MEAM 620 - Sensing

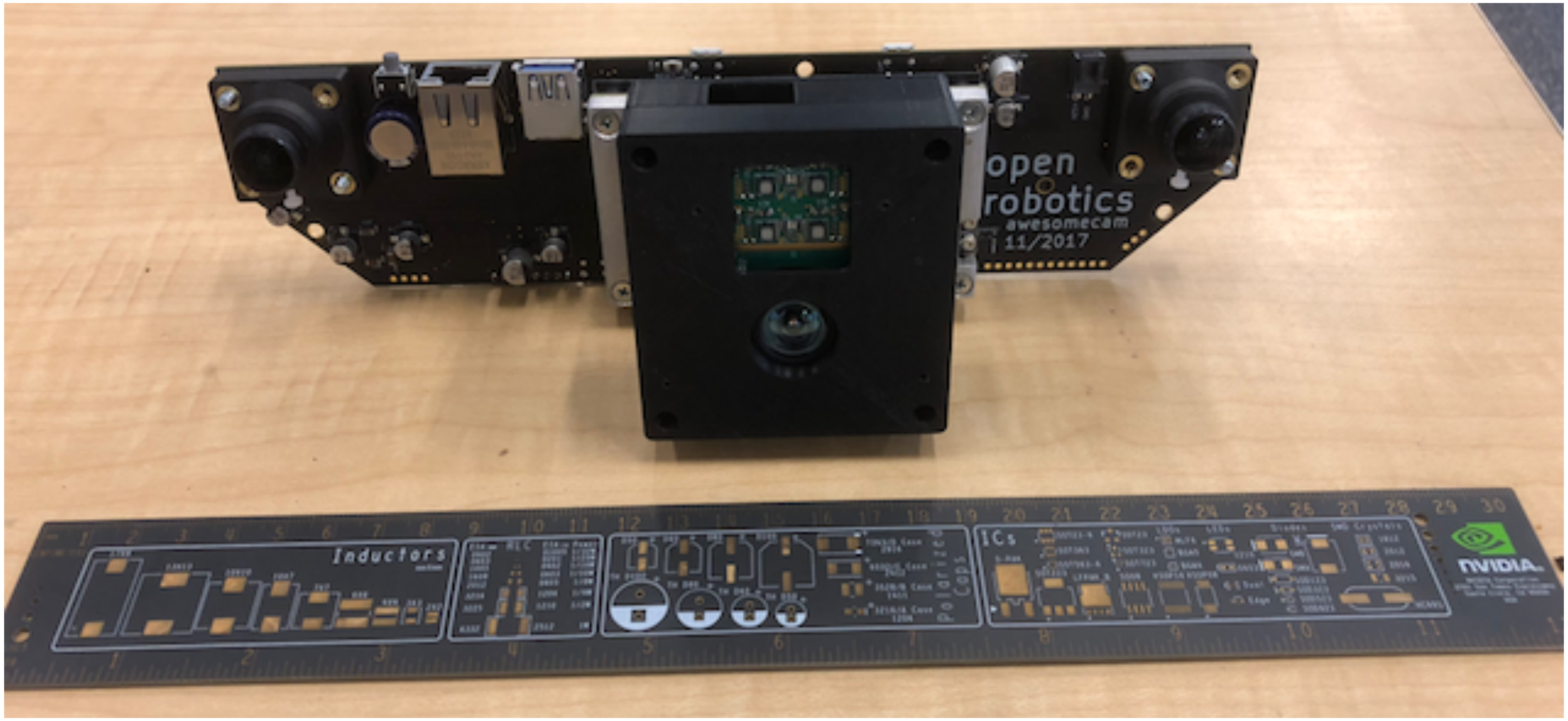
Spring 2020

C.J. Taylor

Dense Stereo

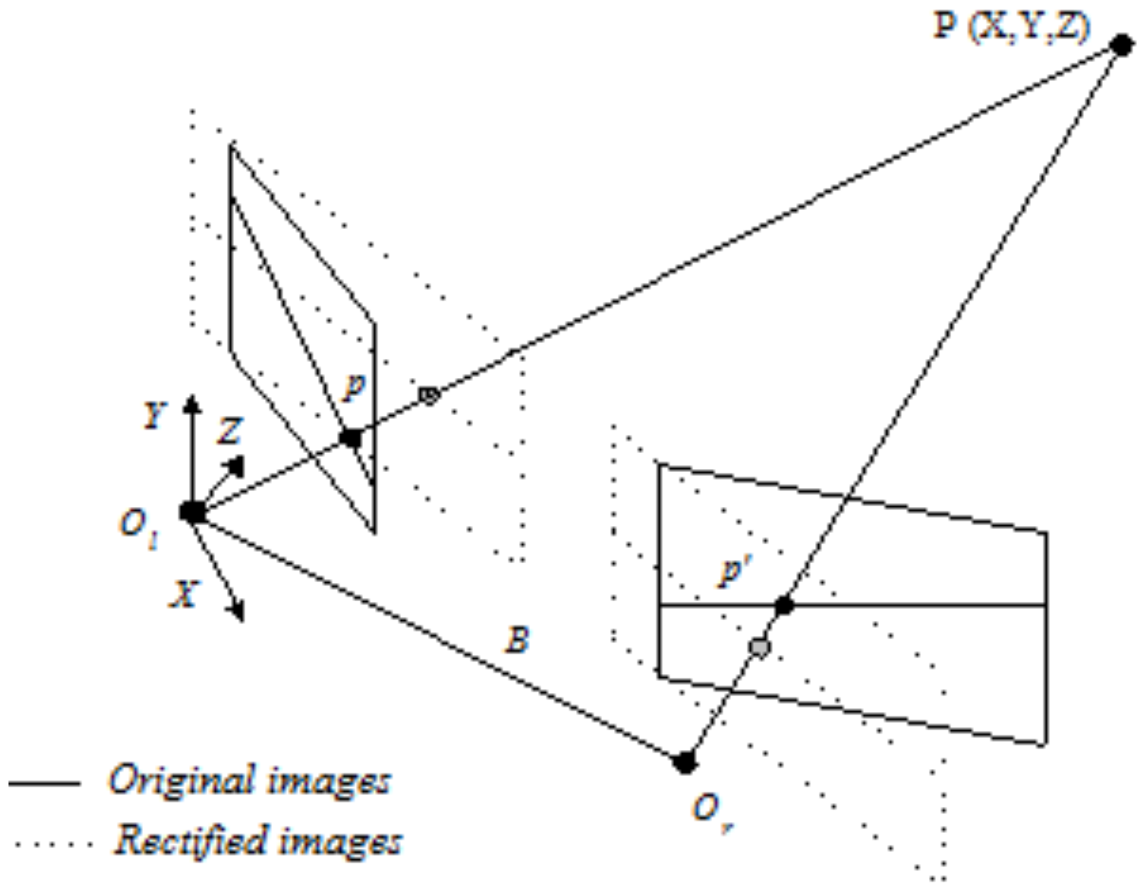
Stereo

- In order to derive 3d measurements of the scene we often construct systems with two cameras separated by a fixed baseline. These systems are referred to as stereo rigs.
- With two cameras we can triangulate features in the scene to determine their depth much as the human visual system does.



