Calibrated Stereo (Part 2) and Feature Matching

Computer Vision I

CSE 252A

Lecture 8

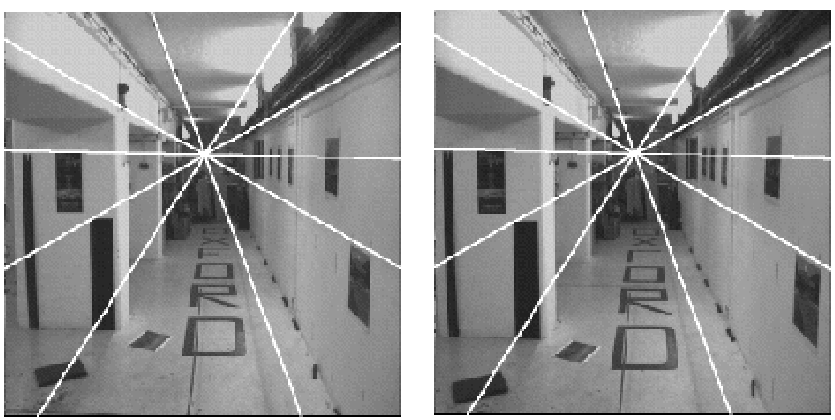
Announcements

- Assignment 1 is due today, 11:59 PM
- Assignment 2 will be released today
  - Due Nov 8, 11:59 PM

Stereo Vision Outline

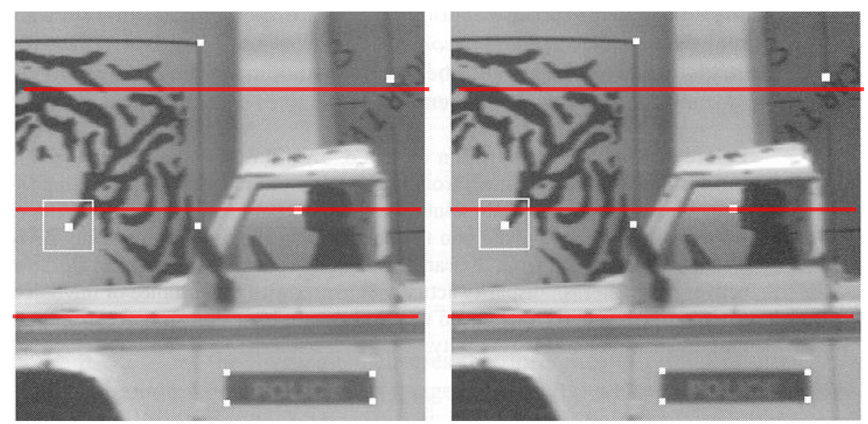
- Offline
  - Calibration of stereo cameras
- Online
  1. Acquire stereo images
  2. Epipolar rectify stereo images
  3. Establish correspondence
  4. Estimate depth

- Epipolar geometry reduces matching complexity from $O(n^4)$ to $O(n^3)$
- But matching requires comparing points across pairs of epipolar lines which may have arbitrary orientation. That can be costly to index.
- Is there a more convenient epipolar geometry



Slanted epipolar lines

vs



Horizontal, row aligned epipolar lines

Establish correspondences

Two Approaches

1. Feature-Based (sparse)
   – From each image, process “monocular” image to obtain cues (e.g., corners, SIFT features)
   – Establish feature correspondence between the two images

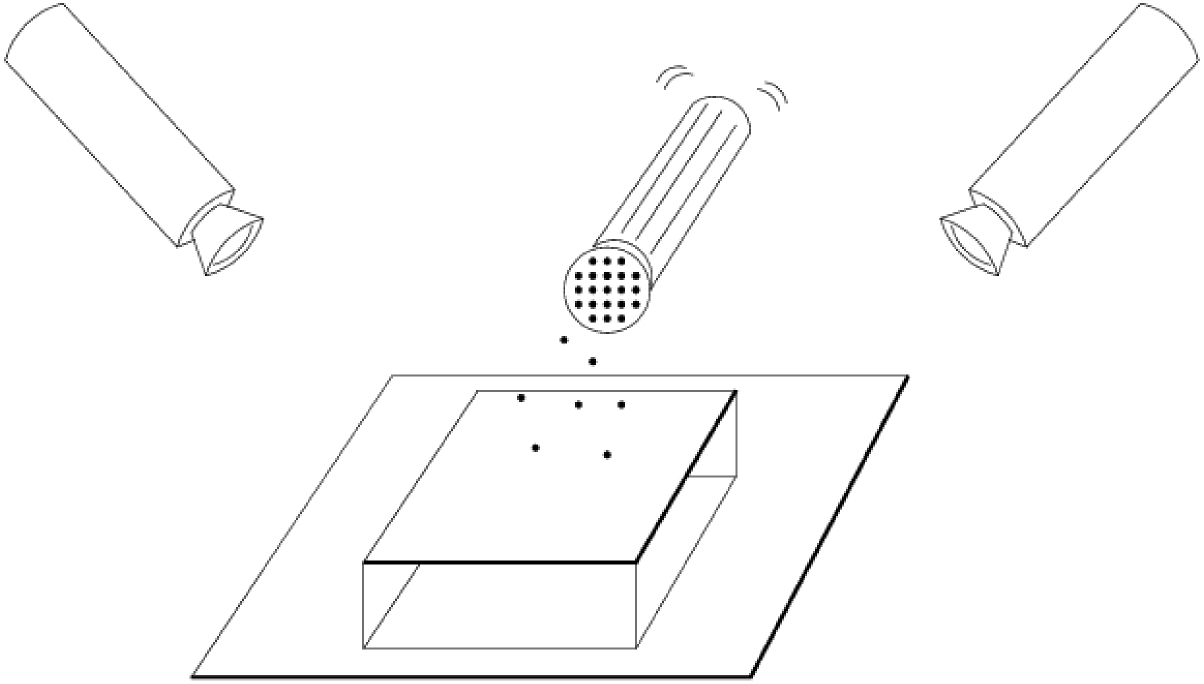
2. Area-Based (dense)
   – Directly compare image regions between the two images

Human Stereopsis: Binocular Fusion

How are the correspondences established?

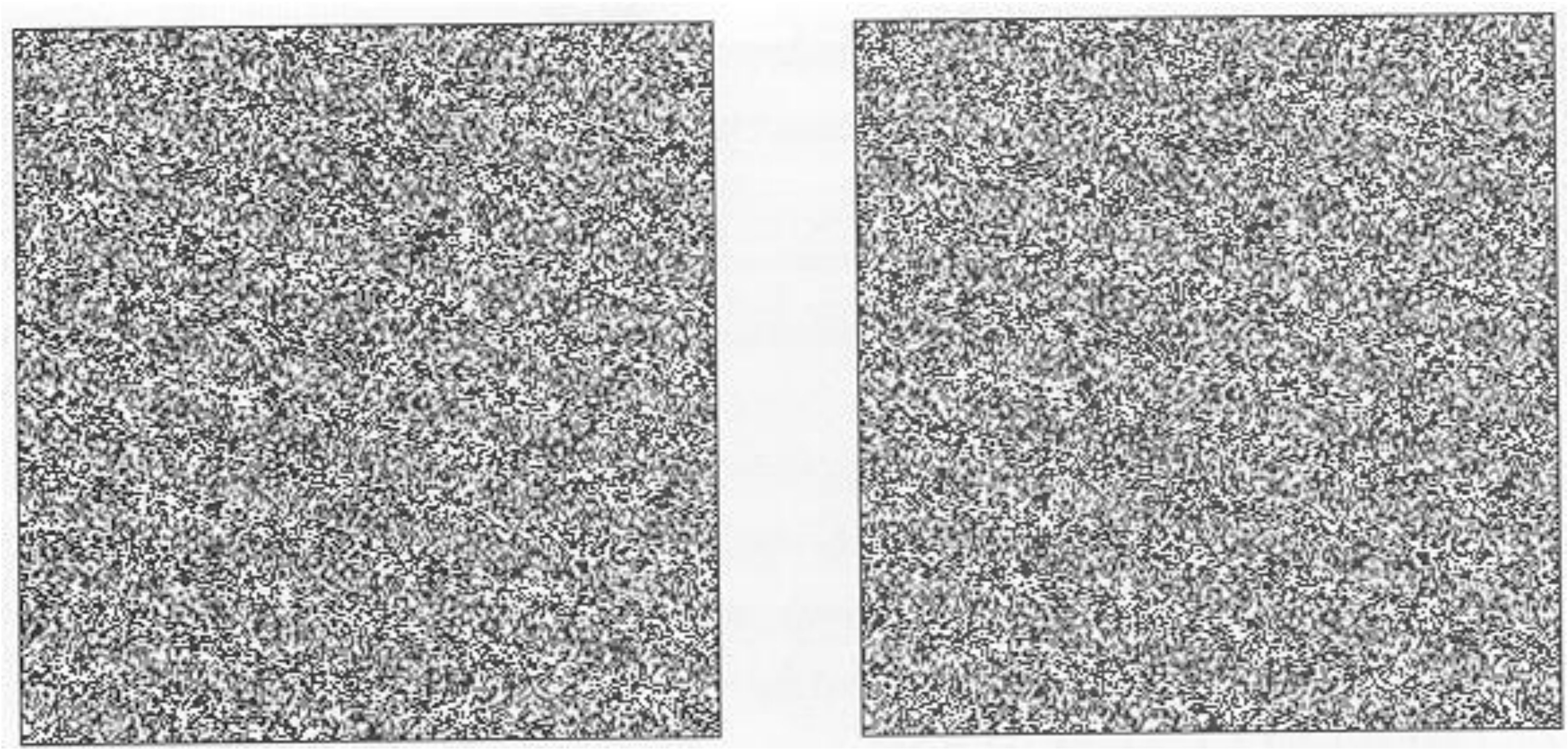
Julesz (1971): Is the mechanism for binocular fusion a monocular process or a binocular one??

- There is anecdotal evidence for the latter (camouflage).

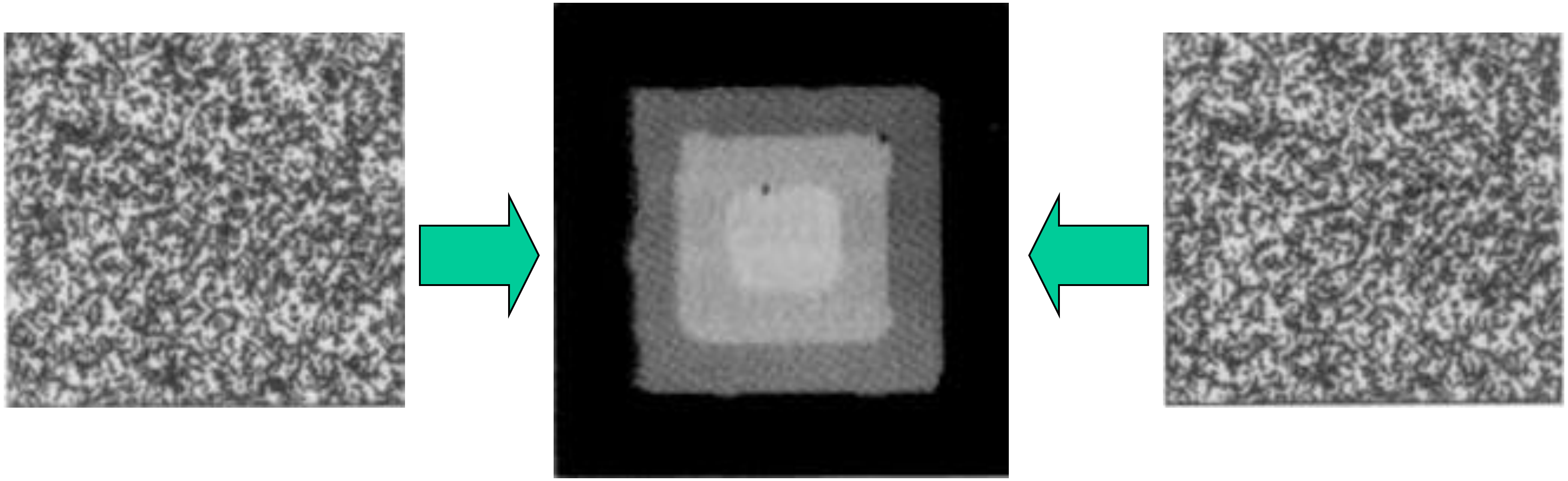


- Random dot stereograms provide an objective answer

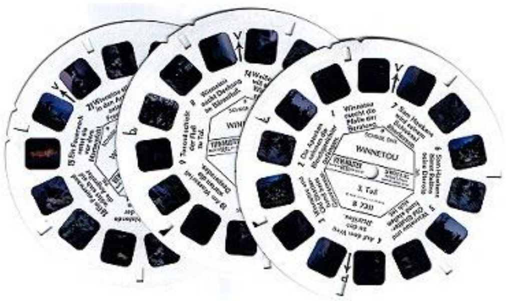
Random Dot Stereograms



Random Dot Stereograms



Stereoscopic 3D



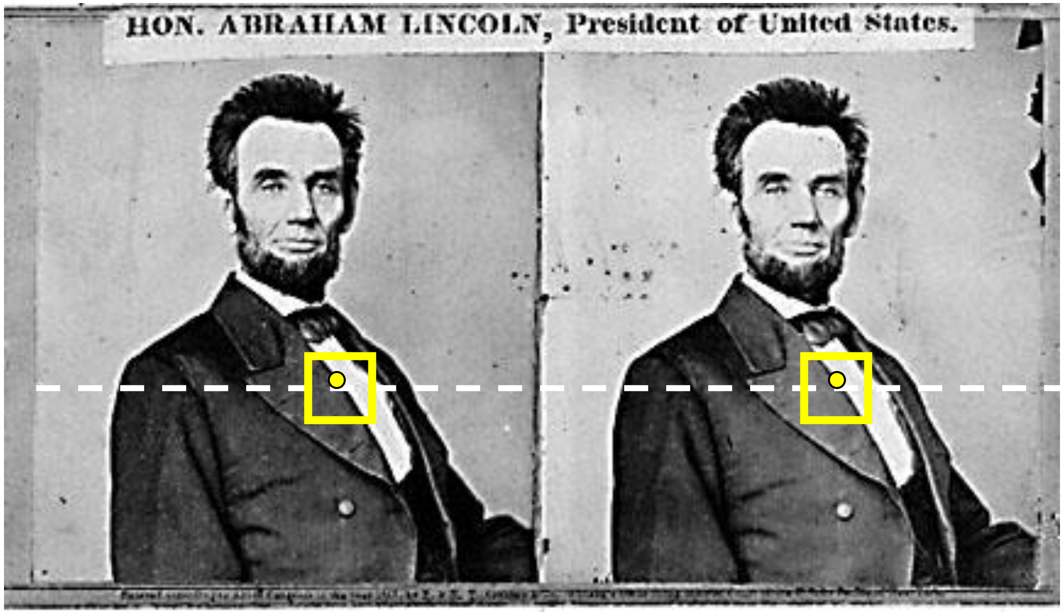
Stereoscopic 3D

Was Rembrandt Stereo Blind?

