Visión por Computadora I

Ing. Maxim Dorogov

(mdorogov@fi.uba.ar)

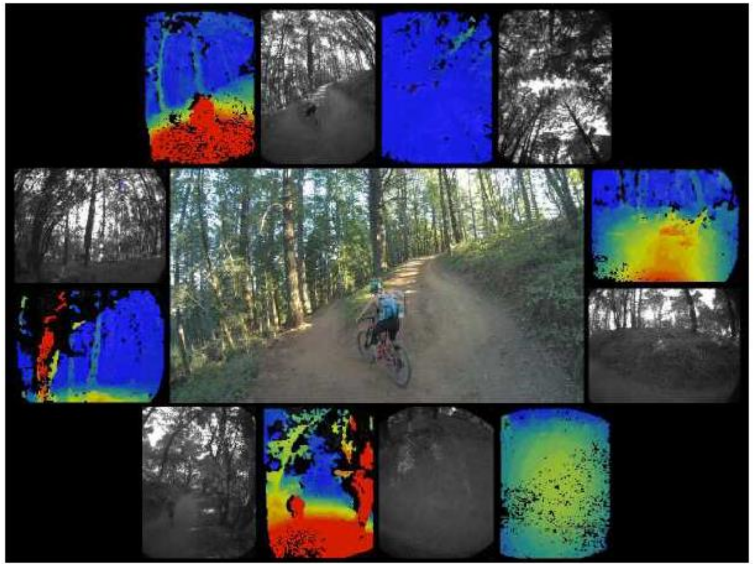
Laboratorio de Sistemas Embebidos -FIUBA

FACULTAD
DE INGENIERIA
Universidad de Buenos Aires

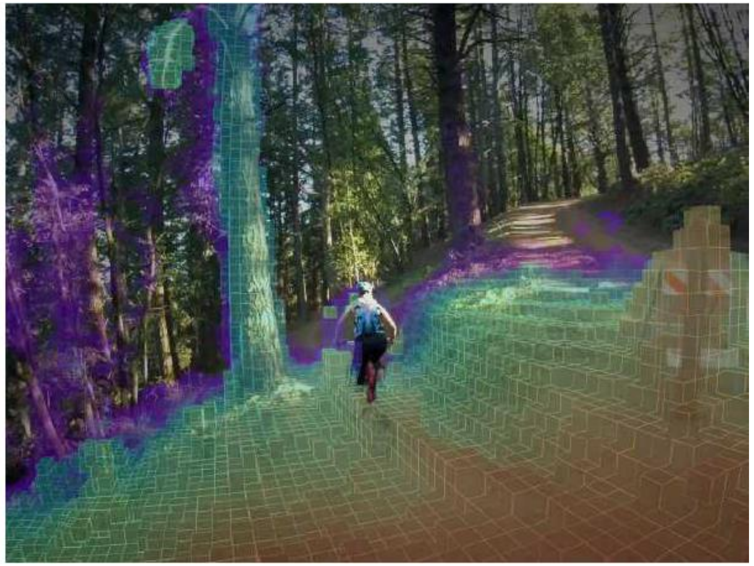
DEPTH ESTIMATION

Why we need depth?

- Scene understanding
- SLAM – Odometry
- Autonomous driving
- Object detection and tracking
- Segmentation
- 3D reconstruction
- Stereo head tracking
- Gaze correction
- Z-keying segmentation



Multiple input images and depth maps.