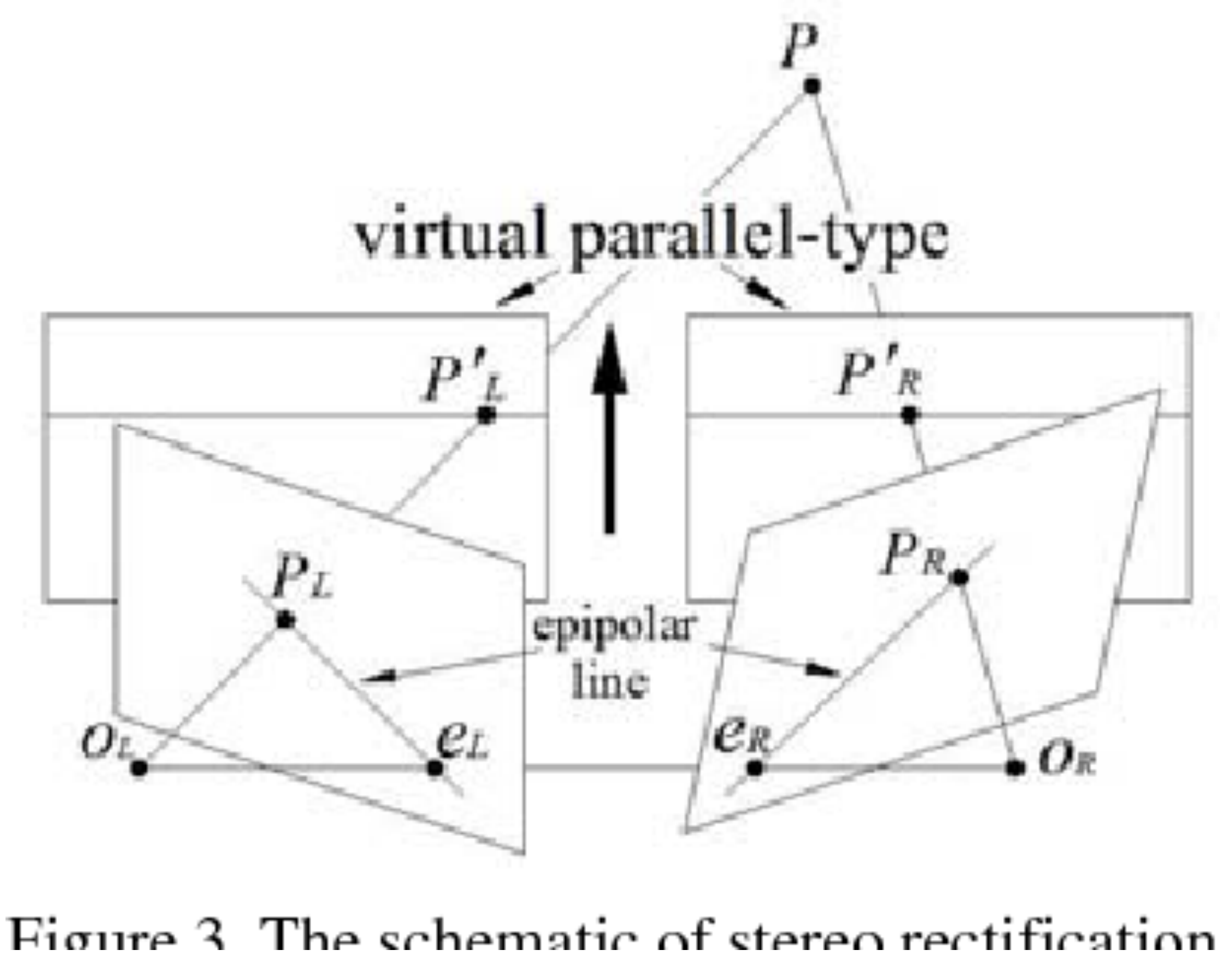
Stereo Rectification

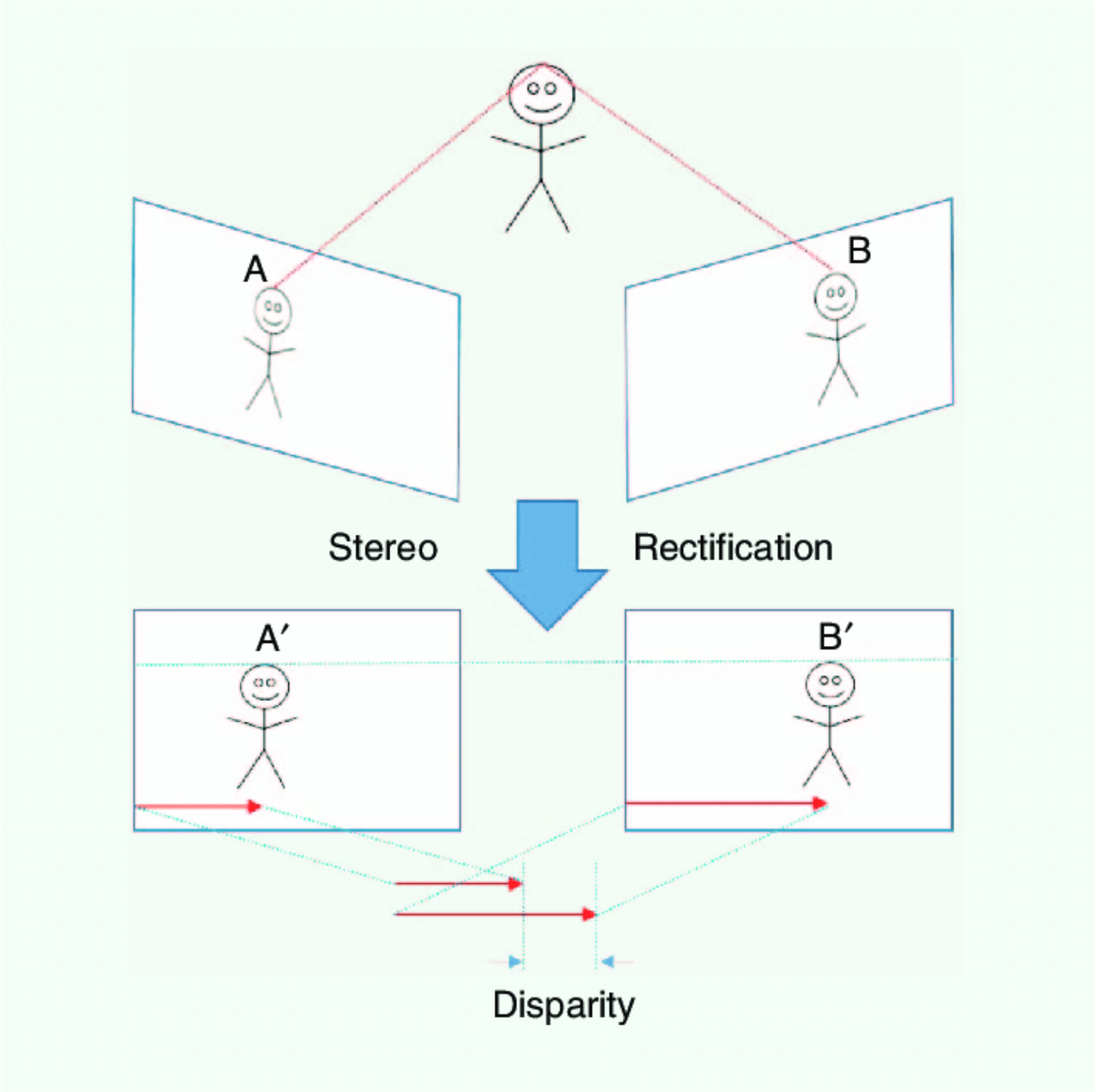
- We will find it convenient to rectify the two images we get from the stereo rig to produce a simpler geometry.
- The rectified images will appear to come from a pair of cameras with identical intrinsic parameters whose optical axes are parallel and where the stereo baseline is aligned with the x-axis
- In this case corresponding features in the world will image on the same row in both images. This is also referred to as placing the epipoles at infinity.



