Tame a Wild Camera:
In-the-Wild Monocular Camera Calibration

Shengjie Zhu, Abhinav Kumar, Masa Hu, and Xiaoming Liu

Tame a Wild Camera:
In-the-Wild Monocular Camera Calibration

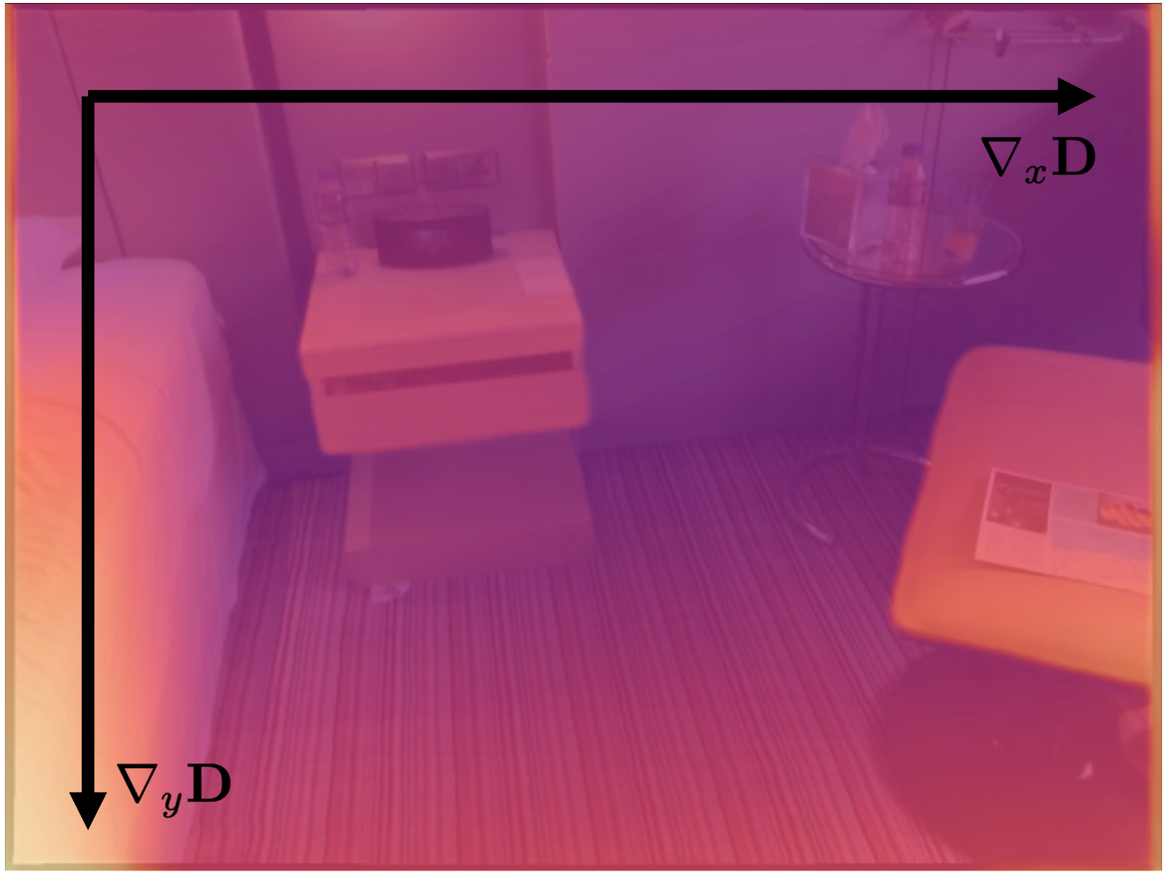
- Perform 4 DoF Monocular Camera Calibration for In-the-Wild Images

