JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

Single Image Depth Prediction Made Better: A Multivariate Gaussian Take

Ce Liu, Suryansh Kumar*, Shuhang Gu, Radu Timofte, Luc Van Gool

Poster: THU-AM-083

*corresponding author

Overview

Task. Single-image depth prediction.

Key Point. Given an image with N pixels, fit the conditional distribution of depth map by N-dimensional Gaussian.

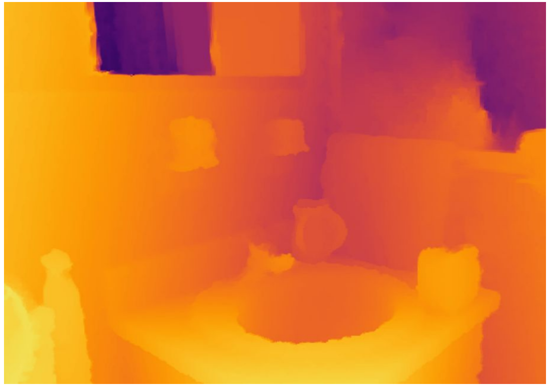
Advantages.

- The likelihood is more general and encapsulates flavors of popular loss functions.
- The formulation could be helpful in broader applications such as uncertainty estimation.

Single Image Depth Prediction (SIDP)

Goal. Predict the depth value for each pixel of input image.

Image

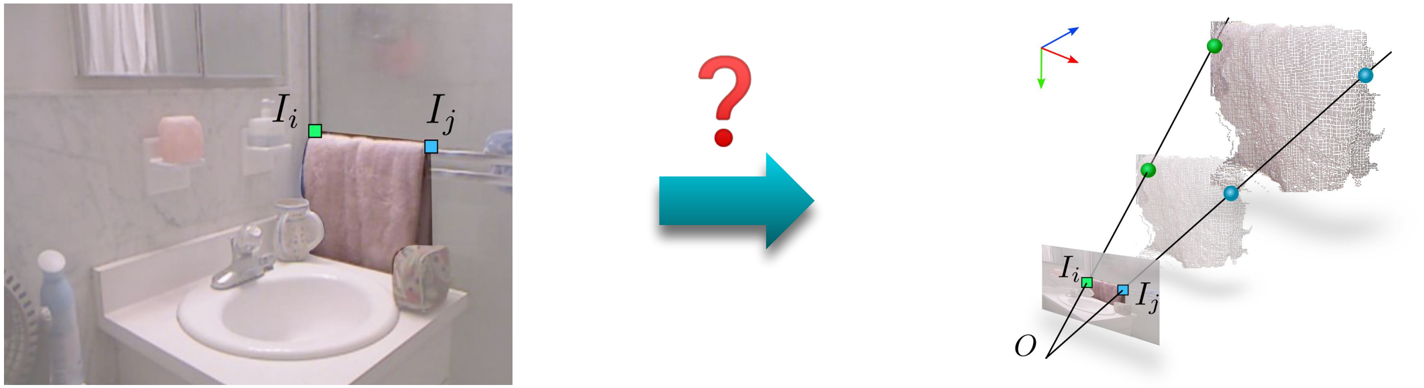


Depth

Applications. VR/AR, novel view synthesis, robotics, …

Scale Ambiguity & Regularity

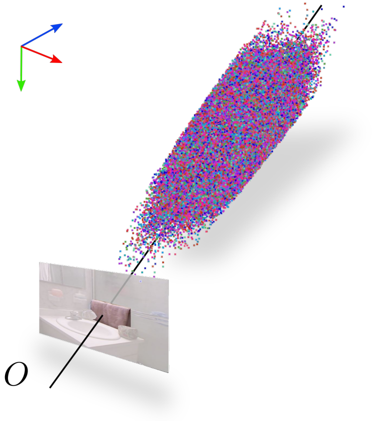
The SIDP problem is **ill-posed.**



Observation. Depth values at nearby pixels often have strong correlation.

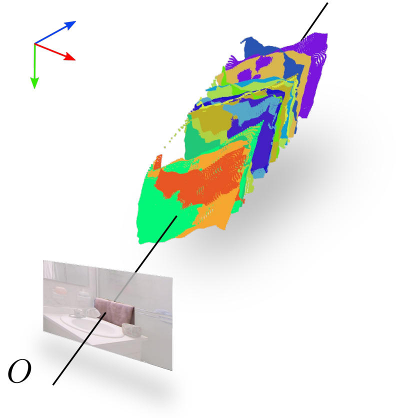
Independent Assumption is Inappropriate

Each depth value follows an independent Gaussian distribution (given the image).



Multivariate Gaussian Distribution

N-pixels follow a *N*-dimensional Gaussian distribution.