Stereo Rectification

- Stereo rectification is simply a transformation of the images that simplifies subsequent subsequent stereo tasks like matching and triangulation.

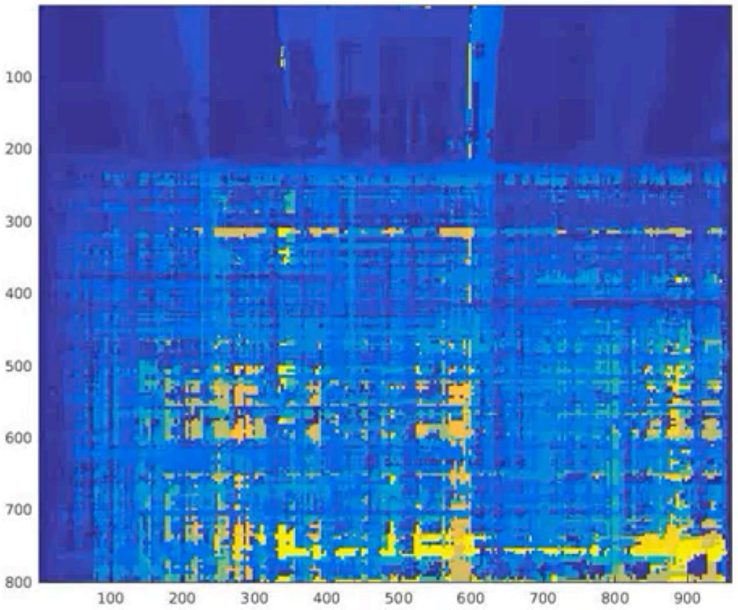




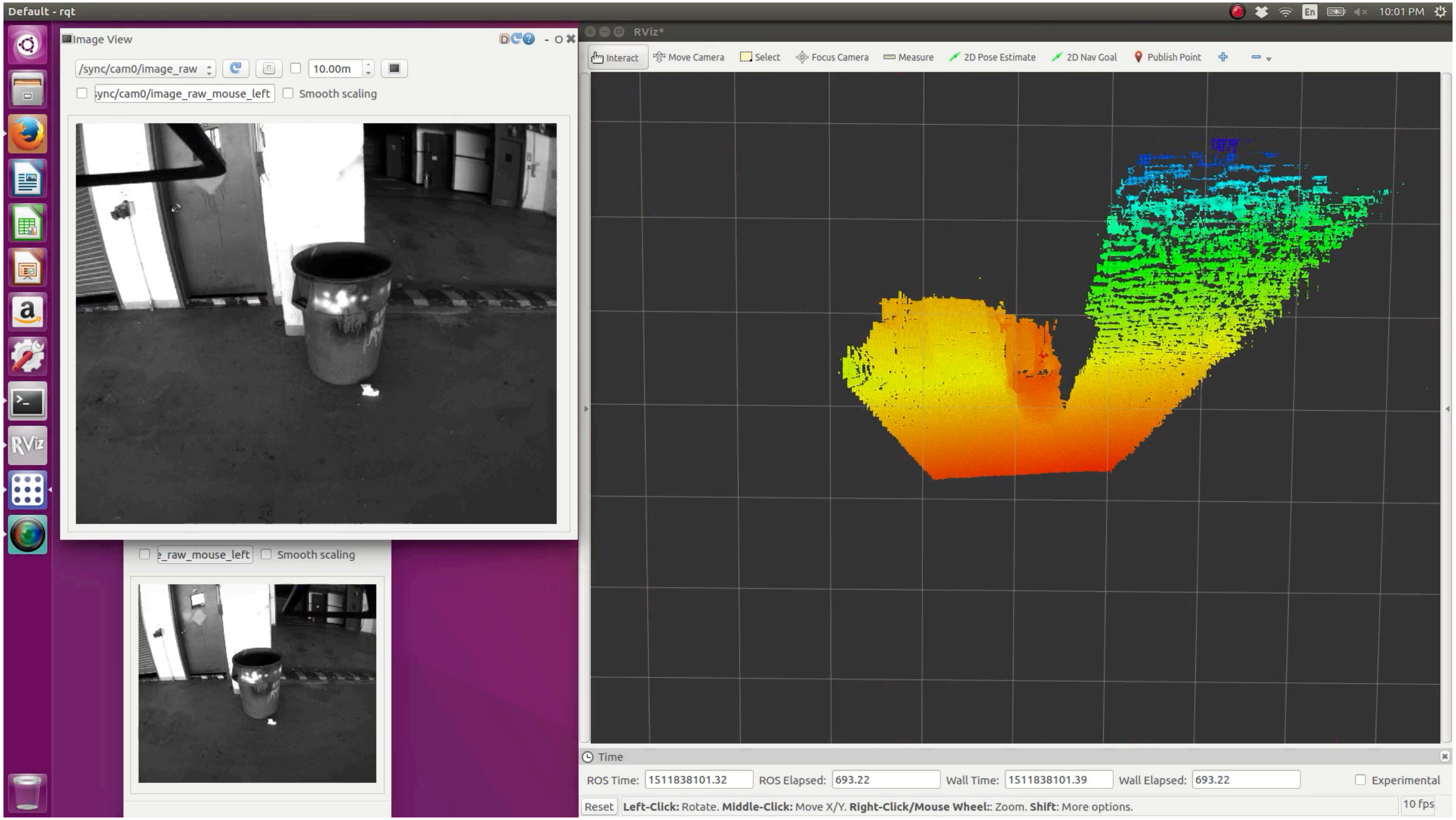
Dense vs Feature based approaches

- So far we have been working with stereo systems that seek to provide measurements of sparse, well defined, features like corners. The idea being that we get on the order of a few hundred stereo measurements for each stereo pair.
- Another way to proceed is to seek to recover a correspondence for all, or almost all, pixels in the two images. This produces a dense set of correspondences which provides a much better understanding of the layout of the environment.

Dense Stereo Example



Dense Stereo Example



Stereo

- The stereo problem is usually broken in to two subproblems
  - The correspondence problem
    - which involves identifying corresponding points in both images
  - The reconstruction problem
    - Recovering the geometry of the scene and/or the relative camera positions

The correspondence problem

- The goal of correspondence algorithms is to find matching locations in the left and right images

Stereo Algorithm

