GAMES
Multi-View Stereo

Qixing Huang

August 6th 2021

Two-View Stereo

Binocular Stereo

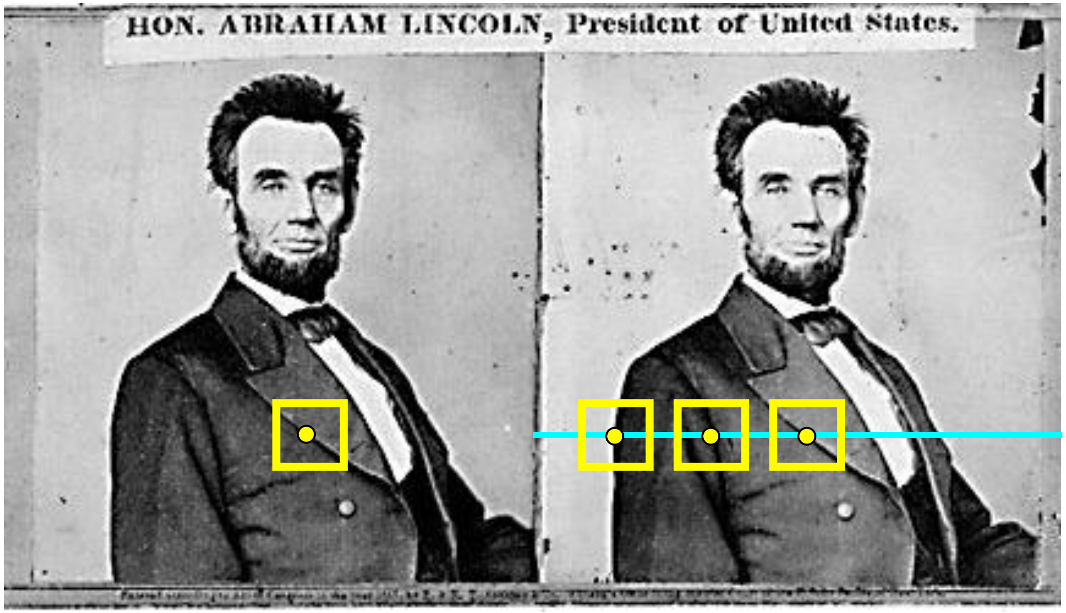
- Given a calibrated binocular stereo pair, fuse it to produce a depth image

image 1

image 2

Dense depth map

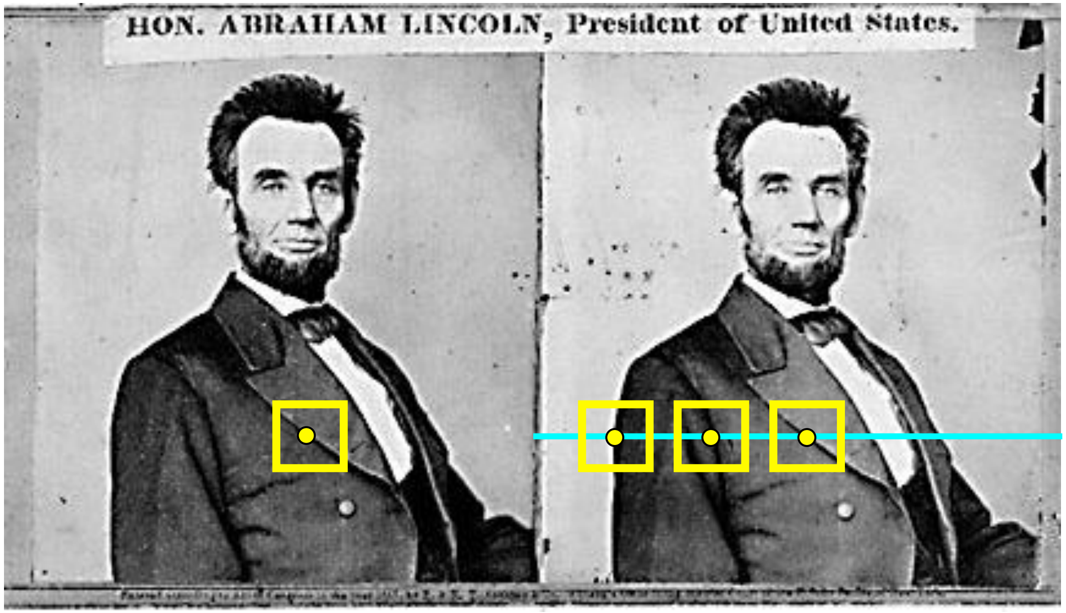
Basic Stereo Matching Algorithm



- For each pixel in the first image
  - Find corresponding epipolar line in the right image
  - Examine all pixels on the epipolar line and pick the best match
  - Triangulate the matches to get depth information

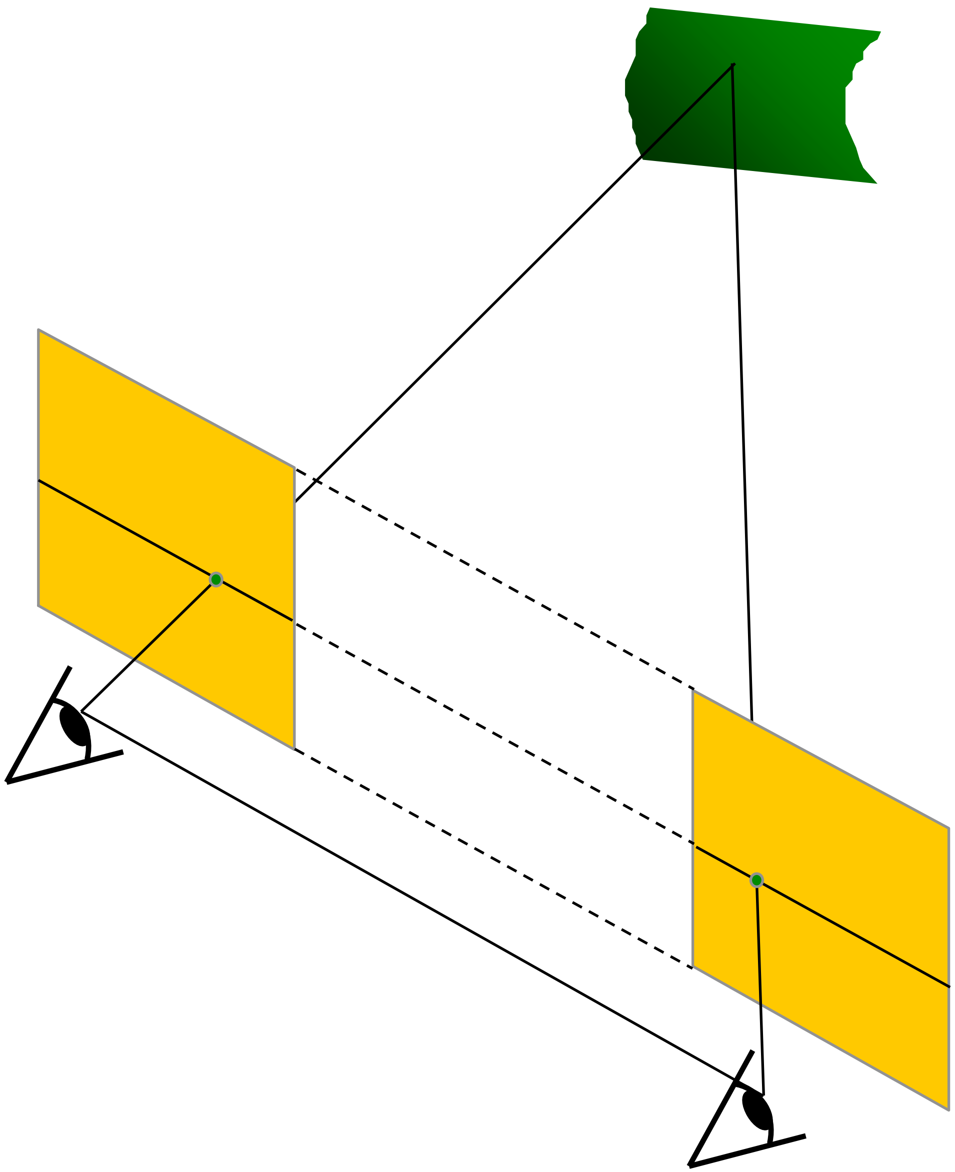
- Simplest case: epipolar lines are corresponding scanlines
  - When does this happen?

