Bi3D: Stereo Depth Estimation via Binary Classifications

Youngju Yoo

Introduction

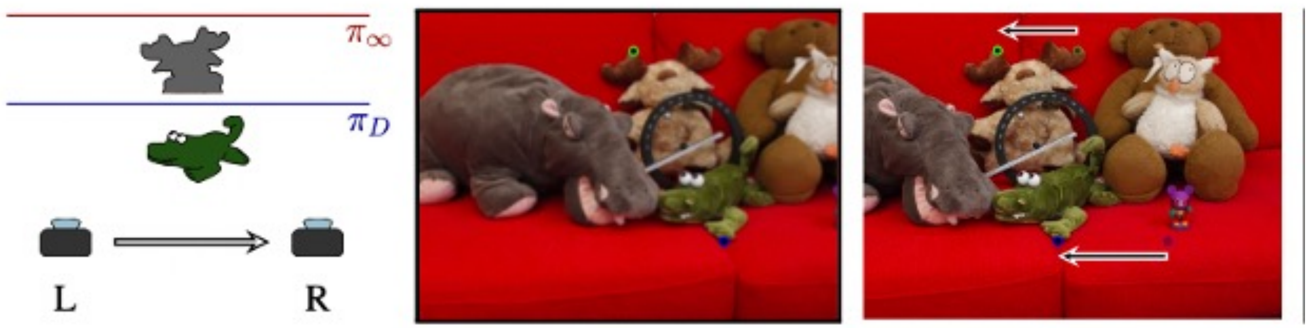
- Bi3D: Estimates depth via a series of **binary classifications**
- Autonomous navigation do not not always require centimeter-accurate depth
- The required accuracy and the range of interest varies with the task.
- Highway driving requires longer range but can deal with a more coarsely quantized depth than parallel parking.
- Rather than testing if objects are at a particular depth D, it classifies them as being closer or farther than D

Existing method

- State-of-the-art stereo algorithms estimate depth with good accuracy, while maintaining real-time execution
- Existing methods :
  - compute depth by testing a number of candidate disparities, and by selecting the most likely under some cost function
  - If an object is outside the range spanned by the candidate disparities, existing methods map it to the candidate disparity with the lowest cost
- Do not offer the flexibility to adapt the depth quantization levels to determine if an object is within a certain distance, or to focus only on a particular range of the scene, without estimating the full depth first.

Method

- Existing algorithm : The direction of the disparity vectors is the same for all the objects in the scene.