- Detail of a 1639 etching

- In Rembrandt's painted self-portraits (left panel) in which the eyes are clearly visible, his left eye frequently looks straight out and the right off to the side. It is the opposite in his etchings (right panel).

Using epipolar & constant Brightness constraints for stereo matching



For each epipolar line

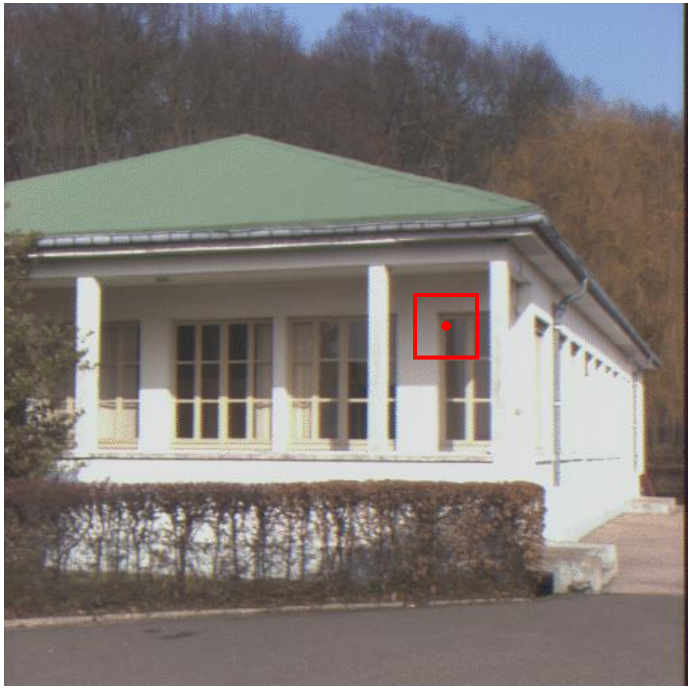
For each pixel in the left image

- compare with every pixel on same epipolar line in right image
- pick pixel with most similar brightness.

This will never work, so: **Match *windows***

(Seitz)

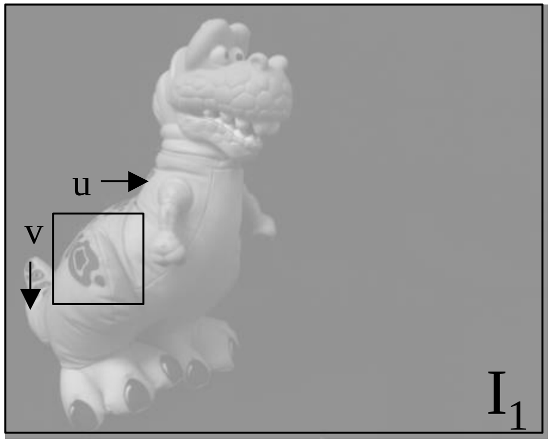
Finding Correspondences

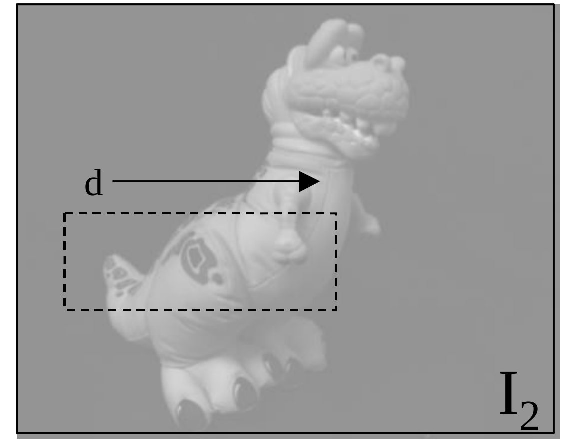


$W(\mathbf{p}_\mathrm{l})$

$W(\mathbf{p}_\mathrm{r})$

Correspondence Search Algorithm

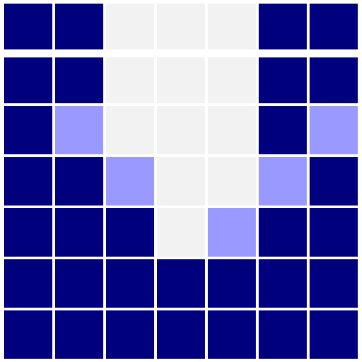




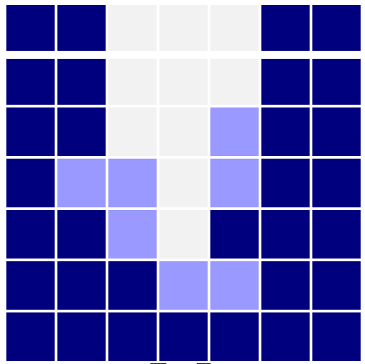

```
For  i = 1:nrows
   for j=1:ncols
            best(i,j) = -1
            for k = mindisparity:maxdisparity
               c = Match_Metric(I1(i,j),I2(i,j+k),winsize)
               if (c > best(i,j))
                        best(i,j) = c
                        disparities(i,j) = k
                end
             end
   end
end
```

O(nrows * ncols * disparities * winx * winy)

Simple match metrics



W_1



W_2

- SSD (Sum of Squared Differences)

$$\sum_{x,y} |W_1(x,y) - W_2(x,y)|^2$$

- NCC (Normalized Cross Correlation)

$$\frac{\sum_{x,y}(W_1(x,y) - \overline{W_1})(W_2(x,y) - \overline{W_2})}{\sigma_{W_1}\sigma_{W_2}}$$

where $\overline{W_i} = \frac{1}{n}\sum_{x,y} W_i$, $\sigma_{W_i} = \sqrt{\frac{1}{n}\sum_{x,y}(W_i - \overline{W_i})^2}$

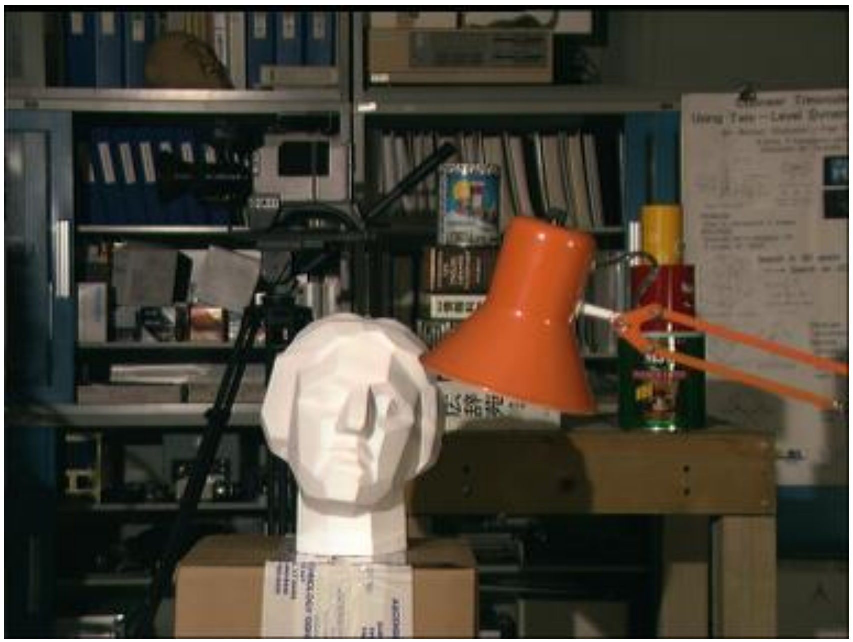
Match Metric Summary

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

These two result in the same matches

Stereo results

– Data from University of Tsukuba

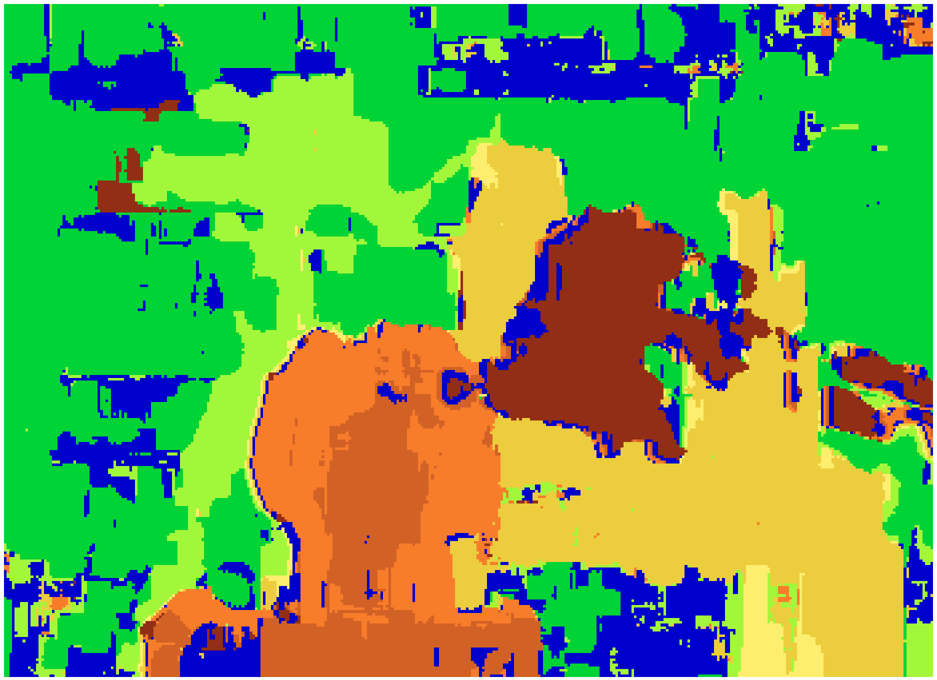


Scene

Ground truth

(Seitz)

Results with greedy algorithm and correlation match metric



Window-based matching
(best window size)

Ground truth

(Seitz)

Results with better method

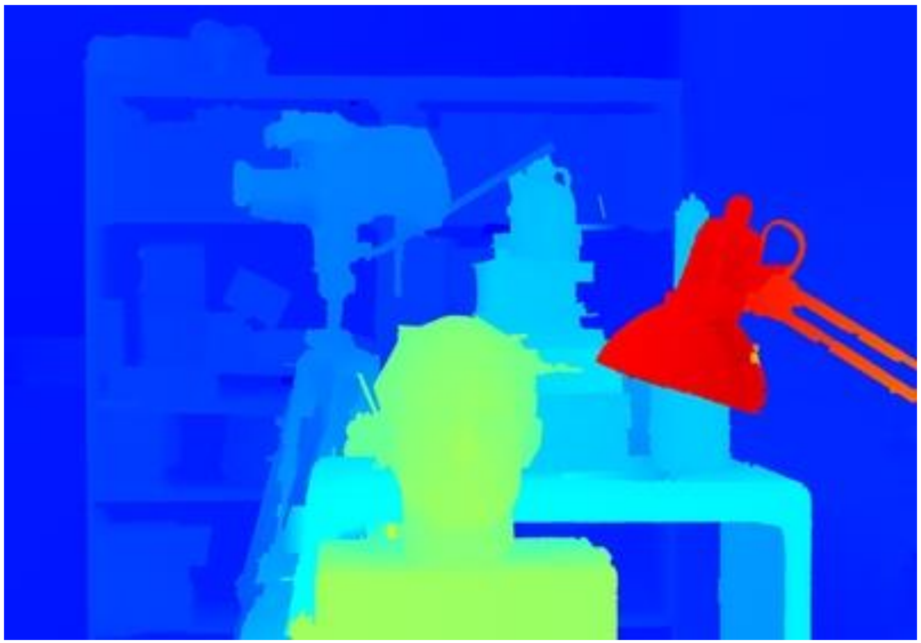
Using global optimization

Boykov et al., Fast Approximate Energy Minimization via Graph Cuts, International Conference on Computer Vision, September 1999.