Monocular 3D sensing for AR / VR Applications

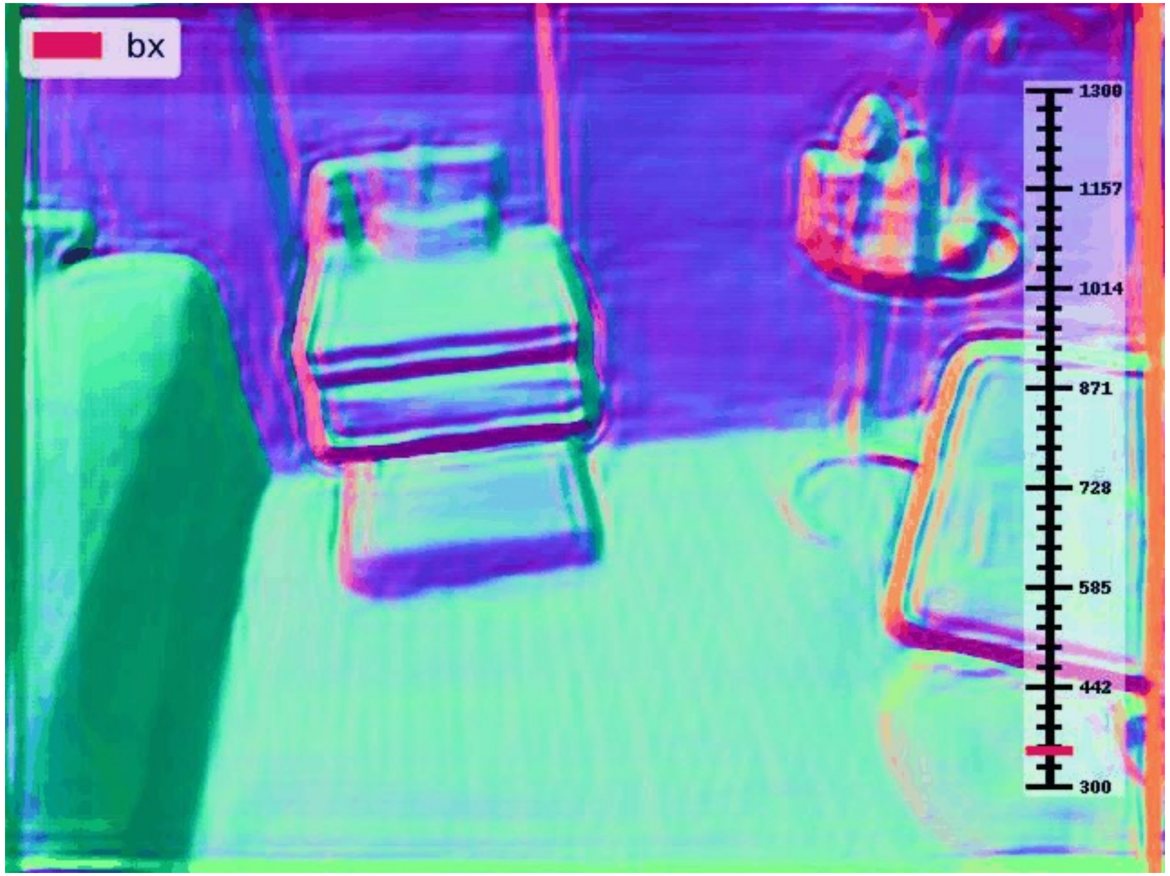
- In-the-wild Monocular 3D Sensing has drawn Rapid Attention
- Question: For an in-the-wild image, how do we know the intrinsic to do the 2D-3D or 3D-2D projection?

