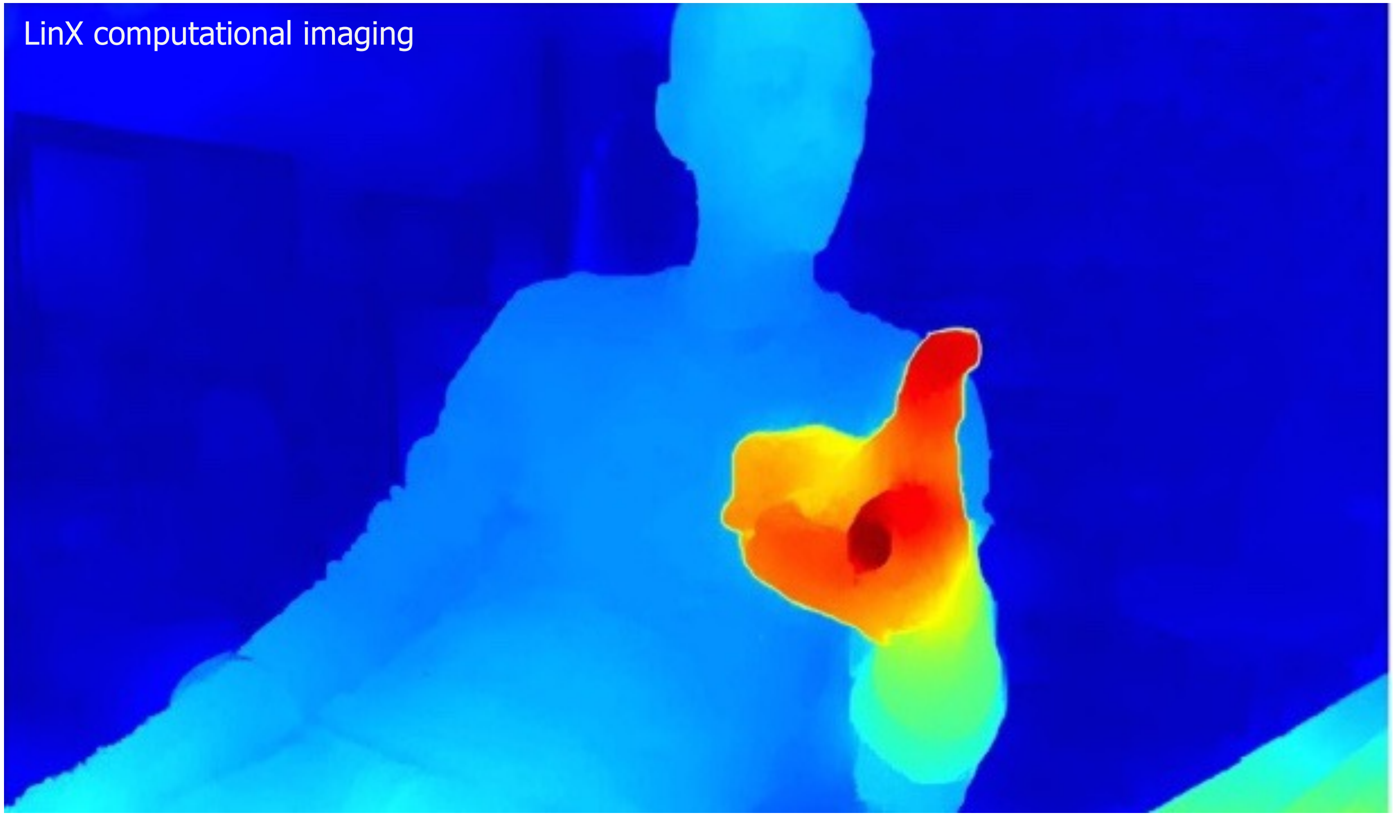
Dense Stereo



Slide credit: Yuri Boykov

Dense Stereo

towards **dense** 3D reconstruction

- (dense) stereo is an example of **dense correspondence**
- another example is dense motion estimation (*optical flow*)

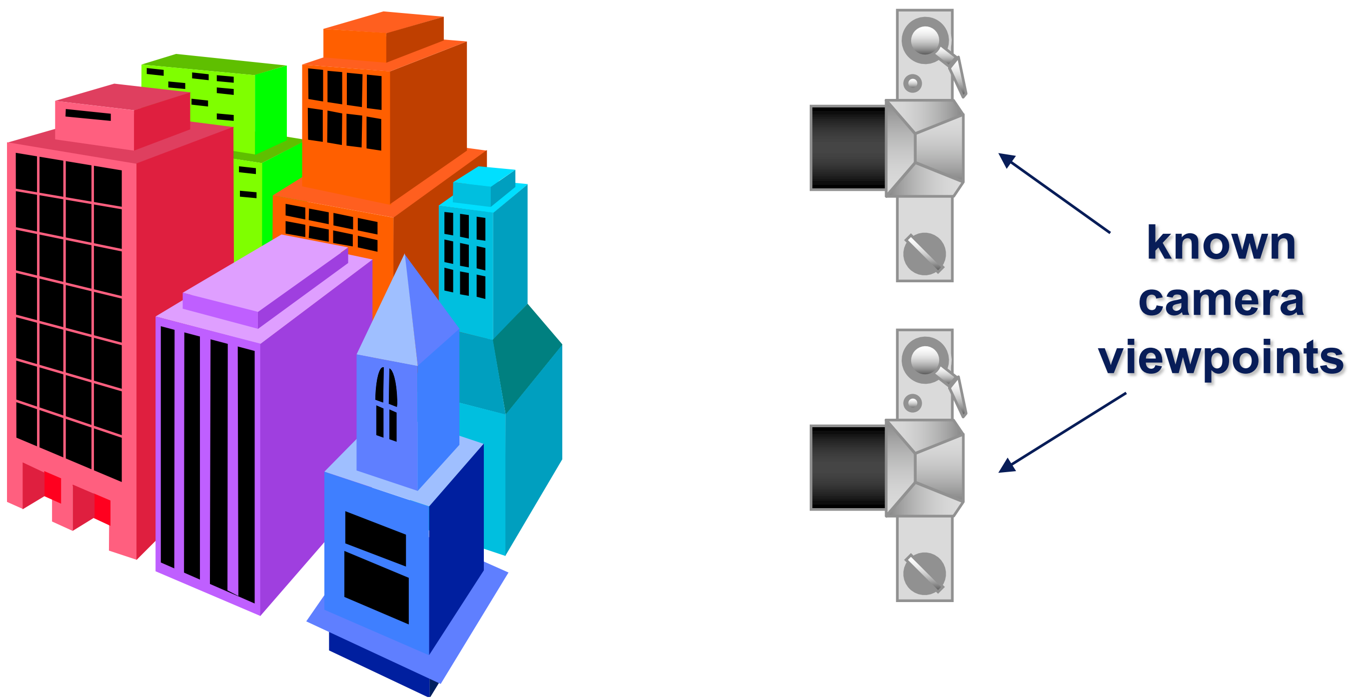
But, **stereo is simpler** since the search for correspondences
is restricted to 1D epipolar lines (versus 2D search for non-rigid motion)

Dense Stereo Correspondence

- window-based (local) stereo
- scan-line stereo correspondence
  - optimization via DP, Viterbi, Dijkstra
- image-grid (global) stereo
  - optimization via graph cuts

Szeliski, Chapter 11

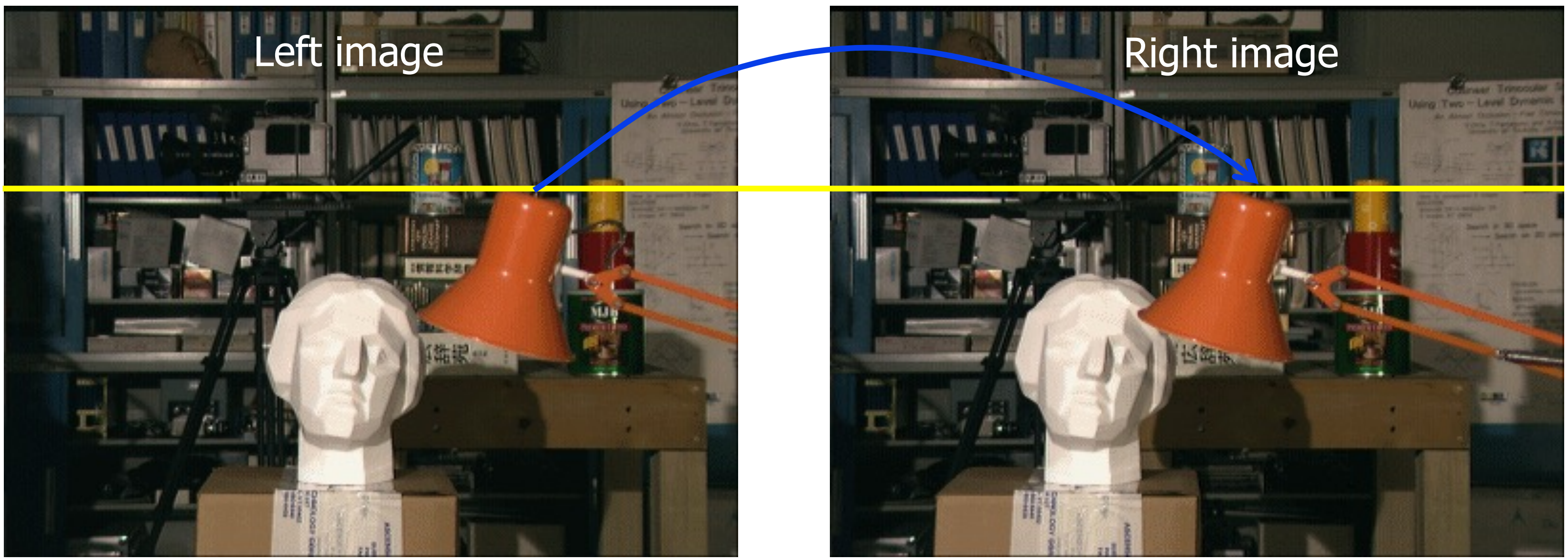
Stereo vision



Two views of the same scene from <u>slightly different</u> point of view

Also called, <u>narrow baseline</u> stereo.