Basic stereo matching algorithm



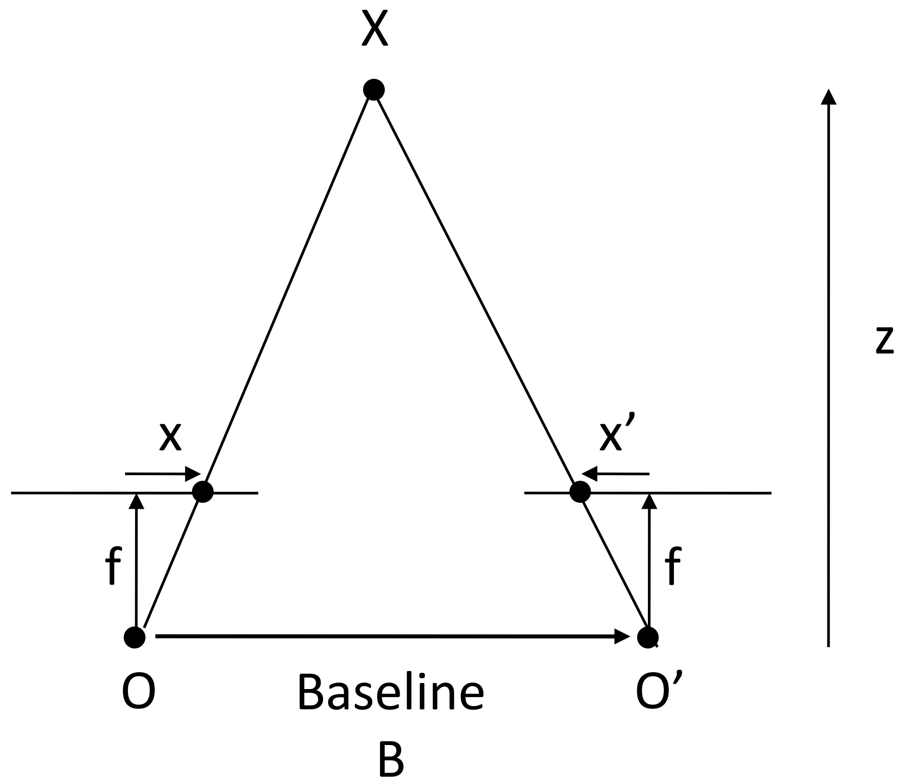
- For each pixel in the first image
  - Find corresponding epipolar line in the right image
  - Examine all pixels on the epipolar line and pick the best match
  - Triangulate the matches to get depth information
- Simplest case: epipolar lines are corresponding scanlines
  - When does this happen?

Simplest Case: Parallel Images



- Image planes of cameras are parallel to each other and to the baseline
- Camera centers are at same height
- Focal lengths are the same
- Then, epipolar lines fall along the horizontal scan lines of the images

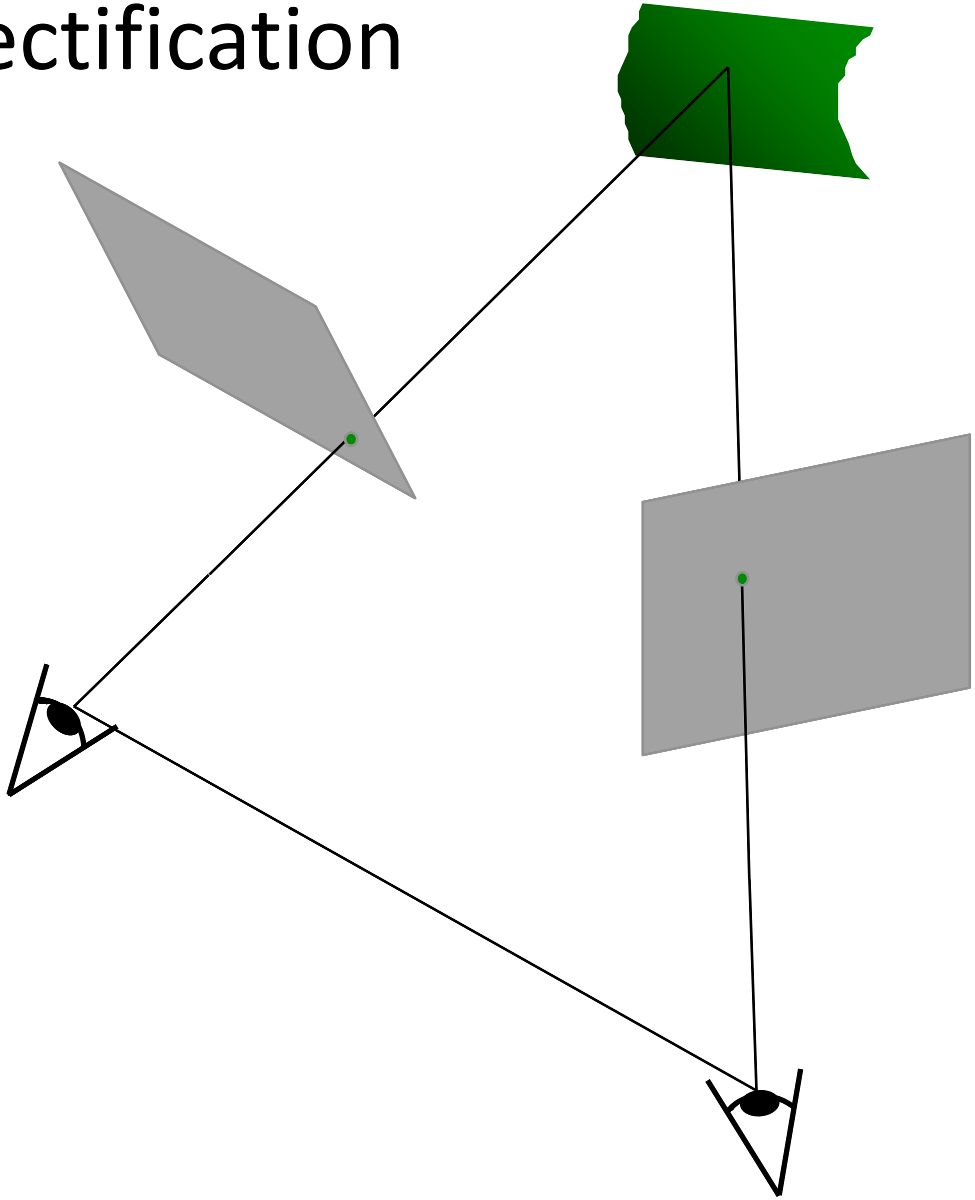
Depth from Disparity



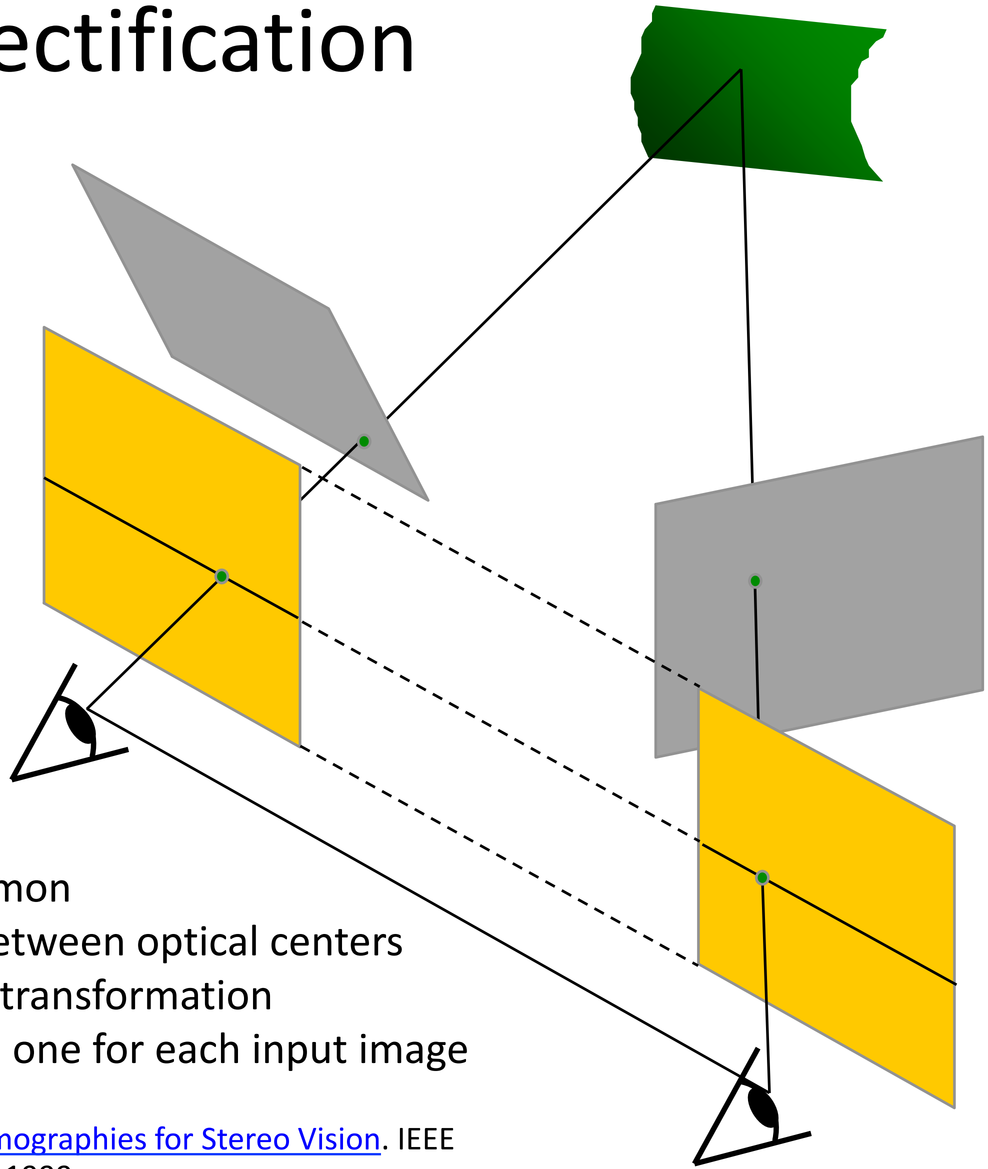
$$disparity = x - x' = \frac{B \times f}{z}$$

Disparity is inversely proportional to depth!