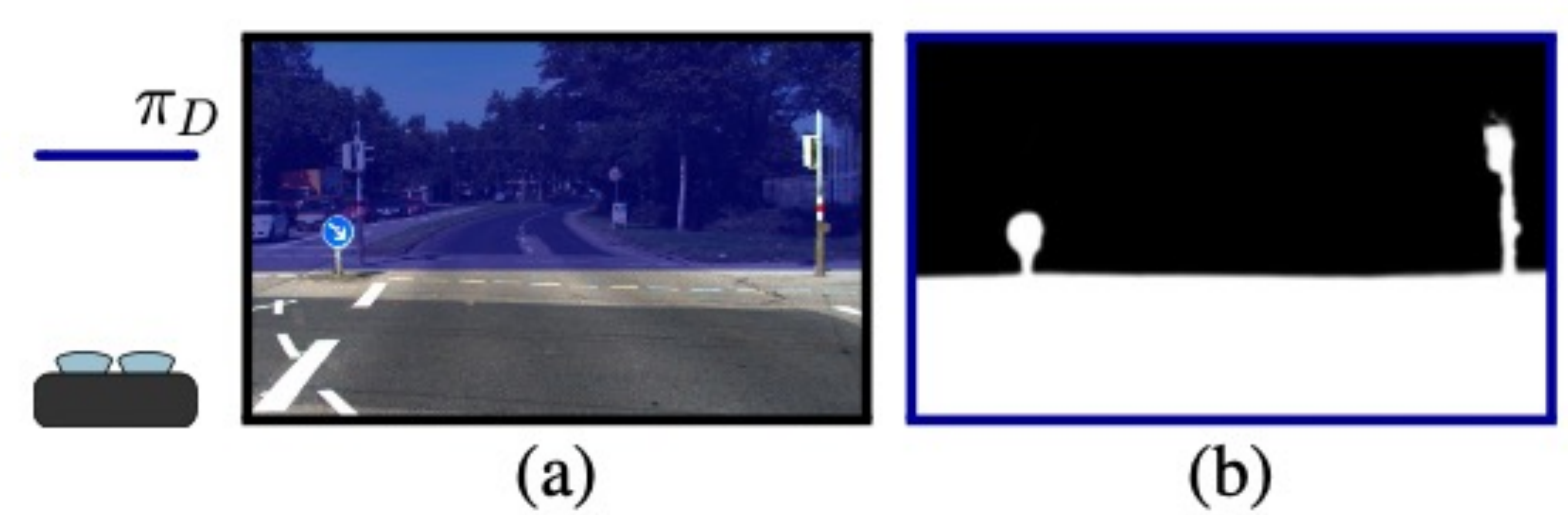
(a) (b) L / R (Animated) (c) R

- After warping image : the disparity of objects on opposite sides of the plane points to opposite directions.

(d) L / Warped R (Animated) (e) Warped R

Bi3D

- Bi3D offers flexible control over the trade-off between latency and depth quantization
- Bi3D can classify objects as being closer or farther than a given distance in a few milliseconds : **Binary depth estimation**



Bi3D

- When a larger time budget is available, Bi3D can compute depth with varying quantization and execution time growing linearly with the number of levels : **quantized depth**

(c)

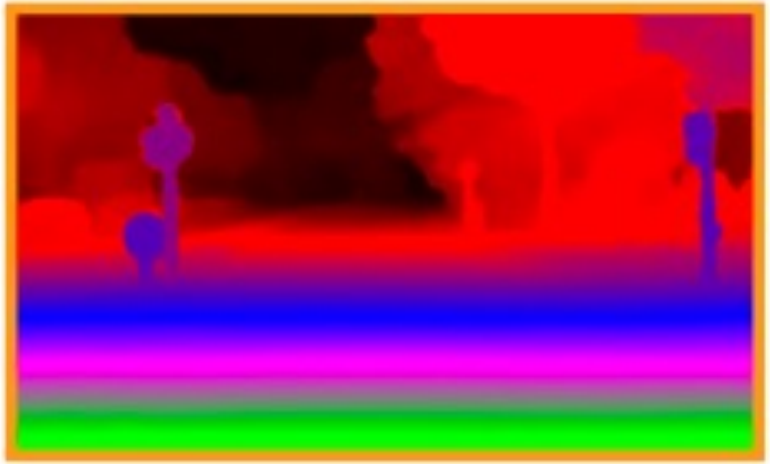
Bi3D

- It can estimate continuous depth in a range while identifying objects outside of this range as closer or farther than the extremes of the range : **selective depth estimation**

(d)

Bi3D

- Bi3D can estimate the full depth with quality comparable with the state-of-the-art.



(e)

Result

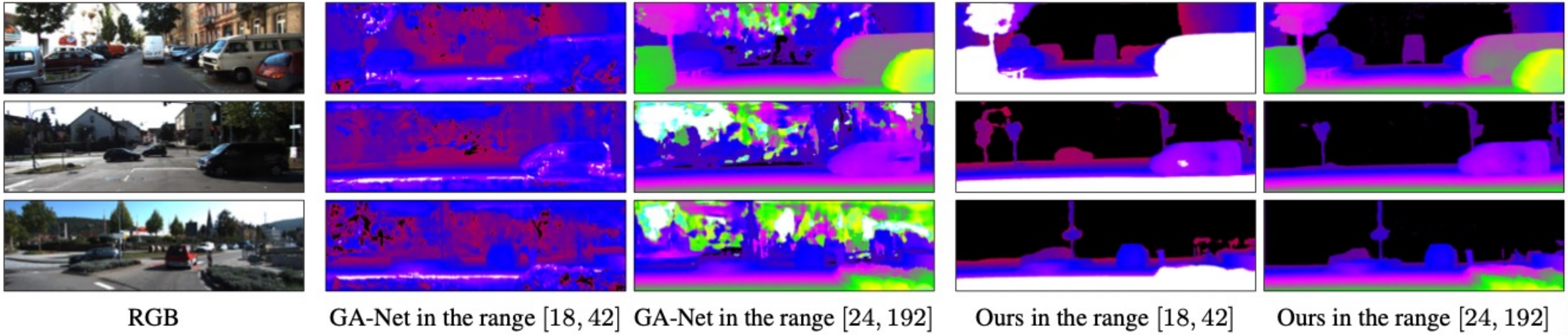


Figure 8: Example of selective depth estimation for three different scenes and with two different disparity ranges, indicated in the labels. Our method predicts the correct depth in the range of interest. White and black pixels are those detected as being in front or behind the selected range, respectively.

