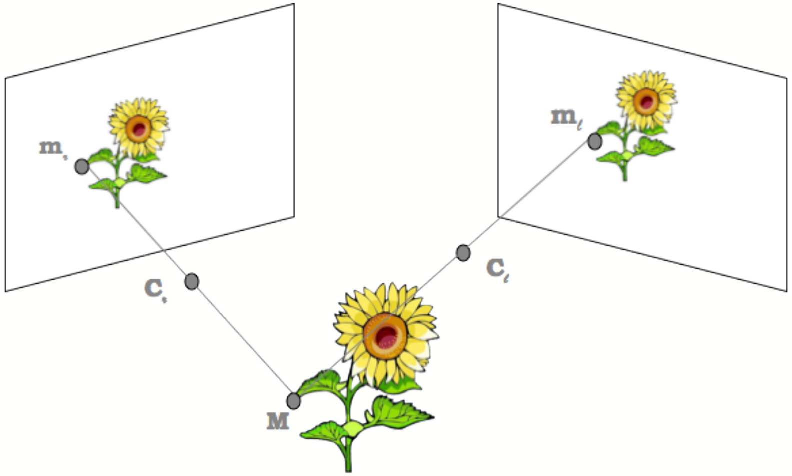
Stereo Matching: an Overview

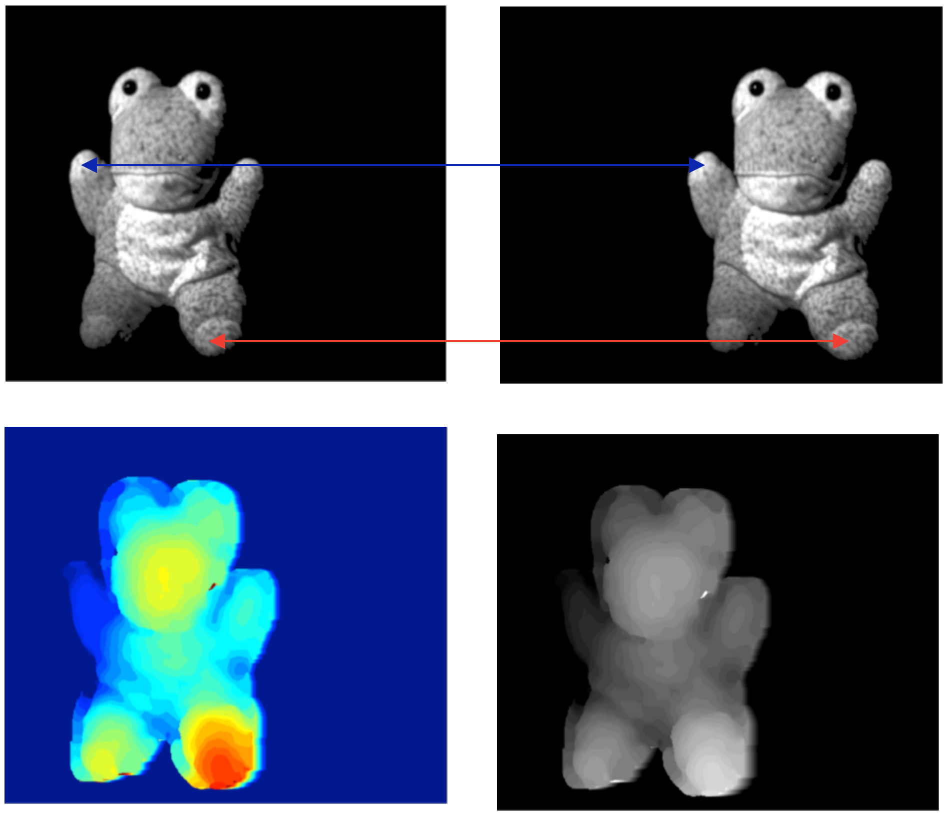
Andrea Fusiello
http://www.diegm.uniud.it/fusiello/

Stereo analysis



- Find the (**corresponding**) points $\mathbf{m_l}$ and $\mathbf{m_r}$ in the two images that are projections of the same 3D point $\mathbf{M}$.
- Epipolar constraint reduces the search to one dimension
- Rectification reduces the search along columns
- The horizontal shift of the point is called **disparity**

Left view



Right view

hot=close

light=close

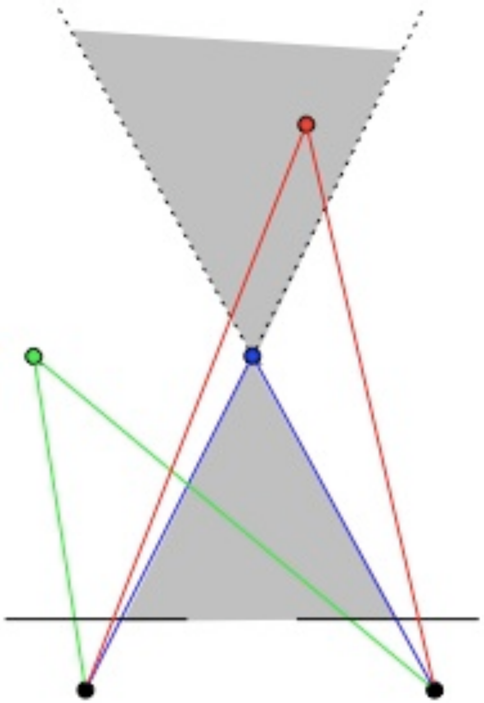
disparity

- The main underlying assumption that allow to search for conjugate points is that iamge patches that are projection of the same surface patch are **similar.**