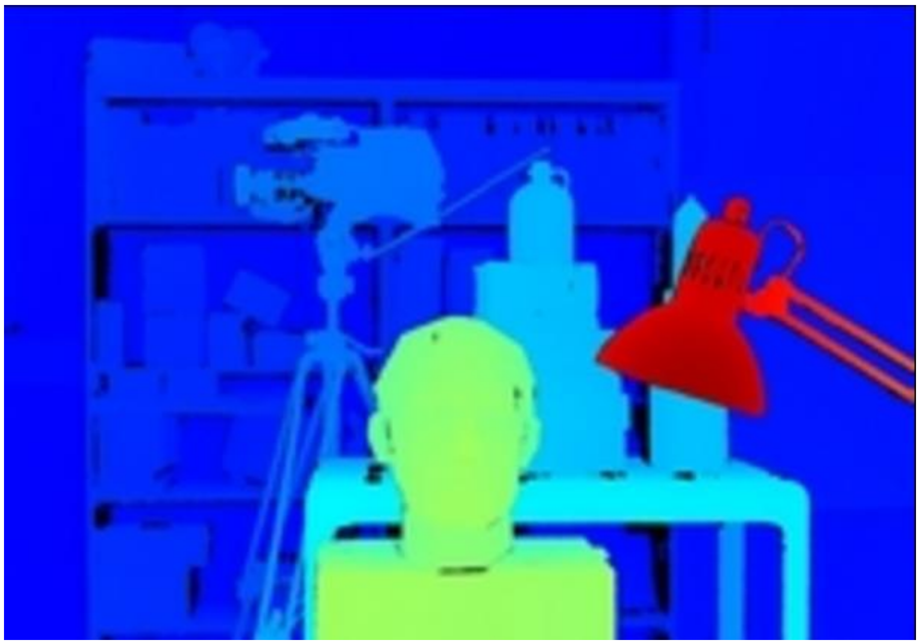
Ground truth

(Seitz)

State of the Art Results



Using neural networks



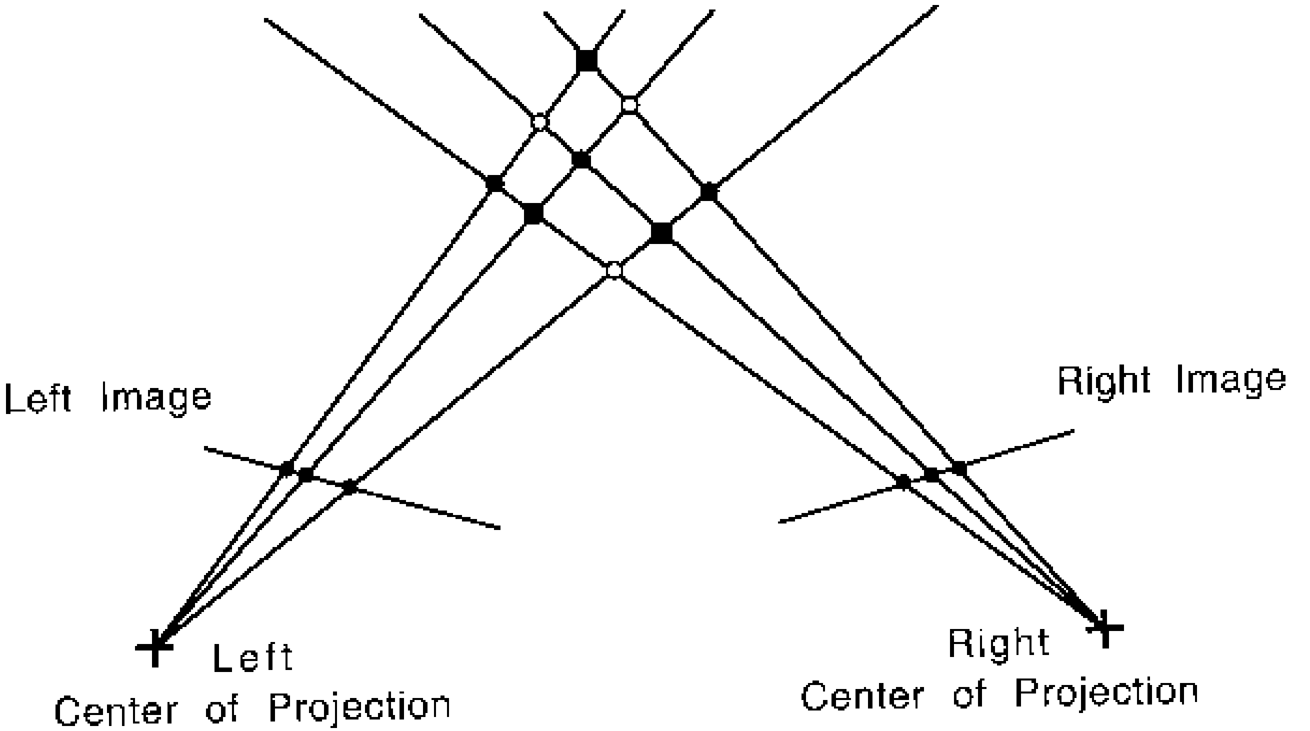
Ground truth

S. Drouyer, S. Beucher, M. Bilodeau, M. Moreaud, and L. Sorbier. Sparse stereo disparity map densification using hierarchical image segmentation. 13th International Symposium on Mathematical Morphology.

Some Issues

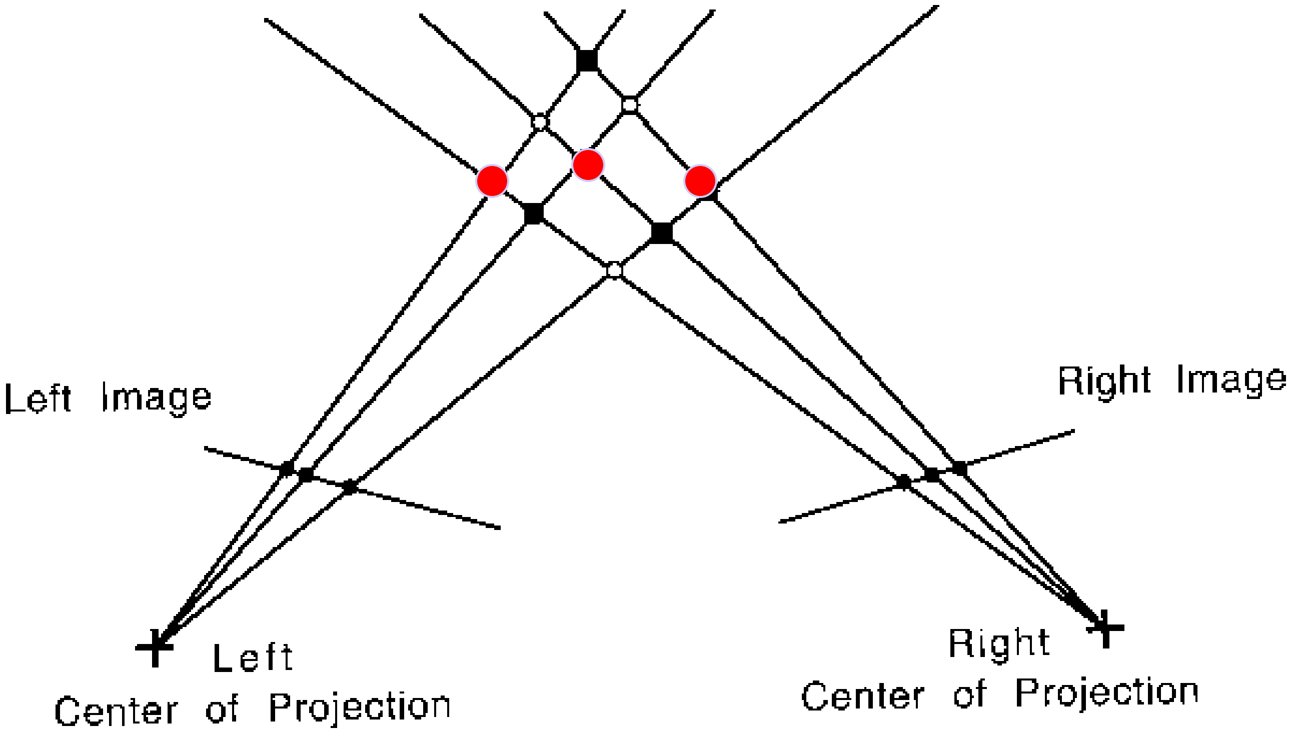
- Epipolar ordering
- Ambiguity
- Window size
- Window shape
- Lighting
- Half occluded regions

A challenge: Multiple Interpretations



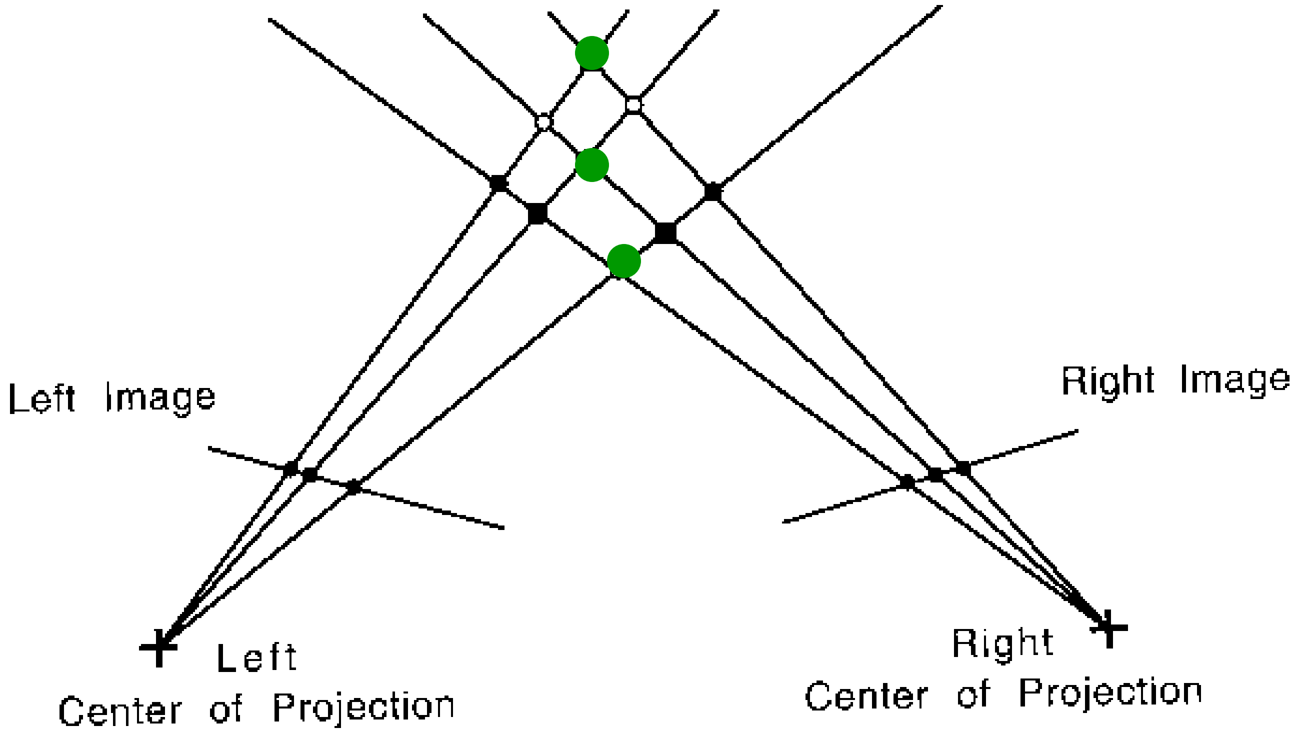
Each feature on left epipolar line match one and only one feature on right epipolar line.

Multiple Interpretations



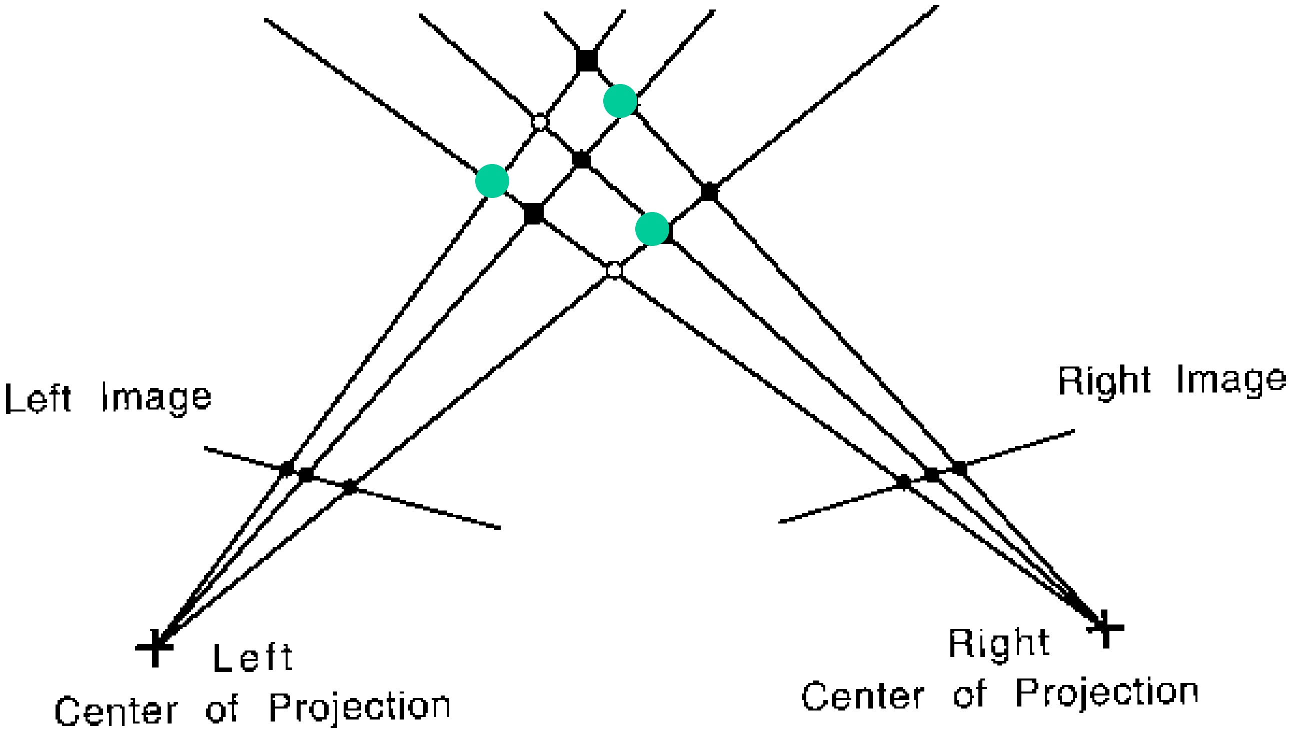
Each feature on left epipolar line match one and only one feature on right epipolar line.

Multiple Interpretations



Each feature on left epipolar line match one and only one feature on right epipolar line.