HITNet: Hierarchical Iterative Tile Refinement Network for Real-time Stereo Matching

Youngju Yoo

Introduction

- Recent research : focused on developing accurate but computationally expensive deep learning approaches.
- Multiple recent stereo matching methods have increased the efficiency without explicitly operating on a full 3D cost volume.
- But they do not directly generalize to passive stereo due to the lack of using a powerful machine learning system.
- **HITNet** : overcomes the computational disadvantages of operating on a 3D volume

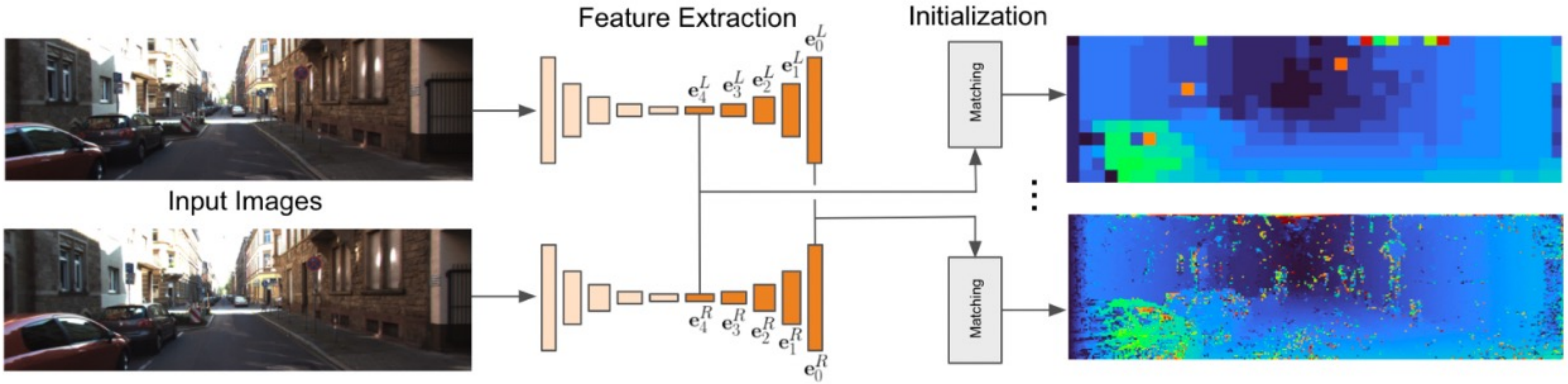
Method

- A **fast multi-resolution initialization step** that computes high resolution matches using learned features.
- An efficient **2D disparity propagation** that makes use of slanted support windows with learned descriptors.

Method

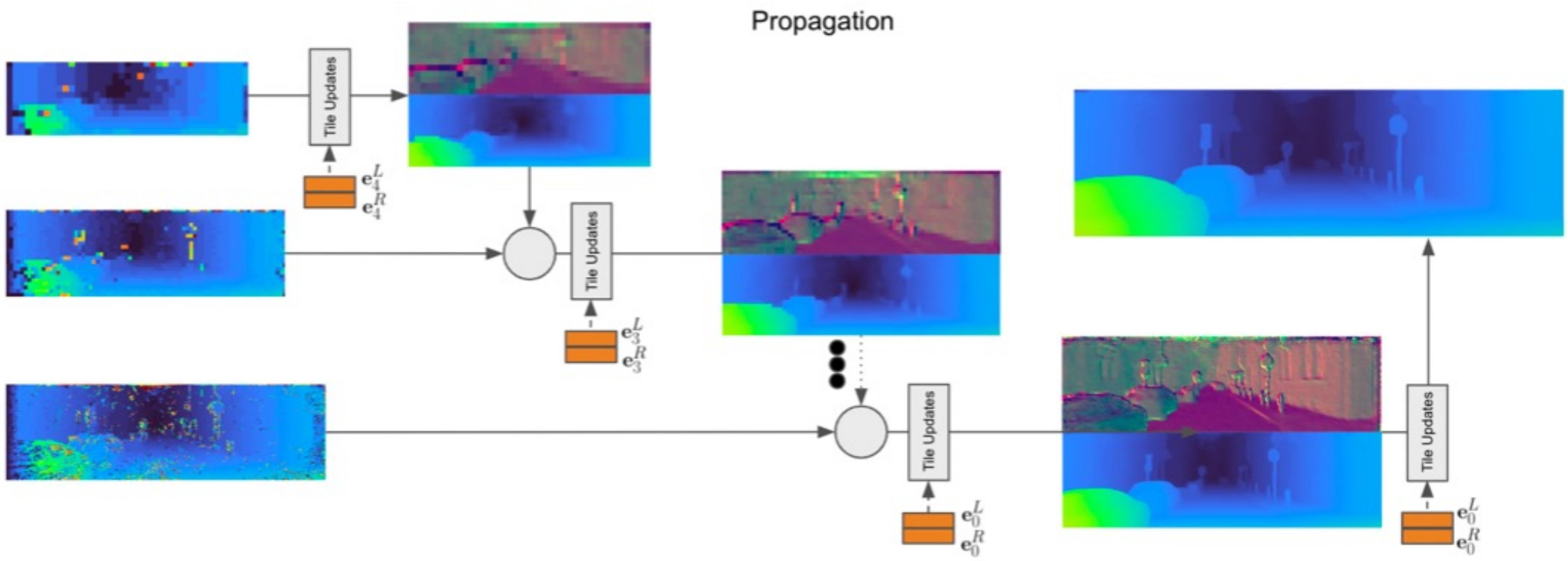
- Initialize disparity maps as fronto parallel tiles at multiple resolutions.
- A matcher evaluates multiple hypotheses and selects the one with the lowest l1 distance between left and right view feature.