- For each epipolar line in the left image locate the corresponding epipolar line in the right image and then find matches using a 1D stereo algorithm

Recovering Depth

- The inverse relationship between disparity and depth means that points that are very distant appear to have approximately zero disparity
- Another important consequence is that the error in the reconstructed depth increases as you go further away since a small difference in disparity translates to a large difference in depth as disparity approaches 0

$$d = (u_l - u_r) = \left(\frac{f}{s_x}\right)\left(\frac{b}{Z}\right)$$

$$Z = \left(\frac{f}{s_x}\right)\left(\frac{b}{d}\right)$$

Correlation based approaches

- A common approach to finding correspondences is to search for local regions that appear similar
- Algorithm
  - For each u_l find a d that minimizes/maximizes $c(u_l, d)$

$$c(u_l, d) = \psi(I_l(u_l - w : u_l + w), I_r((u_l - d) - w : (u_l - d) + w)$$

Match metrics

- Some commonly used match metrics
  - SSD - Sum of Squared differences

$$\psi(x,y) = \sum_i (x_i - y_i)^2$$

  - Cross Correlation

$$\psi(x,y) = \sum_i (x_i y_i)$$

  - Normalized Cross Correlation

$$\psi(x,y) = \frac{1}{(\sigma_x \sigma_y)} \sum_i ((x_i - \bar{x})(y_i - \bar{y}))$$