- This may not be true because of:
  - Occlusions: some points are visible in one image but not in the other
  - Non-lambertian lighting effect: the radiance of non-lambertian surfaces depends on the viewpoint (eg. specular effects)
  - Perspective: the projected shape depends on the viewpoint (eg. Frontal vs slanted)

Constraints

- Similarity constraint
- Epipolar constraint
- Uniqueness constraint: a point in one image has at most one corresponding point in the other image (fails with transparent objects)
- Continuity: disparity is piecewise smooth
- Ordering constraint.
  Fails for points in the forbidden zone



Local vs Global methods

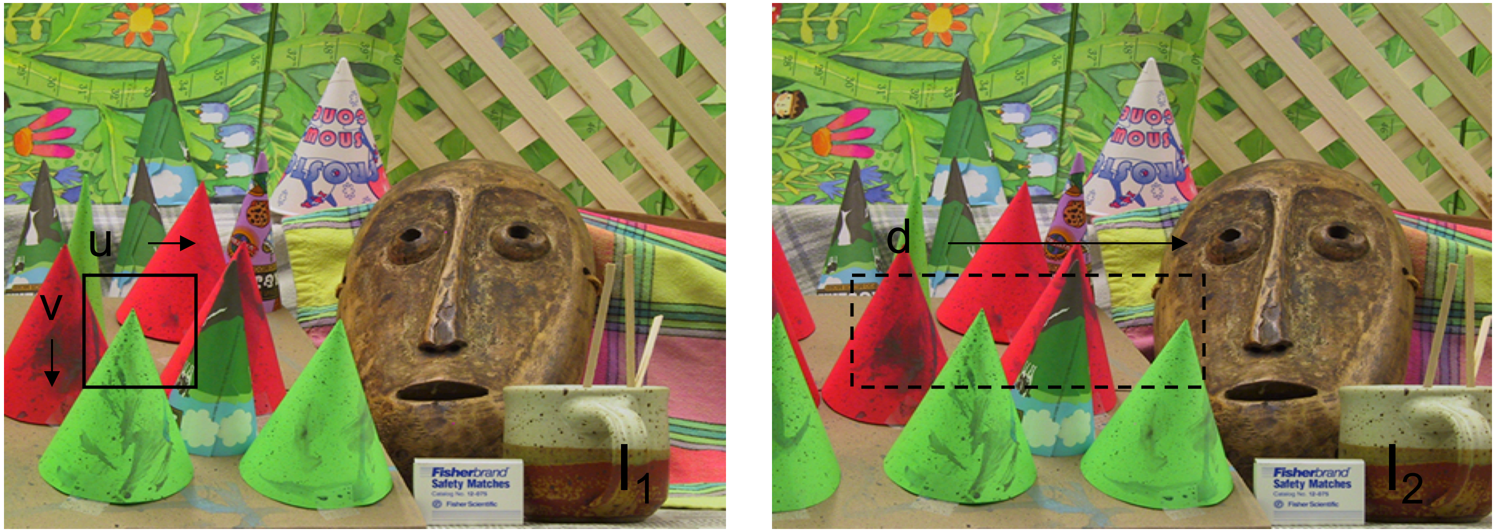
- All methods attempt to match pixels in one image with pixels in the other image by exploiting a number of constraints.

- **Local methods**: use constraints on a small numer of pixels surrounding the pixel of interest.
  - Block matching

- **Global methods**: use constraints on scan-lines or the whole image.
  - Dynamic programming
  - Graph cuts

Block matching

- Estimate disparity at a point by comparing a small region about that point with congruent regions extracted from the other image.
- Three classes of metrics used for the comparision:
  - Correlation (NCC)
  - Intensity difference (SAD, SSD)
  - Rank (rank transform, census transform)



Copied from [BBH03]

- Block matching searches one image for the best corresponding region for a template in the other image.
- Shift the template along the epipolar line in a pre-defined disparity range.

Block-matching costs

MATCH METRIC	DEFINITION
Normalized Cross-Correlation (NCC)	$\dfrac{\sum_{u,v}\left(I_1(u,v)-\bar{I}_1\right)\cdot\left(I_2(u+d,v)-\bar{I}_2\right)}{\sqrt{\sum_{u,v}\left(I_1(u,v)-\bar{I}_1\right)^2\cdot\left(I_2(u+d,v)-\bar{I}_2\right)^2}}$
Sum of Squared Differences (SSD)	$\sum_{u,v}\left(I_1(u,v)-I_2(u+d,v)\right)^2$
Normalized SSD	$\sum_{u,v}\left(\dfrac{\left(I_1(u,v)-\bar{I}_1\right)}{\sqrt{\sum_{u,v}\left(I_1(u,v)-\bar{I}_1\right)^2}}-\dfrac{\left(I_2(u+d,v)-\bar{I}_2\right)}{\sqrt{\sum_{u,v}\left(I_2(u+d,v)-\bar{I}_2\right)^2}}\right)^2$
Sum of Absolute Differences (SAD)	$\sum_{u,v}\left\lvert I_1(u,v)-I_2(u+d,v)\right\rvert$
Rank	$\sum_{u,v}\left(I_1'(u,v)-I_2'(u+d,v)\right)$ $I_k'(u,v)=\sum_{m,n}I_k(m,n)<I_k(u,v)$
Census	$\sum_{u,v}HAMMING\left(I_1'(u,v),I_2'(u+d,v)\right)$ $I_k'(u,v)=BITSTRING_{m,n}\left(I_k(m,n)<I_k(u,v)\right)$