Low-Rank Assumption

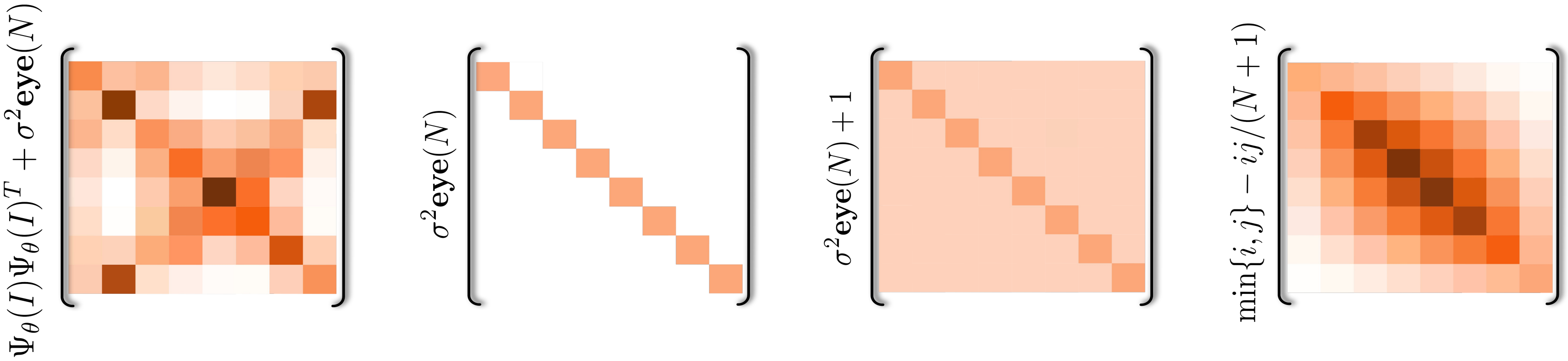
$$\Phi(Z|\theta, I) = \mathcal{N}(\mu_\theta(I), \Sigma_\theta(I, I))$$

$$\Sigma_\theta(I, I) = \Psi_\theta(I)\Psi_\theta(I)^T + \sigma^2 \mathbf{eye}(N)$$

where $\mu_\theta(I) \in \mathbb{R}^{N\times 1}, \Sigma_\theta(I, I) \in \mathbb{R}^{N\times N}, \Psi_\theta(I, I) \in \mathbb{R}^{N\times M}, M \ll N.$

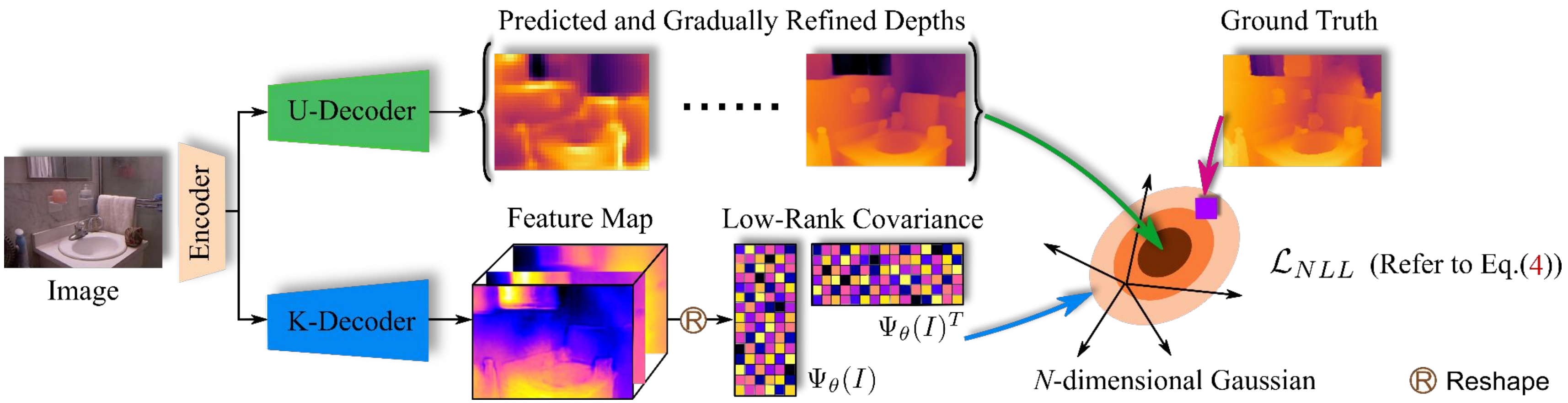
The time complexity reduces from $O(N^3)$ to $O(NM + M^3)$.

Relation to Popular Loss Function



(a) Ours (b) L2 Loss (c) SI Loss (d) Gradient Loss

Network Architecture



© Liu *et al.* **CVPR 2023**

Results. NYU Depth V2

Method	Backbone	SILog ↓	Abs Rel ↓	RMS ↓	δ_1 ↑
DPT-Hybrid	ViT-B	-	0.110	0.357	0.904
AdaBins	EffNet-B5+ ViT-mini	10.570	0.103	0.364	0.903
NeWCRFs	Swin-L	9.102	0.095	0.331	0.922
Ours	Swin-L	**8.323**	**0.087**	**0.311**	**0.933**

Results. KITTI Benchmark

Method	Backbone	SILog ↓	Abs Rel ↓	Sq Rel ↓	iRMS ↓
DORN	ResNet-101	11.80	8.93	2.19	13.22
BTS	DenseNet-161	11.67	9.04	2.21	12.23
NeWCRFs	Swin-L	10.39	8.37	1.83	11.03
Ours	Swin-L	**9.93**	**7.99**	**1.68**	**10.63**

Conclusion

- A formulation with multivariate Gaussian distribution for depth map is introduced.
- The proposed likelihood is more general and encapsulates flavors of popular loss functions.
- The formulation could be helpful in broader applications such as uncertainty estimation.