Binary Match Metrics

- More recently stereo implementations have begun constructing simple binary descriptors for each window which capture the gross variation in intensity
- For example we could construct a binary vector which records whether or not each pixel intensity in the window is greater or lesser than the intensity of the central pixel.
- Two windows can then be compared by computing the Hamming Distance which simply refers to the number of differences between two bit strings.
- This approach is robust to changes in bias and scale between the two images and leads to efficient stereo cost volume computation.
- egs. Census Transform, BRIEF descriptor, ORB descriptor

Matching Issues

- There are a number of problems that afflict basic stereo algorithms many of which are due to the tacit assumption that all of the pixels within a given matching window have the same disparity and are thus translated by the same amount.
- Some of these are listed below
  - Thin objects like wires can pose a problem since image windows include the wire and the background and they have different depths
  - Slanted surfaces are problematic since the disparity varies accross the window leading to a warping in the correspondence
  - Half occluded regions lead to pixels that have no matches

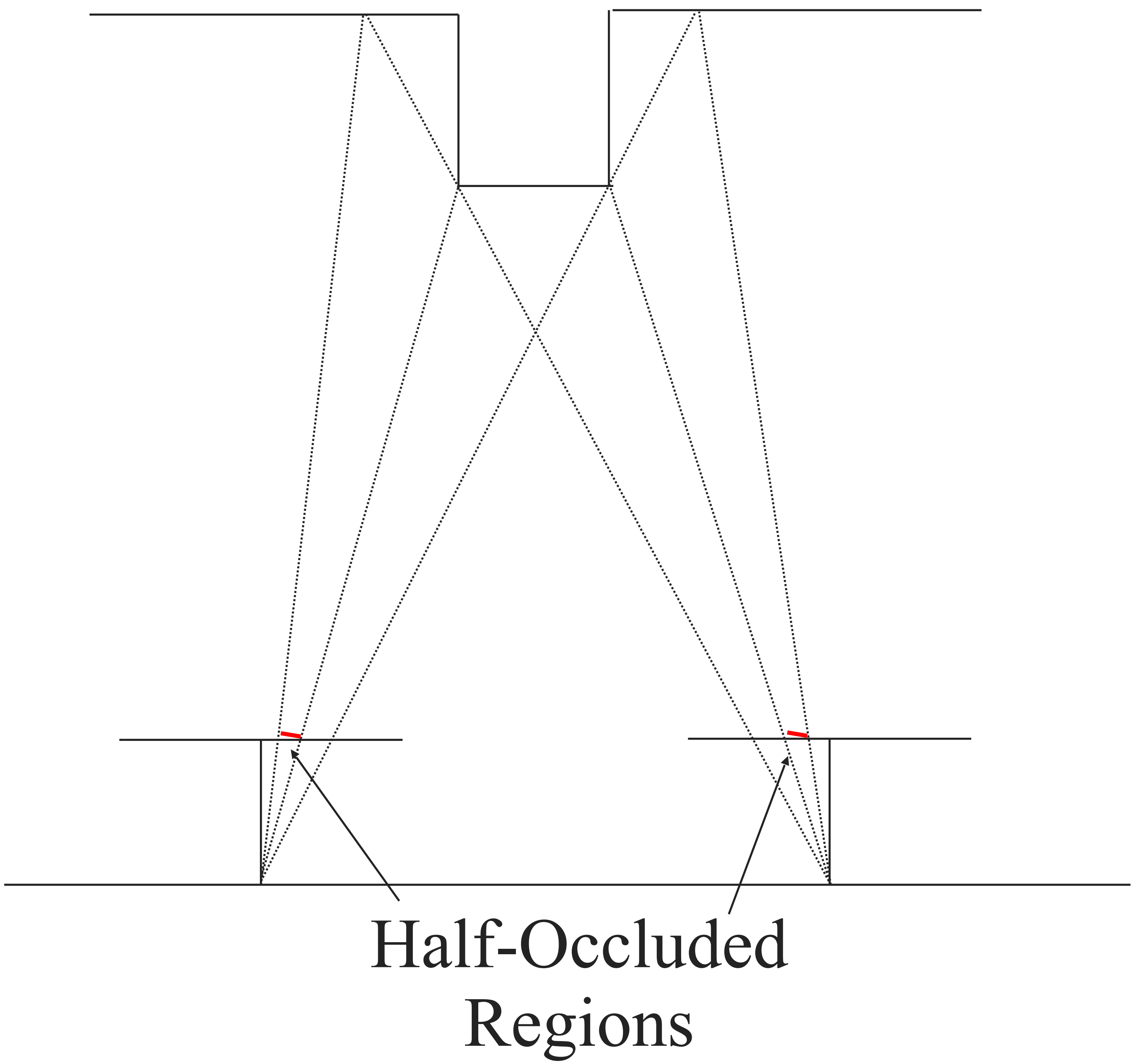
Basic Problems

- Simple approaches based on match metrics can break down in various cases. Chief among these are