Stereo Image Rectification



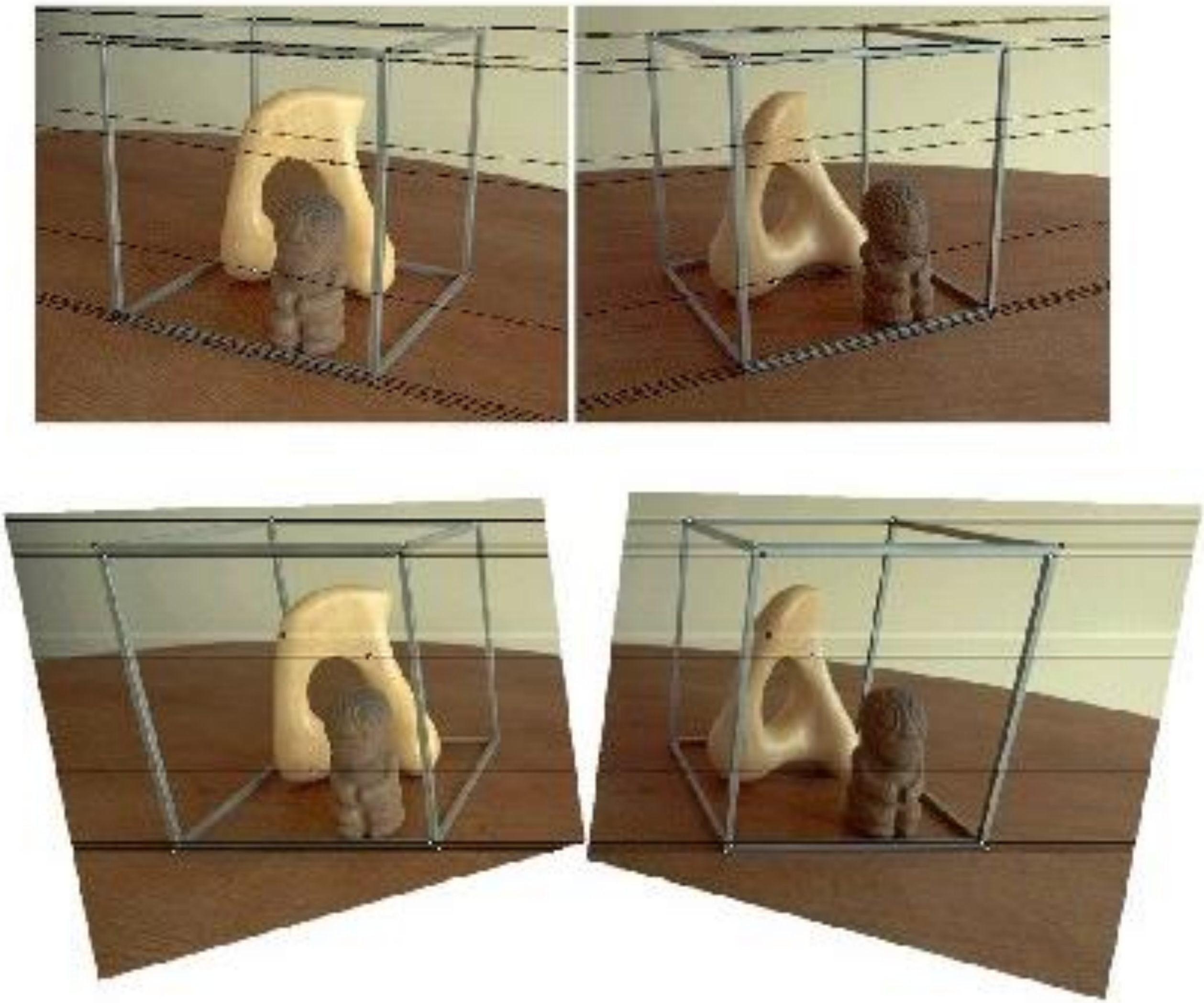
Stereo Image Rectification



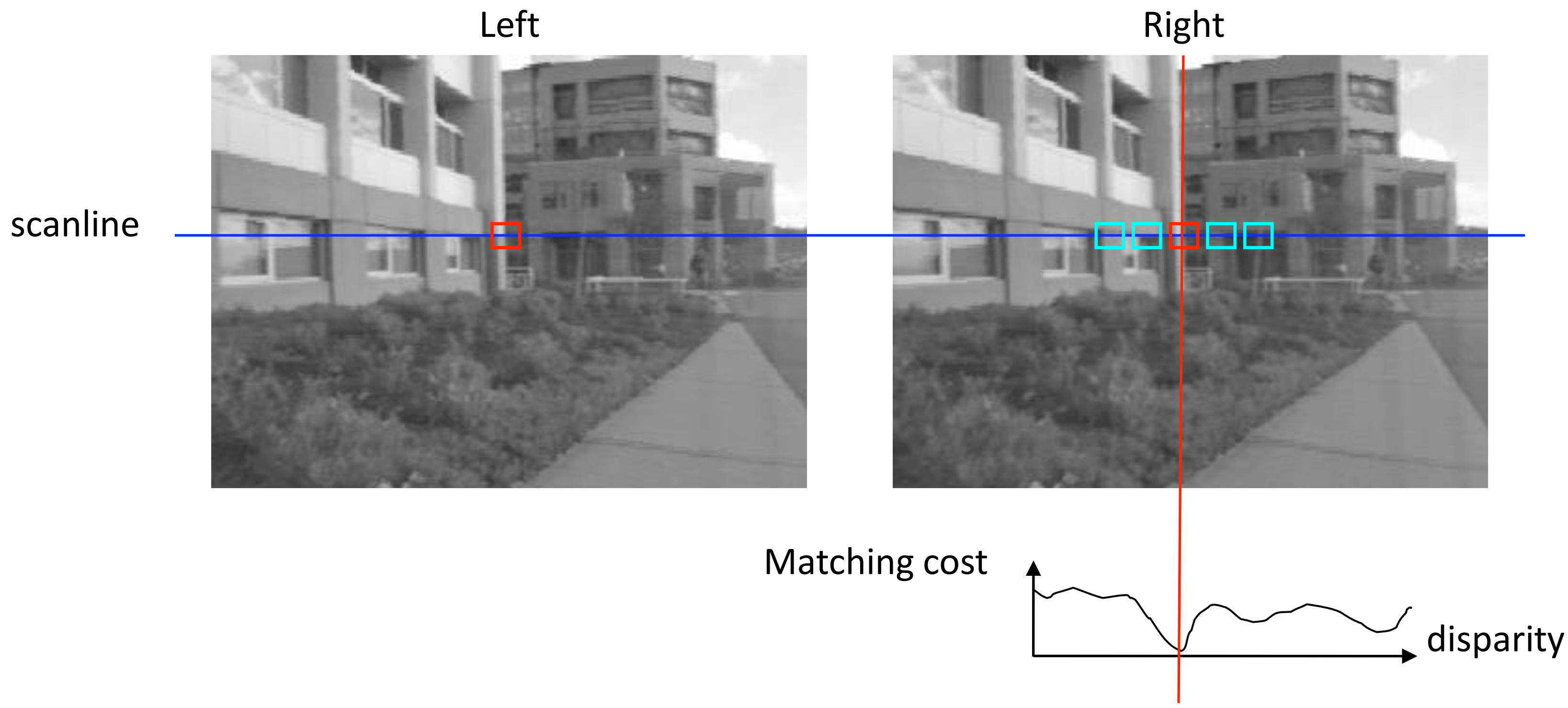
- reproject image planes onto a common
- plane parallel to the line between optical centers
- pixel motion is horizontal after this transformation
- two homographies (3x3 transform), one for each input image reprojection

➢C. Loop and Z. Zhang. [Computing Rectifying Homographies for Stereo Vision](). IEEE Conf. Computer Vision and Pattern Recognition, 1999.