Extracted from [BBH03]

Census transform

- Census transform is defined in a window
- Encode in a bit string whether each pixel of the window is greater or less than the central pixel
- Then compare strings with Hamming distance
- Eliminate sensitivity to absolute intensity and to outliers

89	63	72
67	55	64
58	51	49

00000011

Reliability/accuracy tradeoff

- Reliability: provide sufficient intensity variation inside a window --> large window
- Accuracy: localize the disparity accurately --> small window

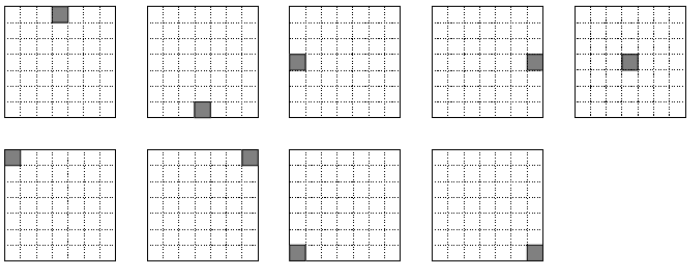
- Solution:
  - Adaptive/shiftable windows [Kanade&Okutomi, Fusiello et al.]
  - Hierarchical approaches [Anandan]

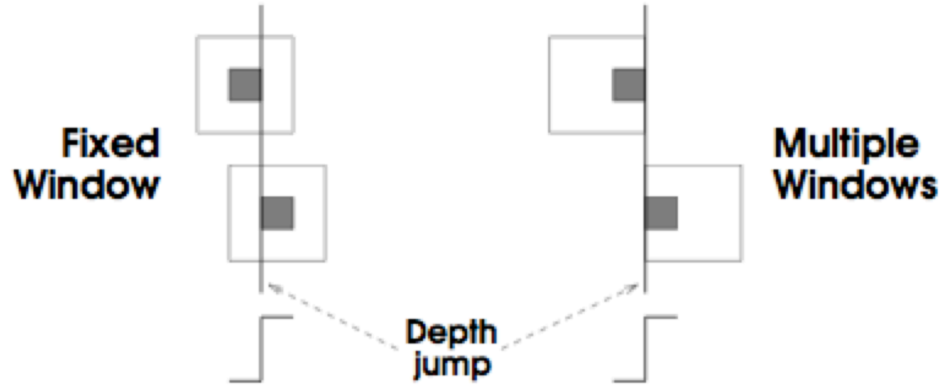
Adaptive windows

- The ideal window should
  - include sufficient intensity variation and
  - not include a depth discontinuity

- Itensity variation is measured in the image
- Depth is unknown
- Solution:
  - Start with a standard window size
  - Adapt the window according to the current disparity estimate
  - Iterate…
- The optimal window for each pixel is computed.
- Computationally expensive, convergence problems.