  - Situations where the intensity is constant leading to regions which are not particularly unique and easily mismatched to constant intensity regions in the other image
  - Repeated structures lead to situations where one patch looks similar to many other locations in the other image.

Half Occluded Regions

- In certain situations there may be points in the scene that are visible from one of the two cameras but not the other. These are referred to as half-occluded regions
- These regions are frequently associated with depth discontinuities in the scene

Half Occluded regions



Half occluded regions cont'd

- Half occluded regions pose a challenge to most simple stereo correspondence algorithms
- Consequently many stereo algorithms have problems at depth discontinuities
- The human visual, by contrast, seems to be sensitive to half-occluded regions and uses them to detect surface boundaries

Left-Right Consistency Constraint

- One approach to detecting half -occluded regions is to run the stereo procedure in both directions. From left to right and right to left, and then to check that correspondences are consistent in both solutions.

Table of matching scores for a line in the image

Left

Right

										10									
										10									
										8									
						10	9	9	7	5	6	8	10						
										8									
										9									
										10									
										10									

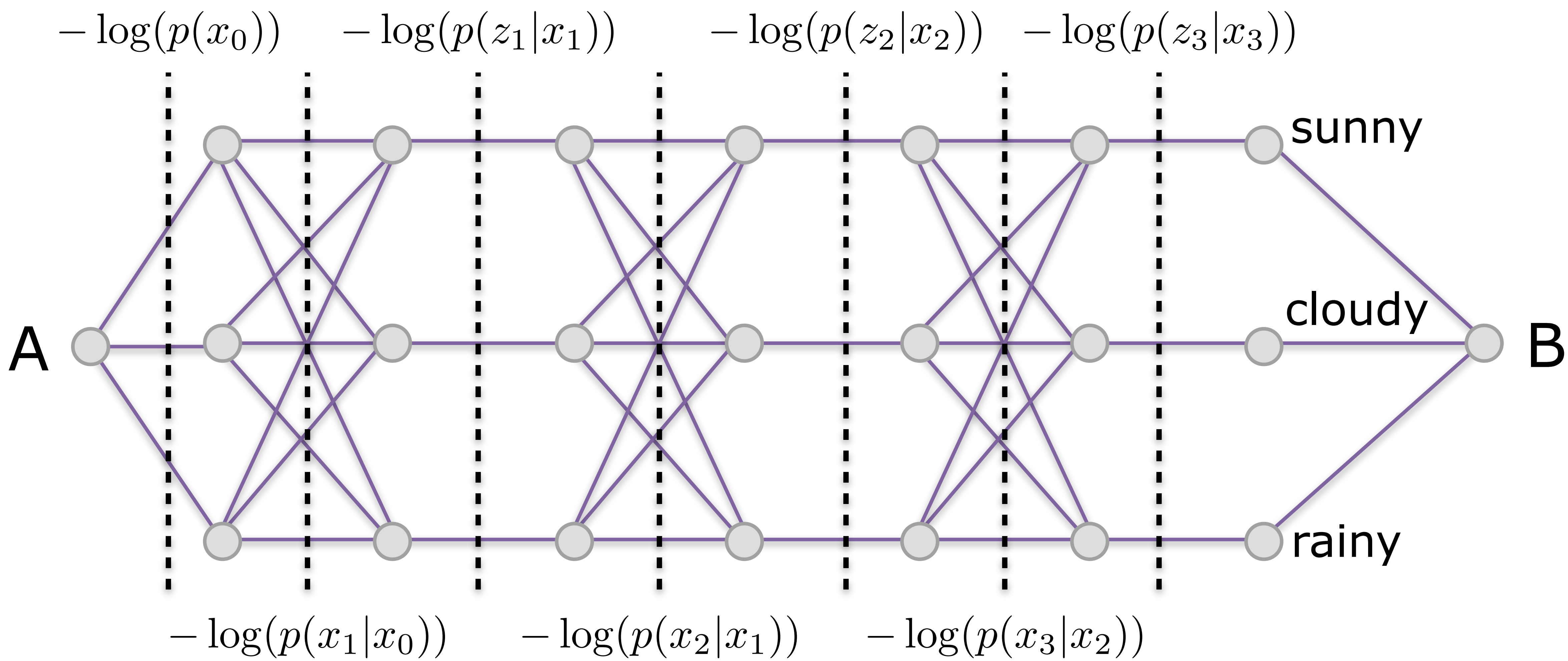
Correspondence

- Since stereo matching is an ill-posed problem global constraints are often employed
- These constraints effectively represent prior knowledge about the scenes