$$\mathbf{h} = [\underbrace{d, d_x, d_y,}_{\text{plane}} \underbrace{\mathbf{p}}_{\text{descriptor}}]$$



Method

- The propagation step takes tile hypotheses as input and outputs refined tile hypotheses based on spatial propagation of information and fusion of information. -> hypotheses update



Limitation

- It needs to be trained on a dataset with ground truth depth.
- Paper planning to investigate self-supervised methods and self-distillation methods to further increase the accuracy and decrease the amount of training data.

STTR : Revisiting Stereo Depth Estimation From a Sequence-to-Sequence Perspective with Transformers

Youngju Yoo

Introduction

- A limited disparity range : many of the best performing approaches are con-strained to a manually pre-specified disparity range. It is not flexible to properties of the physical scene and the camera setup. -> **relaxes the limitation of a fixed disparity range**
- Prior algorithms may not always be valid in occluded region. **-> identifies occluded regions and provides confidence estimates** (to enable rejection of occluded and low-confidence estimates)
- Impose uniqueness constraints during the matching process

STTR

- Capture long-range associations between pixels
- Use **Self-attention and cross-attention** which update the feature representations by considering both image context and position information -> **Use the sequential nature and geometric properties**

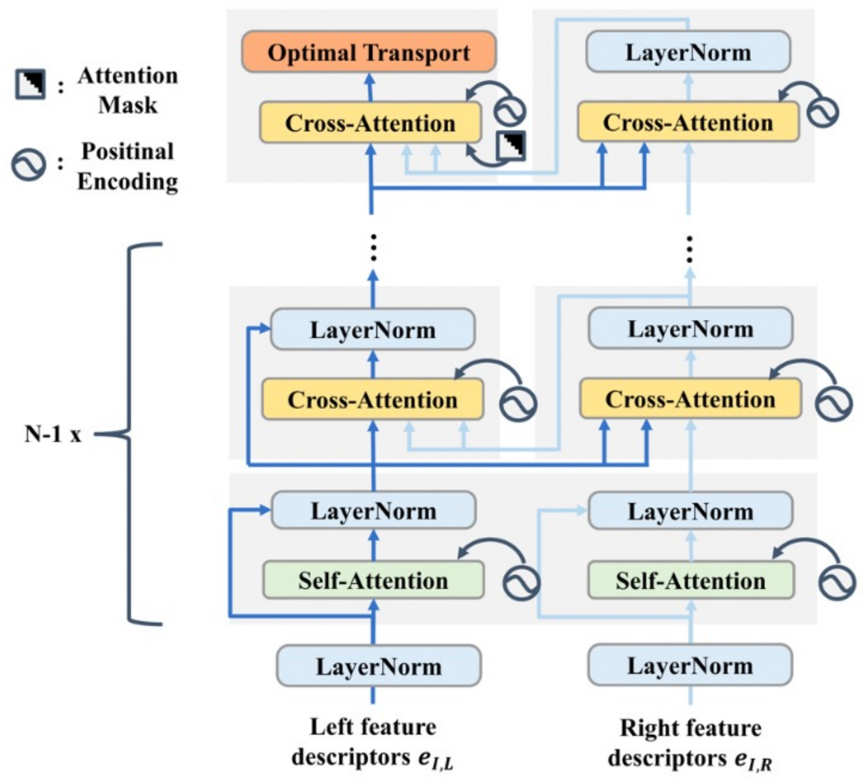


Figure 2. Overview of the Transformer module with alternating self- and cross-attention. Note that in the last cross-attention layer, the optimal transport and attention mask are added.

