ETH zürich

Stero Video Depth Estimation in 2025
Research Proposal

3D Vision Group 18

Hepeng Fan, Qingrui Deng, Tong Su, Yixin Zhou
Supervisor: Haofei Xu
10.03.2025

Monocular Depth Estimation

- ***Depth*** refers to the distance from the **object point to the camera center**
- Two ways of describing depth: **Relative Depth and Metric Depth**

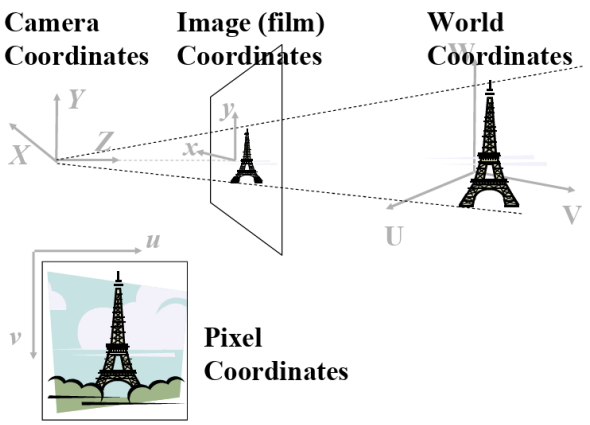
$$\boldsymbol{relative\ depth + scale = metric\ depth}$$

$f($  $) =$ 

single image depth map

1.5	1.8	2.5	2.6
1.1	0.2	0.5	0.3
3.8	5.5	2.1	1.5

relative depth map



$$x = f\frac{X}{Z}$$

$$y = f\frac{Y}{Z}$$

1.5**m**	1.8**m**	2.5**m**	2.6**m**
1.1**m**	0.2**m**	0.5**m**	0.3**m**
3.8**m**	5.5**m**	2.1**m**	1.5**m**

metric depth map

Pipeline for metric depth estimation

raw image

CNN/Transformer/SD...

relative depth

LIDAR

fine-tuning

Stero Video

metric depth

Raw Video

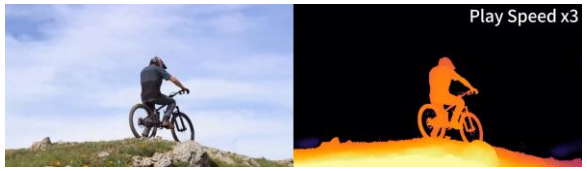
Image Depth Estimation
(Video Stitching)

Consistent Video Depth Estimation

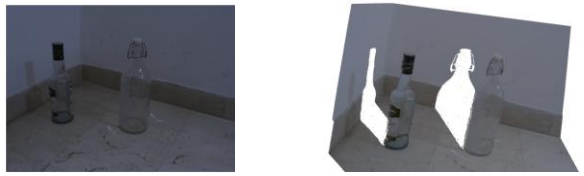
Our Project: Stero **Video Metric** Depth Estimation

Baseline Model

- Video Depth Anything



- FoundationStero



Model Design

Video Depth Anything | FoundationStero

Datasets

KITTI, SceneFlow, RealEstate10k…

metric depth

Model Evaluation

Evaluation Metrics

- Abs.Rel
- End-Point Error
- Temporal Consistency Score

Comparion

SOTA Models: BIDA Stero, Dynamic Stero

Challenges and Potential Solutions

How to improve the temporal consistency of FoundationStero with the help of Video Depth Anything

Solution: Add one temporal-spatial layer to FoundationStero

ETH zürich

Questions & Feedback