lect 15 Binocular Stereo

Given a pair of (2 or more images) calibrated image, find the depth of each pixel.

↘ we know the transformation.

⇒ Reason: Camera is perspective projection

Clues:

objects far from camera moves slow and looks small

1. Basic Stereo matching alg.

① for each pixel, find the epipolar line in another img.

② check all pixels in epipolar line, find the best match

⋮

③ triangulate the matches to find out depths.

(find a corresponding 3D point that could produce this patten)

For simplest case:

epipolar line is scanline, two cameras are parallel

↑ horizontal line.

The epipolar constraint is

$$x'^T E x = 0, \quad E = [t_\times] R, \quad R = I, \quad t = (T, 0, 0) \Rightarrow E = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -T \\ 0 & T & 0 \end{bmatrix}$$

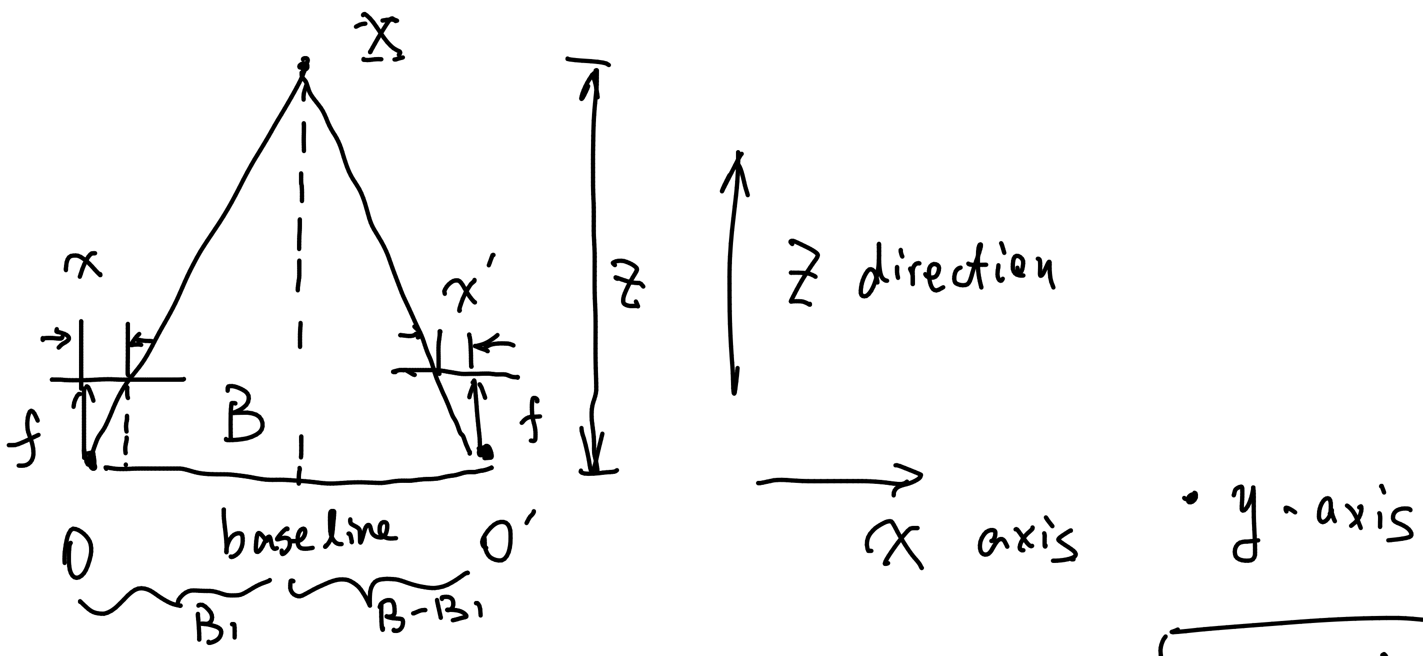
↑ only x-axis

So for point-pair $\begin{bmatrix} u \\ v \\ 1 \end{bmatrix}$ $\begin{bmatrix} u' \\ v' \\ 1 \end{bmatrix}$

$$\hookrightarrow [u', v', 1] \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -T \\ 0 & T & 0 \end{bmatrix} \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = 0$$

$$\Leftrightarrow [u', v', 1] \begin{bmatrix} 0 \\ -T \\ Tv \end{bmatrix} = 0 \Leftrightarrow Tv' = Tv \Rightarrow v = v' \text{ if } T \neq 0$$