Shiftable windows



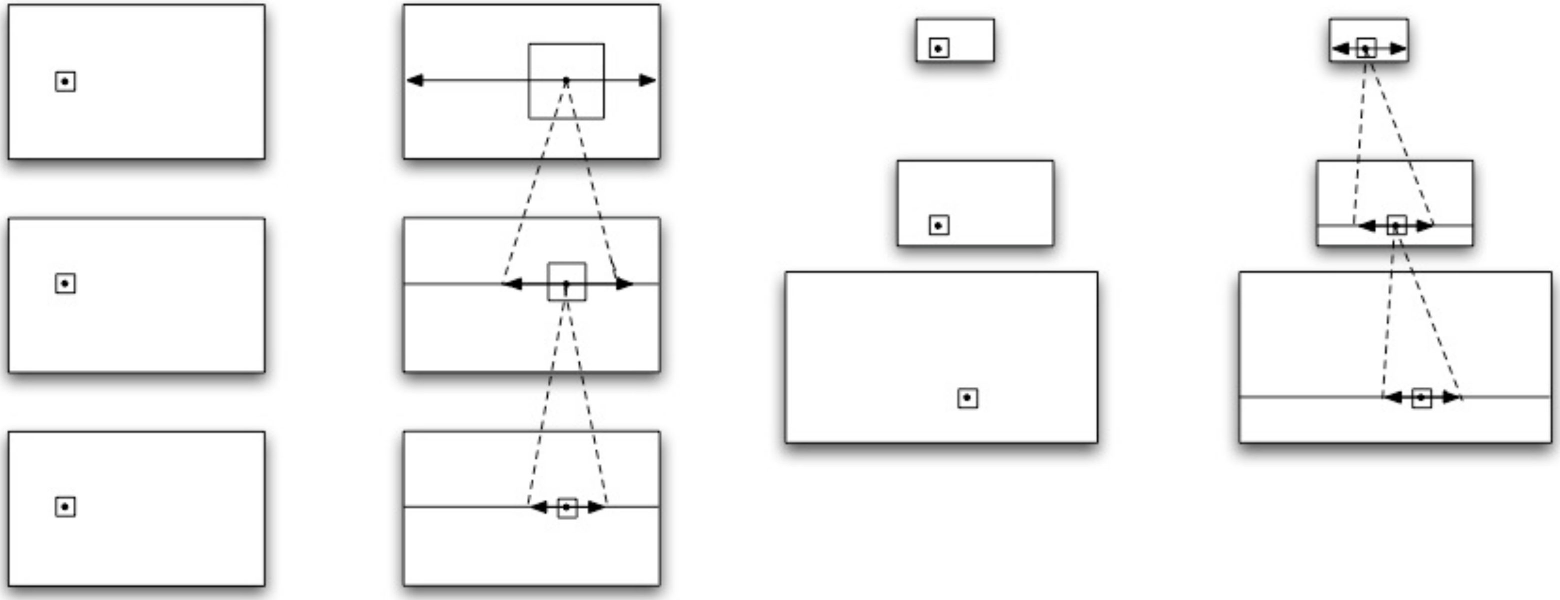


- Nine asymmetric windows are used for each pixel
- The window minimizing the matching score is more likely to cover a region of constant depth
- The effect of trying all the shifted windows around a pixel is the same as taking the matching scores of the non-shifted windows in the same neighborhood
- This method address the accuracy issue; the size if fixed (reliability is taken for granted)

Hierarchical approaches

- At the coarse level large windows provide a reliable but inaccurate disparity estimate
- At the finer levels the accuracy is improved with smaller windows and smaller search areas

Copied from [SKS05]



Fine-to-fine

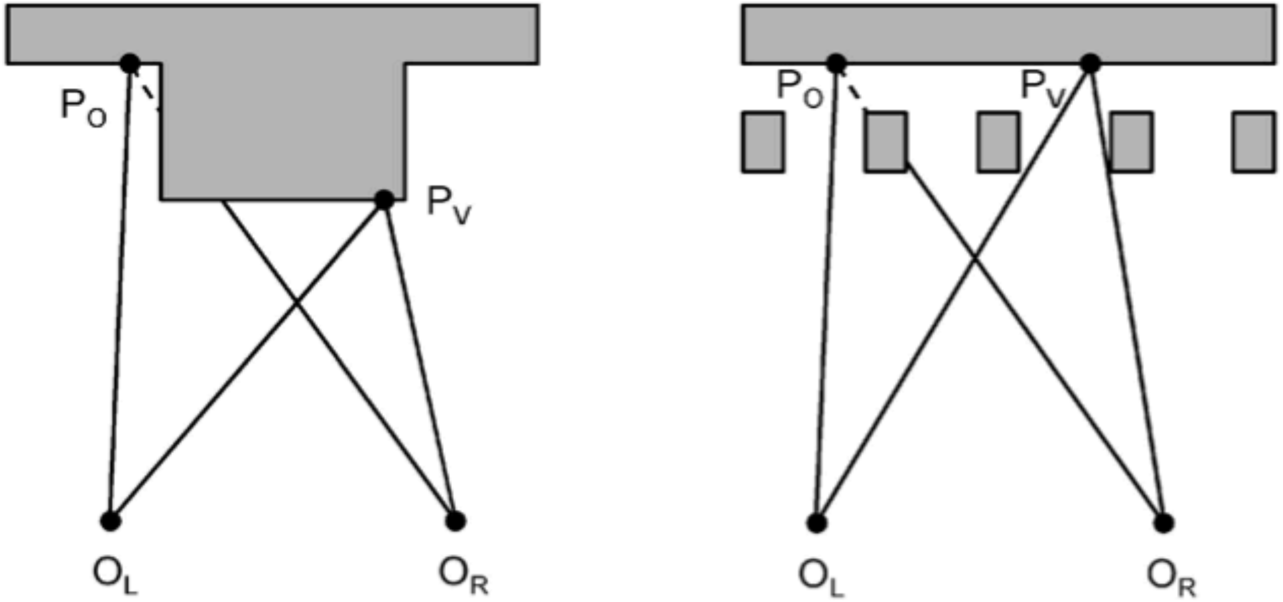
Coarse-to-fine

Post-processing

- Matching reliability indicators:
  - Value of matching score
  - Variance/entropy of intensities
  - Peakness of the matching score (curvature, peak ratio)
  - Neighborhood consistency/smoothness of disparity
  - Left-right consistency (see Occlusion handling)
- Sub-pixel refinment
  - Fitting a curve (spline, parabola) to the matching scores
- Disparity enhancement
  - Median filter, MRF relaxation
  - Hole filling
  - …