Intrinsic can be estimated from Monocular Priors

- **Monocular Depth**



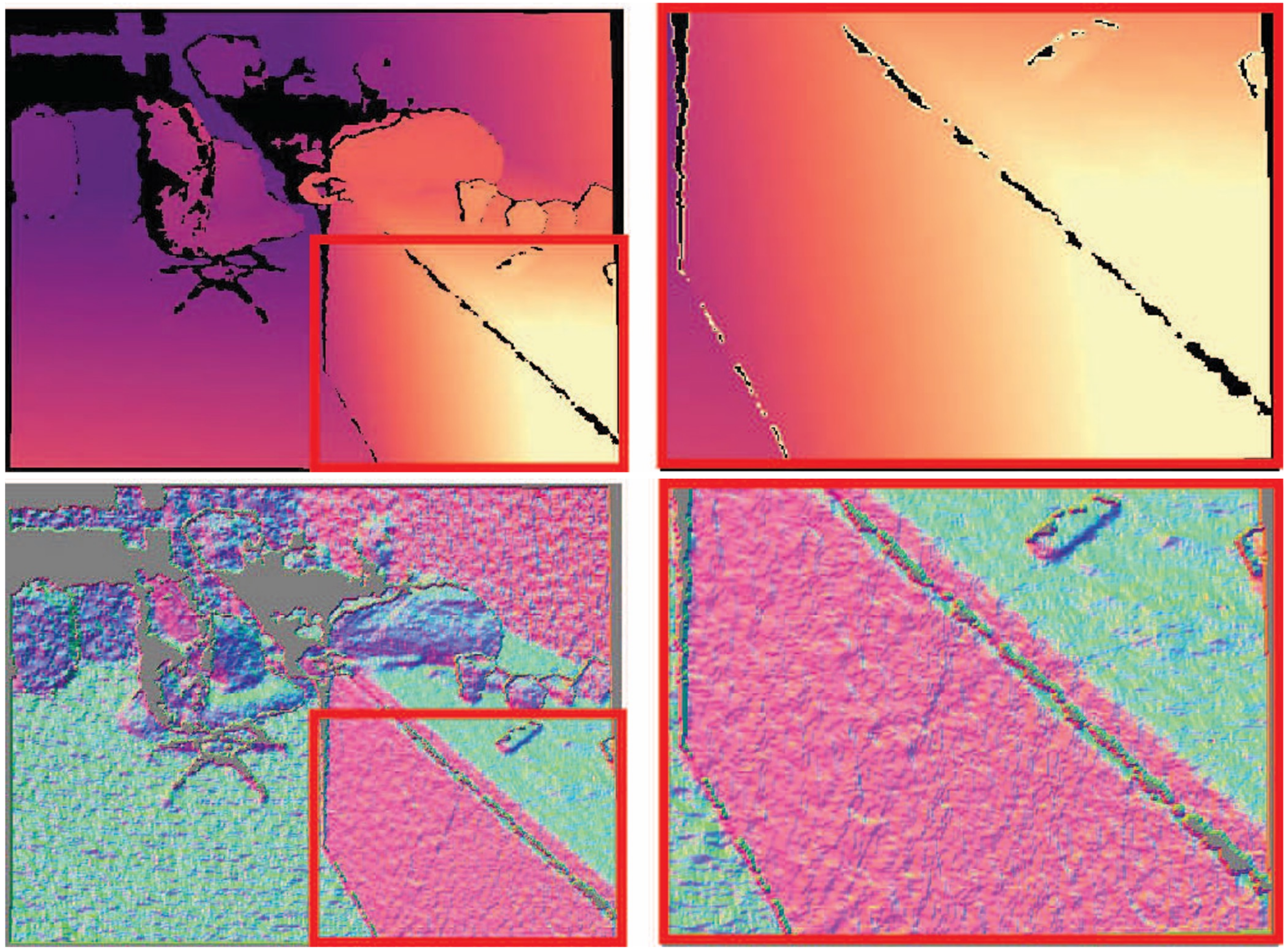
- **Surface Normal**



- Correct intrinsic conclude correct surface normal from monocular depth estimator

Issues to the Depth-Normal Minimal Solver

- Derivative requires an infeasible low-bias and low-variance monocular depth
- Even Groundtruth Depthmap leads to noisy surface normal