Multiple Interpretations



Each feature on left epipolar line match one and only one feature on right epipolar line.

Some Issues

- Epipolar ordering
- **Ambiguity**
- Window size
- Window shape
- Lighting
- Half occluded regions

Ambiguity

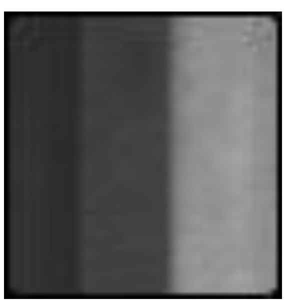
$\mathbf{W}_L$

$\mathbf{W}_1$

$\mathbf{W}_2$

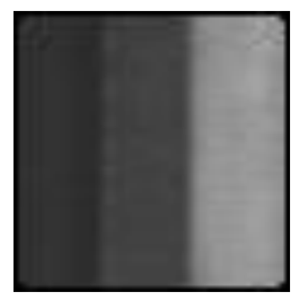
It's a coin toss whether $\mathbf{W}_L$ will match $\mathbf{W}_1$ or $\mathbf{W}_2$

Ambiguity



$\mathbf{W}_1$

$\mathbf{W}_2$



$\mathbf{W}_R$

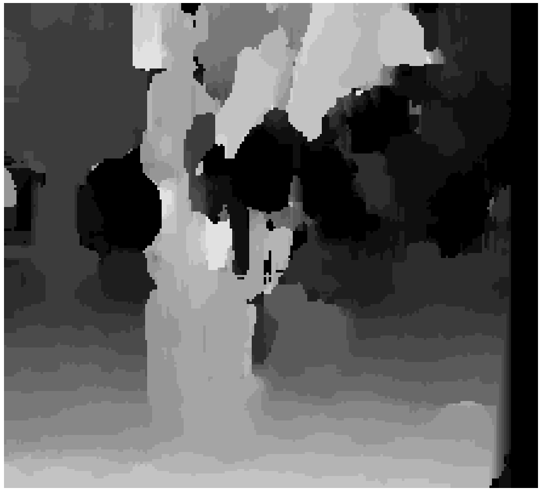
With the Greedy Algorithm $\mathbf{W}_1$ might match $\mathbf{W}_R$, but $\mathbf{W}_2$ might also match $\mathbf{W}_R$

Some Issues

- Epipolar ordering
- Ambiguity
- **Window size**
- Window shape
- Lighting
- Half occluded regions

Window size

W = 3



W = 20

- Effect of window size

Better results with *adaptive window*

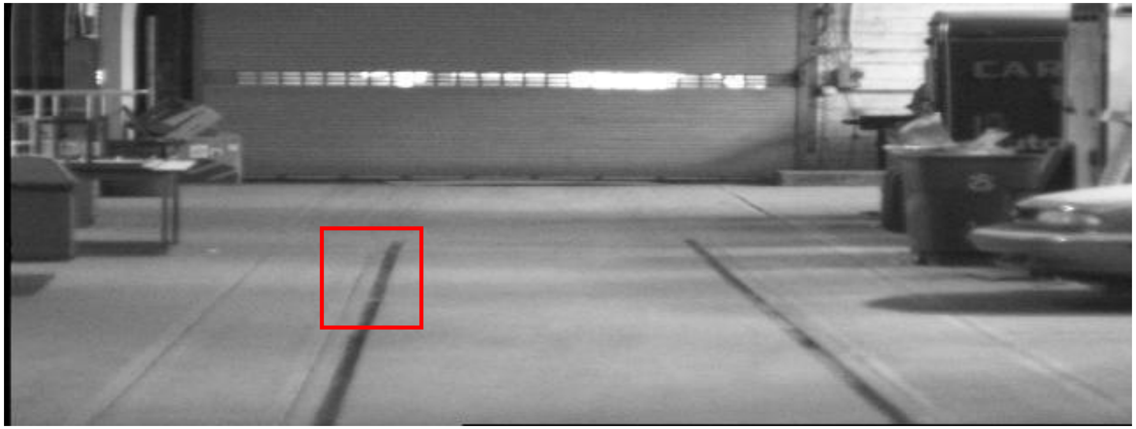
- T. Kanade and M. Okutomi, *A Stereo Matching Algorithm with an Adaptive Window: Theory and Experiment*,, Proc. International Conference on Robotics and Automation, 1991.
- D. Scharstein and R. Szeliski. Stereo matching with nonlinear diffusion. International Journal of Computer Vision, 28(2):155-174, July 1998

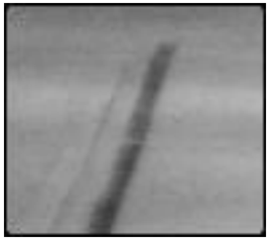
(Seitz)

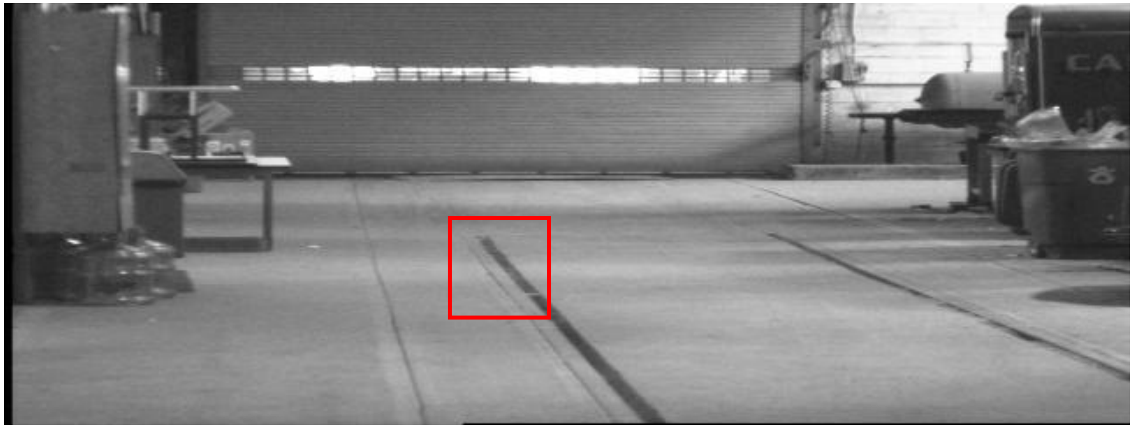
Some Issues

- Epipolar ordering
- Ambiguity
- Window size
- **Window shape**
- Lighting
- Half occluded regions

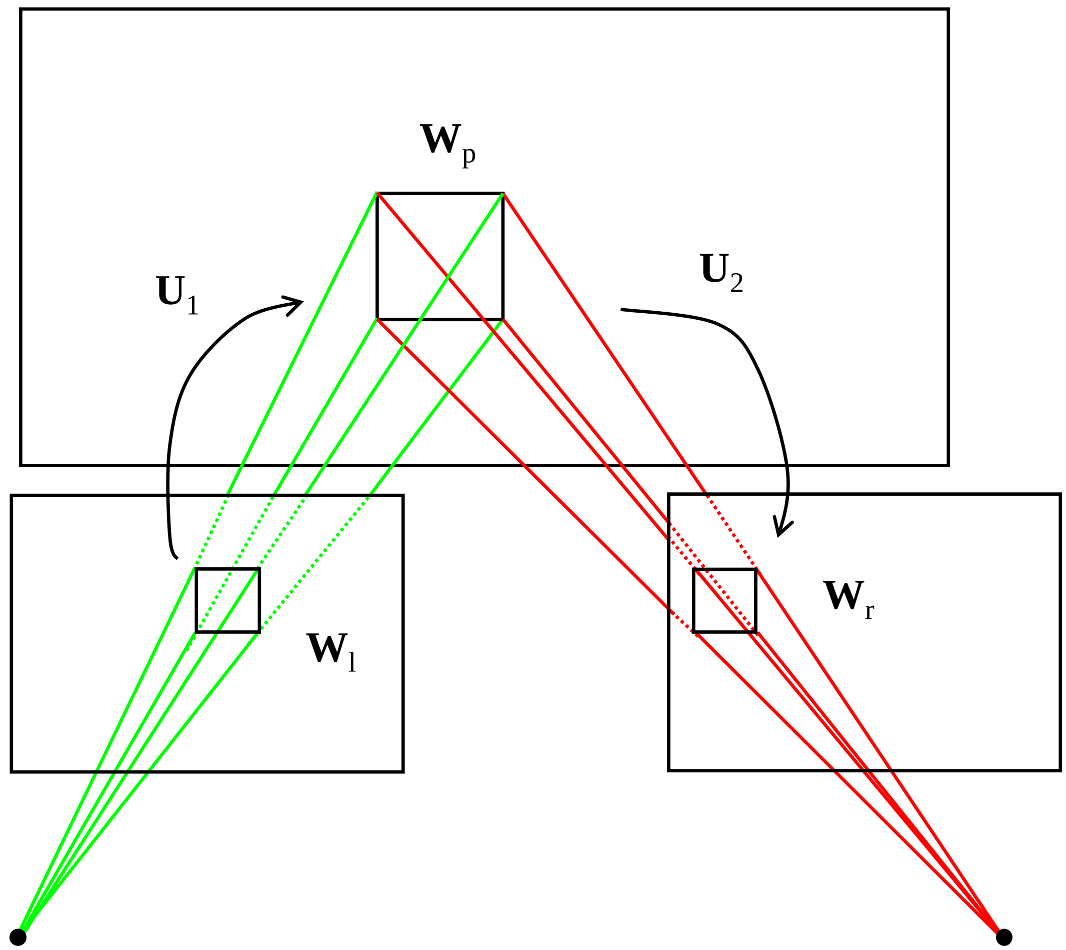
Window Shape and Forshortening







Window Shape: Fronto-parallel Configuration



- When scene plane is parallel to the image planes, a square $\mathbf{w_p}$ in the scene projects to squares in the images $\mathbf{w_l}$ and $\mathbf{w_r}$
- But when scene plane is tilted, $\mathbf{w_p}$ projects to a quadrilaterial in the images

Some Issues

- Epipolar ordering
- Window size
- Ambiguity
- Window shape
- **Lighting**
- Half occluded regions

Lighting Conditions (Photometric Variations)

$W(\mathbf{P}_l)$

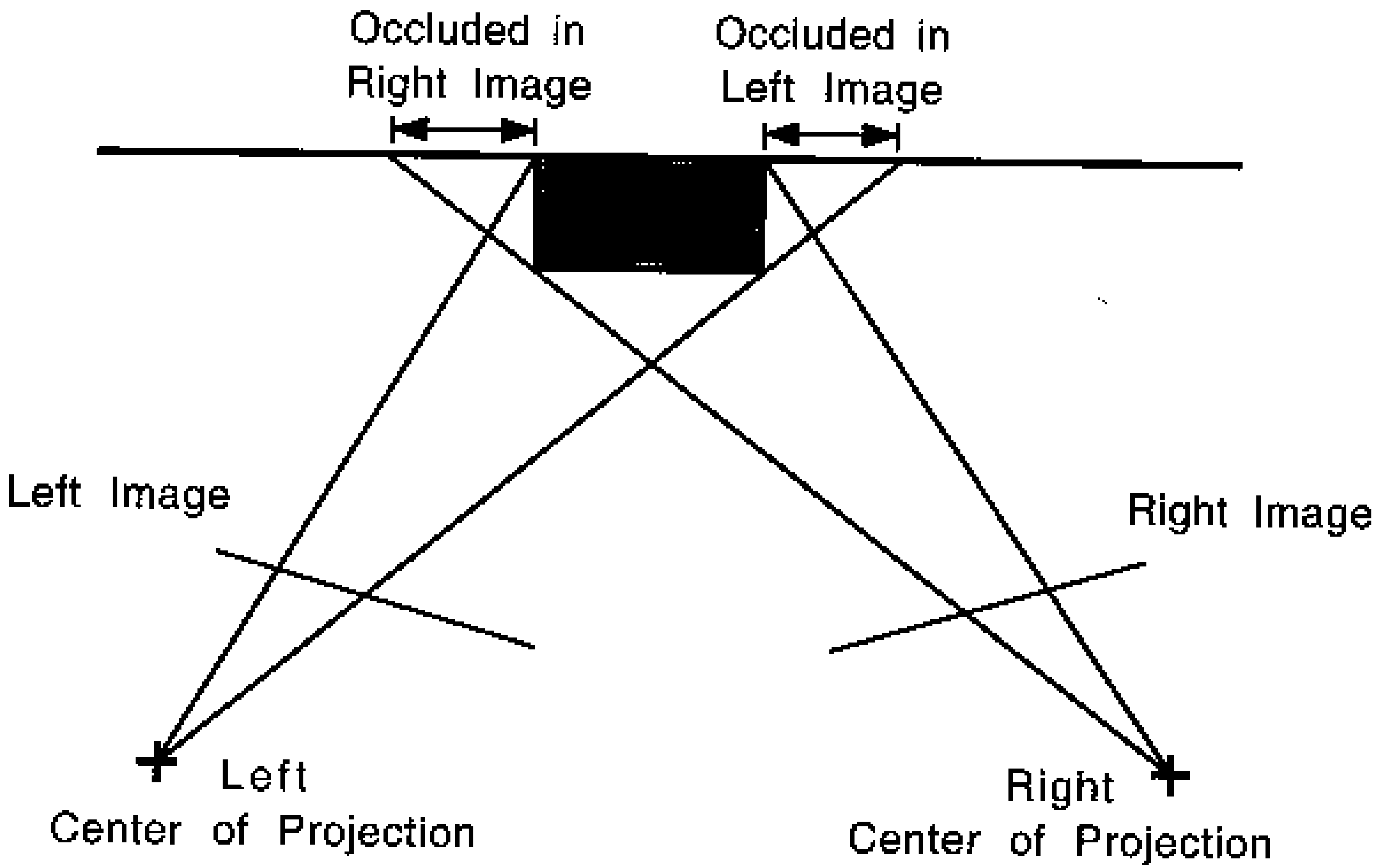
$W(\mathbf{P}_r)$

Does the match metric handle matching across differences of brightness?