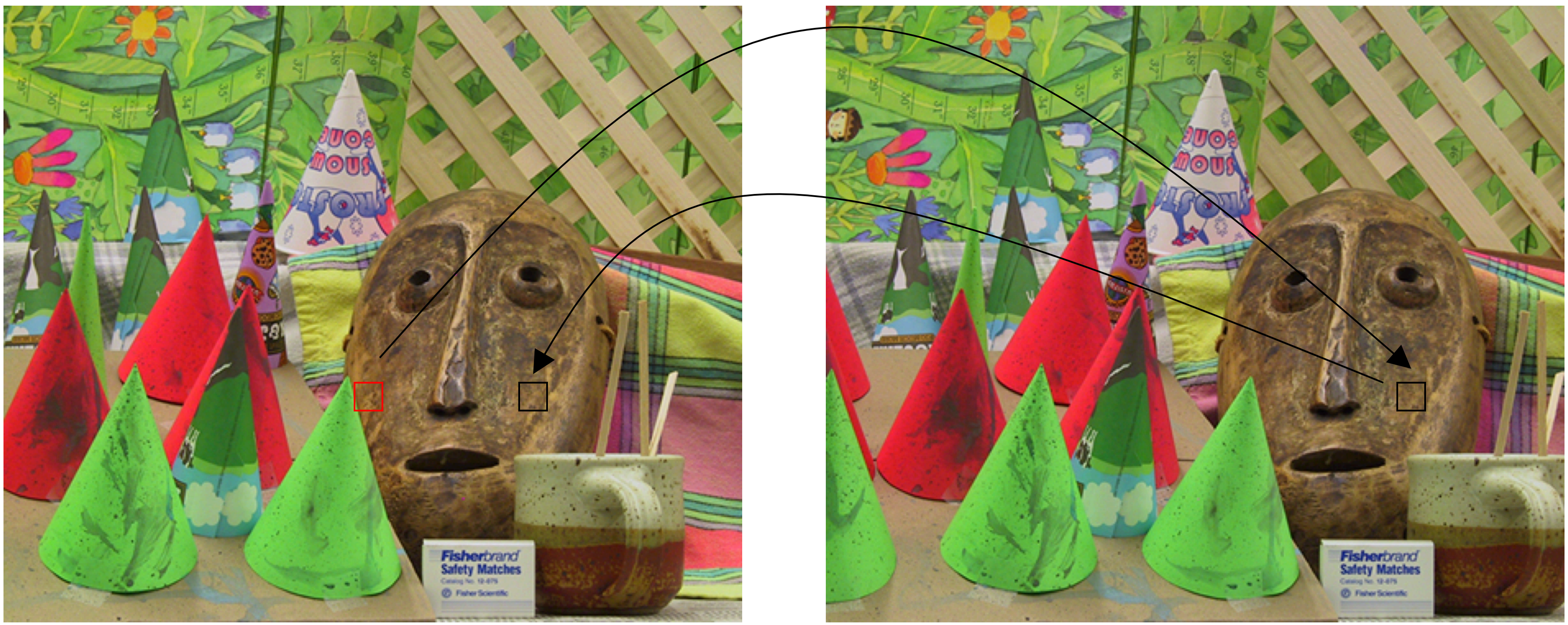
Occlusion handling

- Points visible in one camera but not in the other.
- In the case of narrow occluding objects the ordering constraint fails



Extracted from [BBH03]

- Left-right consistency



- Ordering (assuming no narrow occlusors)

Other local methods

- Gradient-based (Optical Flow) methods
  - Determine small disparities (1/2 pixel) by relating motion and image brightness
- Feature based methods
  - Block matching and gradient-based methods are sensitive to depth discontinuities and uniform regions
  - Solution: limit the correspondence search to reliable features in the images (e.g. Harris corners)
  - Matches are reliable but sparse
  - Segmentation based: first segment the images and then match the segmented regions
  - Produces dense maps but it is sensitive to the original segmentation

Global Methods

- Exploit nonlocal constraints to reduce sensitivity to regions that fail to match (uniform texture, occlusions…)
- Make explicit smoothness assumptions and solve an optimization problem
- Greater computational complexity

Disparity Space Image (DSI)

- 3D image C(x,y,d) is the matching cost of pixel (x,y) in the reference image with pixel (x,y+d) in the other image