Rectification Example

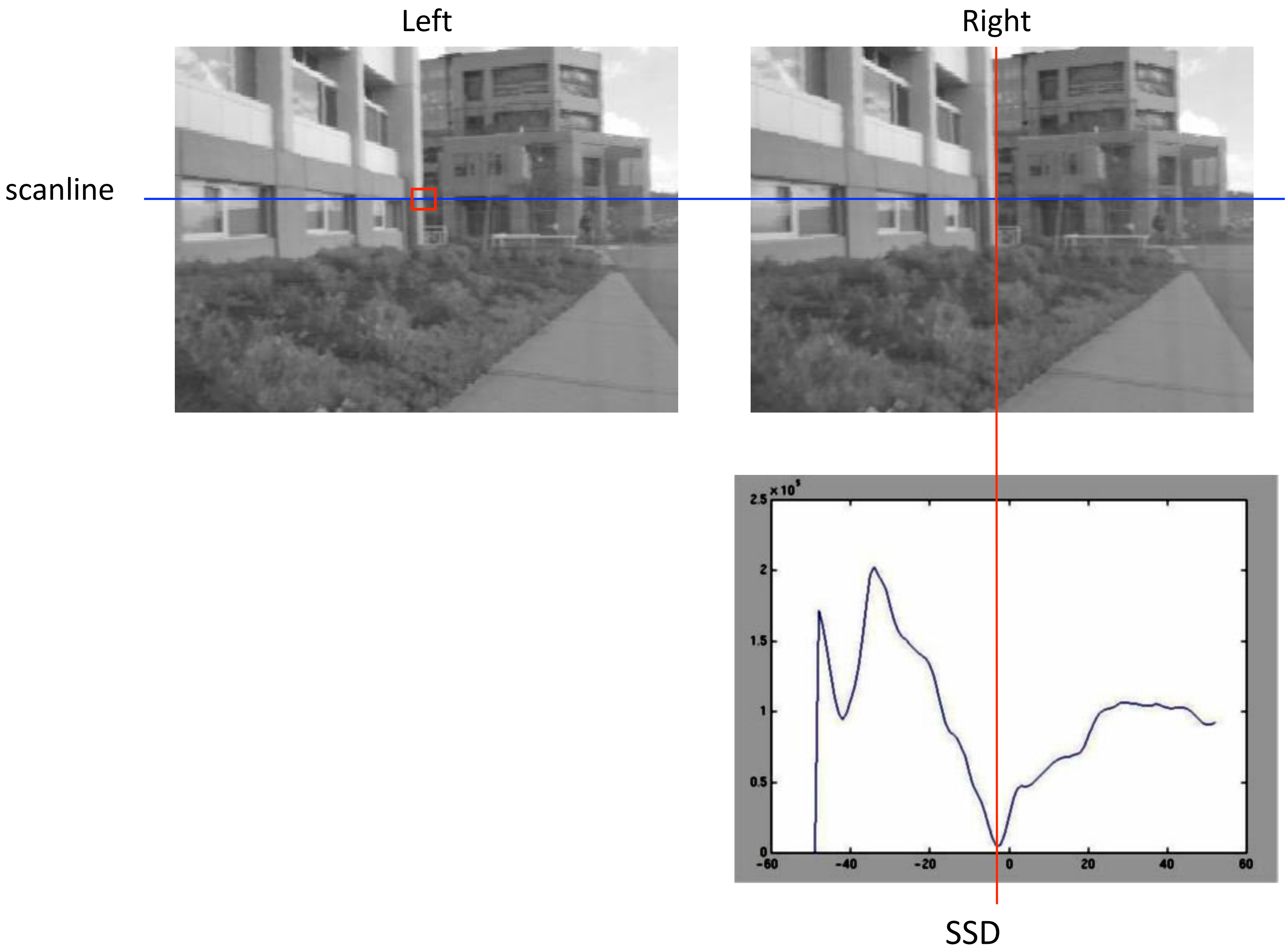


Correspondence search



- Slide a window along the right scanline and compare contents of that window with the reference window in the left image
- Matching cost: SSD or normalized correlation

Correspondence search

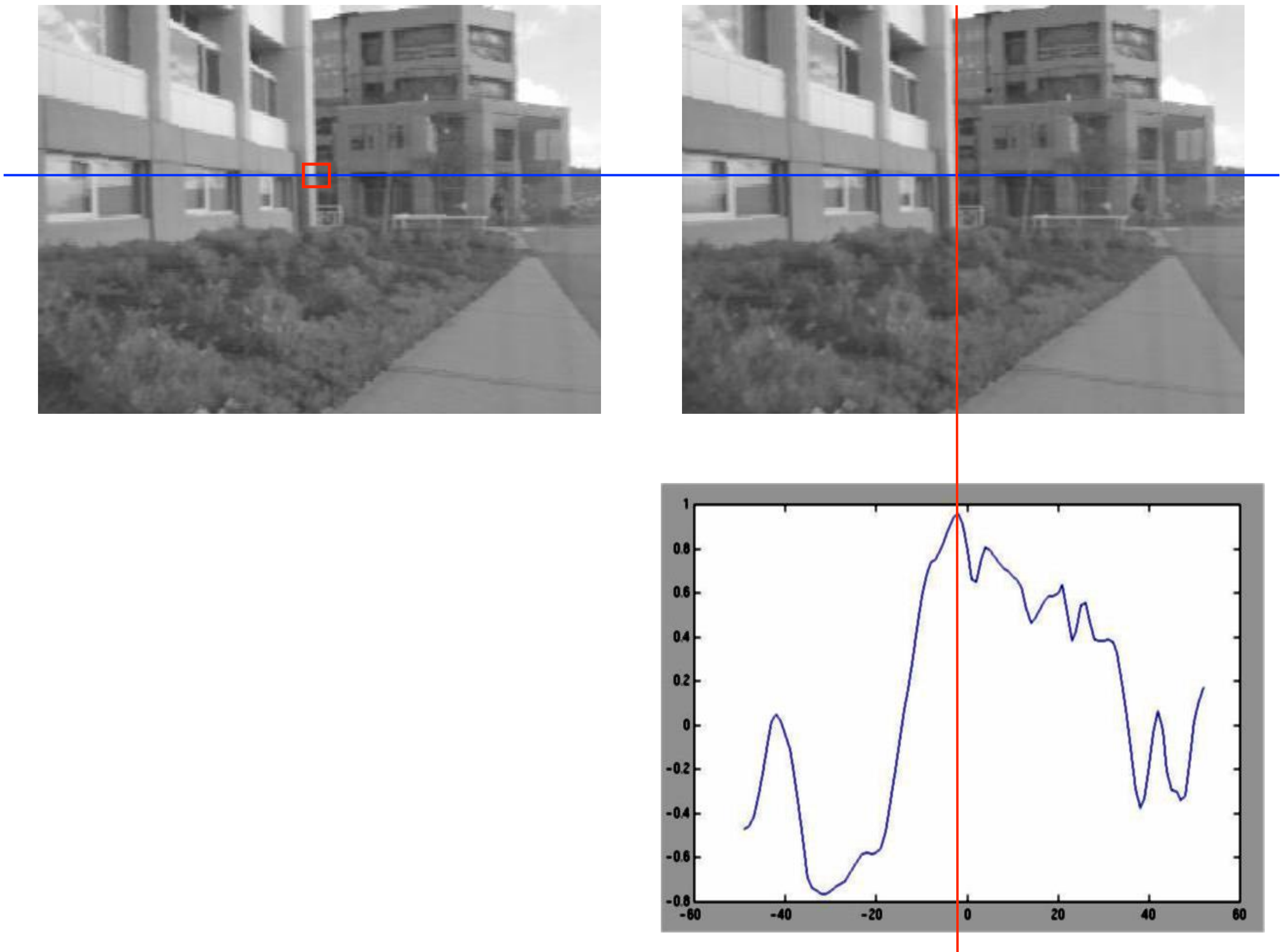


Correspondence search

Left

Right

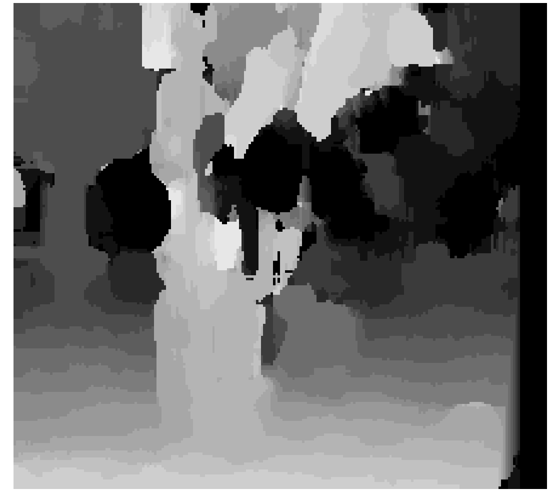
scanline



Norm. corr

Effect of window size

W = 3



W = 20

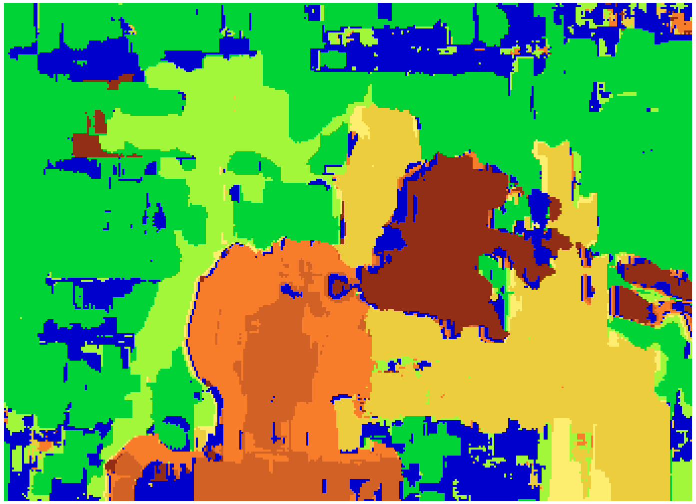
– Smaller window
  - \+ More detail
  - – More noise