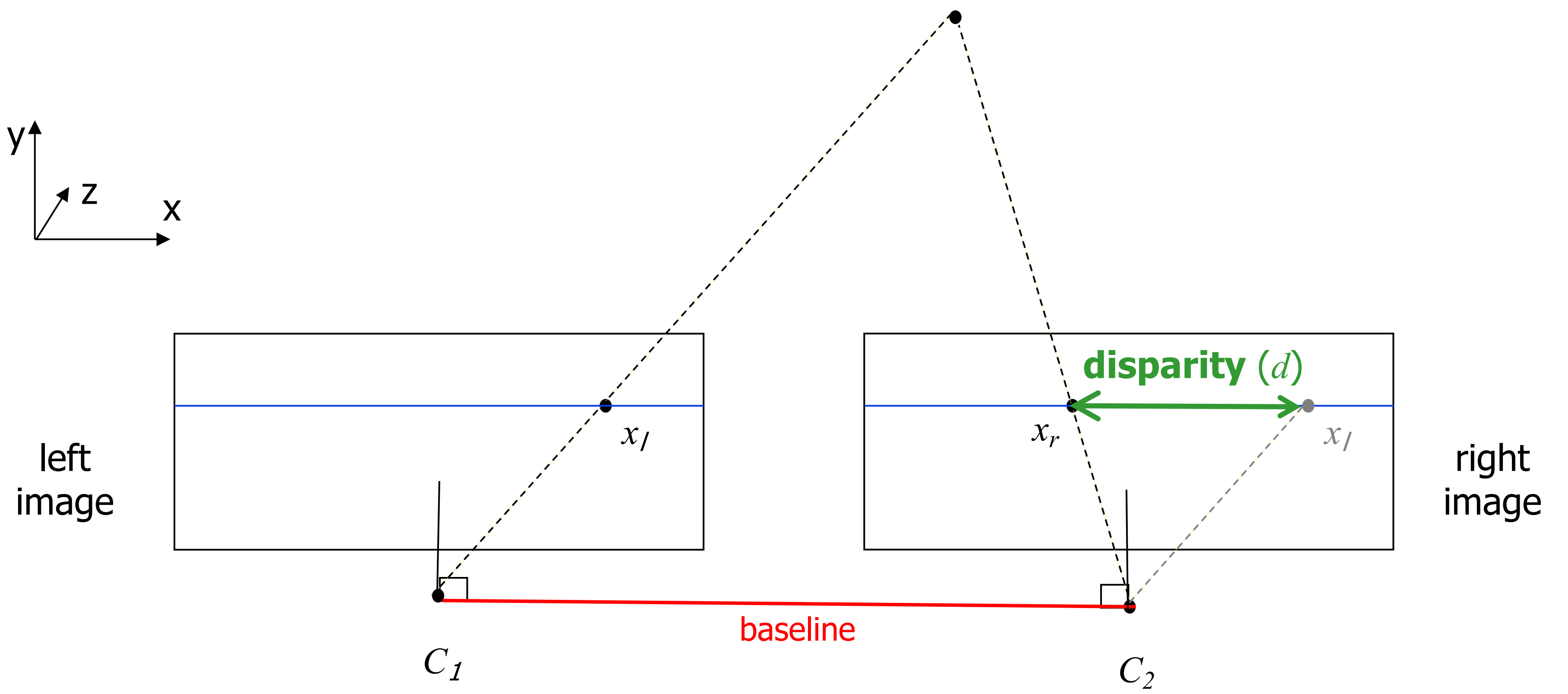
Stereo as a *correspondence* problem



(After rectification) all correspondences are along the same horizontal scan lines

(epipolar lines)

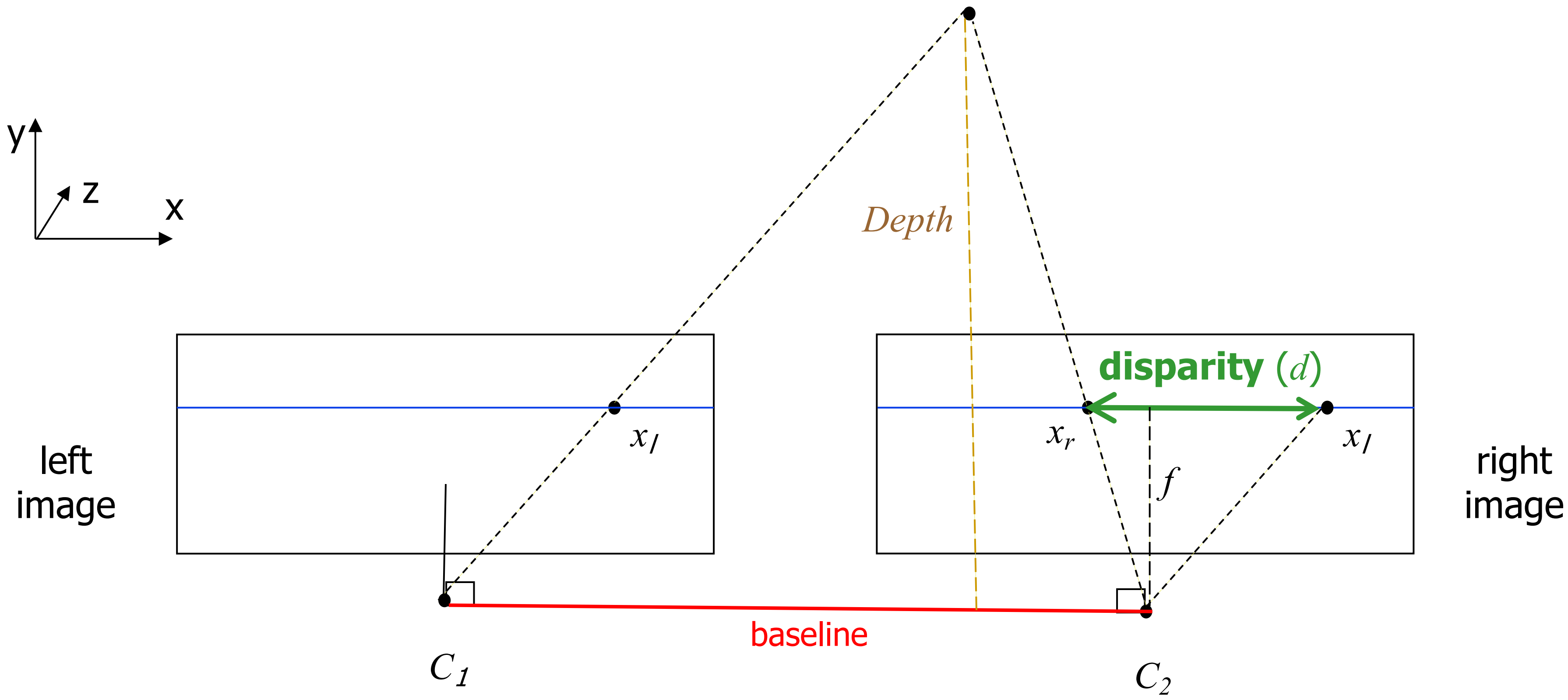
Stereo Cameras



epipolar lines are parallel to the x axis

difference between the x-coordinates of x_l and x_r is called the disparity

Stereo Cameras



Depth = $|C_1C_2| \cdot f / d$

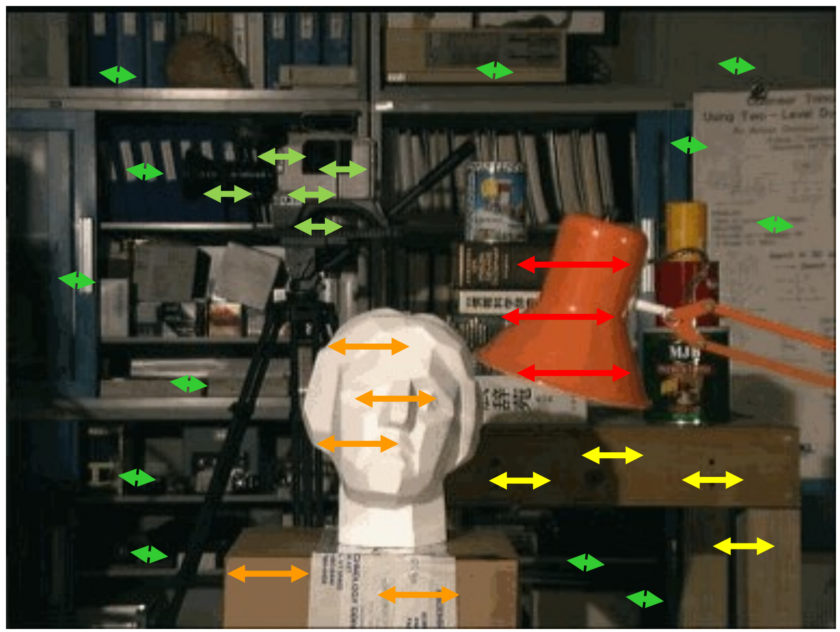
Stereo

Correspondences are described by shifts
along horizontal scan lines (<u>epipolar lines</u>)

which can be represented by scalars (**disparities)**

Stereo

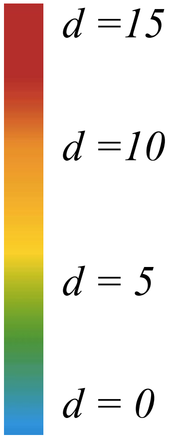
closer objects (smaller depths) correspond to **larger disparities**



Correspondences are described by shifts along horizontal scan lines (<u>epipolar lines</u>)

which can be represented by scalars (**disparities)**

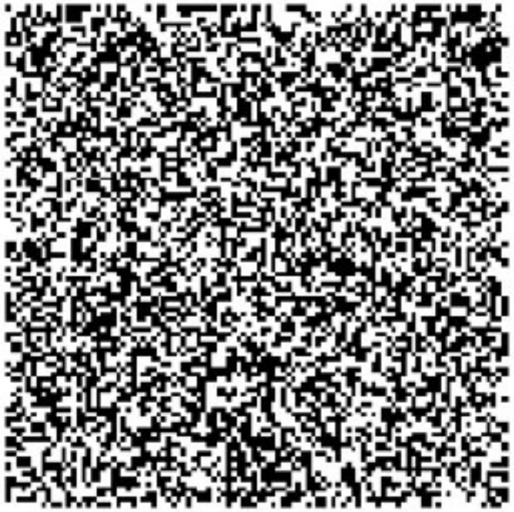
Stereo



- If x-shifts (disparities) are known for all pixels in the left (or right) image then we can visualize them as a **disparity map** – scalar valued function $d(p)$
- larger disparities correspond to closer objects

Stereo Correspondence problem

- Human vision can solve it
  (even for "random dot" stereograms)



- Can computer vision solve it?