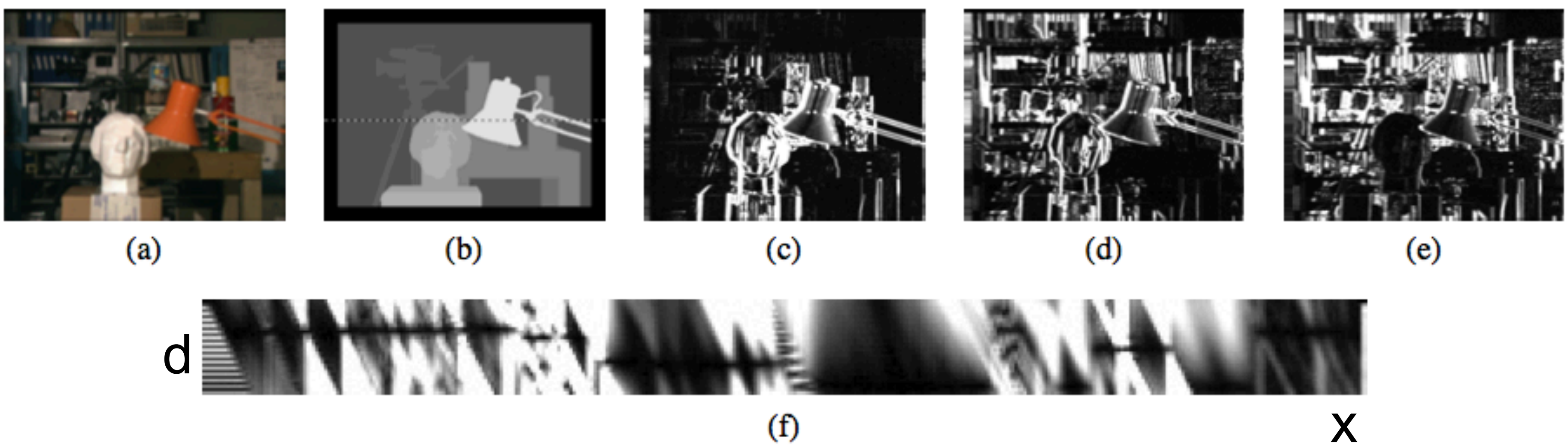


Figure 1: *Slices through a typical disparity space image (DSI): (a) original color image; (b) ground-truth disparities; (c–e) three* (x, y) *slices for* $d = 10, 16, 21$*; (e) an* (x, d) *slice for* $y = 151$ *(the dashed line in Figure (b)). Different dark (matching) regions are visible in Figures (c–e), e.g., the bookshelves, table and cans, and head statue, while three different disparity levels can be seen as horizontal lines in the* (x, d) *slice (Figure (f)). Note the dark bands in the various DSIs, which indicate regions that match at this disparity. (Smaller dark regions are often the result of textureless regions.)*

Extracted from [SS02]

Stereo matching in the DSI

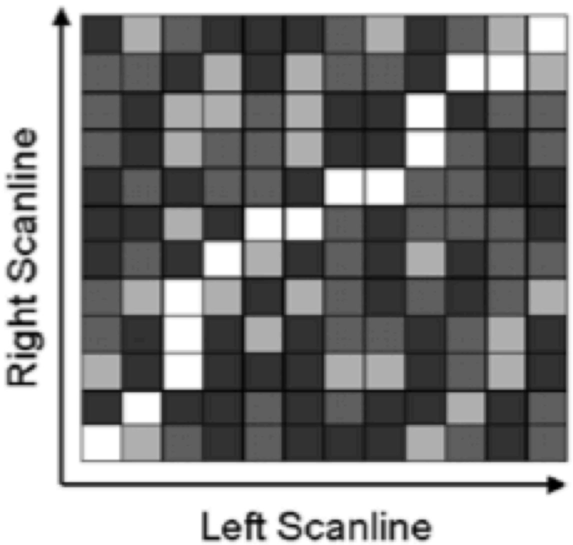
- The goal of a stereo correspondence algorithm is to produce a disparity map d(x,y)
- This can be seen as a surface embedded in the DSI
- The surface must have some optimality properties:
  - Lowerst cost
  - Piecewise smoothness
  - …

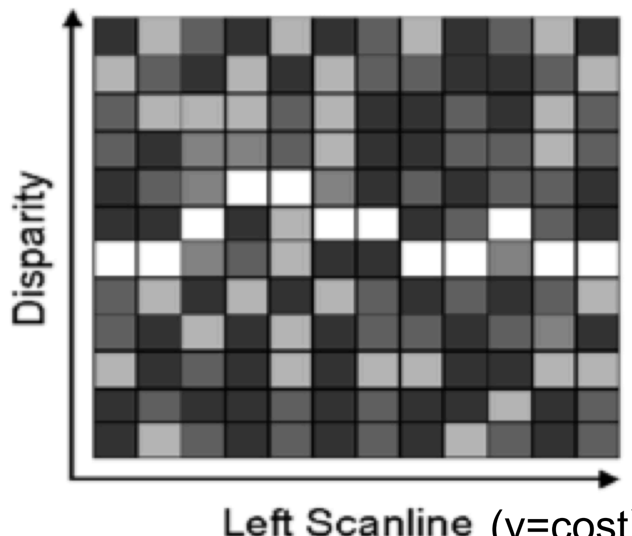
Dynamic programming

- Use ordering and smoothness constraints to optimize correspondences in each scan-line independently.
- The original (2D) problem is decomposed in several simpler ones (1D)
- Vertical coherence is lost (though it may be incorporated)

- Compute the minimum-cost path through a (x,d) slice of the DSI [Intille&Bobick]; or equivalently

- Compute the minimum-cost path through the matrix of all pairwise matching costs between two corresponding scanlines [Otha&Kanade, Cox et al.]

