Introduction to Depth Estimation

Pre-Machine Learning Era

Chinchuthakun Worameth

Department of Transdisciplinary Science and Engineering
Tokyo Institute of Technology

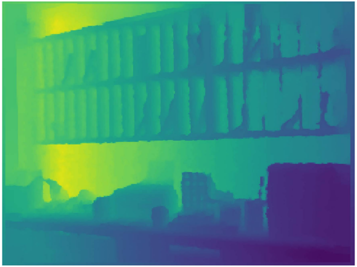
May 27, 2021

Table of Contents

What is depth estimation?

A task of predicting depth information/generating a corresponding **depth map** from images.

Raw image [1]



Depth map [1]

What is depth estimation? (con't)

Approaches can be classified into:

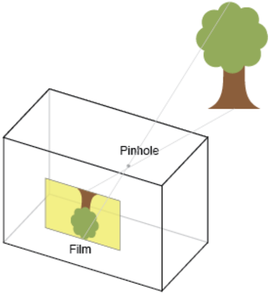
- **Active**: Emit waves to the scene and measure the time taken by it, e.g. **Time of flight (ToF)**.

$$d = \frac{ct}{2}$$

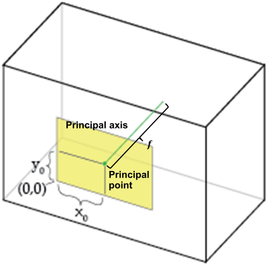
- **Passive**: Measure distance by using image(s). Trading off between accuracy and processing time.
  - **Monocular**: Single image or video sequence.
  - **Stereo**: 2 images
  - **Multiview**: > 2 images

Pinhole camera model

- Project points in **real world coordinate** to **image coordinate**
- Can be defined by intrinsic parameters: **focal length** f, **principal point offset** (x_0, y_0), and **axis skew** s.



Overview [2]



Intrinsic parameters [2]

Pinhole camera model (con't)

Intrinsic/Calibration matrix

$$x = KX, K = \begin{bmatrix} f & s & x_0 \\ 0 & af & y_0 \\ 0 & 0 & 1 \end{bmatrix}$$

Assume $a = 1$ and $s = 0$, by defining image coordinate as below, we have

$$x = f\frac{X}{Z} \text{ and } y = f\frac{Y}{Z}$$

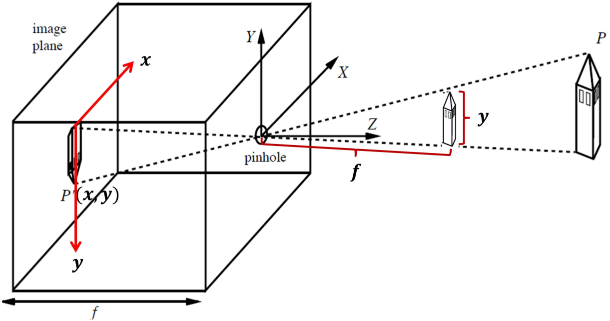
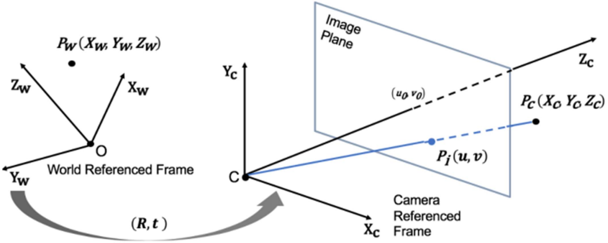


Image from [3]

Pinhole camera model (con't)

We also need **Extrinsic matrix** $[R|t]$ because we need to consider the position of camera in the real world. Therefore, we can transform real world coordinate P to image coordinate p by using **camera matrix** P