Global Constraints

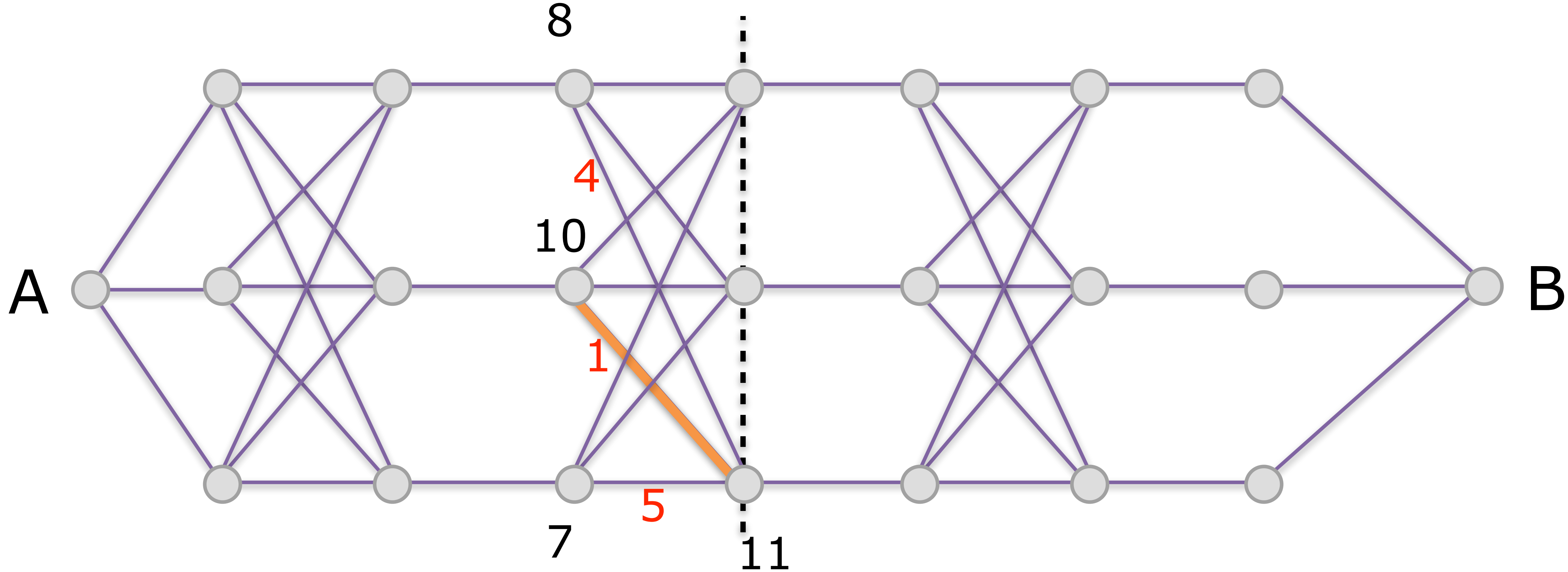
- Constraints
  - Uniqueness of match
  - The ordering constraint
  - Disparity gradient constraint
- Matching techniques
  - Dynamic Programming
  - Relaxation labeling

Dynamic Programming



- This diagram shows how we can recast the optimization problem as one of finding the shortest path between 2 points, A and B, in a trellis graph.
- At each stage of the trellis graph the edge weights model one aspect of the sum being minimized, $-\log(p(x_0))$, $-\log p(x_1|x_0)$, $-\log p(z_1|x_1)$ etc.
- In this diagram the edge weights in the last phase linking to node B are all 0.

Viterbi Algorithm



- The optimization problem can be solved efficiently by considering each stage in the trellis in turn from left to right. At each stage each node computes the shortest distance to the start node by considering the shortest distances in the previous stage. We need only record the shortest path distance for each node and its parent in the previous stage.
- When the algorithm reaches the final stage we can trace back through the parent pointers to recover the optimal path/state assignment.
- This is an example of *Dynamic Programming* where our optimization is split over a number of stages and the partial results from one stage of the optimization are used in the next.