– Larger window
  - \+ Smoother disparity maps
  - – Less detail

Results with window search

Data

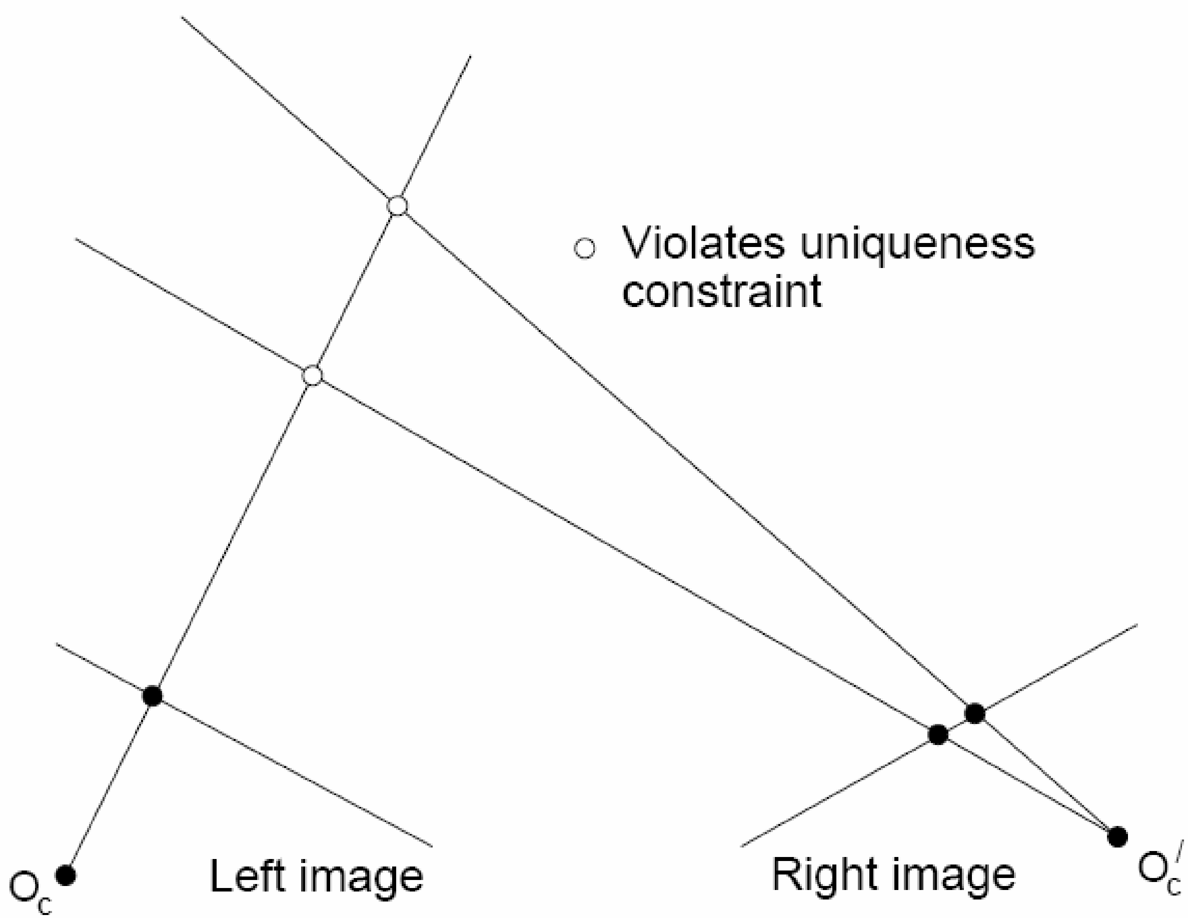
Window-based matching



Ground truth

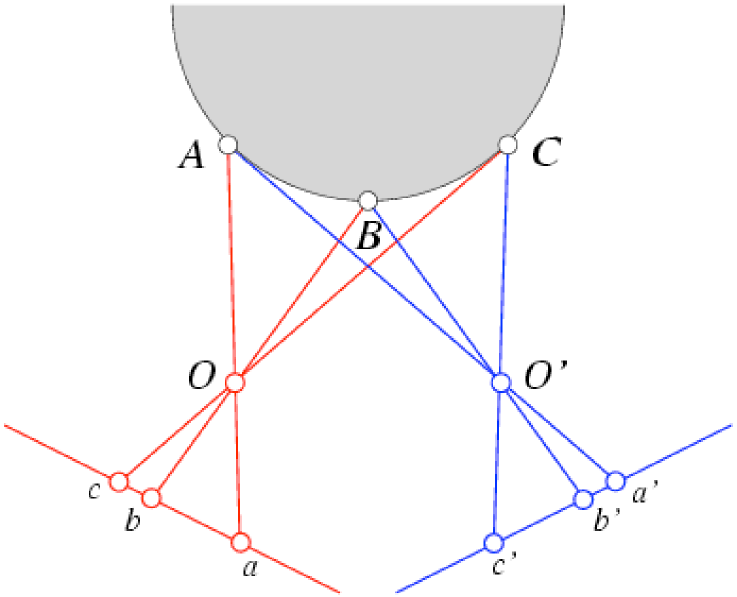
Non-local constraints

- Uniqueness
  - For any point in one image, there should be at most one matching point in the other image



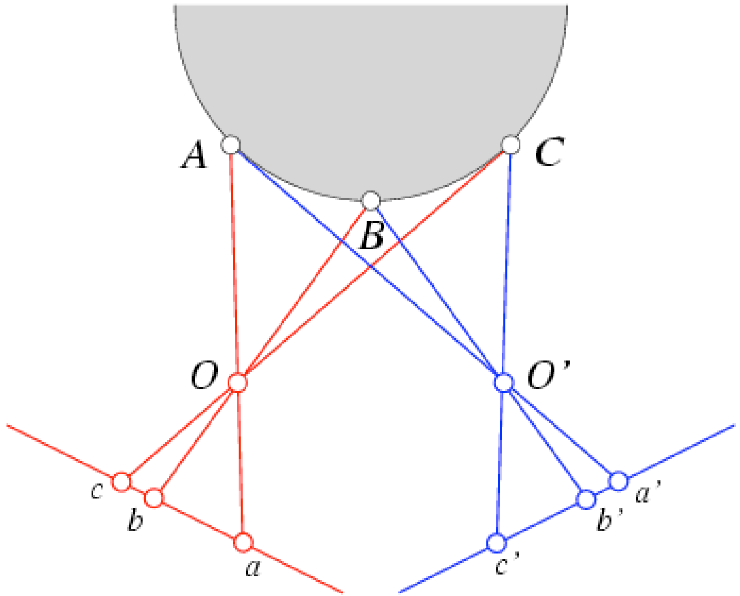
Non-local constraints

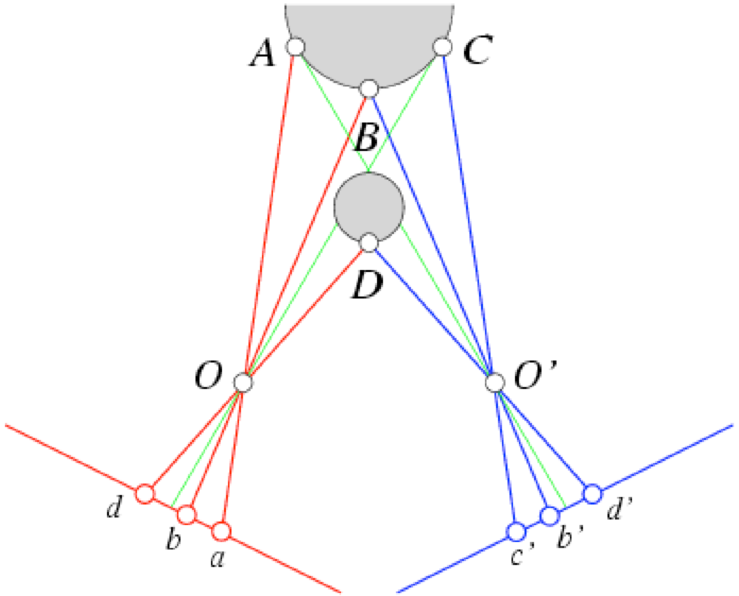
- Uniqueness
  - For any point in one image, there should be at most one matching point in the other image
- Ordering
  - Corresponding points should be in the same order in both views



Non-local constraints

- Uniqueness
  - For any point in one image, there should be at most one matching point in the other image
- Ordering
  - Corresponding points should be in the same order in both views





Ordering constraint doesn't hold

Consistency Constraints

- Uniqueness
  - For any point in one image, there should be at most one matching point in the other image
- Ordering
  - Corresponding points should be in the same order in both views
- Smoothness
  - We expect disparity values to change slowly (for the most part)

MRF Formulation:

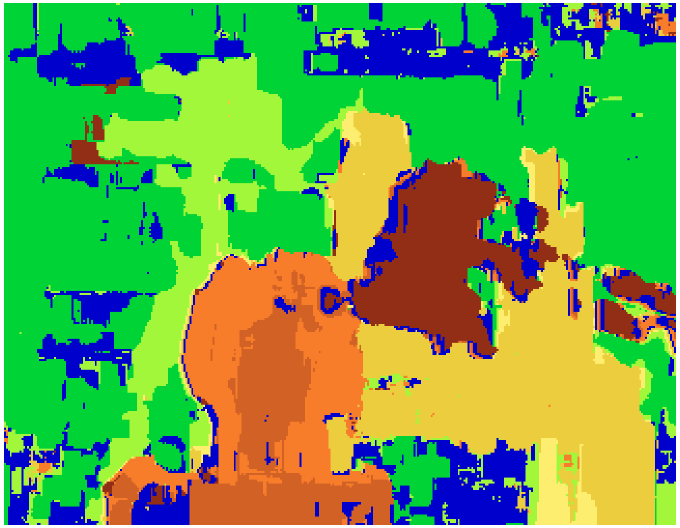
$$E(d) = E_d(d) + \lambda E_s(d)$$

Pixel matching score

Consistency Scores

Comparsion

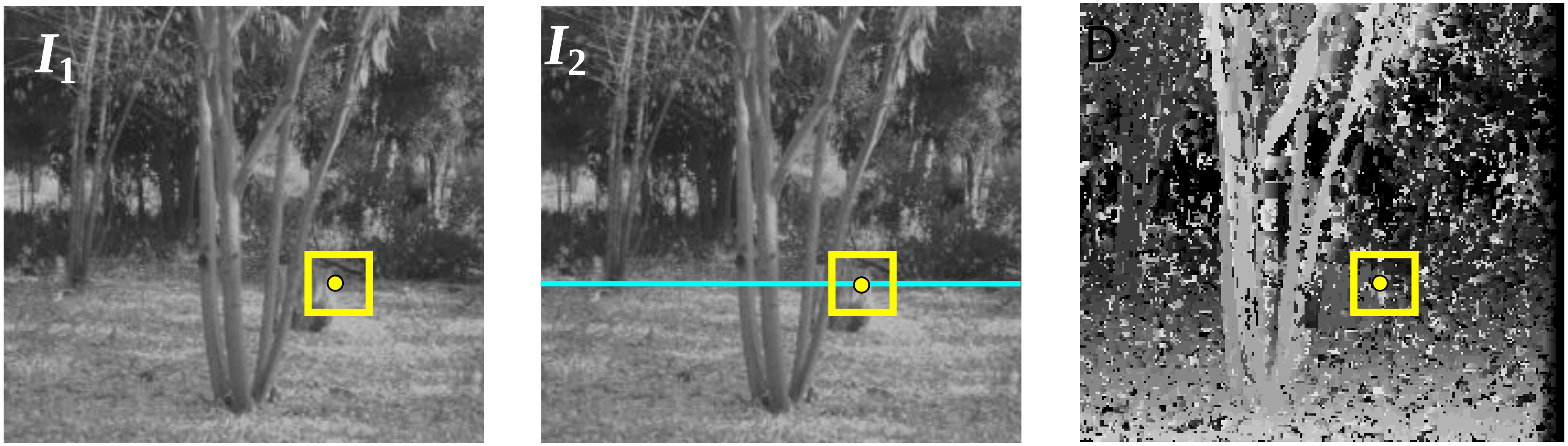
Window-Based Search:



Graph Cut:

Ground Truth

Stereo matching as energy minimization



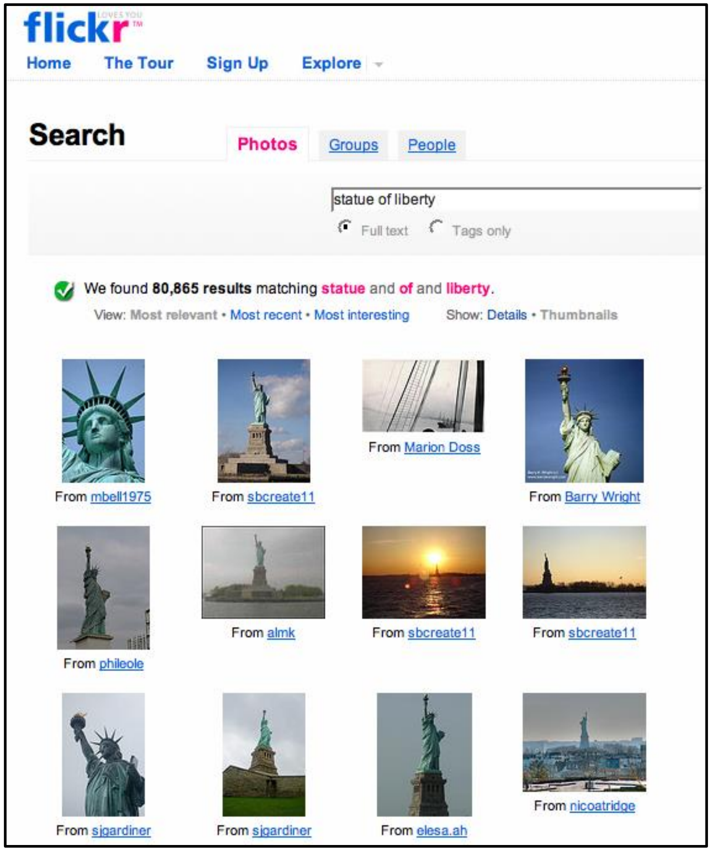
- Graph-cuts can be used to minimize such energy

Y. Boykov, O. Veksler, and R. Zabih, [Fast Approximate Energy Minimization via Graph Cuts](), PAMI 2001

Multi-View Stereo

Multi-View Stereo from Internet Collections

[Goesele et al. 07]