Maybe

see *Middlebury Stereo Database*
for the state-of-the art results
http://cat.middlebury.edu/stereo/

Stereo

- Window based
  - Matching rigid windows around each pixel
  - Each window is matched independently
- Scan-line based approach
  - Finding coherent correspondences for each scan-line
  - Scan-lines are independent
    - DP, shortest paths
- Muti-scan-line approach
  - Finding coherent correspondences for all pixels
    - Graph cuts

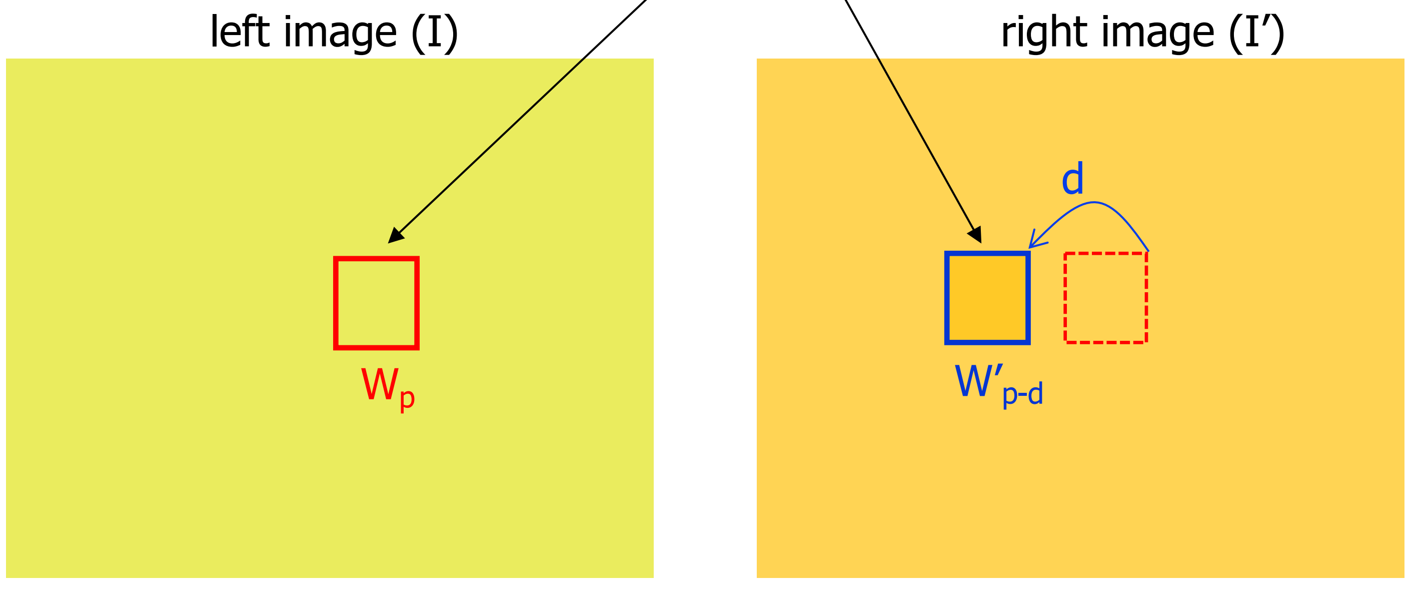
Stereo Correspondence problem
Window based approach

- For any given point p in left image consider window (or image patch) W_p around it
- Find matching window W_q on the same scan line in the right image that looks most similar to W_p

naïve implementation
has |I|*|d|*|W| operations

SSD (sum of squared differences) approach

computing SSD(p,d) $= \sum_{(x,y)\in W_p} (I(x,y) - I'(x-d,y))^2$

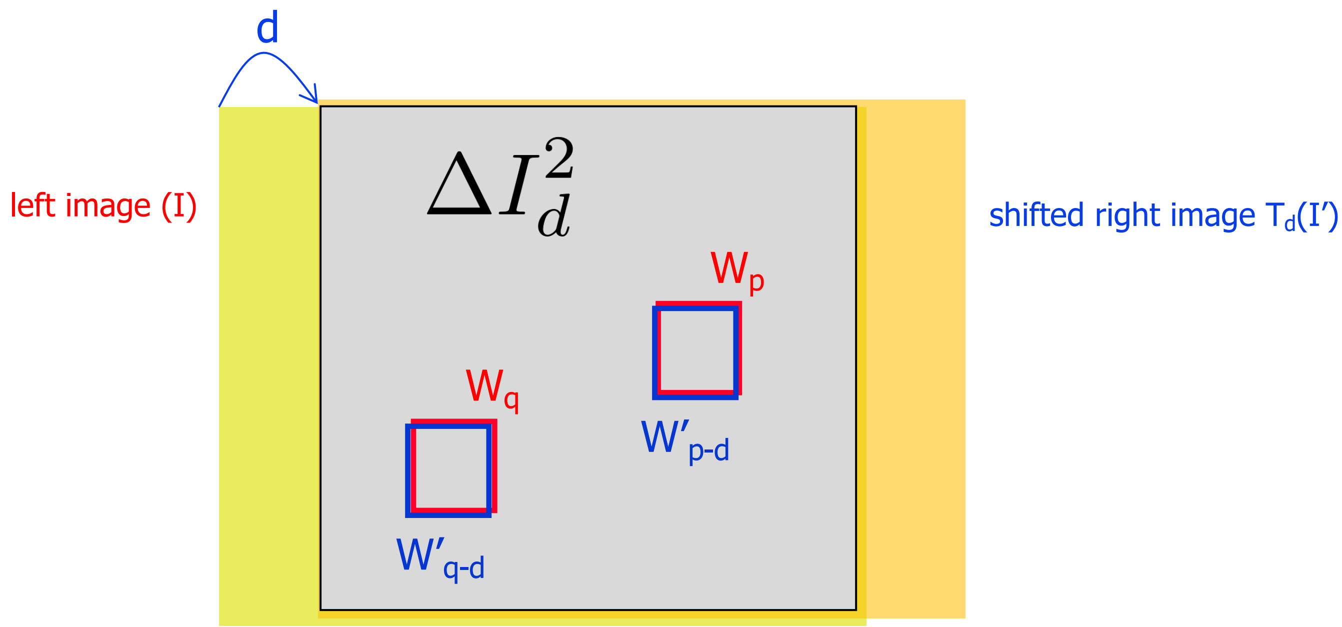


for any pixel p compute SSD between windows W_p and W'_{p-d}
for all disparities d (in some interval [*min_d, max_d*])

then $\boxed{\hat{d}_p = \arg\min_d SSD(p,d)}$

computing SSD efficiently

- For each fixed d, compute <u>sq. differences image</u> ΔI_d^2



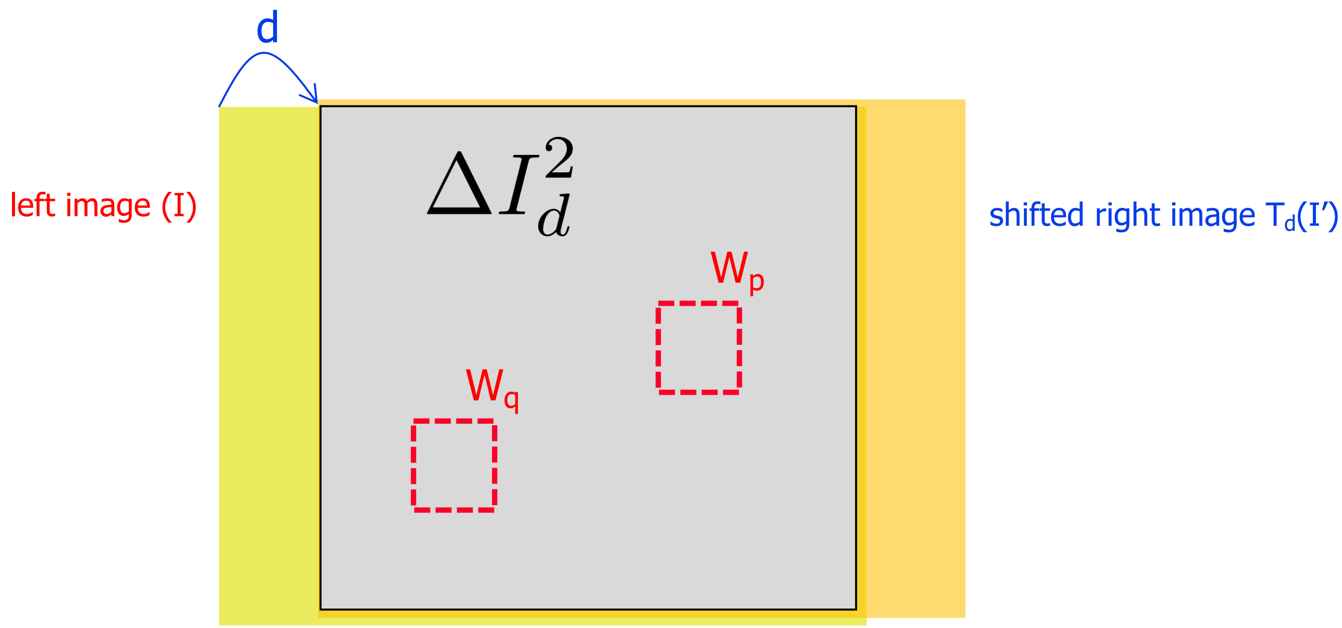
squared differences between the left image I and the shifted right image $T_d(I')$

$$\Delta I_d^2(x,y) \;:=\; (I(x,y) - I'(x-d,y))^2$$

Then, SSD(p,d) between W_p and W'_{p-d} is $\sum_{(x,y)\in W_p} \Delta I_d^2(x,y)$

computing SSD efficiently

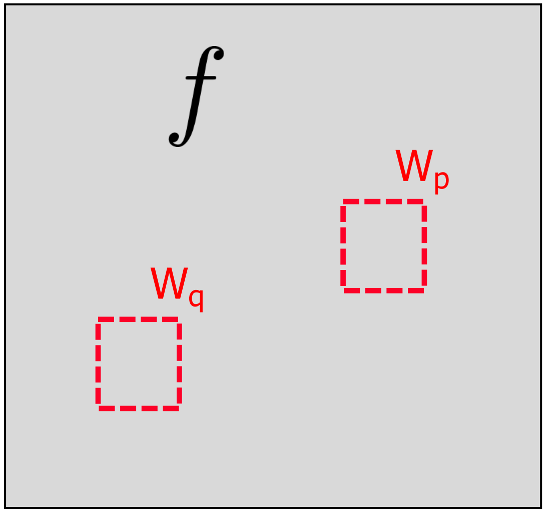
Need to sum pixel values at all possible windows $\square$



Then, SSD(p,d) between W_p and W'_{p-d} is $\sum_{(x,y)\in W_p} \Delta I_d^2(x,y)$

computing SSD efficiently

How to sum pixel values at all possible windows $\square$ in any given image f <u>efficiently</u> ?