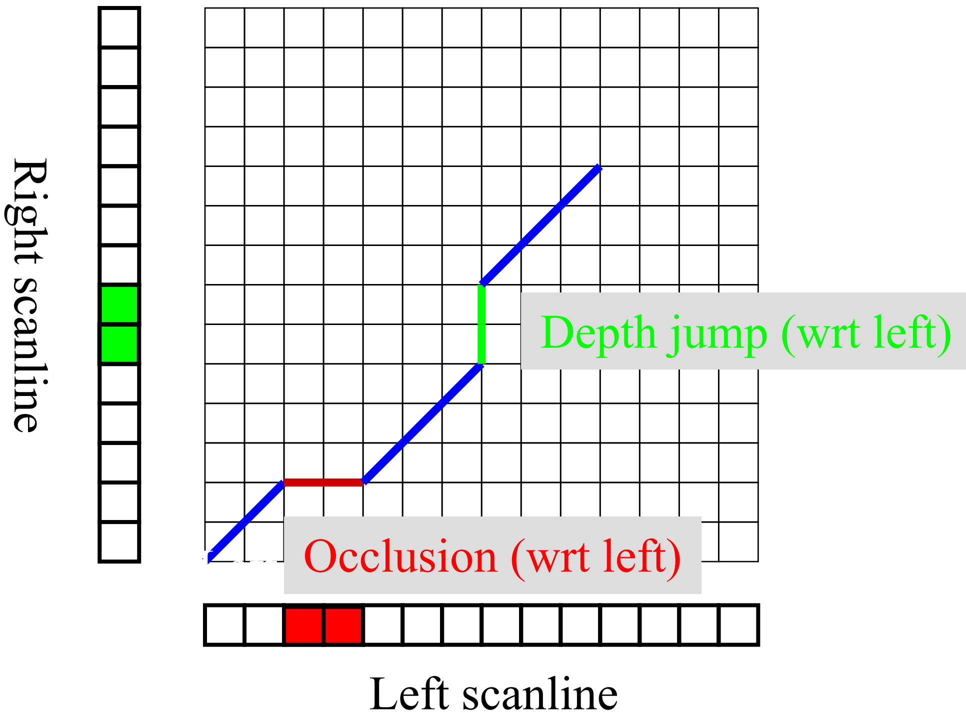


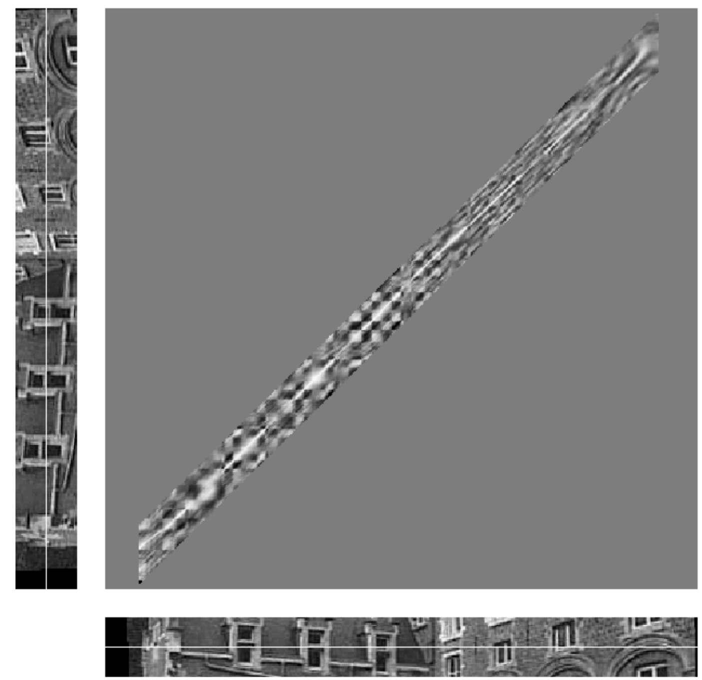


(y=cost)

Extracted from [BBH03]

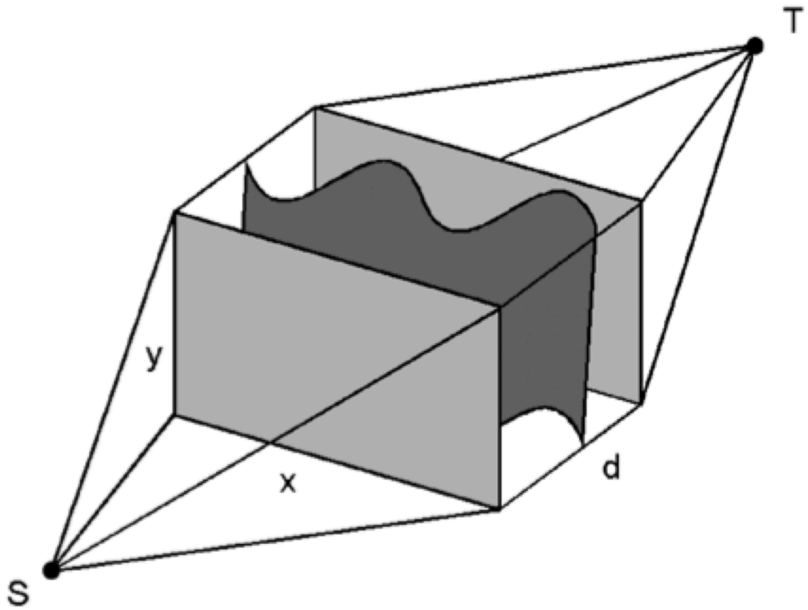
The lighther the lower the matching cost http://www.soe.ucsc.edu/~taoswap/GroupMeeting/Stereo_DanKong_2004_11_24.ppt





http://www.cs.unc.edu/~marc/tutorial/node111.html

Graph cuts



Extracted from [BBH03]