Challenges

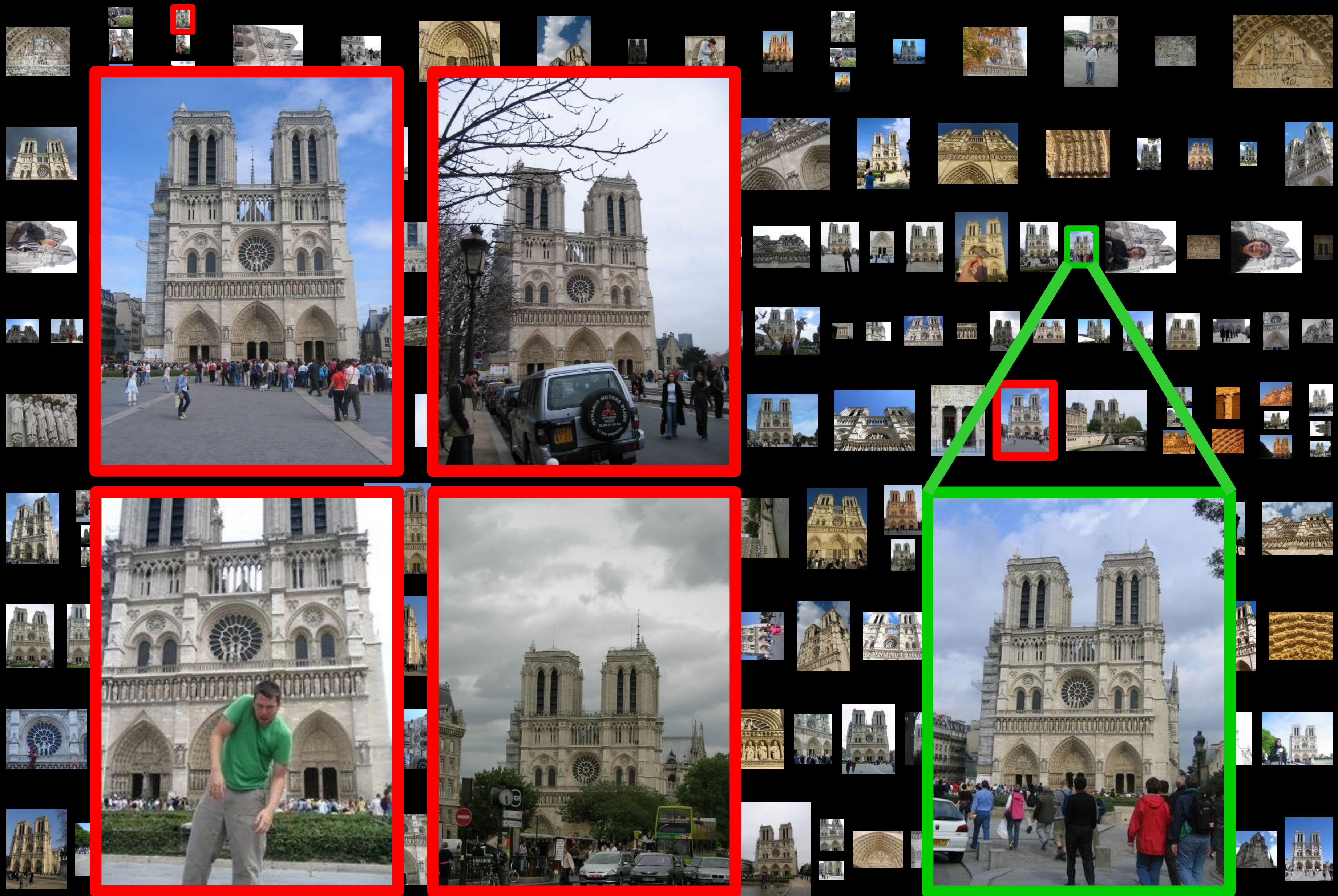
- Appearance variation

- Resolution

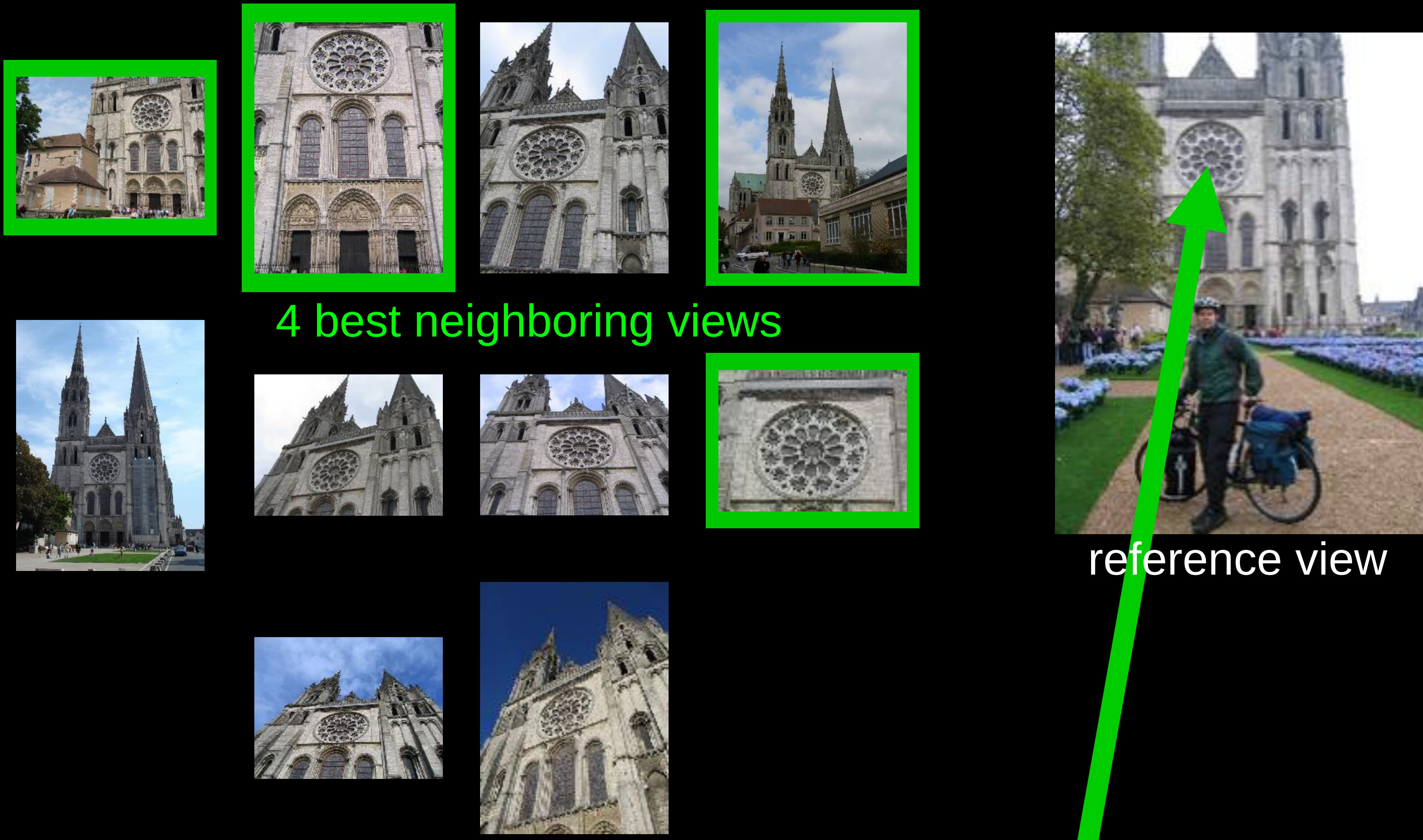
- Massive collections

82754 results for photos matching **notre** and **dame** and **paris**

Law of Nearest Neighbors

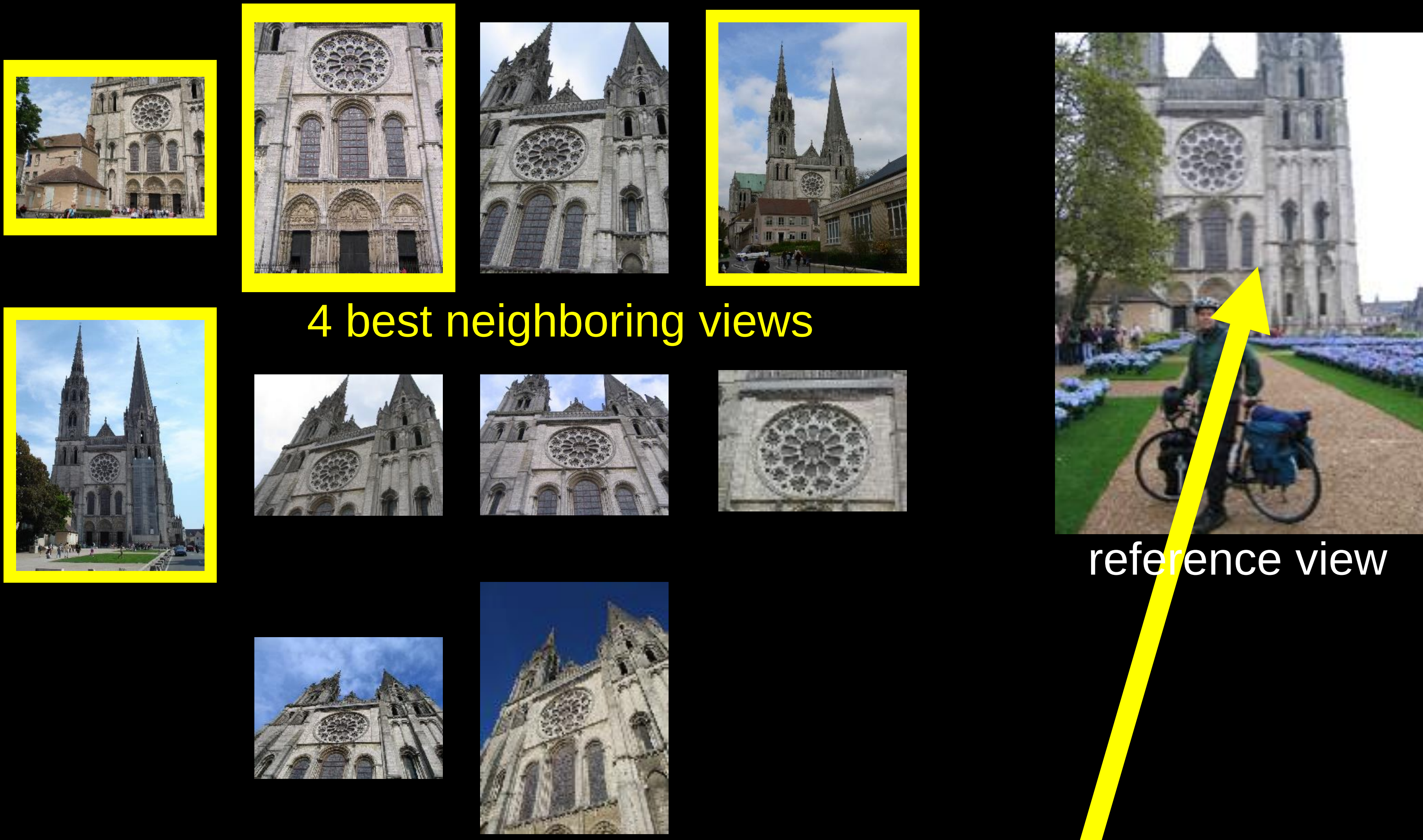


206 *Flickr* images taken by 92 photographers



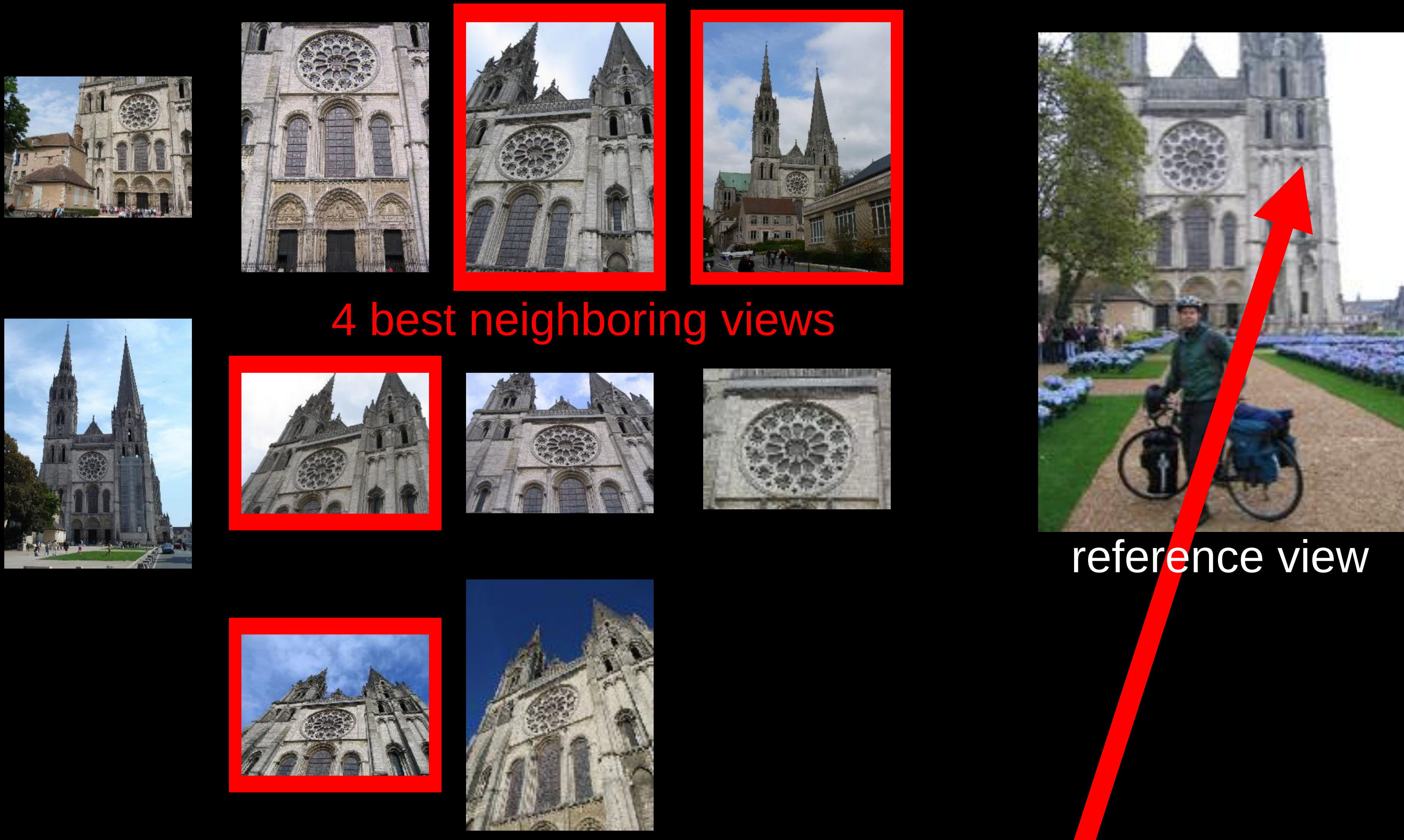
Local view selection

- Automatically select neighboring views for each point in the image
- Desiderata: good matches AND good baselines