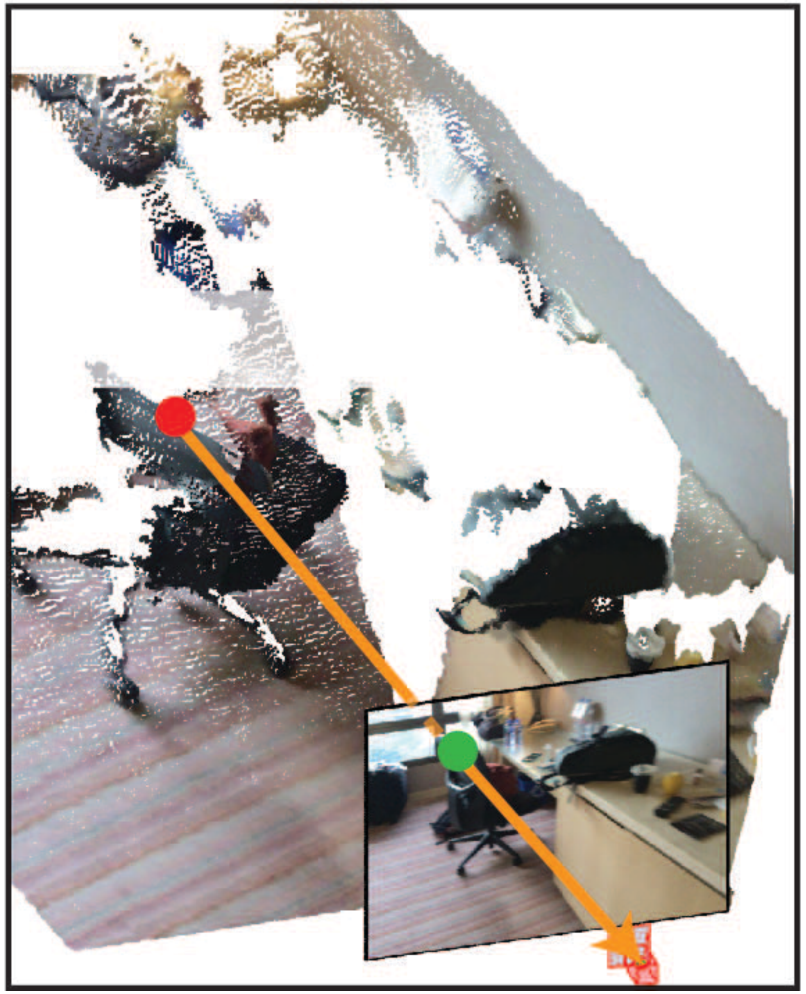
- Derivative leads to noisy solution

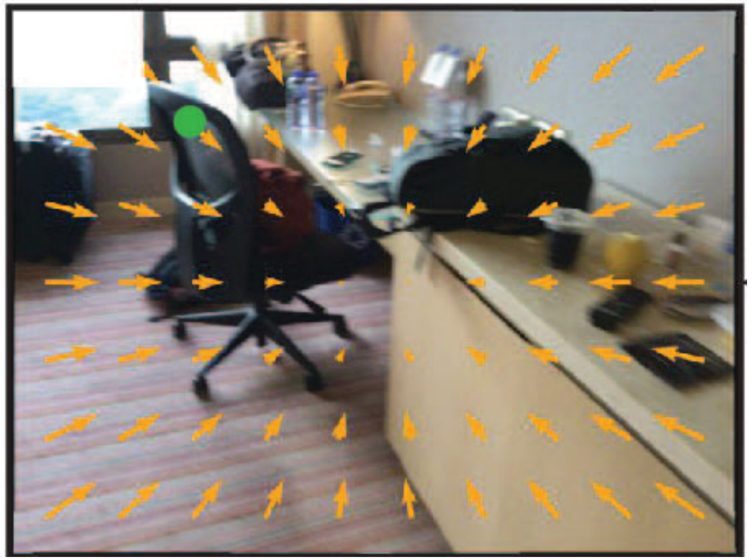
$$\mathbf{n}^\intercal \nabla_x (d \cdot \mathbf{v}) = 0, \quad \mathbf{n}^\intercal \nabla_y (d \cdot \mathbf{v}) = 0.$$

$$n_1 \nabla_x \left(d \cdot \frac{x - b_x}{f_x}\right) + n_2 \frac{y - b_y}{f_y} \nabla_x(d) + n_3 \nabla_x(d) = 0$$

Alternatives to Depth-Normal Minimal Solver

- We propose the Incidence Field, defined as the incidence ray between a 3D point and a 2D pixel