Some Issues

- Epipolar ordering
- Ambiguity
- Window size
- Window shape
- Lighting
- **Half occluded regions**

Half occluded regions



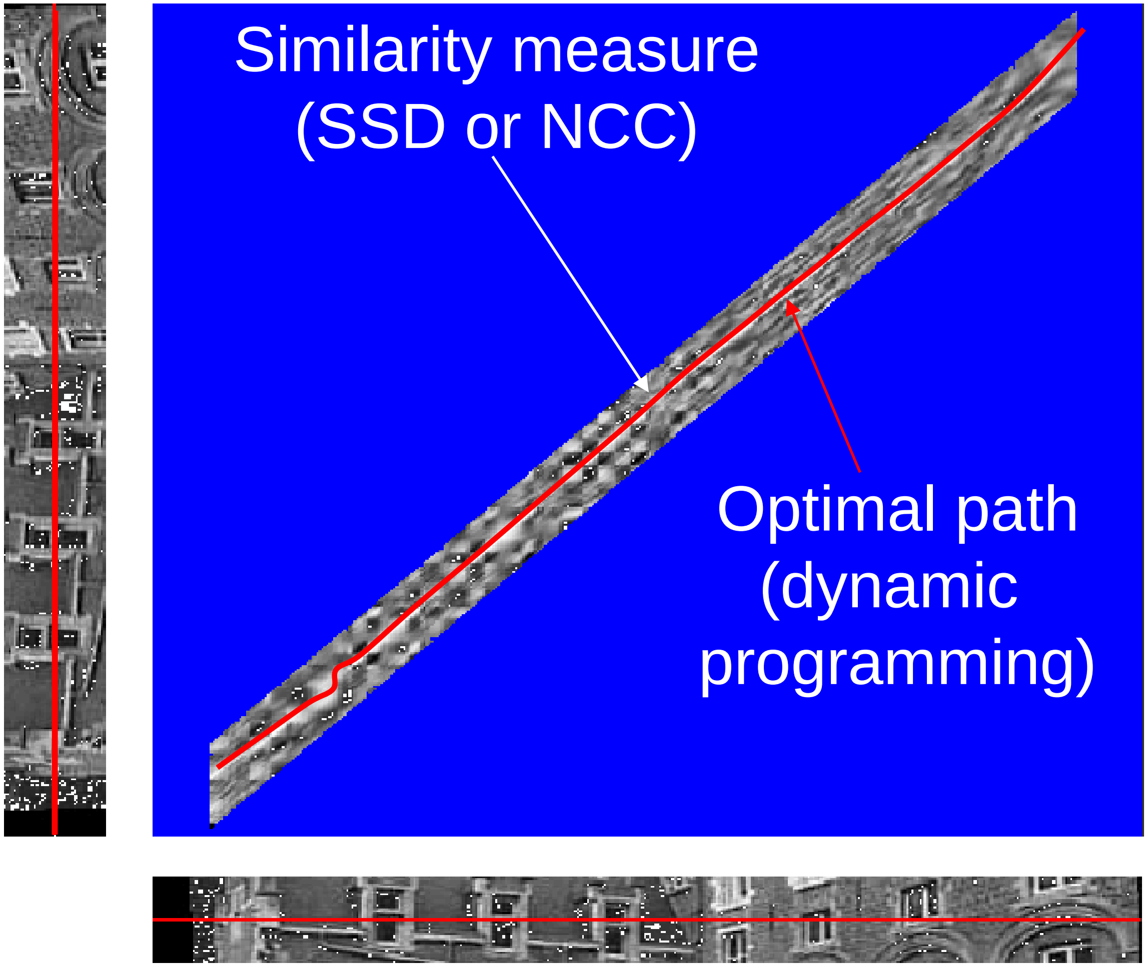
- Half occluded regions are visible in one camera, but not in the other
- They can be a cue for a depth change

Summary of Stereo Constraints

CONSTRAINT	BRIEF DESCRIPTION
1-D Epipolar Search	Arbitrary images of the same scene may be rectified based on epipolar geometry such that stereo matches lie along one-dimensional scanlines. This reduces the computational complexity and also reduces the likelihood of false matches.
Monotonic Ordering	Points along an epipolar scanline appear in the same order in both stereo images, assuming that all objects in the scene are approximately the same distance from the cameras.
Image Brightness Constancy	Assuming Lambertian surfaces, the brightness of corresponding points in stereo images are the same.
Match Uniqueness	For every point in one stereo image, there is at most one corresponding point in the other image.
Disparity Continuity	Disparities vary smoothly (i.e. disparity gradient is small) over most of the image. This assumption is violated at object boundaries.
Disparity Limit	The search space may be reduced significantly by limiting the disparity range, reducing both computational complexity and the likelihood of false matches.
Fronto-Parallel Surfaces	The implicit assumption made by area-based matching is that objects have fronto-parallel surfaces (i.e. depth is constant within the region of local support). This assumption is violated by sloping and creased surfaces.
Feature Similarity	Corresponding features must be similar (e.g. edges must have roughly the same length and orientation).
Structural Grouping	Corresponding feature groupings and their connectivity must be consistent.

(From G. Hager)

Stereo matching



Constraints
- epipolar
- ordering
- uniqueness
- disparity limit
- disparity gradient limit

Trade-off
- Matching cost (data)
- Discontinuities (prior)

(Cox et al. CVGIP'96; Koch'96; Falkenhagen´97;
Van Meerbergen,Vergauwen,Pollefeys,VanGool IJCV'02)

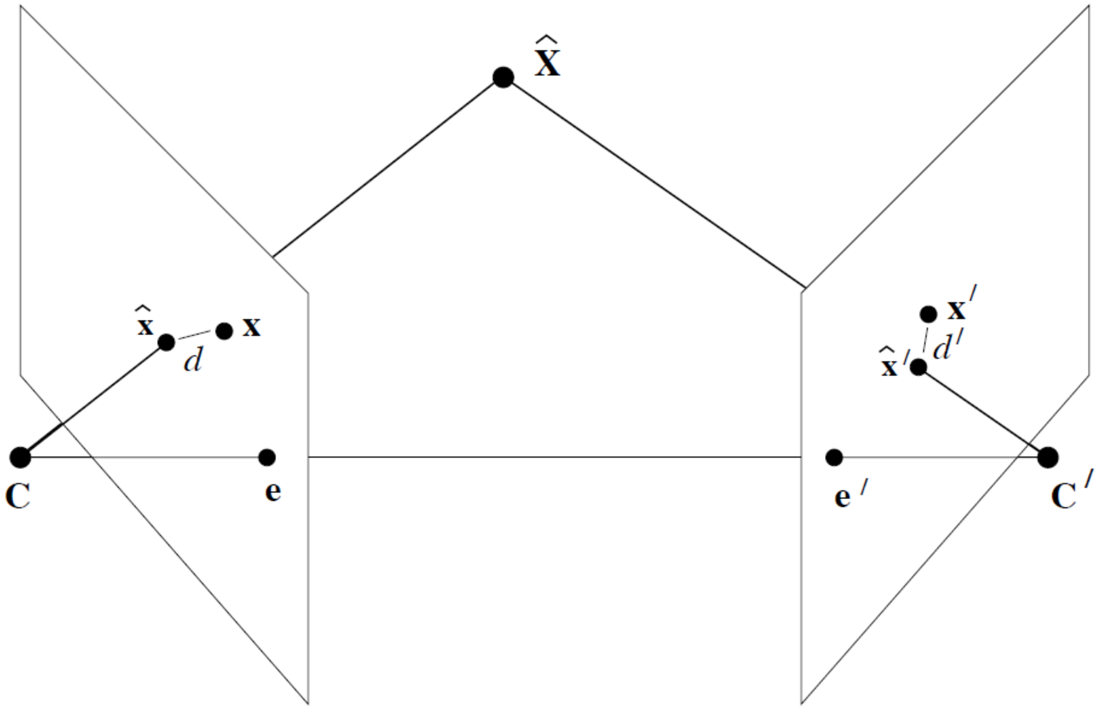
(From Pollefeys)

Estimate depth

Reconstruction: General 3D case

Given two image measurements x and x′, estimate scene point $\widehat{\mathrm{X}}$

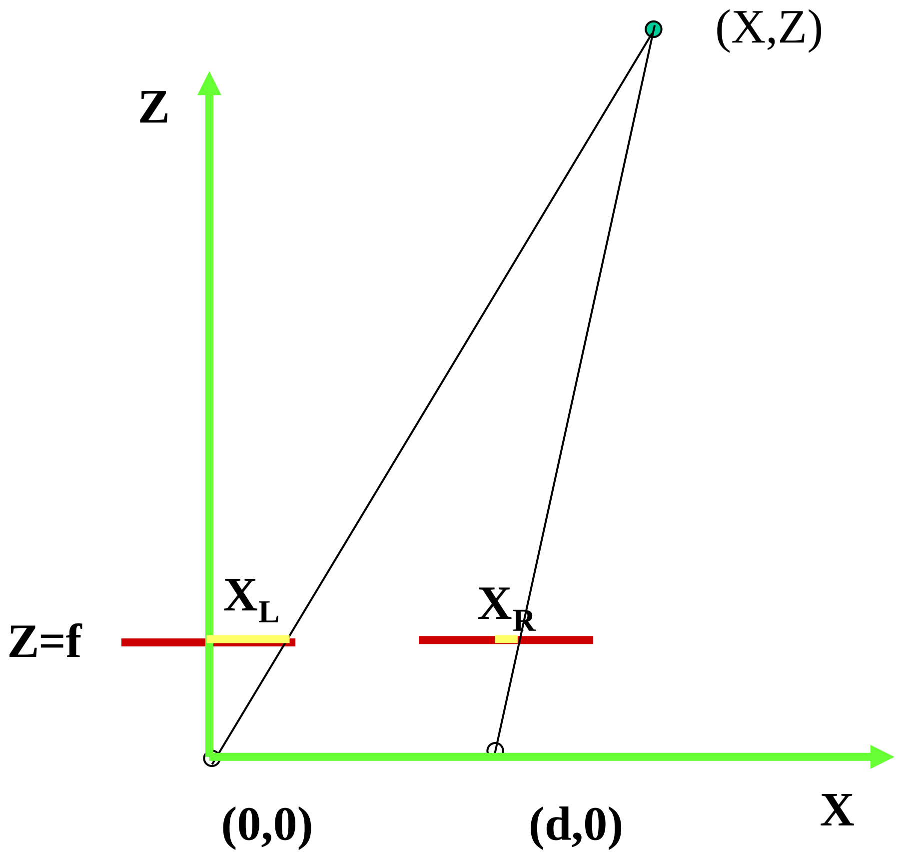
$\hat{\mathbf{x}} = \mathsf{P}\widehat{\mathbf{X}}$



$\hat{\mathbf{x}}' = \mathsf{P}'\widehat{\mathbf{X}}$

Estimate $\widehat{\mathrm{X}}$ that minimizes $d(\mathbf{x}, \hat{\mathbf{x}})^2 + d(\mathbf{x}', \hat{\mathbf{x}}')^2$

Binocular Stereo: Estimating Depth
2-D world with 1-D image plane



$X_L = f(X/Z)$ $X_R = f((X-d)/Z)$

Two measurements: X_L, X_R
Two unknowns: X,Z

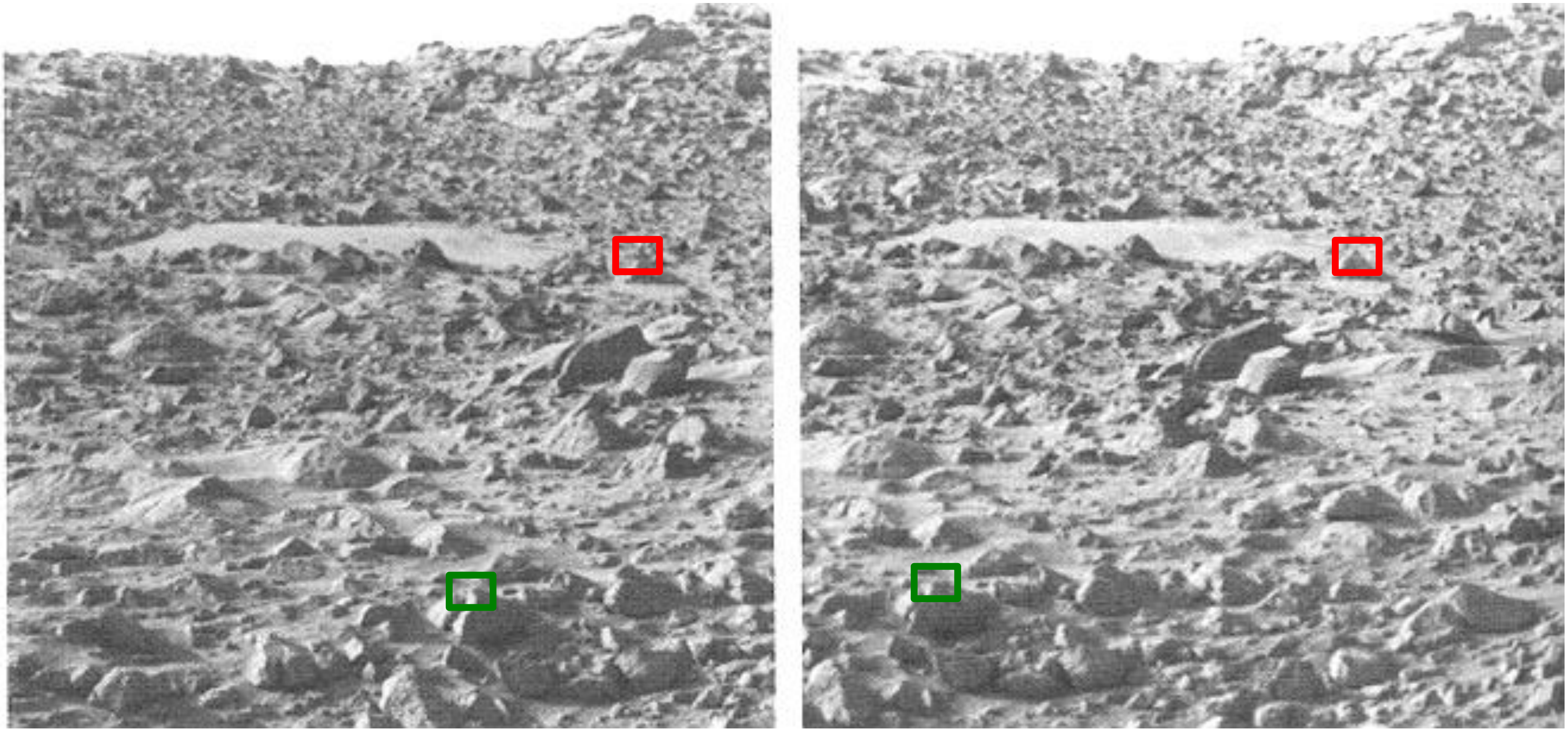
Constants:
- **Baseline: d**
- **Focal length: f**

$$X = \frac{d\,X_L}{(X_L - X_R)}$$

$$Z = \frac{d\,f}{(X_L - X_R)}$$

Disparity: $(X_L - X_R)$

(Adapted from Hager)



Faraway points – small disparity

Infinitely far, zero disparity

Nearby points – large disparity

$$Z = \frac{df}{x_L - x_R}$$

More on stereo …

The Middlebury Stereo Vision Research Page
https://vision.middlebury.edu/stereo/

Recommended reading

D. Scharstein and R. Szeliski.
A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms. IJCV 47(1/2/3):7-42, April-June 2002. PDF file (1.15 MB) - includes current evaluation. Microsoft Research Technical Report MSR-TR-2001-81, November 2001.

