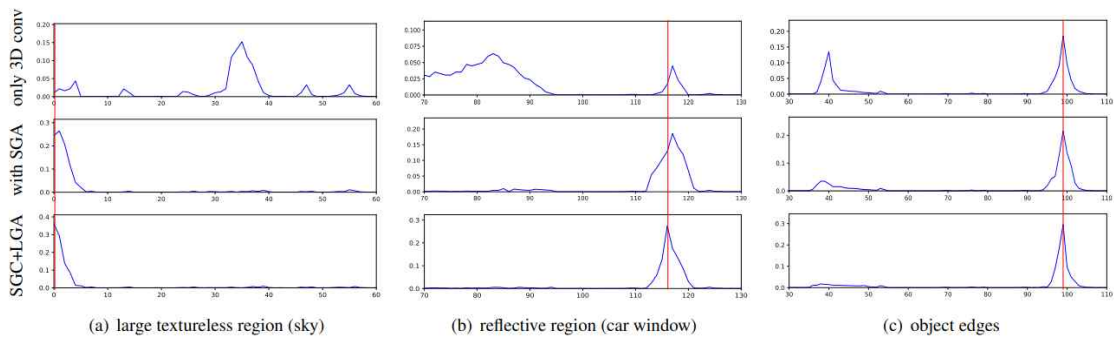
1) 실험 대상은 Scene Flow Dataset과 KITTI Benchmarks 이용.

Table 1: Evaluations of GA-Nets with different settings. Average end point error (EPE) and threshold error rate are used for evaluations.

Feature Extraction Stacked Block	Cost Aggregation			Scene Flow		KITTI 2015 Error Rates (%)
	Densely Concatenate	SGA Layer	LGA Layer	EPE Error	Error Rates (%)	
✓				1.26	13.4	3.39
✓				1.19	13.0	3.31
✓				1.14	12.5	3.25
✓		+1		1.05	11.7	3.09
✓		+2		0.97	11.0	2.96
✓		+3		0.90	10.5	2.85
✓		+4		0.89	10.4	2.83
✓	✓	+3	✓	<b>0.84</b>	<b>9.9</b>	<b>2.71</b>

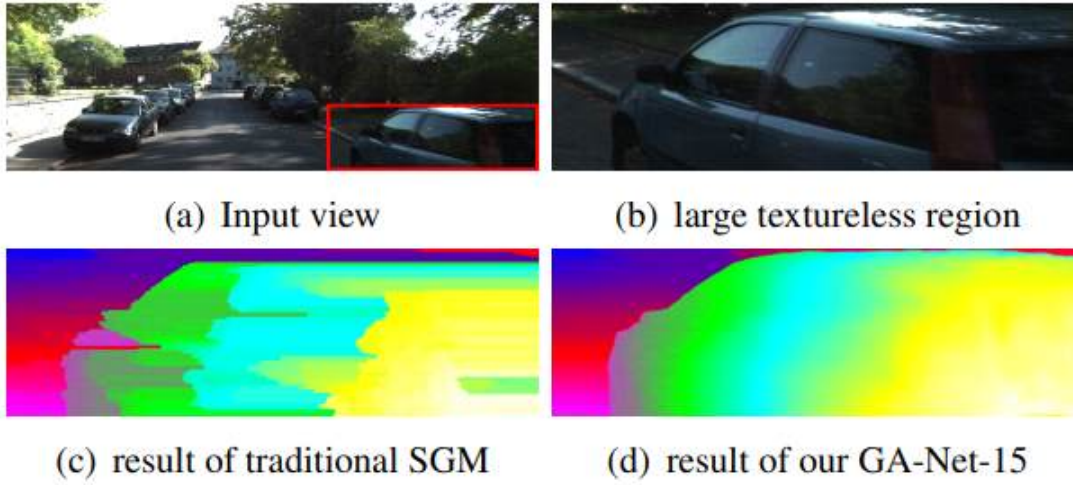
#### 7. SGM과 3D Conv 비교

1) 빨간 선은 Ground Truth와 매칭 지점



결론 → SGA+LGA 방식이 noise를 잘 잡음.(The SGA layers successfully suppress these noise)

## 8. Traditional's SGA vs GA-Net's SGA



## 9. 결과

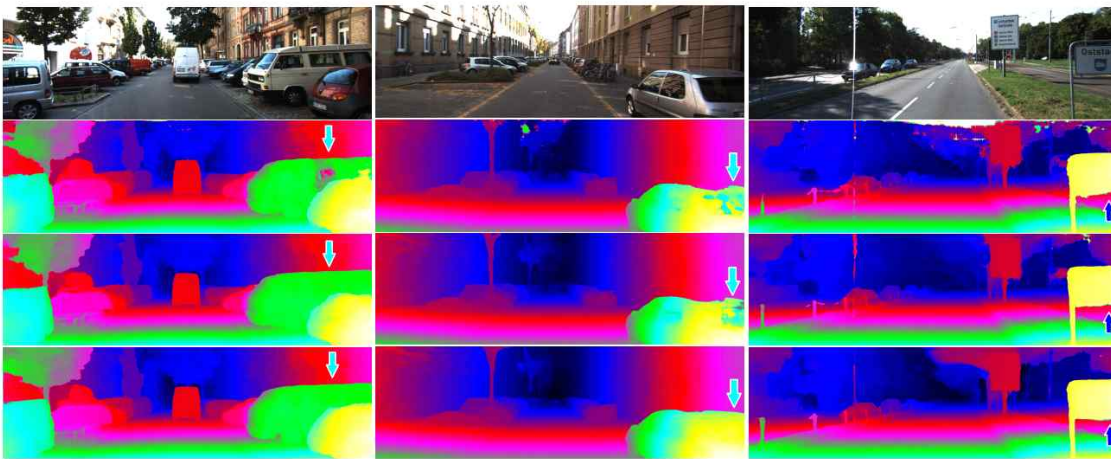


Figure 6: Results visualization and comparisons. *First row*: input image. *Second row*: Results of GC-Net [13]. *Third row*: Results of PSMNet [3]. *Last row*: Results of our GA-Net. Significant improvements are pointed out by blue arrows. The guided aggregations can effectively aggregate the disparity information to the large textureless regions (e.g. the cars and the windows) and give precise estimations. It can also aggregate the object knowledge and preserve the depth structure very well (last column).

첫 번째 줄: 입력 영상

두 번째 줄: GC-Net 결과

세 번째 줄: PSMNet 결과

네 번째 줄: GA-Net 결과

봐야할 부분 : 영상 속 파란 화살표 부분