- The DSI becomes a graph; capacity of edges defined as a function of the cost of adjacent nodes
- The min cut is analogous to the best path along a pair of scanlines determined by DP but extended to 3D

Stereo Matching approaches

APPROACH	REFERENCES	BRIEF DESCRIPTION
LOCAL METHODS		
Block Matching	[1], [7], [26], [89]	Search for maximum match score or minimum error over small region, typically using variants of cross-correlation or robust rank metrics.
Gradient-Based Optimization	[51], [44]	Minimize a functional, typically the sum of squared differences, over a small region.
Feature Matching	[8], [11], [23], [62], [72], [84]	Match dependable features rather than intensities themselves.
GLOBAL METHODS		
Dynamic Programming	[5], [9], [10], [19], [36], [60]	Determine the disparity surface for a scanline as the best path between two sequences of ordered primitives. Typically, order is defined by the epipolar ordering constraint.
Intrinsic Curves	[80], [81]	Map epipolar scanlines to intrinsic curve space to convert the search problem to a nearest-neighbors lookup problem. Ambiguities are resolved using dynamic programming.
Graph Cuts	[13], [14], [45], [65], [79], [92]	Determine the disparity surface as the minimum cut of the maximum flow in a graph.
Nonlinear Diffusion	[52], [68], [71]	Aggregate support by applying a local diffusion process.
Belief Propagation	[77]	Solve for disparities via message passing in a belief network.
Correspondenceless Methods	[27], [30], [48]	Deform a model of the scene based on an objective function.

Extracted from [BBH03]

Scharstein&Szeliski's taxonomy

- Building blocks of stereo algorithms
  - Matching cost computation
    - SSD, SAD, NCC, …
  - Cost aggregation
    - Summing or averaging over a support region of DSI
  - Disparity computation/optimization
    - Local: Winner-Take-All
    - Global: Energy minimization (MRF, GC), DP, …
  - Disparity refinement (post processing)
    - sub-pixel, occlusion detection, …

Method	Matching cost	Aggregation	Optimization
SSD (traditional)	squared difference	square window	WTA
Hannah [51]	cross-correlation	(square window)	WTA
Nishihara [82]	binarized filters	square window	WTA
Kass [63]	filter banks	-none-	WTA
Fleet *et al*. [40]	phase	-none-	phase-matching
Jones and Malik [57]	filter banks	-none-	WTA
Kanade [58]	absolute difference	square window	WTA
Scharstein [95]	gradient-based	Gaussian	WTA
Zabih and Woodfill [129]	rank transform	(square window)	WTA
Cox *et al*. [32]	histogram eq.	-none-	DP
Frohlinghaus and Buhmann [41]	wavelet phase	-none-	phase-matching
Birchfield and Tomasi [12]	shifted abs. diff	-none-	DP
Marr and Poggio [73]	binary images	iterative aggregation	WTA
Prazdny [89]	binary images	3D aggregation	WTA
Szeliski and Hinton [114]	binary images	iterative 3D aggregation	WTA
Okutomi and Kanade [84]	squared difference	adaptive window	WTA
Yang *et al*. [127]	cross-correlation	non-linear filtering	hier. WTA
Shah [103]	squared difference	non-linear diffusion	regularization
Boykov *et al*. [22]	thresh. abs. diff.	connected-component	WTA
Scharstein and Szeliski [97]	robust sq. diff.	iterative 3D aggregation	mean-field
Zitnick and Kanade [132]	squared difference	iterative aggregation	WTA
Veksler [124]	abs. diff - avg.	adaptive window	WTA
Quam [90]	cross-correlation	-none-	hier. warp
Barnard [6]	squared difference	-none-	SA
Geiger *et al*. [46]	squared difference	shiftable window	DP
Belhumeur [9]	squared difference	-none-	DP
Cox *et al*. [31]	squared difference	-none-	DP
Ishikawa and Geiger [55]	squared difference	-none-	graph cut
Roy and Cox [92]	squared difference	-none-	graph cut
Bobick and Intille [18]	absolute difference	shiftable window	DP
Boykov *et al*. [23]	squared difference	-none-	graph cut
Kolmogorov and Zabih [65]	squared difference	-none-	graph cut
Fusiello *et al*.	squared difference	shiftable windows	WTA

Adapted from [SS02]

References

- [BBH03] Brown, M. Z., Burschka, D., and Hager, G. D. 2003. Advances in Computational Stereo. *IEEE Trans. Pattern Anal. Mach. Intell.* 25, 8 (Aug. 2003), 993-1008.
- [SS02] Scharstein, D. and Szeliski, R. 2002. A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms. *Int. J. Comput. Vision* 47, 1-3 (Apr. 2002), 7-42.

- The other citations can be found within.