- Linear Constraint raised from Incidence Field

$$\frac{x - b_x}{f_x} = v_x, \quad \frac{y - b_y}{f_y} = v_y$$

- Incidence Field is a direct parametrization of intrinsic

- Definition of Incidence Field:

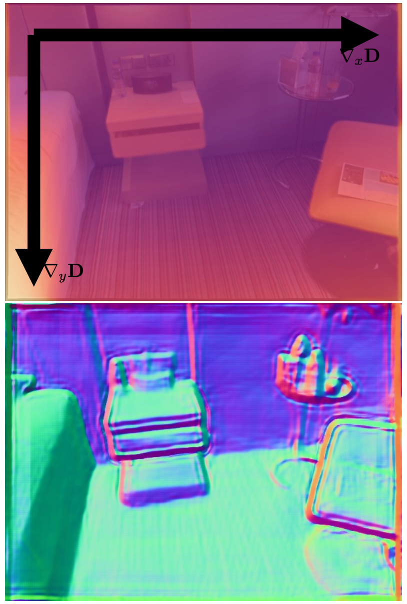
$$\mathbf{P} = d \cdot \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} = d \cdot \mathbf{K}^{-1} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = d \cdot \begin{bmatrix} \frac{x-b_x}{f_x} \\ \frac{y-b_y}{f_y} \\ 1 \end{bmatrix} = d \cdot \begin{bmatrix} v_x \\ v_y \\ 1 \end{bmatrix} = d \cdot \mathbf{v}$$

Is Incidence Field a Learnable Monocular 3D Prior?

- Incidence Field seems to be irrelevant to the content of the image

Relate Incidence Field to Monocular 3D Priors

- **Monocular Depth & Surface Normal**



- Intrinsic:

$$\mathbf{K} = \begin{bmatrix} f_x & 0 & b_x \\ 0 & f_y & b_y \\ 0 & 0 & 1 \end{bmatrix}$$

- Intrinsic and Incidence

$$\mathbf{P} = d \cdot \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} = d \cdot \mathbf{K}^{-1} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = d \cdot \begin{bmatrix} \frac{x-b_x}{f_x} \\ \frac{y-b_y}{f_y} \\ 1 \end{bmatrix} = d \cdot \begin{bmatrix} v_x \\ v_y \\ 1 \end{bmatrix} = d \cdot \mathbf{v}$$

3D Point — Intrinsic — Incidence Vector

- From Relationship between Monocular Depth and Surface Normal:

$$\mathbf{n}^\intercal \nabla_x (d \cdot \mathbf{v}) = 0 \qquad \mathbf{n}^\intercal \nabla_y (d \cdot \mathbf{v}) = 0$$