Myron Z. Brown, Darius Burschka, and Gregory D. Hager. Advances in Computational Stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(8):993-1008, 2003.

Some Challenges & Problems

- Photometric issues:
  - specularities
  - strongly non-Lambertian BRDFs

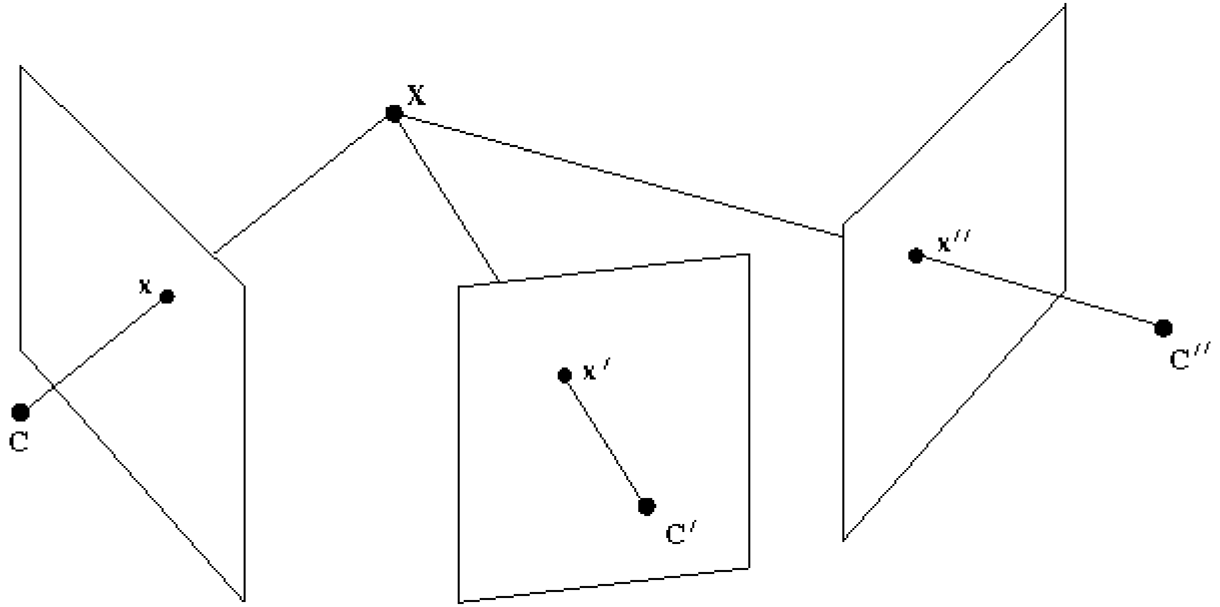
- Surface structure
  - lack of texture
  - repeating texture within horopter bracket

- Geometric ambiguities
  - as surfaces turn away, difficult to get accurate reconstruction (affine approximate can help)
  - at the occluding contour, likelihood of good match but incorrect reconstruction

Variations on Binocular Stereo

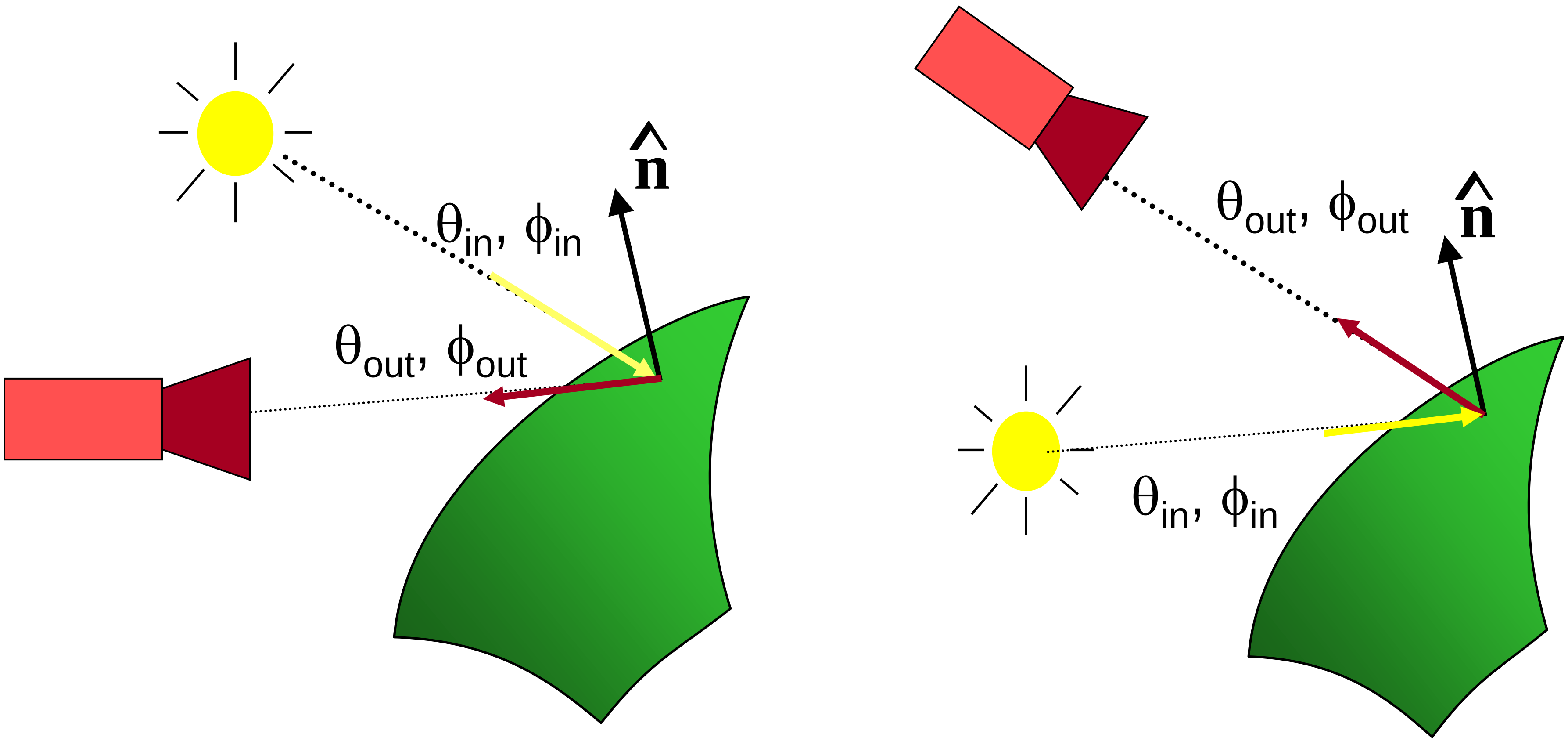
1. Trinocular Stereopsis
2. Helmholtz Reciprocity Stereopsis

Trinocular Epipolar Constraints



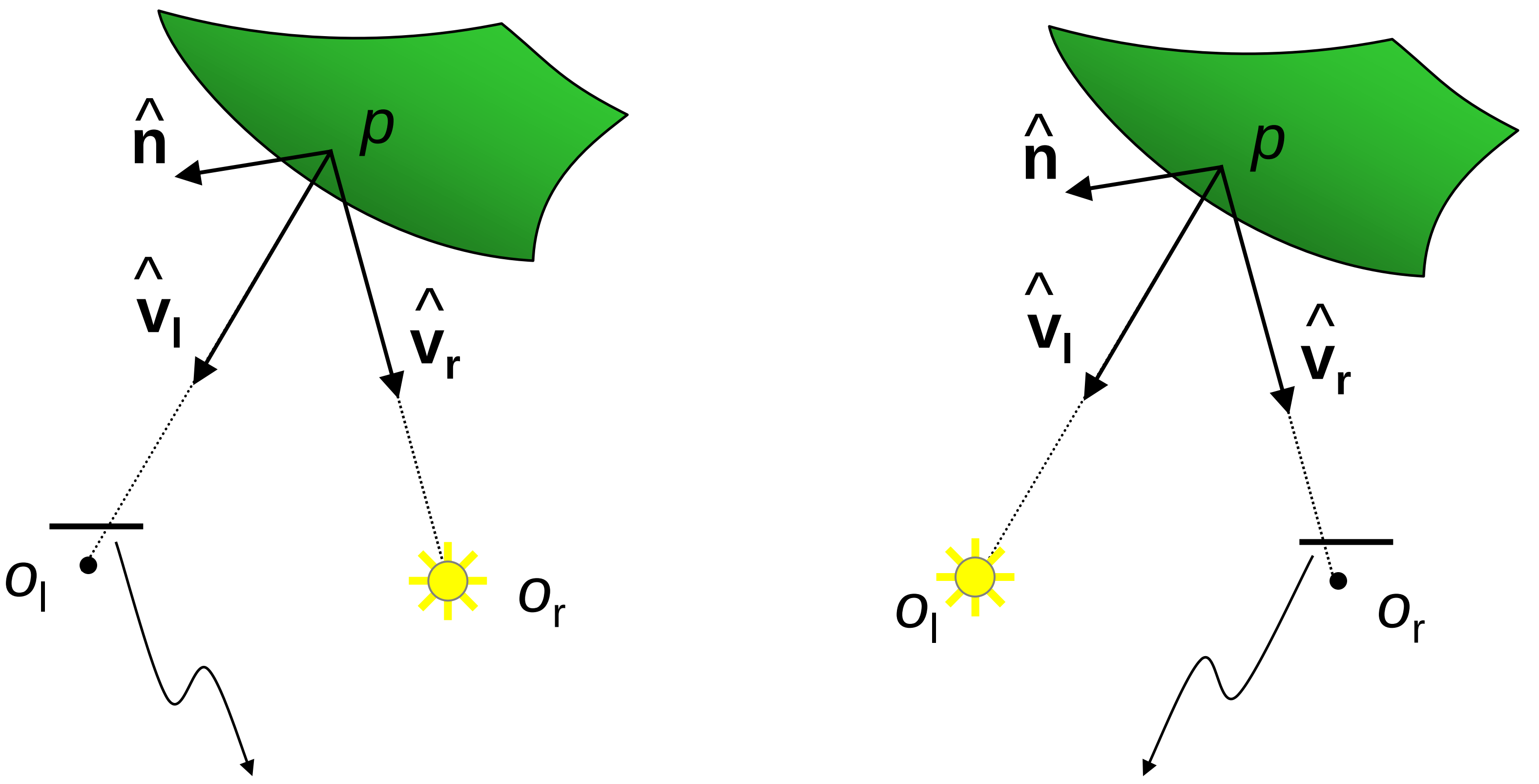
Helmholtz reciprocity

$$\rho(\theta_{in}, \phi_{in} ; \theta_{out}, \phi_{out}) = \rho(\theta_{out}, \phi_{out} ; \theta_{in}, \phi_{in})$$



[Helmholtz, 1910], [Minnaert, 1941], [Nicodemus et al, 1977]

Point Source Illumination



$$\mathrm{i}_l = \rho(\hat{\mathrm{v}}_\mathrm{r}, \hat{\mathrm{v}}_\mathrm{l}) \frac{\hat{\mathrm{n}} \cdot \hat{\mathrm{v}}_\mathrm{r}}{\left|o_r - p\right|^2}$$

=

$$\mathrm{i}_r = \rho(\hat{\mathrm{v}}_\mathrm{l}, \hat{\mathrm{v}}_\mathrm{r}) \frac{\hat{\mathrm{n}} \cdot \hat{\mathrm{v}}_\mathrm{l}}{\left|o_l - p\right|^2}$$