Then,



$$\frac{Z}{f} = \frac{B_1}{x} = \frac{B - B_1}{-x'} \Rightarrow Z = \frac{Bf}{x - x'}$$

x' is negative

$x - x'$ is called disparity

depth is inversly propotional to disparity

$$Z = \frac{Bf}{x - x'}$$

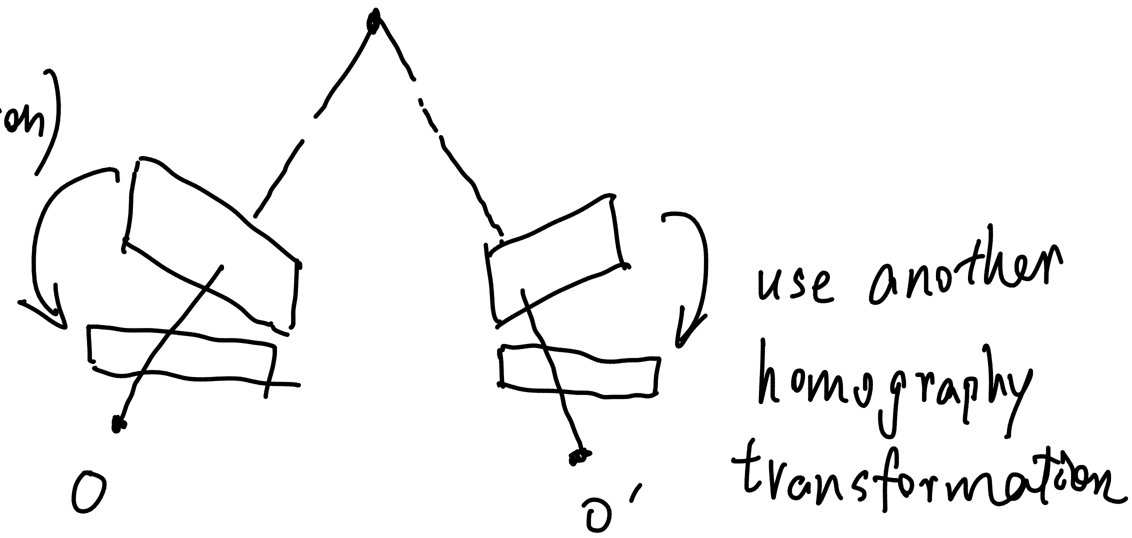
General Case.

two cameras are not parallel.

! Convert the case to the simplest case (parallel camera).

Use another projection to pretend there are two cameras parallel.

(Stereo Image Rectification)



For corresponding pixel search,

Use a small window and find the best match

But, this method will fail often. especially scan two textureless images, occlusion, repetition, Non-Lambertian surface, specralation

↑ some objects may disappear or shadowed in another image also, when

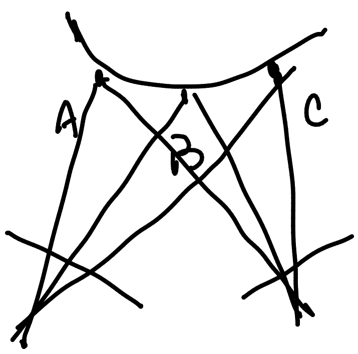
- window size ↑ loss details, false alarm, smooth
- window size ↓ noisy more detail

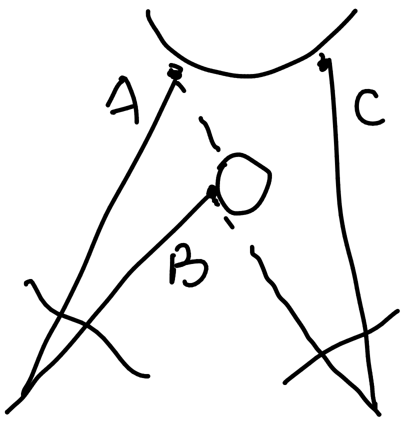
We can use non-local features

① choose unique features

② Order constrain. say (a, b, c) are located one by one from left to right. (a', b', c') will order (c', b', a').

But may fail





③ Smothness.

Scanline Stereo.