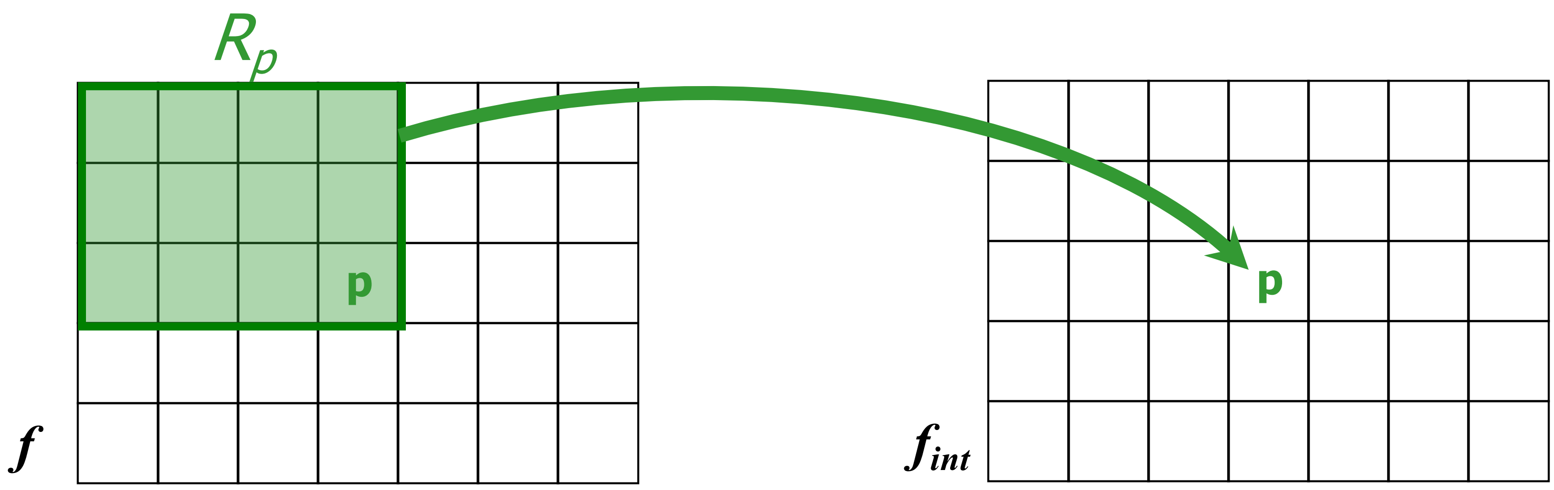
For SSD we just have special case $f(x,y) \equiv \Delta I_d^2(x,y)$

standard general trick:

"Integral Images"

$$f_{int}(p) := \sum_{q \in R_p} f(q)$$

- Define integral image $f_{int}(p)$ as the sum (integral) of image f over pixels in rectangle $R_p := \{q \mid \text{“}q \leq p\text{”}\}$



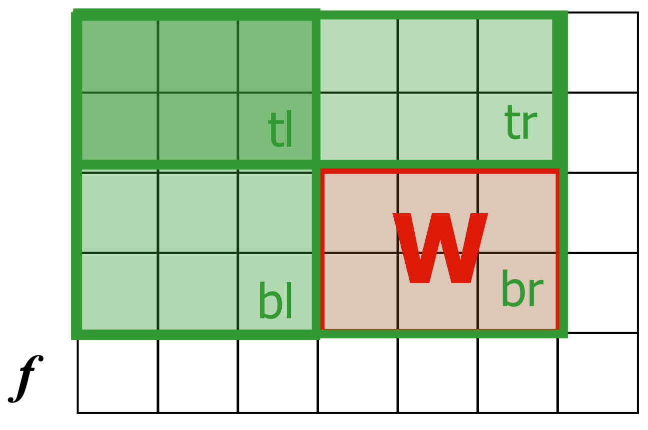
- Can compute $f_{int}(p)$ for all p in two passes over image f (How?)

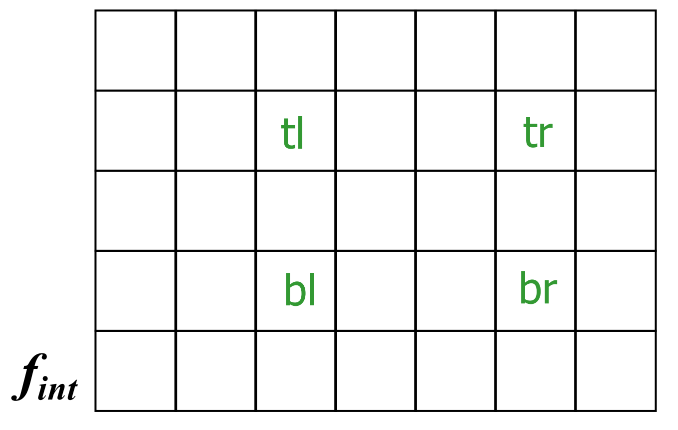
standard general trick:

"Integral Images"

$$f_{int}(p) := \sum_{q \in R_p} f(q)$$

- Define integral image $f_{int}(p)$ as the sum (integral) of image f over pixels in rectangle $R_p := \{q \mid \text{“}q \leq p\text{”}\}$

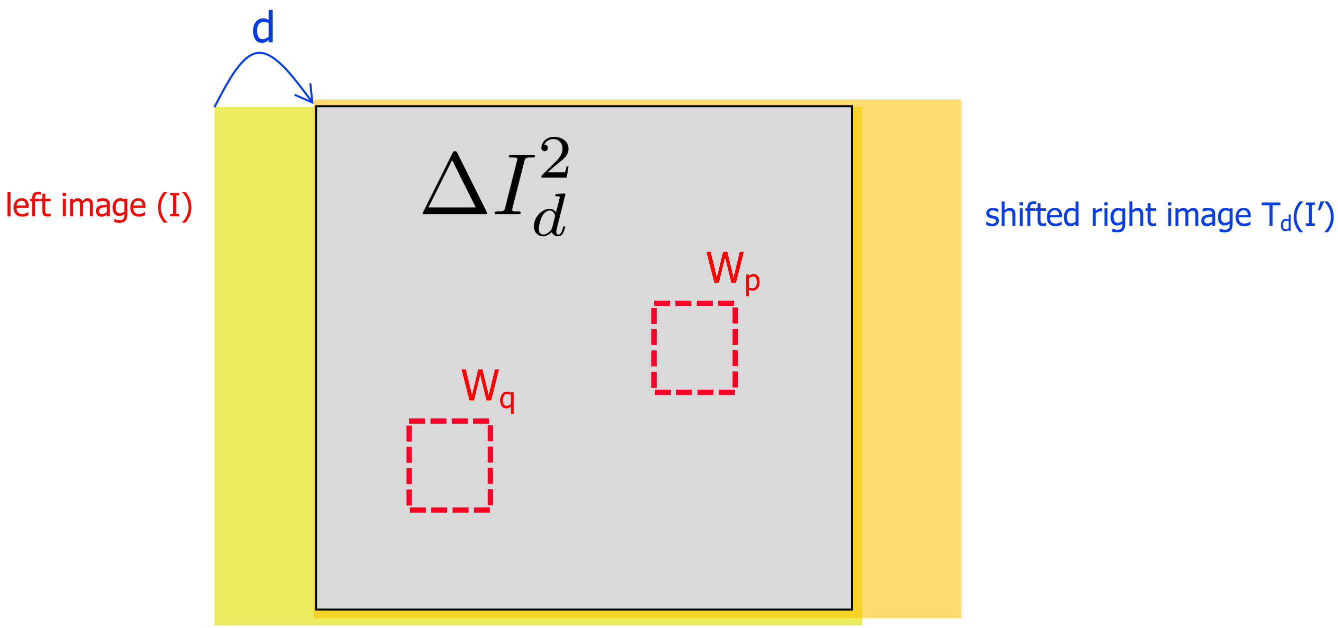




- Now, for any W the sum (integral) of f inside that window can be computed as $\sum_{q \in W} f(q) = f_{int}(\text{br}) - f_{int}(\text{bl}) - f_{int}(\text{tr}) + f_{int}(\text{tl})$

computing SSD efficiently

For SSD we have special case $f(x,y) \equiv \Delta I_d^2(x,y)$



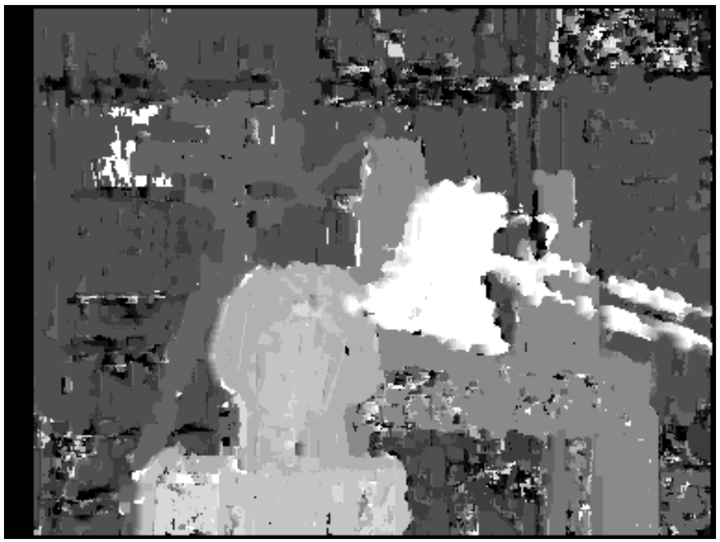
Now, the sum of ΔI_d^2 at any window $\square$ takes 4 operations <u>independently of window size</u>

=> O(|I|*|d|) window-based stereo algorithm

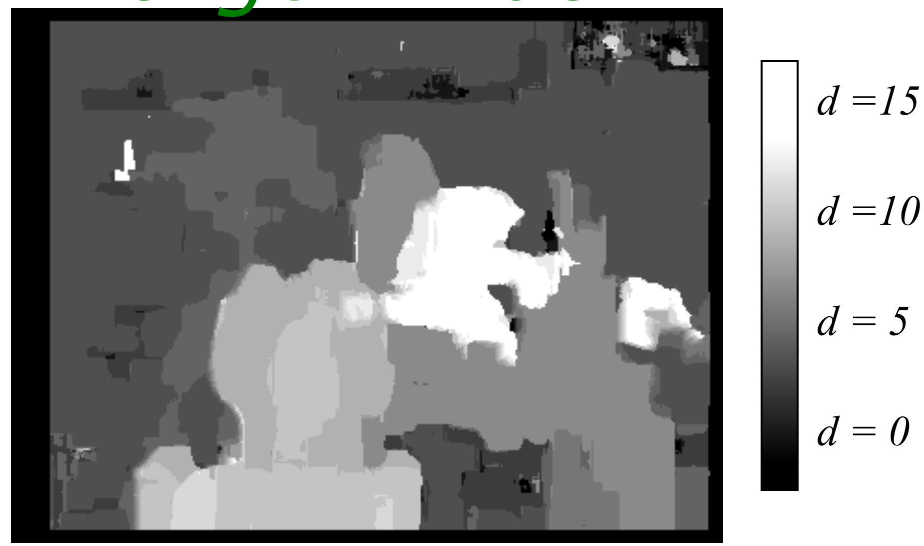
Problems with Fixed Windows

disparity maps $\hat{d}_p = \arg\min_d SSD(p,d)$ for:

small window



large window



- better at boundaries
- noisy in low texture areas

- better in low texture areas
- blurred boundaries

Q: what do we implicitly assume when using low SSD(d,p) at a window around pixel p as a criteria for "good" disparity d ?

window algorithms

- Maybe variable window size (pixel specific)?
  - What is the right window size?
  - Correspondences are still found <u>independently</u> at each pixel (no coherence)

- All window-based solutions can be though of as "local" solutions - but very fast!