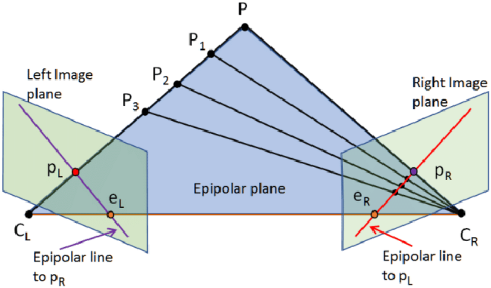
$$x = PX, P = K[R|t]$$



Epipolar Geometry

Geometry relationship of a scene from two points of view, shifted by $[R|t]$

- **Epipolar plane**
- **Base line**: $\overline{C_L C_R}$
- **Epipole**: e_L and e_R
- **Epipolar line**: $\overline{p_L e_L}$ and $\overline{p_R e_R}$



Epipolar constraints: Corresponding points (sometimes called **conjugate pair pixels**) must be observed on the same epipolar line.

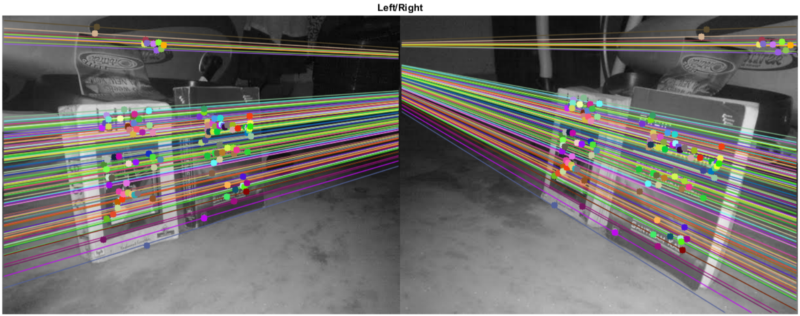


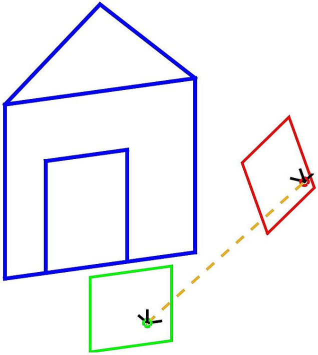
Image from [4]

How to get depth map?

1. **Camera calibration**: determining camera matrix for each camera
   - **Zhang's algorithm** [5] can recover parameters by using $>=2$ images of planar calibration objects, e.g. chess board, in different orientations.
2. **Image rectification**: recovering a simpler and more linear image from raw image
3. **Image matching**: generating **disparity map**
4. **Depth estimation**: calculating depth map from disparity map

Image rectification

One simple method from [6] is shown below.



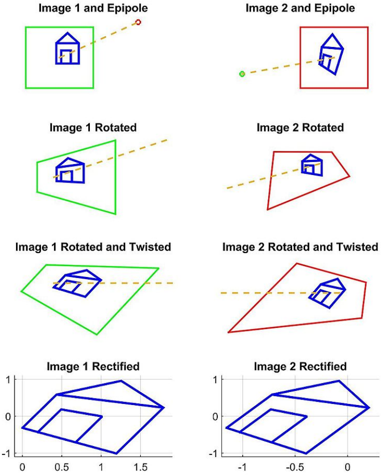
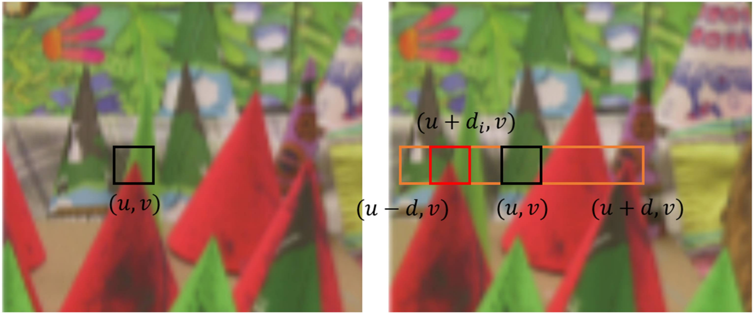


Image matching: Block matching

One of the most intuitive method is **Block matching**. We call d a **disparity** of image patch centered at (u, v).