Normal Assisted Stereo Depth Estimation

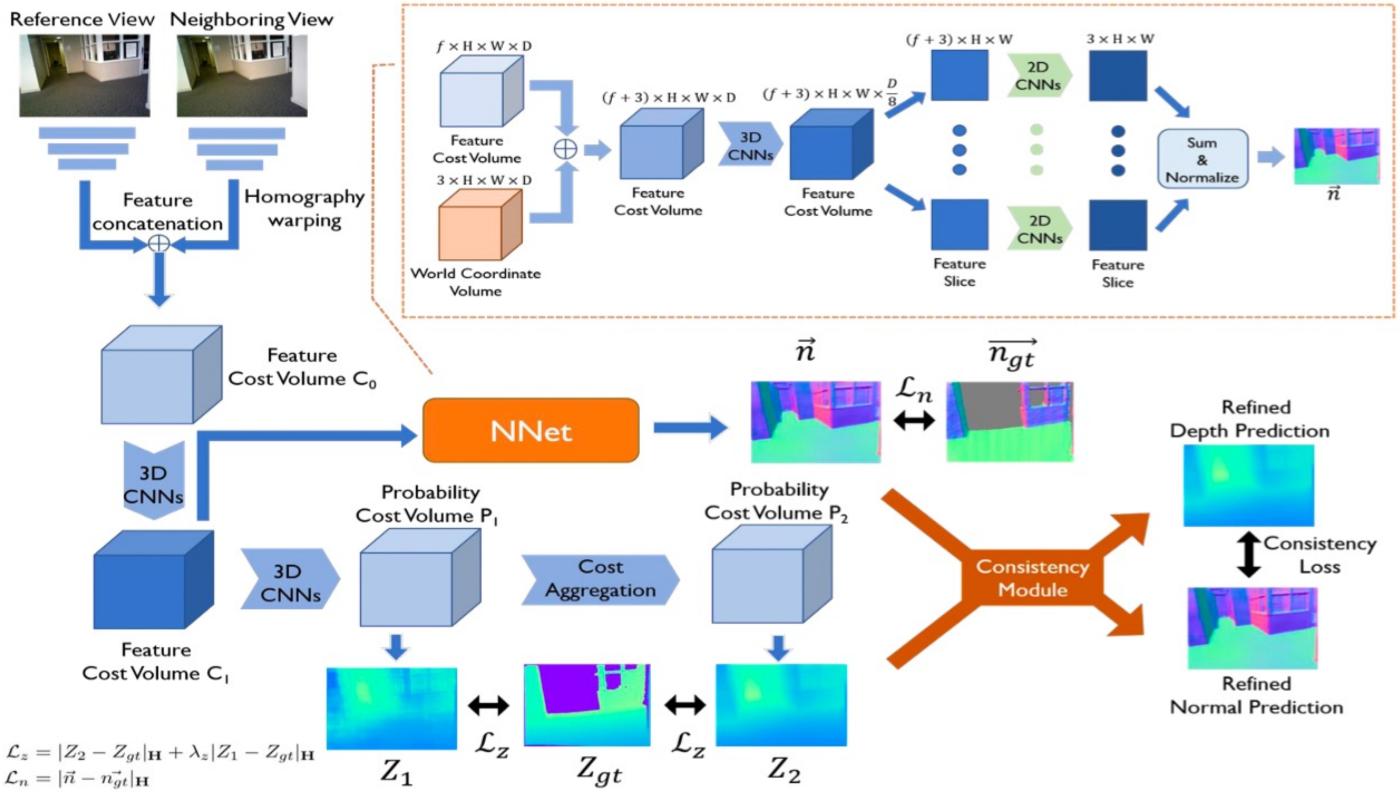
Youngju Yoo

Introduction

- Accurate stereo depth estimation methods have demonstrated competitive performance with limited number of views.
- when building cross-view correspondences is hard, these methods still cannot produce satisfying results.
- The lack of geometric constraints leads to bumpy depth prediction especially in areas with low texture or that are textureless.
- **Couple the learning of a multi-view normal estimation module and a multi-view depth estimation module.**

method

- The first module consists of joint estimation of depth and normal maps from the cost volume built from multi-view image features.
- The subsequent module refines the predicted depth by enforcing consistency between the predicted depth and normal maps using the proposed consistency loss.



Result

- Joint learning can improve both the prediction of normal and depth, and the accuracy & smoothness can be further improved by enforcing the consistency.