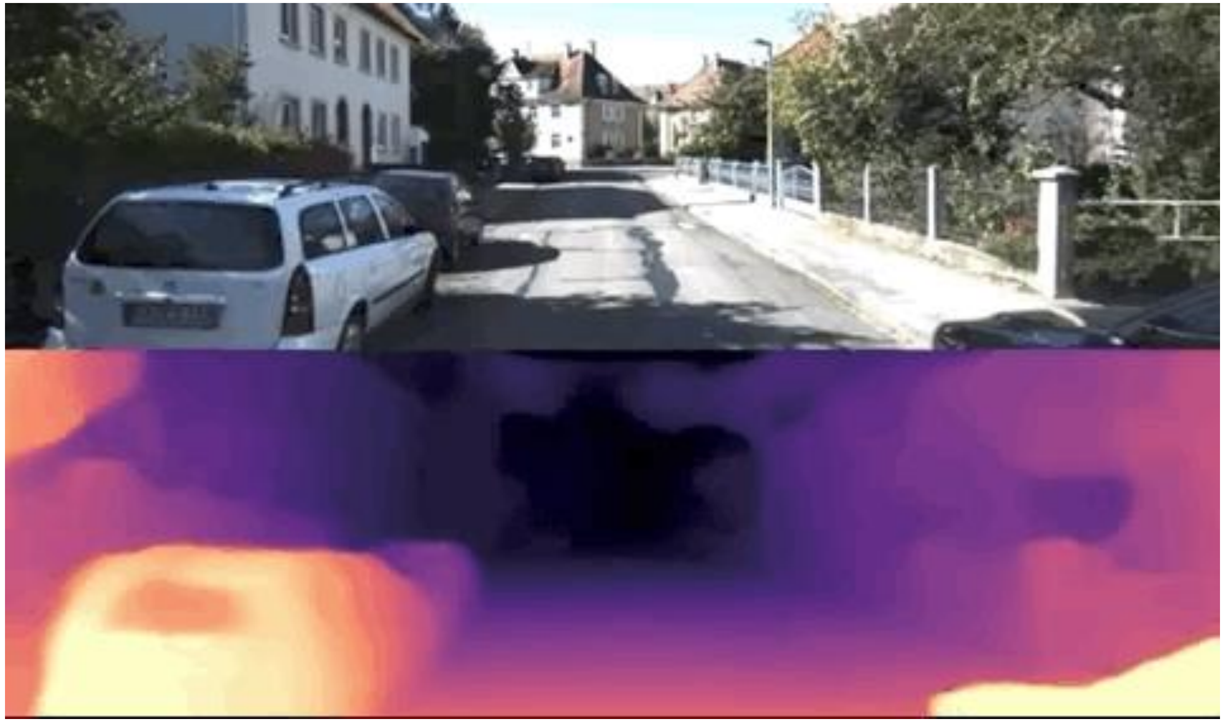


Fully integrated 3D map

Most active research area in computer vision…

DEPTH ESTIMATION

More examples:



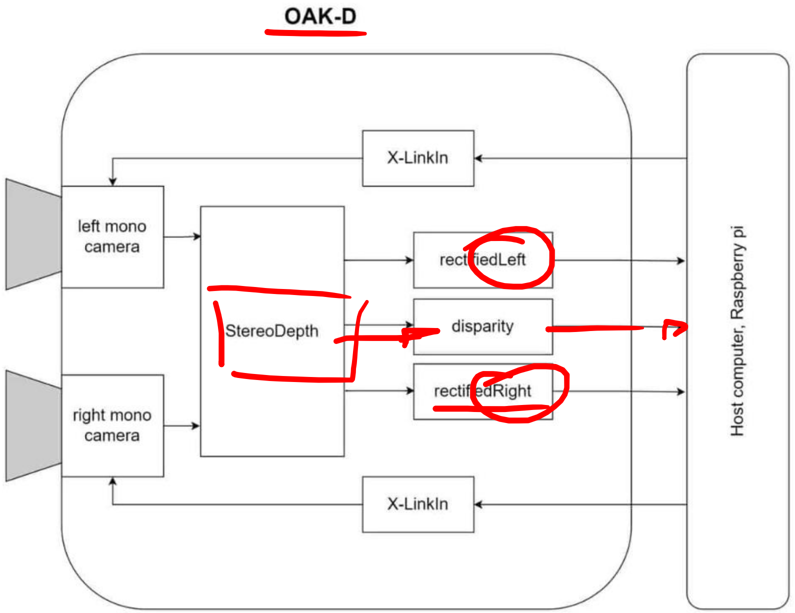
Real time depth estimation with disparity map

DEPTH ESTIMATION

Hardware:

Intel Real Sense RGB-D camera.

Luxonis OAK-D camera.



OAK-D block diagram.

Custom stereo configurations with standard cameras.

DEPTH ESTIMATION

Why stereo?