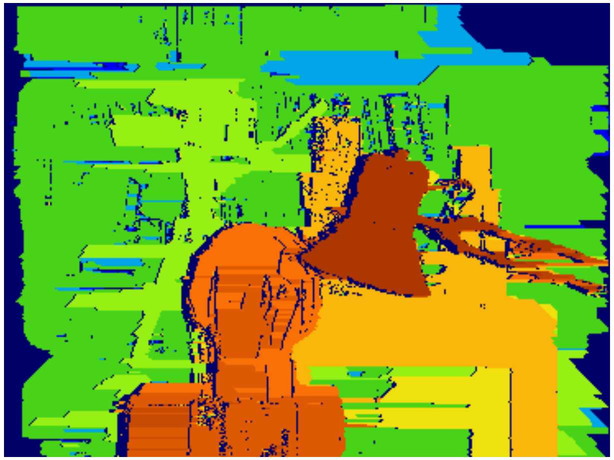
- How to go to "global" solutions?

need priors
to compensate
for local data ambiguity

  - use *objectives* (a.k.a. *energy* or *loss* functions)
    - *regularization* (e.g. spatial coherence)
  - optimization

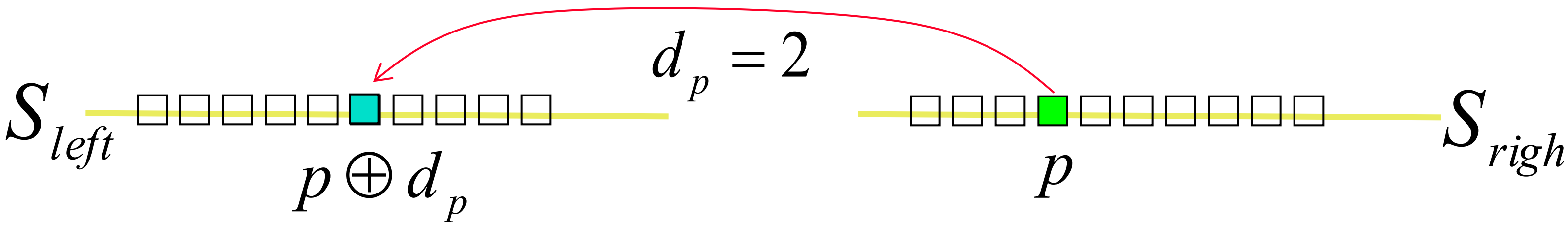
Stereo Correspondence problem
Scan-line approach

- Scan-line stereo
  - coherently match pixels in each scan line
  - DP or shortest paths work (easy 1D optimization)
  - Note: scan lines are still matched <u>independently</u>
    - – streaking artifacts

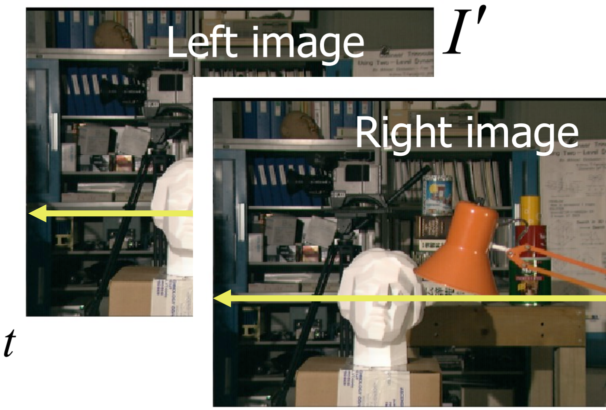


NOTE: *Loss*, *Energy*, *Cost*, and *Objective* function
mean the same thing and are used indiscriminately in different contexts

DP for scan-line stereo



Viterbi algorithm can be used to optimize the following **loss function** over *disparities* $\mathbf{d} = \{d_p \mid p \in S\}$ for pixels p on a given scan-line S_{right}



I' I

$$E(\mathbf{d}) = \sum_{p\in S} \underbrace{D_p(d_p)}_{=\ |I_p - I'_{p\oplus d_p}|} + \sum_{p\in S} \underbrace{V(d_p, d_{p+1})}_{=\ w\,|d_p - d_{p+1}|} = \sum_{\{p,q\}\in N} E(d_p, d_q)$$

photo consistency

spatial coherence

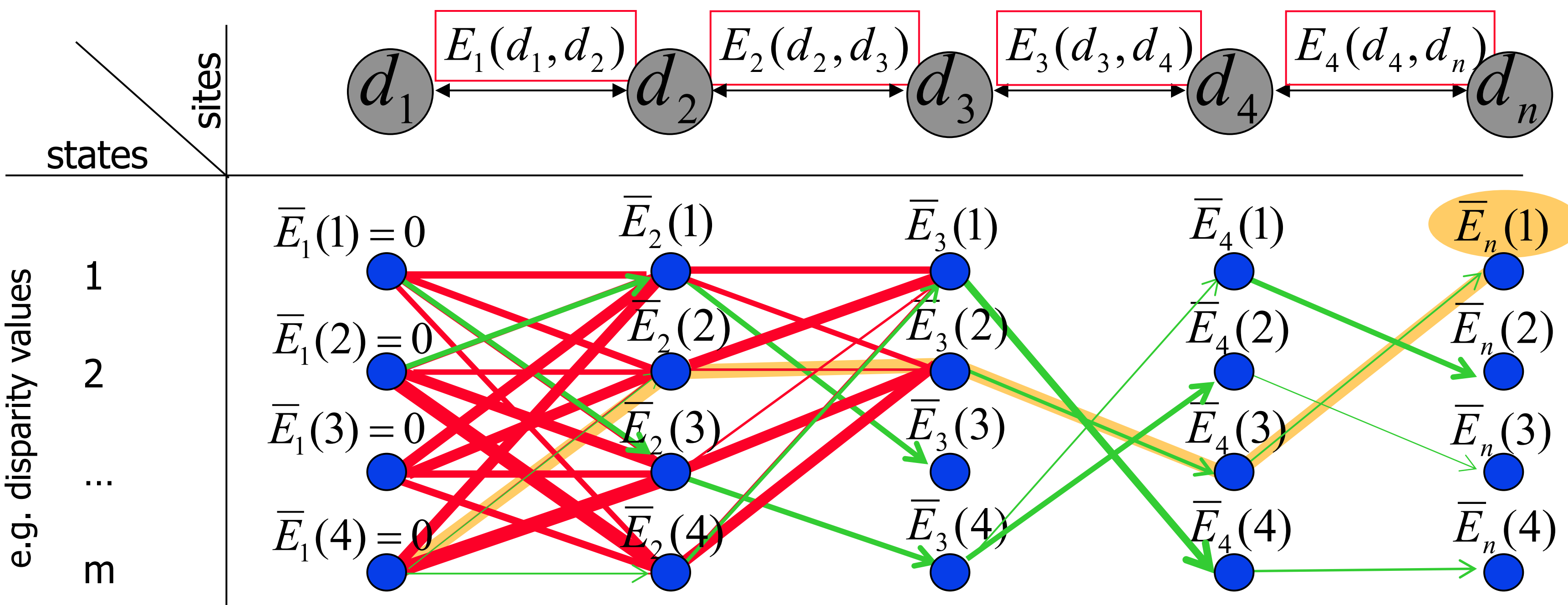
Such pairwise loss can be optimized in $O(nm^2)$ on non-loopy graphs (e.g., chains)

Dynamic Programming (DP)
Viterbi Algorithm

$d_i \in \{0, 1, ..., m-1\}$

Consider **pair-wise interactions** between sites (pixels) on a **chain** (scan-line)

$$E(d_1, ..., d_n) \;=\; E_1(d_1, d_2) + E_2(d_2, d_3) + \cdots + E_{n-1}(d_{n-1}, d_n)$$



$\bar{E}_p(k)$ - internal "loss counter" at "site" p and "state" k

$$\bar{E}_{p+1}(k) = \min_i \left(\bar{E}_p(i) + E_p(i,k) \right)$$

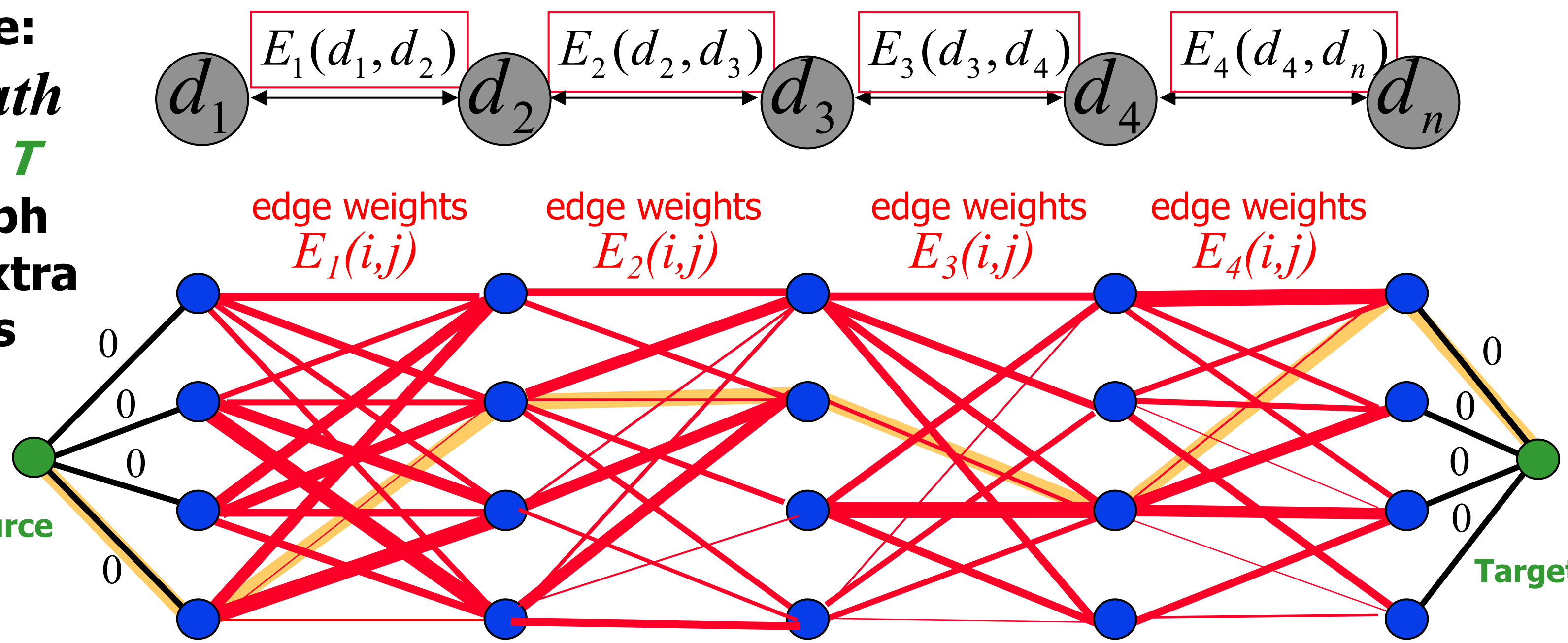
Complexity: $O(nm^2)$, worst case = best case

Dynamic Programming (DP)
Shortest paths Algorithm

Consider **pair-wise interactions** between sites (pixels) on a **chain** (scan-line)

$$E_1(d_1,d_2)+E_2(d_2,d_3)+\ldots+E_{n-1}(d_{n-1},d_n)$$

Alternative:
shortest path
from *S* to *T*
on the graph
with two extra
terminals



Complexity: $O(nm^2+nm\,log(nm))$ - worst case

But, the best case could be better than Viterbi. Why?

Coherent disparity map on 2D grid?