→ Instead of choosing a window, match a line of pixels.

$$\rightarrow E(D) = \sum_i \left(W_1(i) - W_2(i+D(i))\right)^2 + \lambda \sum_{\|i-j\|<r} \rho\left(D(i) - D(j)\right) \leftarrow \min E(D)$$

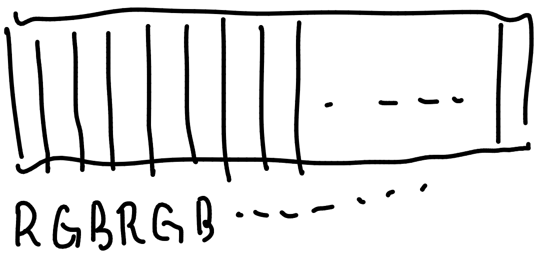
First term: window should be similar

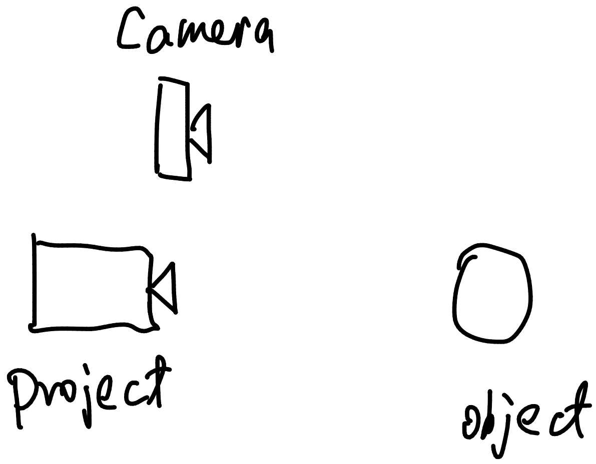
Second term: smooth: neighbours should be similar

Active Stereo with structured light.

Manually generate some unique features

① Projector project elaborately designed lights. say





strips of unique light. are same

Say [kinect] use infrared light. But the object should not be far away, because of resolution, power. And, not for specular surface.

Aligning range Images

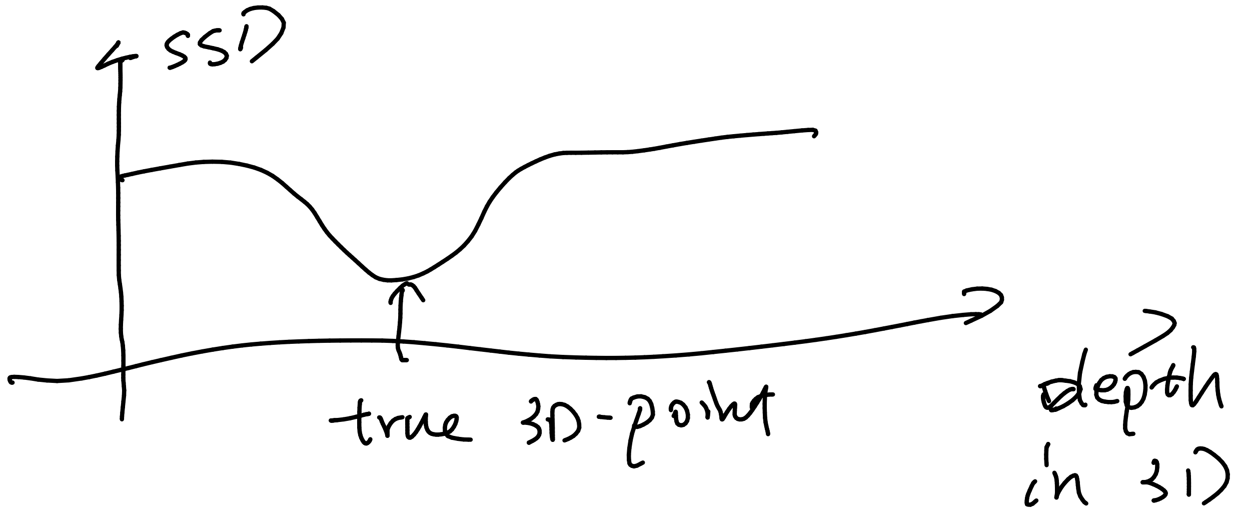
Build a 3-D model. only view from one direction is not sufficient.

↳ Multi-view Stereo.

Basic idea:

Similar method for two-camera case. this time, we will check the error on all related views.

For "search" putative image points' pair,
we search on the 3D-point, and find out relative image pixel in each camera. see, whether it matches.



In reality, we don't search purely on depth, we search on 1/depth. (inverse depth)