Depth and structure are ambiguous from a single view…

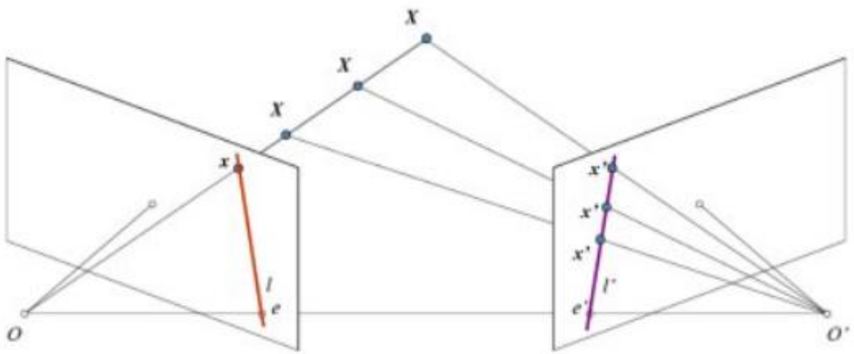
DEPTH ESTIMATION

Stereo Matching:

- Is the process of taking two or more images and finding matching pixels in the images and converting their 2D positions into 3D depths.
- Matching pixels are pixels of different images that are projections of the same 3D point.

Disparity Map

- Difference in the position between matching pixels from left and rigth cameras. Depth of a point in a scene is inversely proportional to the difference of corresponding image points.



3D point projected into different 2D coordinates

Stereo disparity between two images

DEPTH ESTIMATION

Assumptions and limitations

- Cameras are leveled, matching pixels in both images have the same Y coordinate.
- Images are coplanar
- No optical distortion
- Textures must be present in the scene
- Works well only for a specific distance range: *The disparity reduces when the object moves further away from the cameras and the images look identical.*

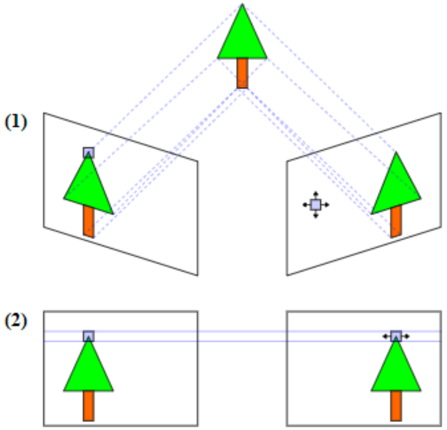
Radial distortion causes straight lines to appear curved

Stereo Calibration

1. Individual camera calibration
2. Calculate the rotation and translation between the two cameras [`cv2.stereoCalibrate()`] and get the Essential and Fundamental matrix.
3. Using the camera intrinsics and the rotation and translation between the cameras, we can now apply stereo rectification.
   - `cv2.stereoRectify()`

Stereo rectification applies rotations to make both camera image planes be in the same plane.

The cameras positions are fixed, the transformations need to be calculated only once. Hence we calculate the mappings that transform a stereo image pair with `cv2.initUndistortRectifyMap()` and store them for further use.



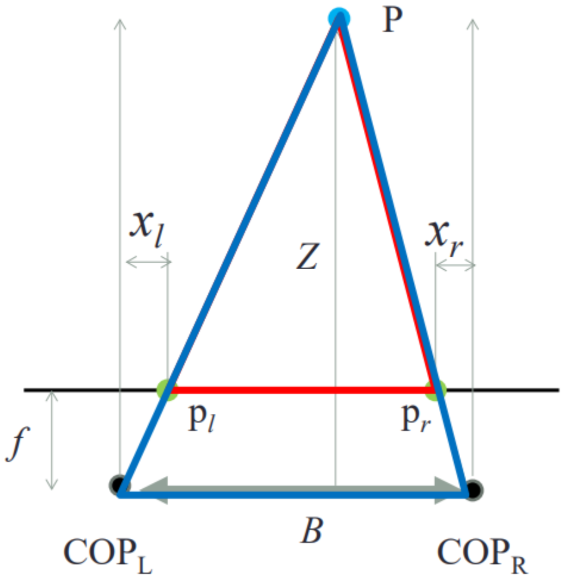
DEPTH ESTIMATION

Depth and disparity

(p_l, p_r, P) and (COP_L, COP_R, P) are similar triangles. Combining this information with the basic monocular perspective projection we get:

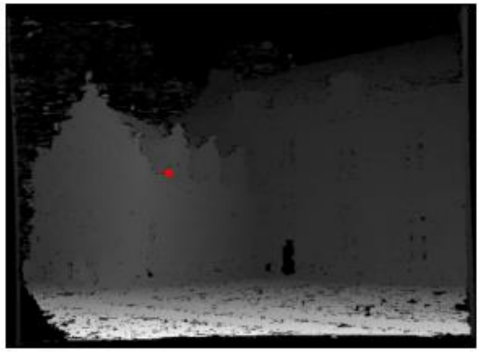
$$Z = f\frac{B}{x_l - x_r}$$

- $x_l - x_r$ is the disparity for the given pixel P
- $Z \propto \frac{1}{x_l - x_r}$ **:** Depth is inversely proportional to disparity, thus measurements are limited to nearby objects.