- Scan-line stereo generates streaking artifacts
- Can't use Viterbi to find globally optimal solutions on loopy graphs (e.g. grids) ☹

(Note: other algorithms apply, e.g. *graph cuts*, *belief propagation*, *TRWS*, etc)

- Regularization problems in vision is an interesting domain for optimization algorithms

(Note: it is known that *gradient descent* does not work well for such problems)

Note: we will go over **graph cut** methods that can find globally optimal solutions for certain energies/losses on arbitrary (loopy) graphs

Estimating (optimizing) disparities: over **points** vs. **scan-lines** vs. **grid**

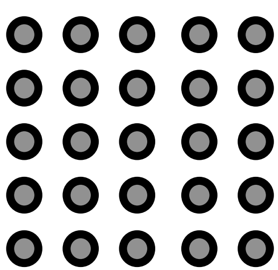
Consider energy (loss) function over disparities

$\mathbf{d} = \{d_p \mid p \in G\}$ for pixels p on grid G

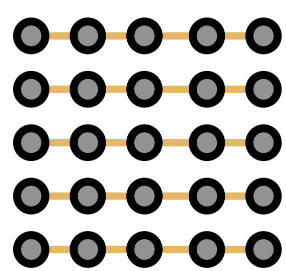
$$E(\mathbf{d}) = \sum_{p \in G} \underbrace{D_p(d_p)}_{= \, |I_p - I'_{p \oplus d_p}|} + \sum_{\{p,q\} \in N} \underbrace{V(d_p, d_q)}_{= \, w\,|d_p - d_q|}$$

photo consistency — spatial coherence

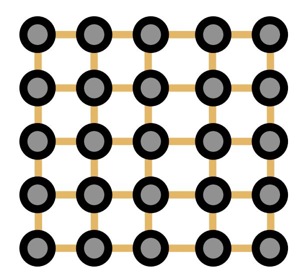
Consider three different <u>neighborhood systems N</u>:



$N = \varnothing$



$N = \{\{p, p \pm 1\} : p \in G\}$



$N = \{\{p,q\} \subset G : |pq| \leq 1\}$

Estimating (optimizing) disparities: over **points** vs. **scan-lines** vs. **grid**

Consider energy (loss) function over disparities

$\mathbf{d} = \{d_p \mid p \in G\}$ for pixels p on grid G

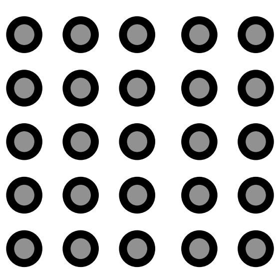
$$E(\mathbf{d}) = \sum_{p \in G} \underbrace{D_p(d_p)}_{= |I_p - I'_{p \oplus d_p}|} + \cancel{\sum_{\{p,q\} \in N} \underbrace{V(d_p, d_q)}_{= w\,|d_p - d_q|}}$$

photo consistency

spatial coherence

smoothness term disappears

CASE 1



$N = \varnothing$

Q: how to optimize $E(\mathbf{d})$ in this case?

$$\forall p \in G \qquad \boxed{\hat{d}_p = \arg\min_d D_p(d)} \qquad O(nm)$$

Q: How does this relate to window-based stereo?

Estimating (optimizing) disparities: over **points** vs. **scan-lines** vs. **grid**

Consider energy (loss) function over disparities
$\mathbf{d} = \{d_p \mid p \in G\}$ for pixels p on grid G

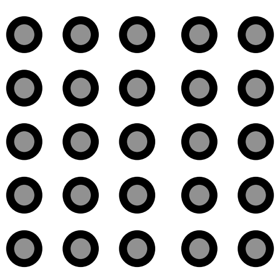
$$E(\mathbf{d}) = \sum_{p\in G} \underbrace{D_p(d_p)}_{=\,\boxed{|I_p - I'_{p\oplus d_p}|}} + \cancel{\sum_{\{p,q\}\in N} \underbrace{V(d_p,d_q)}_{=\,\boxed{w\,|d_p - d_q|}}}$$

photo consistency

~~spatial coherence~~

smoothness term disappears

CASE 1



$N = \varnothing$

Nodes/pixels do not interact (are independent).
Optimization of the sum of **unary terms**,
e.g. $\sum_{p\in G} D_p(d_p)$, is trivial: $O(nm)$

Estimating (optimizing) disparities: over **points** vs. **scan-lines** vs. **grid**

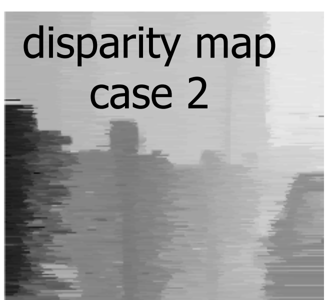
Consider energy (loss) function over disparities

$\mathbf{d} = \{d_p \mid p \in G\}$ for pixels p on grid G

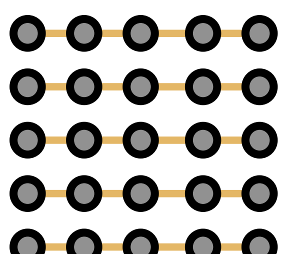
$$E(\mathbf{d}) = \sum_{p \in G} \underbrace{D_p(d_p)}_{= \, |I_p - I'_{p \oplus d_p}|} + \sum_{\{p,q\} \in N} \underbrace{V(d_p, d_q)}_{= \, w\,|d_p - d_q|}$$

photo consistency

spatial coherence



CASE 2



$N = \{\{p, p \pm 1\} : p \in G\}$

Pairwise coherence is enforced, but only between pixels on the same scan line.

Q: how do we optimize such $E(\mathbf{d})$?

$O(nm^2)$

Estimating (optimizing) disparities: over **points** vs. **scan-lines** vs. **grid**

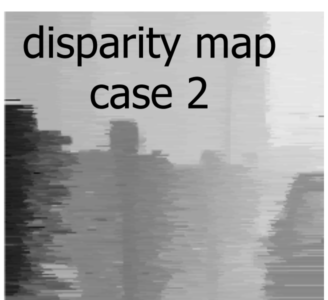
Consider energy (loss) function over disparities

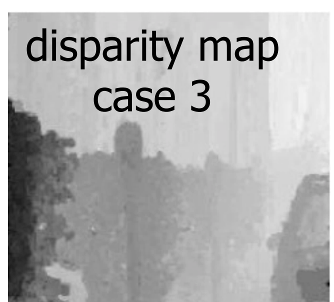
$\mathbf{d} = \{d_p \mid p \in G\}$ for pixels p on grid G

$$E(\mathbf{d}) = \sum_{p \in G} \underbrace{D_p(d_p)}_{= \, |I_p - I'_{p \oplus d_p}|} + \sum_{\{p,q\} \in N} \underbrace{V(d_p, d_q)}_{= \, w\,|d_p - d_q|}$$

photo consistency

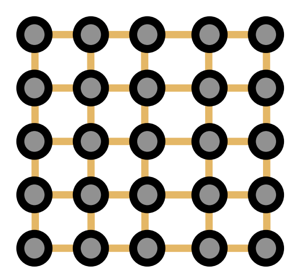
spatial coherence





Pairwise smoothness of the disparity map is enforced both horizontally and vertically.

CASE 3



$$N = \{\{p,q\} \subset G : |pq| \leq 1\}$$

NOTE: *depth map* regularity/smoothness should be isotropic since 3D scene surface is independent of scan-lines (epiplar lines) orientation.

Estimating (optimizing) disparities: over **points** vs. **scan-lines** vs. **grid**

Consider energy (loss) function over disparities

$\mathbf{d} = \{d_p \mid p \in G\}$ for pixels p on grid G

$$E(\mathbf{d}) = \sum_{p \in G} \underbrace{D_p(d_p)}_{= \, |I_p - I'_{p \oplus d_p}|} + \sum_{\{p,q\} \in N} \underbrace{V(d_p, d_q)}_{= \, w\,|d_p - d_q|}$$

photo consistency

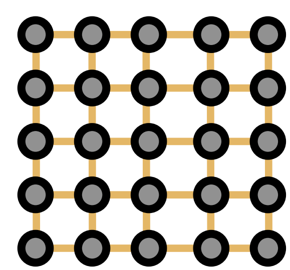
spatial coherence

How to optimize "pairwise" loss on loopy graphs?

NOTE 1: **Viterbi does not apply**, but its extensions (e.g. *message passing*) provide approximate solutions on loopy graphs.

NOTE 2: "*Gradient descent*" can find only local minima for a continuous relaxation of $E(\boldsymbol{d})$ combining non-convex photo-consistency (1st term) and convex *total variation* of $\boldsymbol{d}$ (2nd term).

CASE 3



$N = \{\{p,q\} \subset G : |pq| \leq 1\}$

Graph cut
for spatially coherent stereo on 2D grids

One can **globally minimize** the following energy of disparities $\mathbf{d} = \{d_p \mid p \in G\}$ for pixels p on grid G

$$E(\mathbf{d}) = \sum_{p \in G} \underbrace{D_p(d_p)}_{\substack{\shortparallel \\ \boxed{|I_p - I'_{p \oplus d_p}|}}} + \sum_{\{p,q\} \in N} \underbrace{V(d_p, d_q)}_{\substack{\shortparallel \\ \boxed{w_{pq} \cdot |d_p - d_q|}}}$$

photo consistency

spatial coherence

Unlike shortest paths or Viterbi,
s/t **graph cut algorithms** can
globally minimize certain types of
pairwise energies **on loopy graphs**.

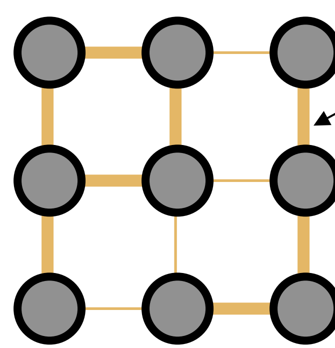
Graph cut
for spatially coherent stereo on 2D grids