How to compare similarity between 2 image patches?

Image matching: Block matching (con't)

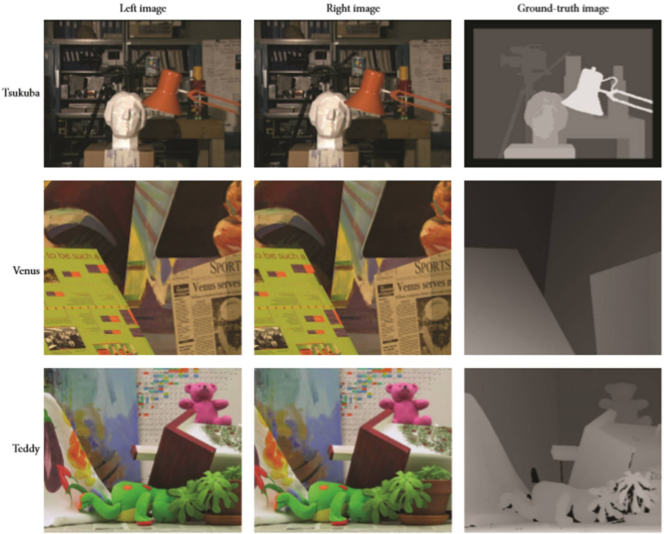
- **Correlation**: Normalized Cross-Correlation (NCC)

$$C_{NCC}(d) = \frac{\sum_{u,v \in W_m(x,y)} (I_L(u,v) - \bar{I}_L) \cdot (I_L(u-d,v) - \bar{I}_R)}{\sqrt{\sum_{u,v \in W_m(x,y)} (I_L(u,v) - \bar{I}_L)^2 \cdot (I_L(u-d,v) - \bar{I}_R)^2}}$$

- **Intensity**: SAD (L_1-norm), SSD (L_2-norm)
- **Rank**: Census transform
  - Encode an image patch into a string
  - Calculate similarity by **Hamming distance** (XOR)
  - More robust against outliner and less dependence on absolute intensity

$$\begin{bmatrix} 7 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 2 & 9 \end{bmatrix} \rightarrow 01110010$$

Disparity map



Depth from disparity

$x_L = f\frac{X}{Z}$ and $x_R = f\frac{X+T_x}{Z}$ implies

$$Z = \frac{fT_x}{d}$$

T_x is base line and d is disparity

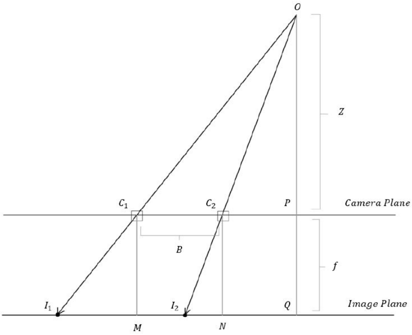


Image from [1]

References I

Chawit Chaijirawiwat.
Monocular Depth Estimation via Transfer Learning and Multi-Task Learning with Semantic Segmentation.
Bachelor's thesis, Tokyo Institute of Technology, Tokyo, July 2019.

The Perspective Camera - An Interactive Tour ←.

CS280: Computer Vision.

Epipolar Geometry, November 2017.

Z. Zhang.
A flexible new technique for camera calibration.
IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(11):1330–1334, November 2000.

References II

Andrea Fusiello, Emanuele Trucco, and Alessandro Verri.
A compact algorithm for rectification of stereo pairs.
Machine Vision and Applications, 12(1):16–22, July 2000.

Richard Szeliski.
Computer Vision: Algorithms and Applications.
2 edition, March 2021.

Pablo Revuelta Sanz, Belén Ruiz Mezcua, and José M. Sánchez Pena.
Depth Estimation - An Introduction.
IntechOpen, July 2012.

Richard Hartley and Andrew Zisserman.
Multiple view geometry in computer vision.
Cambridge University Press, Cambridge, UK ; New York, 2nd ed edition, 2003.