Matching Constraint

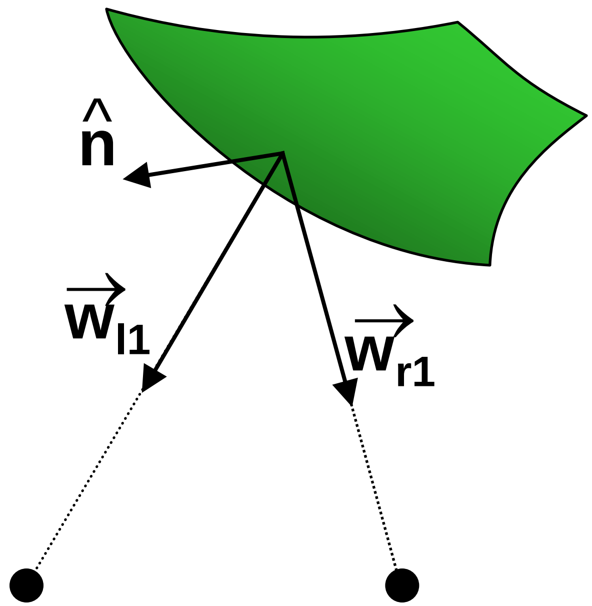
$$\left(i_l \frac{\hat{\mathrm{v}}_\mathrm{l}}{\left|o_l - p\right|^2} - i_r \frac{\hat{\mathrm{v}}_\mathrm{r}}{\left|o_r - p\right|^2} \right) \cdot \hat{\mathrm{n}} = 0$$

$\underbrace{\frac{\hat{\mathrm{v}}_\mathrm{l}}{\left|o_l - p\right|^2}}_{\vec{\mathbf{w}}_\mathbf{l}}$, $\underbrace{\frac{\hat{\mathrm{v}}_\mathrm{r}}{\left|o_r - p\right|^2}}_{\vec{\mathbf{w}}_\mathbf{r}}$

measured (i_l, i_r)

computed from geometric calibration ($\vec{\mathbf{w}}_\mathbf{l}$, $\vec{\mathbf{w}}_\mathbf{r}$)

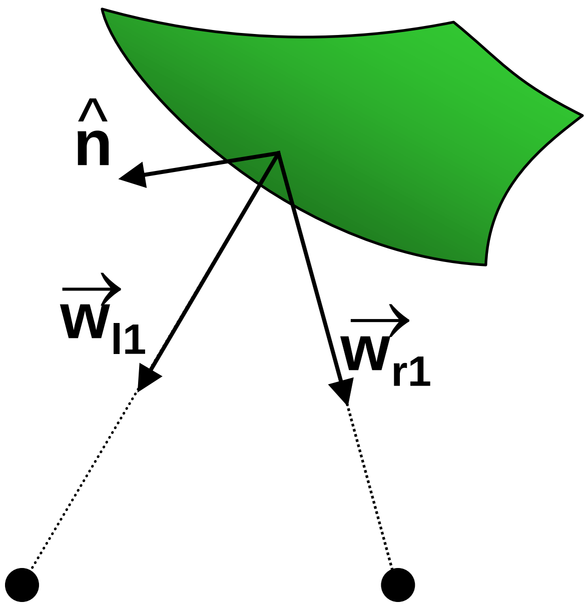
Using Multiple Helmholtz Stereo Pairs



$$\begin{pmatrix} i_{l1}\vec{w}_{l1}^{T} - i_{r1}\vec{w}_{r1}^{T} \\ i_{l2}\vec{w}_{l2}^{T} - i_{r2}\vec{w}_{r2}^{T} \\ \vdots \end{pmatrix}\hat{n} = 0$$

- **Multiple views (at least three pairs) yield a matrix constraint equation**
- **Matrix must be Rank 2**
- **Search for depth where rank constraint is satisfied**

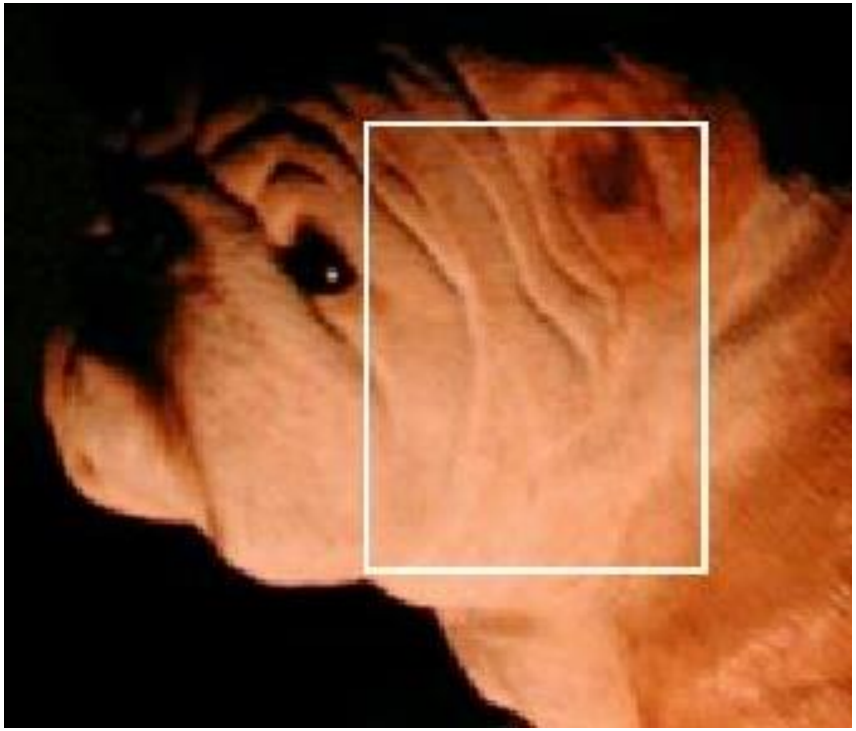
Finding the Normal at each point

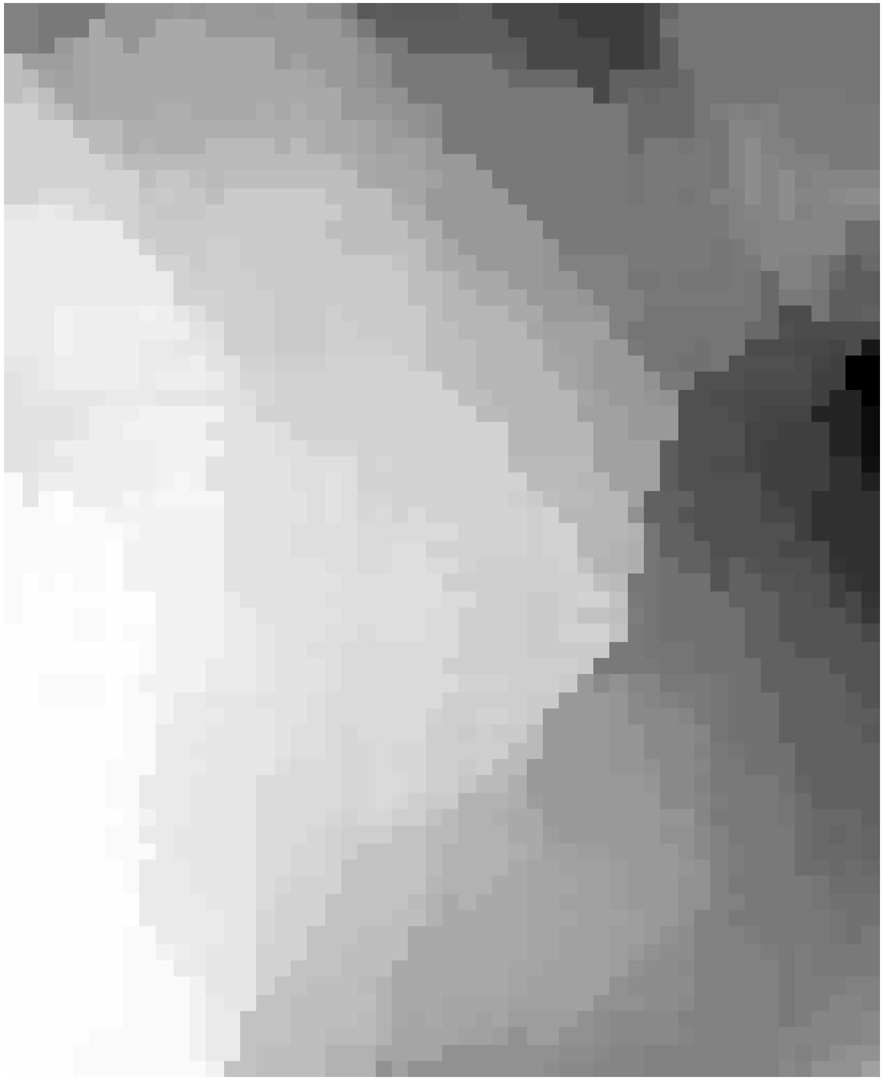


$$\begin{pmatrix} i_{l1}\vec{w}_{l1}^{T} - i_{r1}\vec{w}_{r1}^{T} \\ i_{l2}\vec{w}_{l2}^{T} - i_{r2}\vec{w}_{r2}^{T} \\ \vdots \end{pmatrix} \hat{n} = 0$$

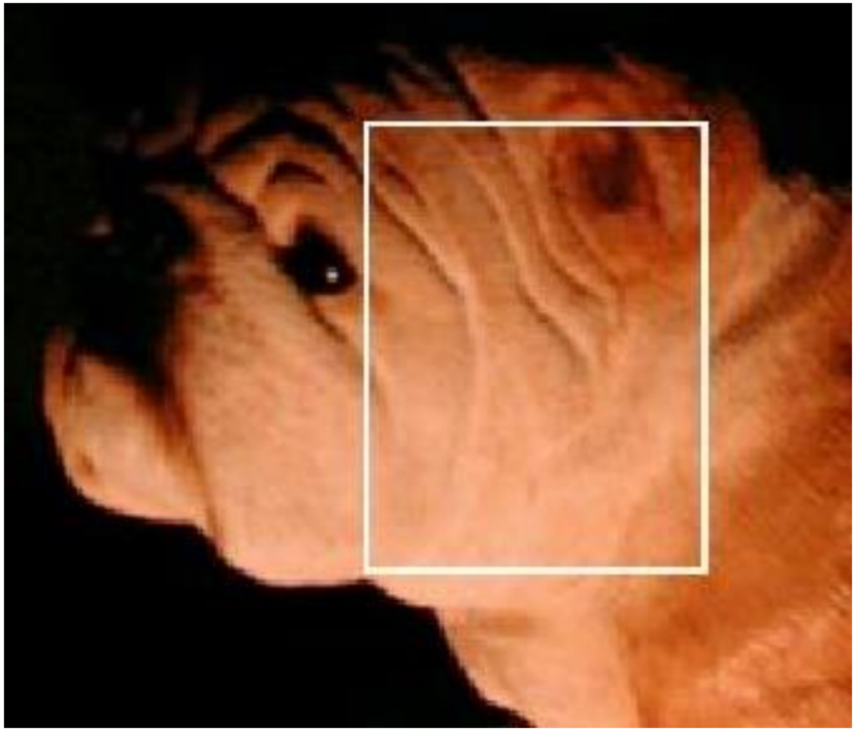
- **Additionally, the surface normal $\hat{n}$ must lie in the null space of the matrix**