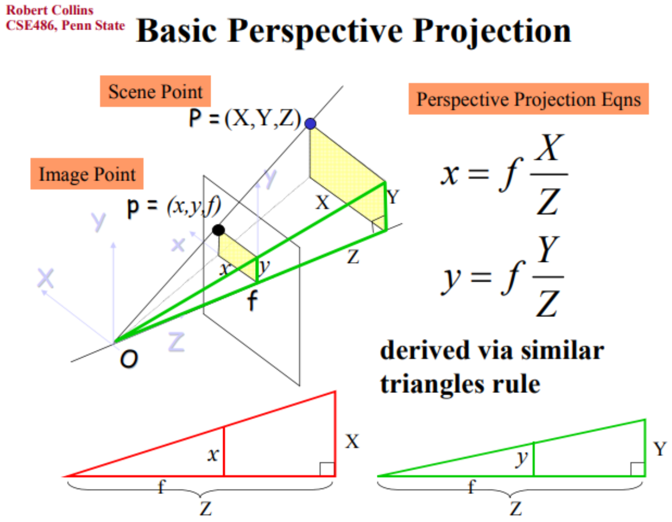
Simplified stereo setting with two cameras

Left image



Disparity map

Right image

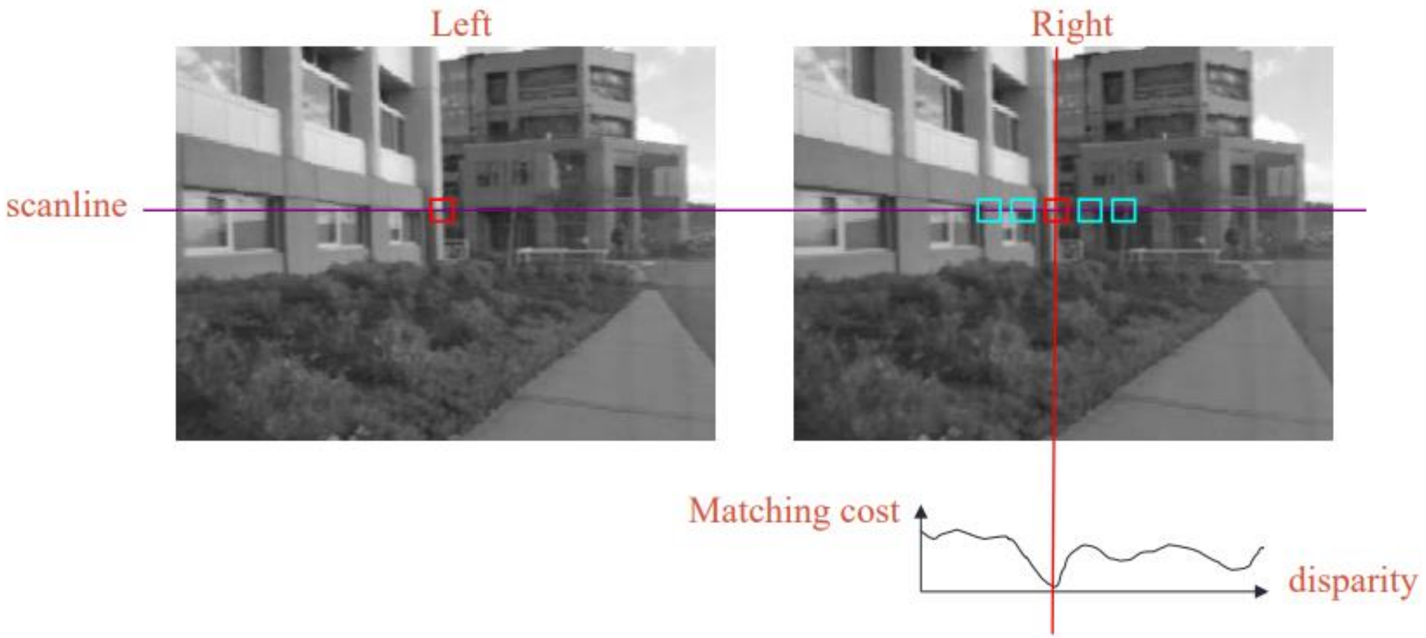


DEPTH ESTIMATION

Finding matching pixels

Block matching algorithm:

- The block pairs with the best score (SSD, SAD, CCOR, etc…) are the best match for the given pixel.
- A straightforward solution is to repeat the process for all the pixels on the image.



Block matching over a scanline with SAD (Sum of Absolute Differences) matching cost.

DEPTH ESTIMATION

Finding matching pixels

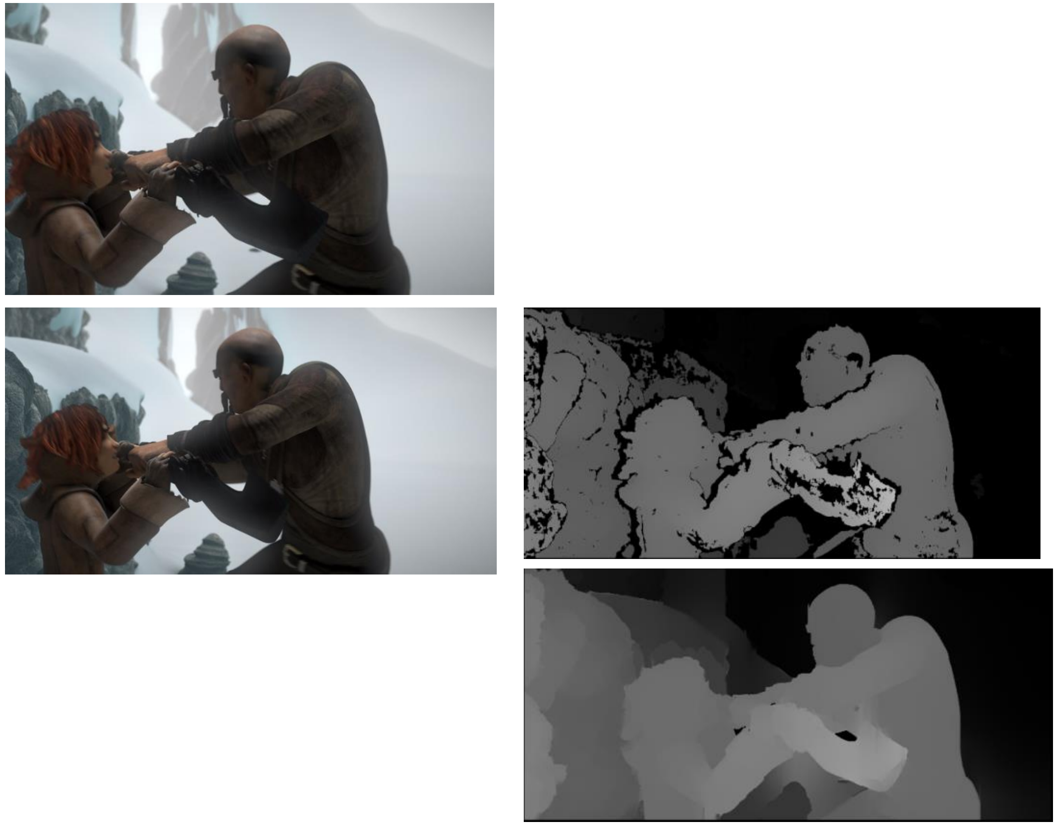
Practically it's impossible to get a one-to-one correspondence:

- Multiple matches for a given block.
- Repeating texture or texture-less regions
- Occlusion or disappearing regions

Dynamic Programming

Dynamic programming is a standard method used to enforce the one-to-one correspondence for a scanline. OpenCV provides some implementations for the Block Matching algorithm: `StereoBM` and `StereoSGBM`:

- Calculates a disparity map for a pair of rectified stereo images.
- Its more robust and less noisy than naïve matching.
- You may need to fine tune the parameters to get better and smooth results.



Input images and `StereoBM` raw output and post-filtered