- This leads to a linear constraints includes intrinsic

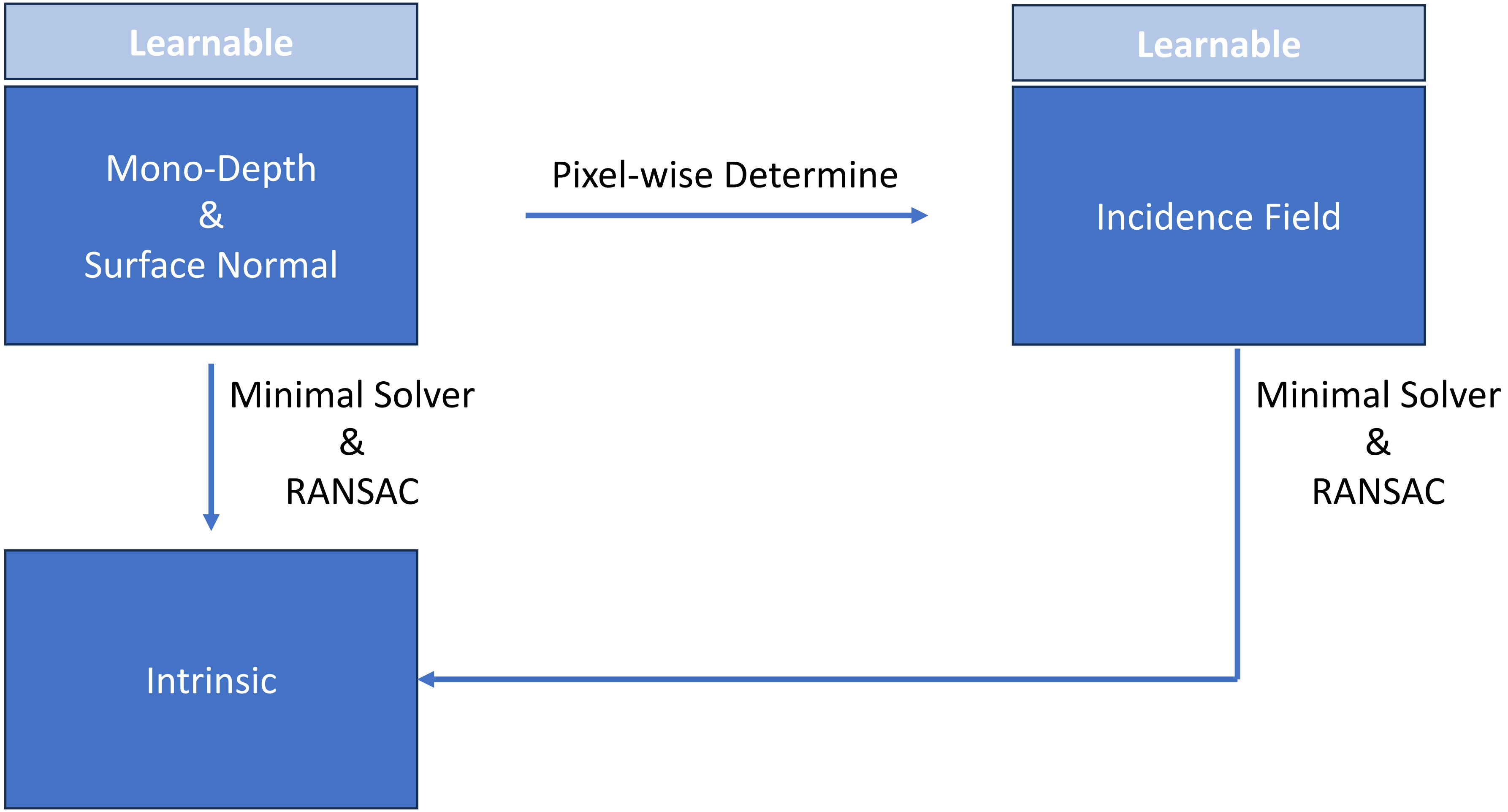
$$n_1 \nabla_x (d \cdot \frac{x - b_x}{f_x}) + n_2 \frac{y - b_y}{f_y} \nabla_x (d) + n_3 \nabla_x (d) = 0 \qquad n_1 \frac{x - b_x}{f_x} \nabla_y (d) + n_2 \nabla_y (d \cdot \frac{y - b_y}{f_y}) + n_3 \nabla_y (d) = 0$$

- Replace to incidence vector

$$n_1 \nabla_x (d) v_x + n_1 (\frac{d}{f_x}) + n_2 v_y \nabla_x (d) + n_3 \nabla_x (d) = 0 \qquad n_1 \nabla_y (d) v_x + n_2 (\frac{d}{f_y}) + n_2 v_y \nabla_y (d) + n_3 \nabla_y (d) = 0$$

- DoF-2 Incidence vector **v** is pixel-wise determined by monocular depth and surface normal

Intrinsic can be calibrated from Monocular Image



Applications

- Dolly-Zoom Video

Applications

- Improve Omni3D (CVPR’23) in-the-wild 3D Detection Results

Applications

Restored $\mathbf{I}_1$ | Cropped $\mathbf{I}_1$ | Original Image $\mathbf{I}$ | Cropped $\mathbf{I}_2$ | Restored $\mathbf{I}_2$

- Detect Image Spatial Manipulation
- Improve Press Image Genuinity

End