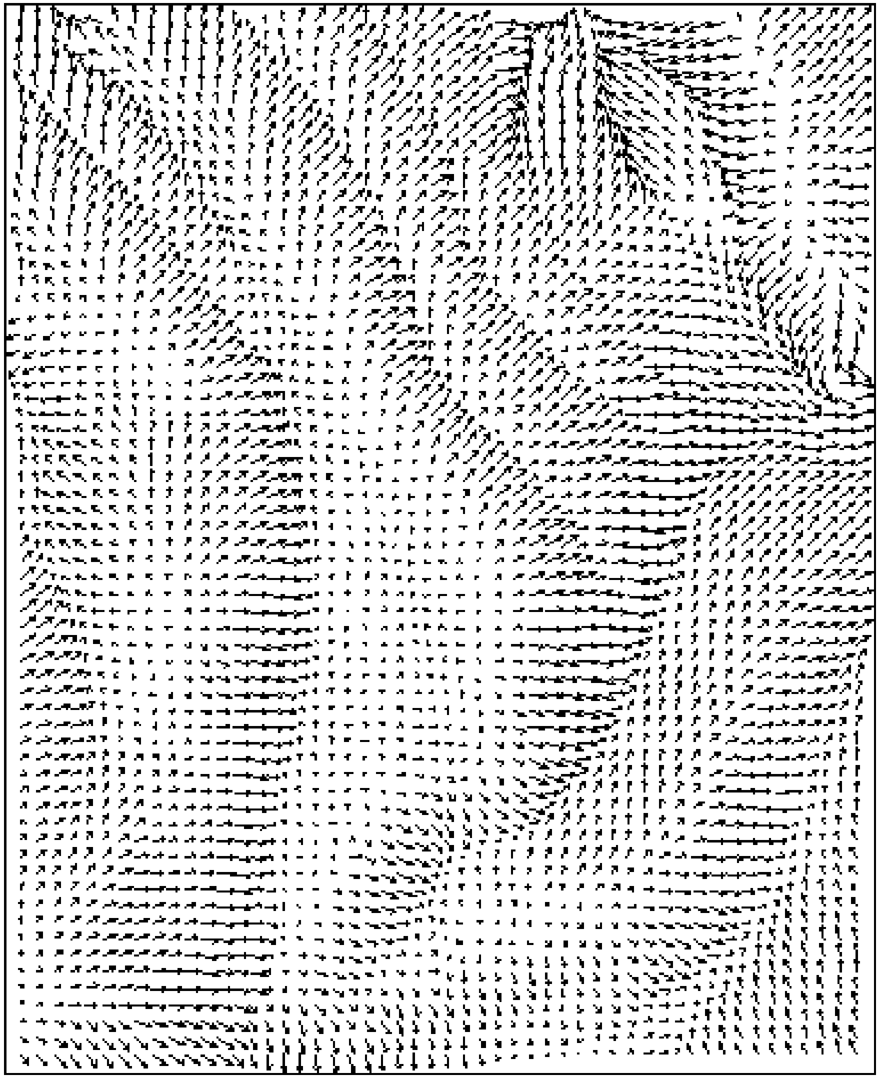
Bulldog: Disparity



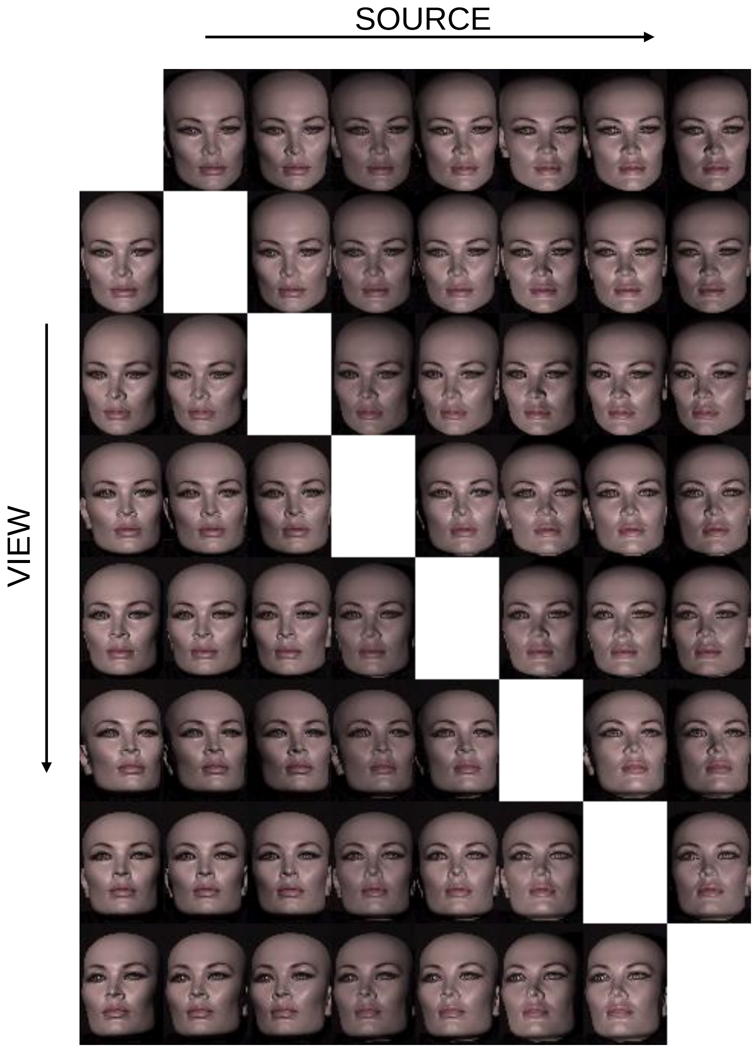


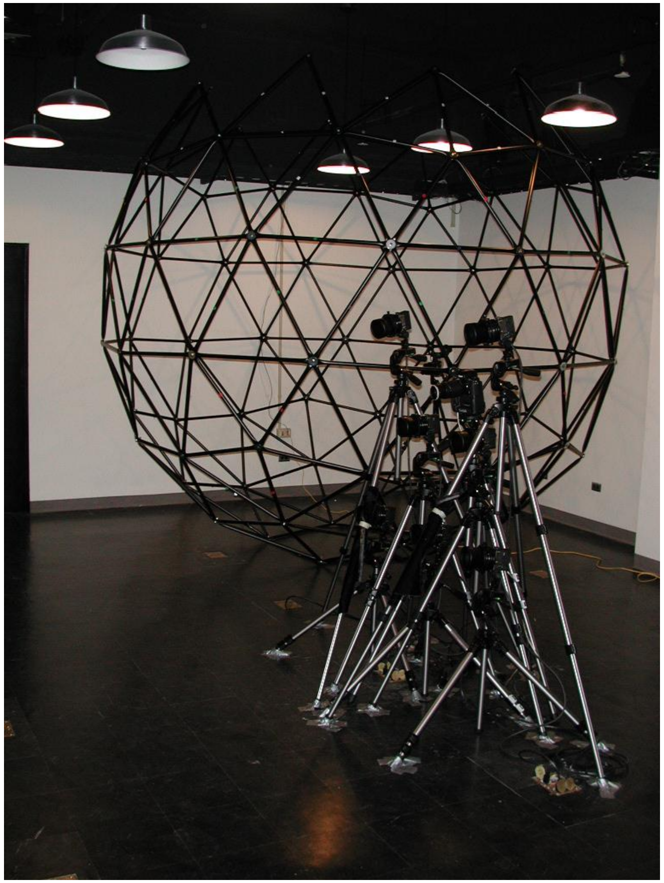
Bulldog: Normal Field



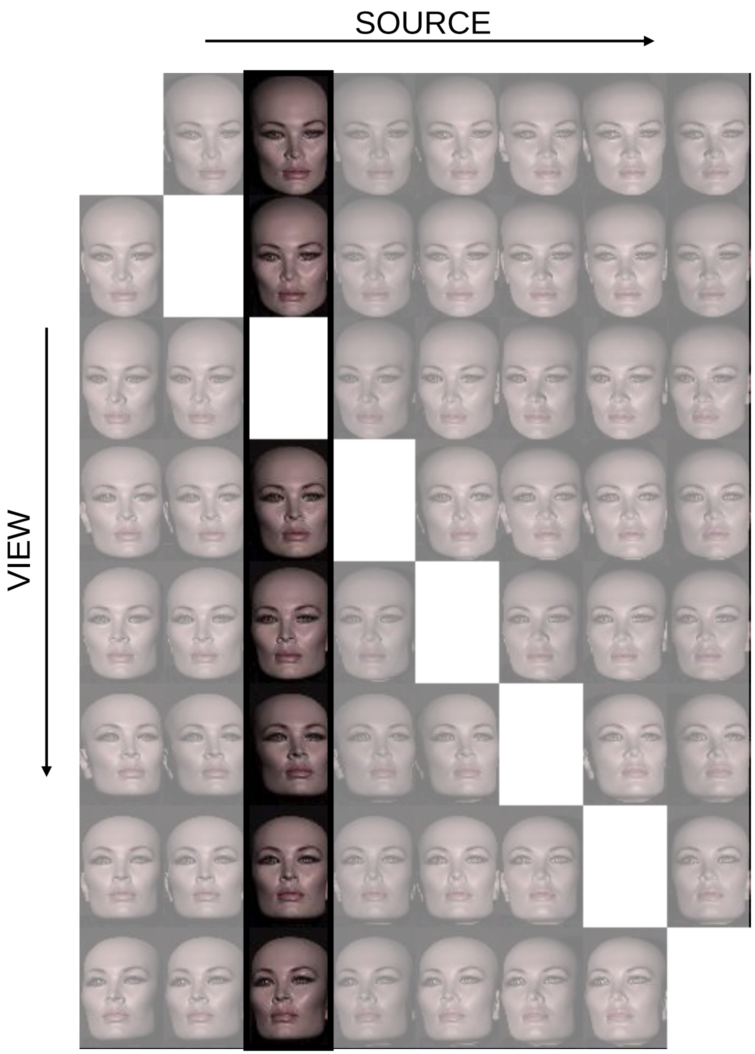


Reciprocal Images: Typical Dataset





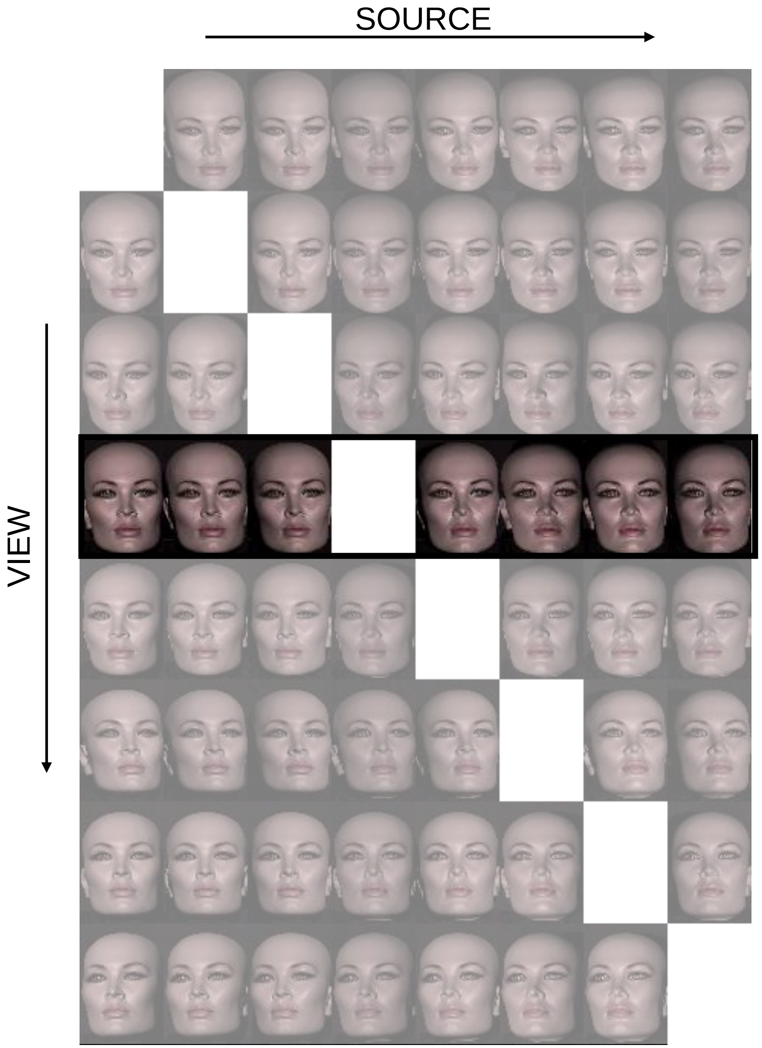
Reciprocal Images: Typical Dataset



Conventional Stereo

- Constant brightness
- No structure in textureless regions

Reciprocal Images: Typical Dataset



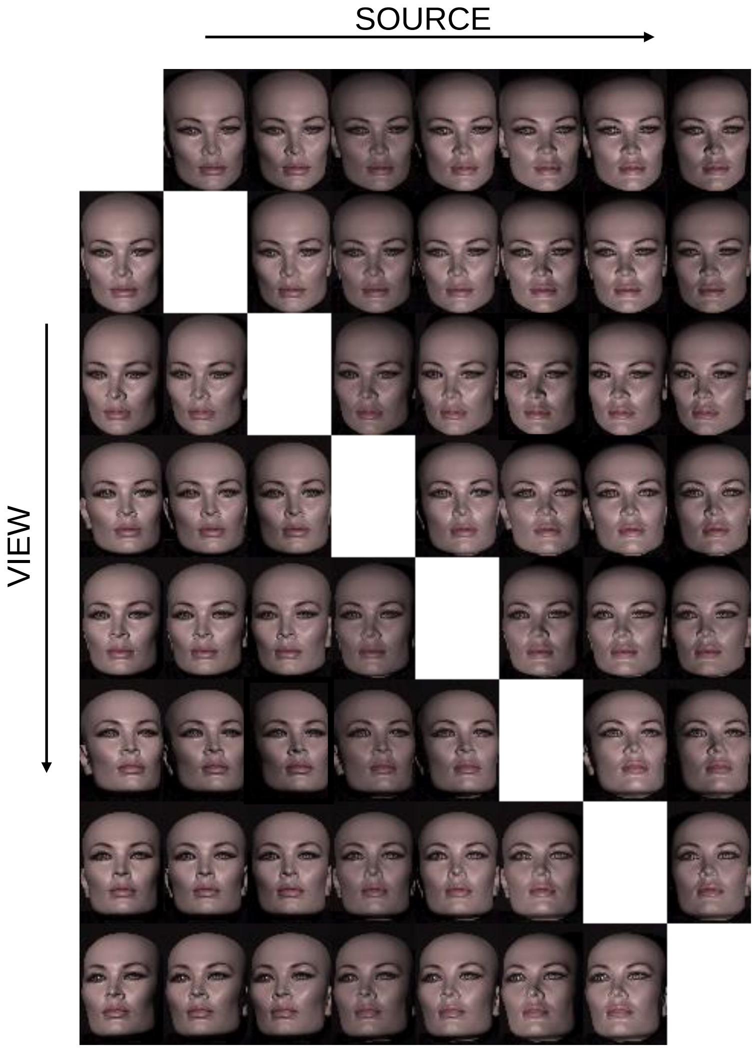
Conventional Stereo

- Constant brightness
- No structure in textureless regions

Photometric Stereo

- Needs reflectance model
- No direct depth estimates

Reciprocal Images: Typical Dataset



Conventional Stereo

- Constant brightness
- No structure in textureless regions

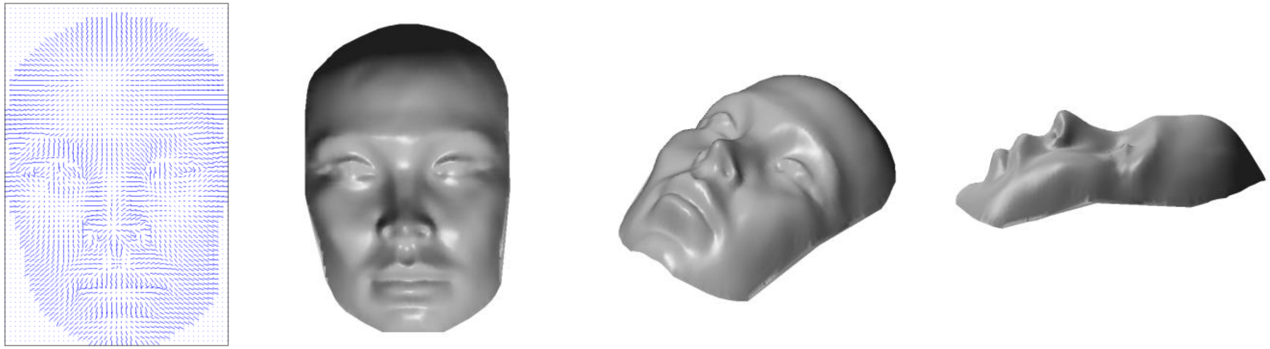
Photometric Stereo

- Needs reflectance model
- No direct depth estimates

Helmholtz Stereo

- No assumed reflectance
- Gives depth and surface normals

Metric Reconstruction



Next Lecture

- Uncalibrated stereo and feature extraction