One can **globally minimize** the following loss over disparities $\mathbf{d} = \{d_p \mid p \in G\}$ for pixels p on grid G

$$E(\mathbf{d}) = \sum_{p \in G} \underbrace{D_p(d_p)}_{\substack{\| \\ |I_p - I'_{p \oplus d_p}|}} + \sum_{\{p,q\} \in N} \underbrace{V(d_p, d_q)}_{\substack{\| \\ w_{pq} \cdot |d_p - d_q|}}$$

photo consistency

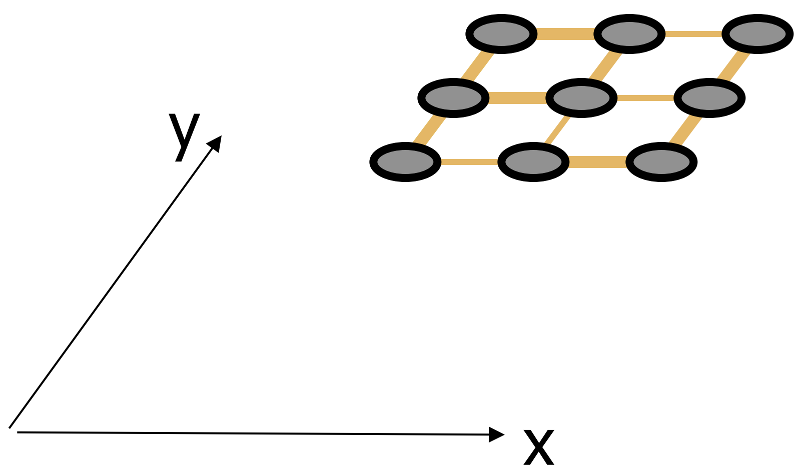
spatial coherence



w_{pq} - weights of neighborhood edges

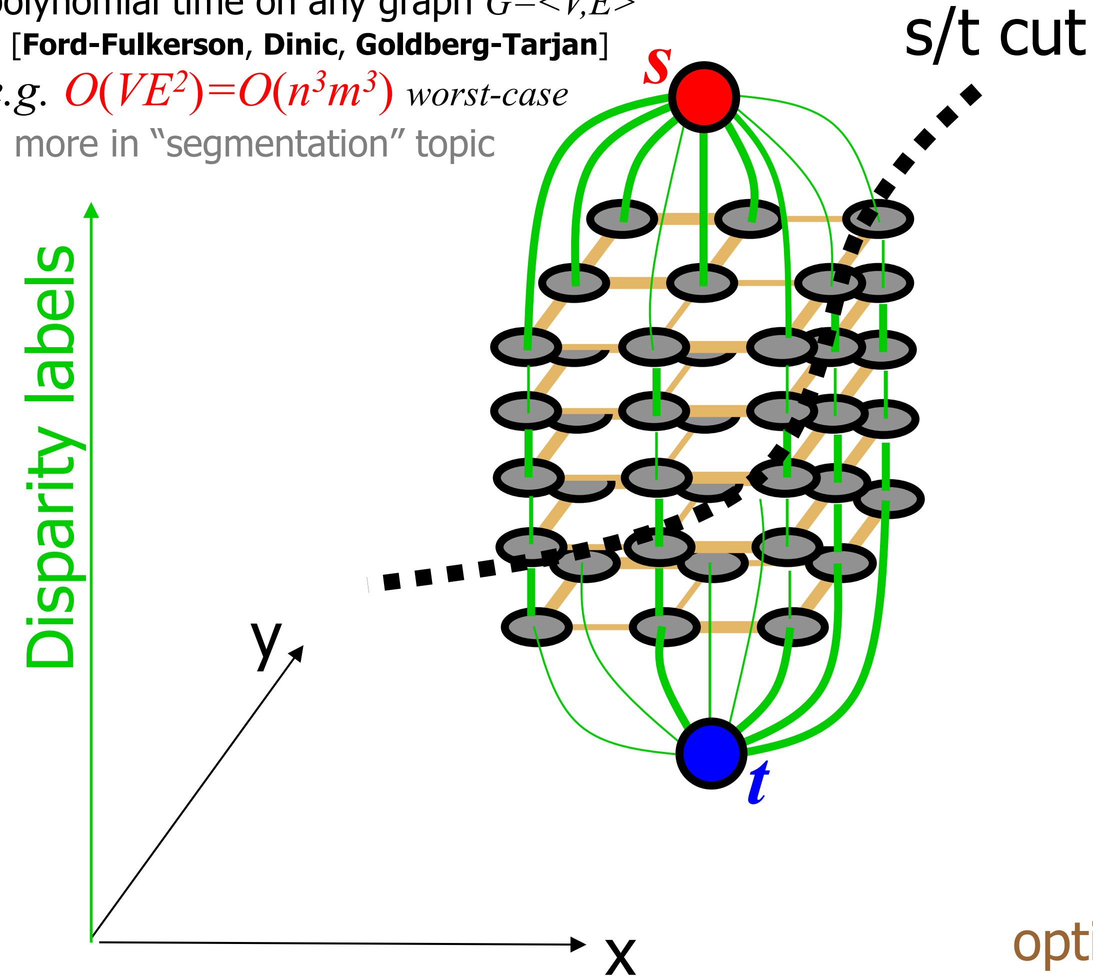
(e.g. may be assigned according to local intensity contrast)

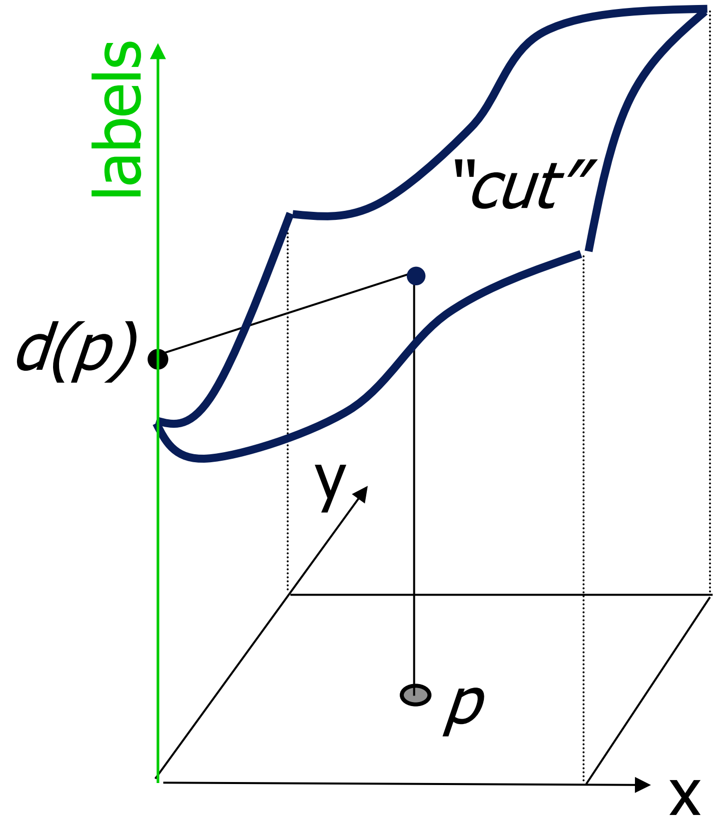
Multi-scan-line stereo with *s-t* graph cuts [Roy&Cox'98, Ishikawa 98]



Multi-scan-line stereo with *s-t* graph cuts [Roy&Cox'98, Ishikawa 98]

Minimum s/t cuts can be found in low-order polynomial time on any graph $G=\langle V,E \rangle$
[**Ford-Fulkerson**, **Dinic**, **Goldberg-Tarjan**]
e.g. $O(VE^2)=O(n^3m^3)$ *worst-case*
more in "segmentation" topic



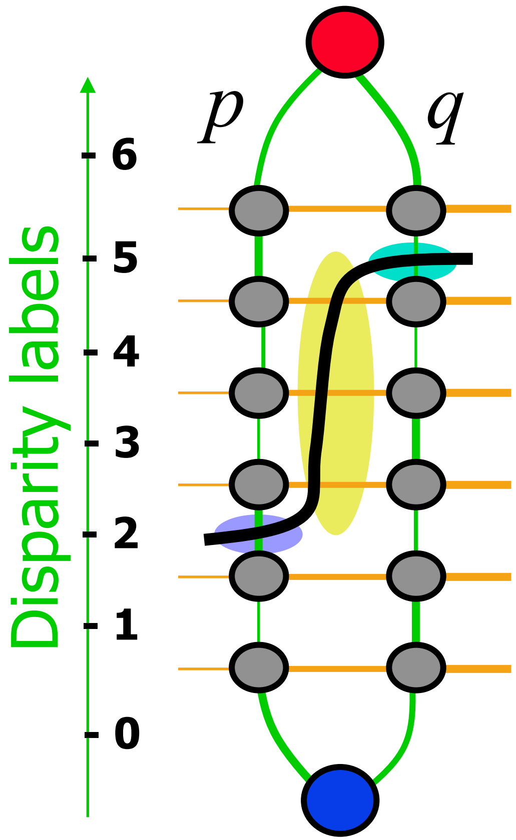


minimum cut will define
optimal disparity map **d** = $\{d_p\}$

Ishikawa&Geiger 98

What loss function do we minimize this way?

Concentrate on one pair of neighboring pixels $\{p, q\} \in N$



cost of vertical edges

$$E(d_p, d_q) = \boxed{D_p(2)} + \boxed{D_q(5)} + \ldots$$

$$+ \quad \boxed{w_{pq} \cdot |3|} \quad + \ldots$$

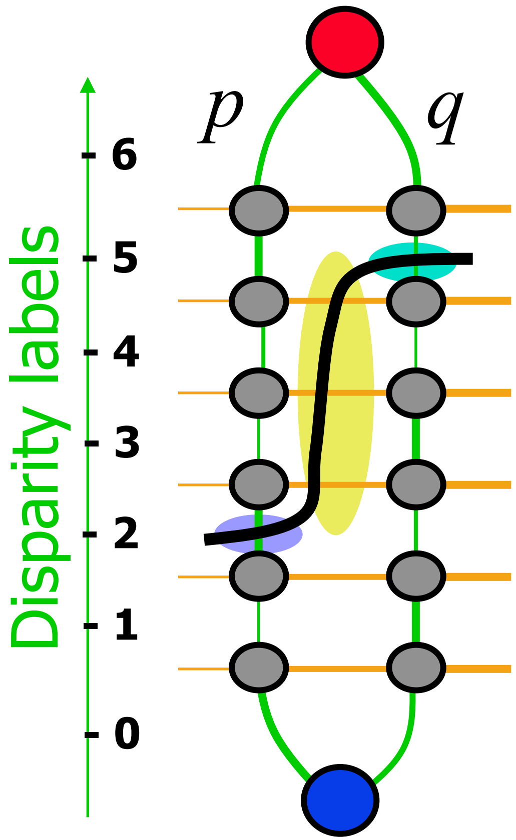
cost of horizontal edges

assuming cut has no folds (optional slides later shows how to make sure)

Ishikawa&Geiger 98

What loss function do we minimize this way?

Concentrate on one pair of neighboring pixels $\{p,q\} \in N$



cost of vertical edges

$$E(d_p, d_q) = \boxed{D_p(d_p)} + \boxed{D_q(d_q)} + \ldots$$

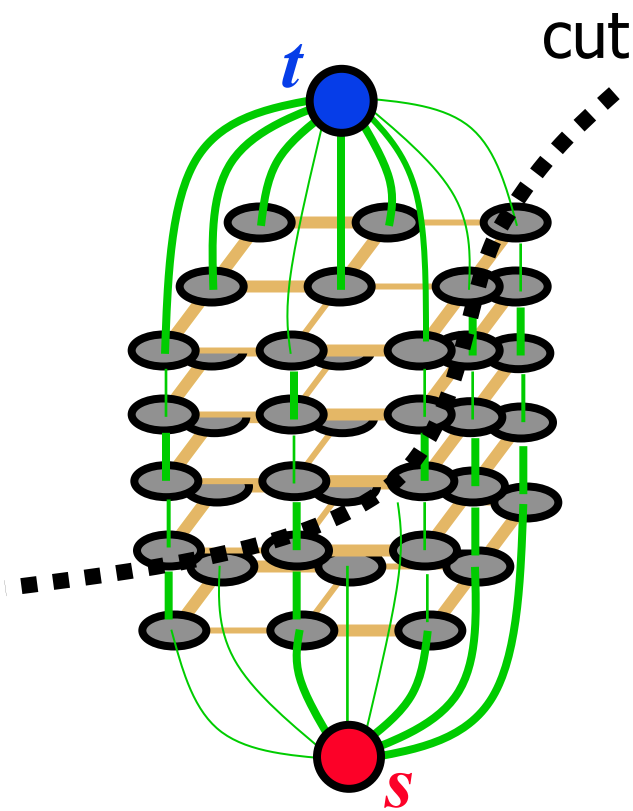
$$+ \boxed{w_{pq} \cdot | d_p - d_q |} + \ldots$$

cost of horizontal edges

Ishikawa&Geiger 98

What loss function do we minimize this way?

The combined loss over the entire grid G is

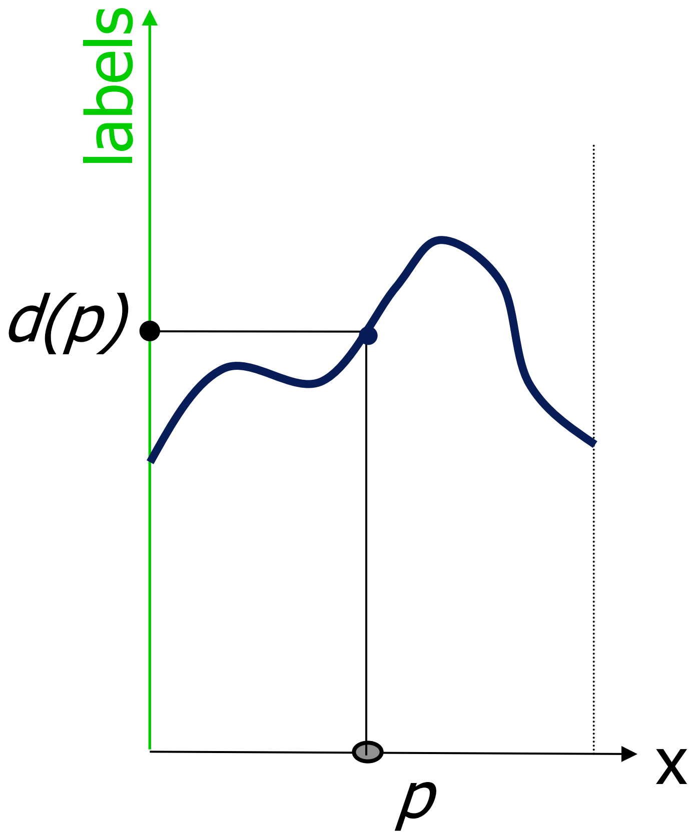


(**photo consistency**, e.g. SSD)
cost of vertical edges

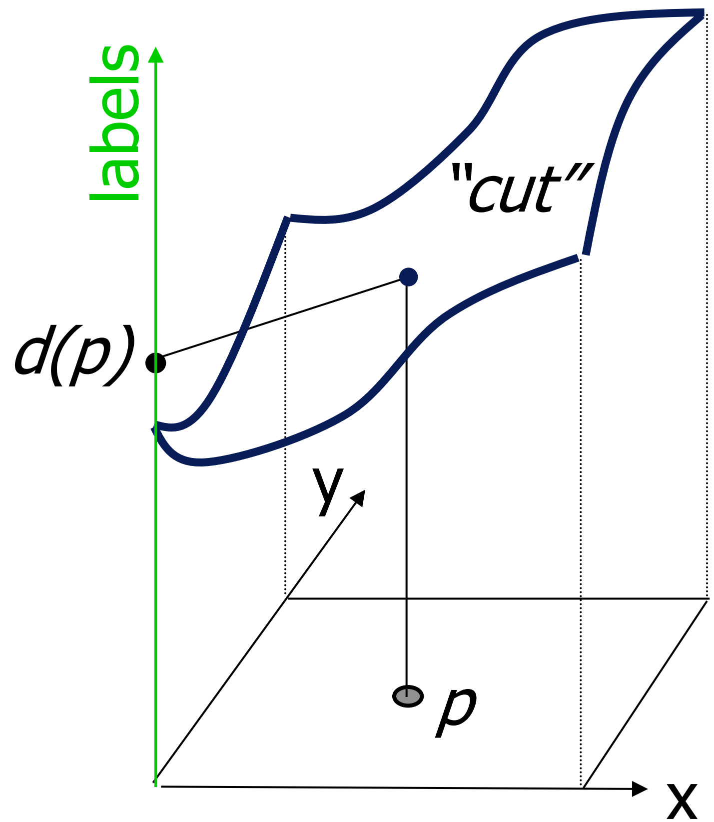
$$E(\mathbf{d}) = \sum_{p \in G} D_p(d_p) + \sum_{\{p,q\} \in N} w_{pq} \cdot |d_p - d_q|$$

cost of horizontal edges
(**spatial consistency**)

Scan-line stereo vs. Multi-scan-line stereo (on whole grid)



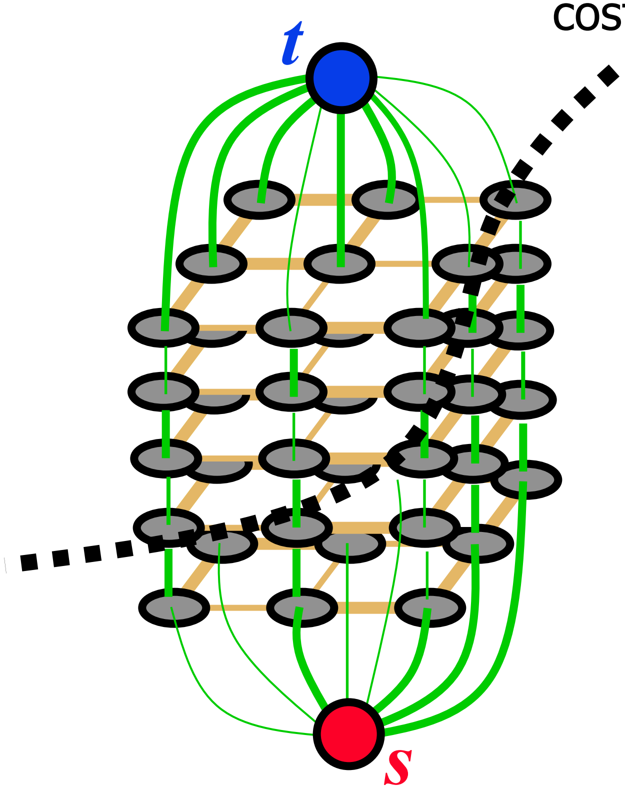
Dynamic Programming
(single scan line optimization)

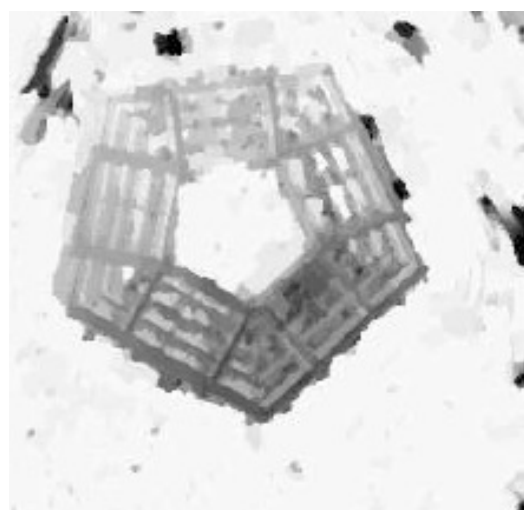


s-t Graph Cuts
(grid optimization)

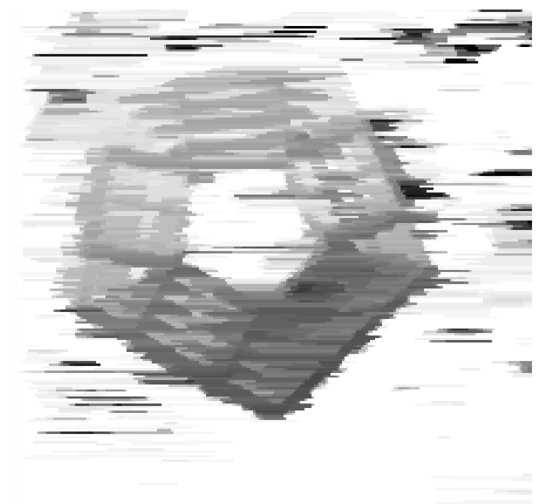
Some results from Roy&Cox

minimum
cost s/t cut



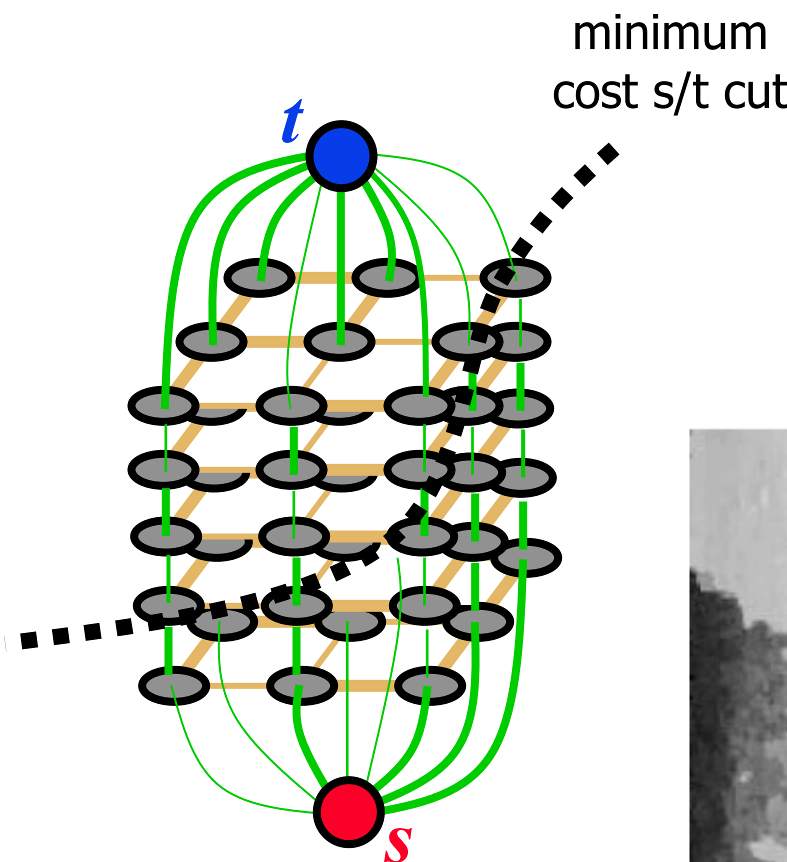


multi scan line stereo
(graph cuts)



single scan-line stereo
(DP)

Some results from Roy&Cox



multi scan line stereo
(graph cuts)

single scan-line stereo
(DP)

Simple Examples: Stereo with only 2 depth layers

binary stereo

essentially,
depth-based binary
segmentation

[Kolmogorov et al. CVPR 2005, IJCV 2008]

Simple Examples:
Stereo with only 2 depth layers

background substitution

essentially,
depth-based binary
segmentation

[Kolmogorov et al. CVPR 2005, IJCV 2008]