Local view selection

- Automatically select neighboring views for each point in the image
- Desiderata: good matches AND good baselines

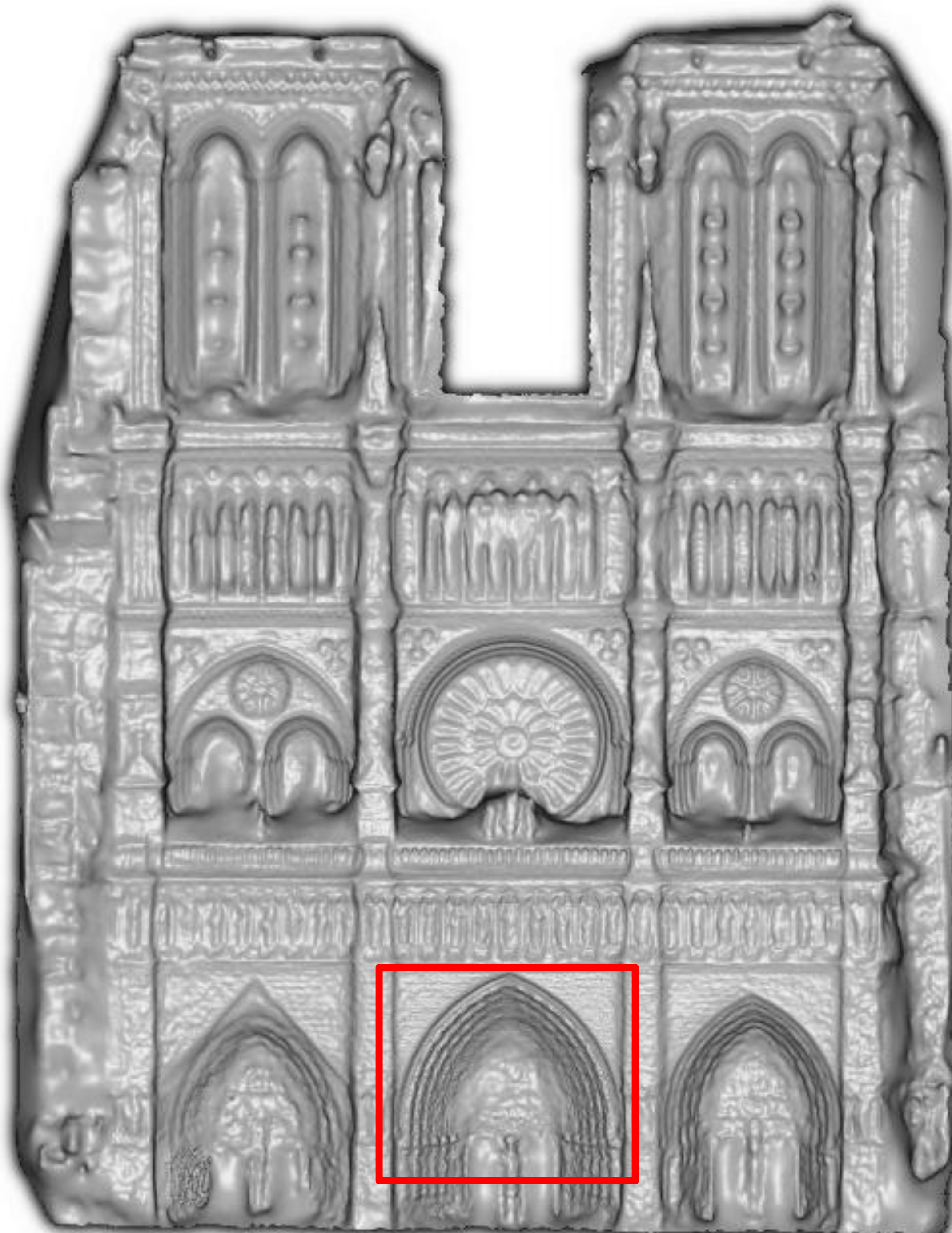


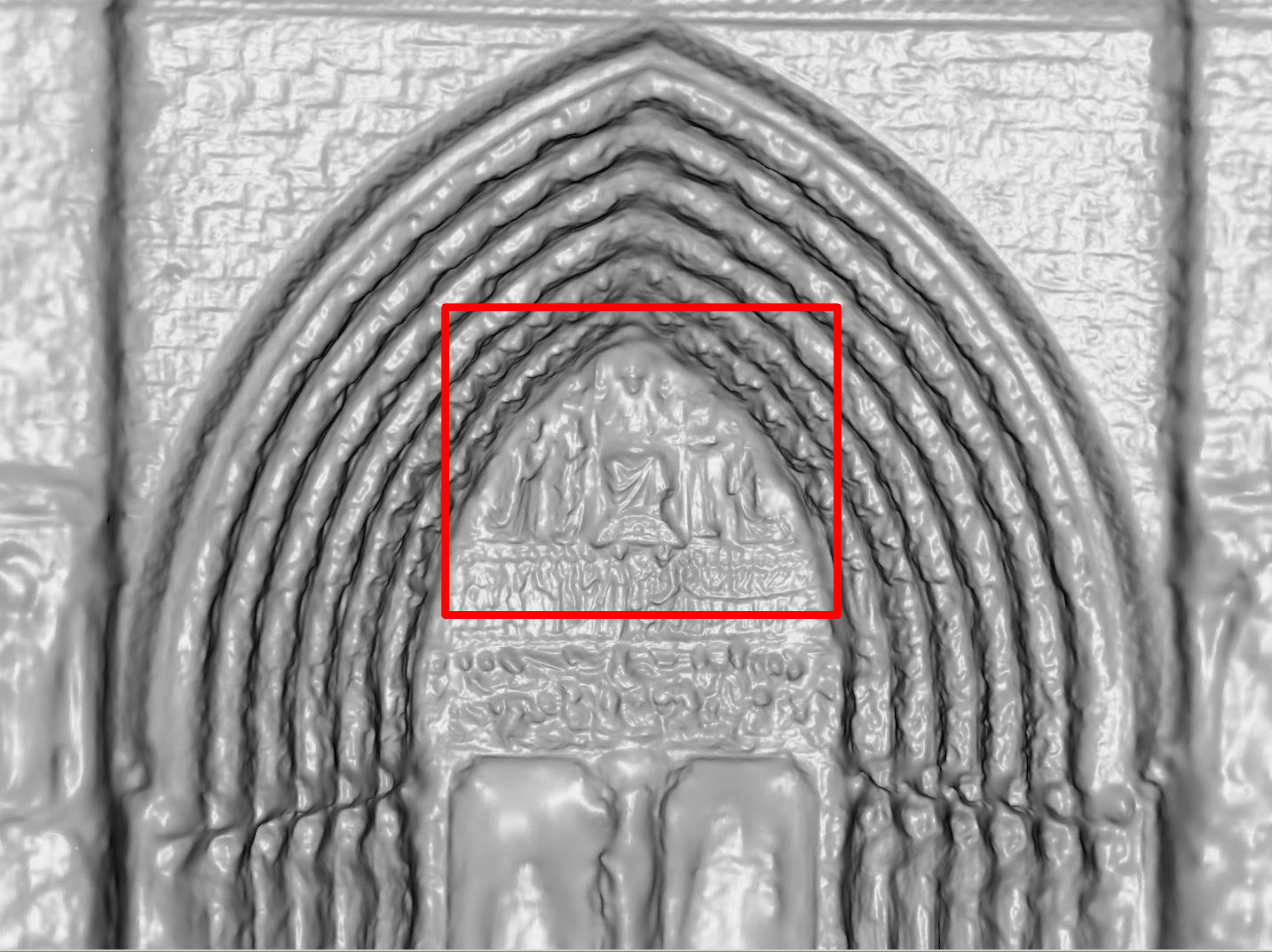
Local view selection

- Automatically select neighboring views for each point in the image
- Desiderata: good matches AND good baselines

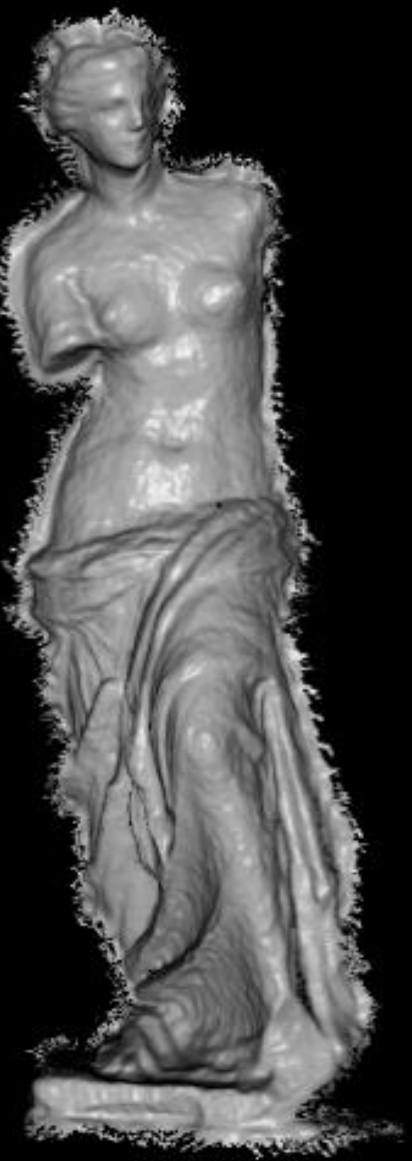
Notre Dame de Paris

653 images
313 photographers



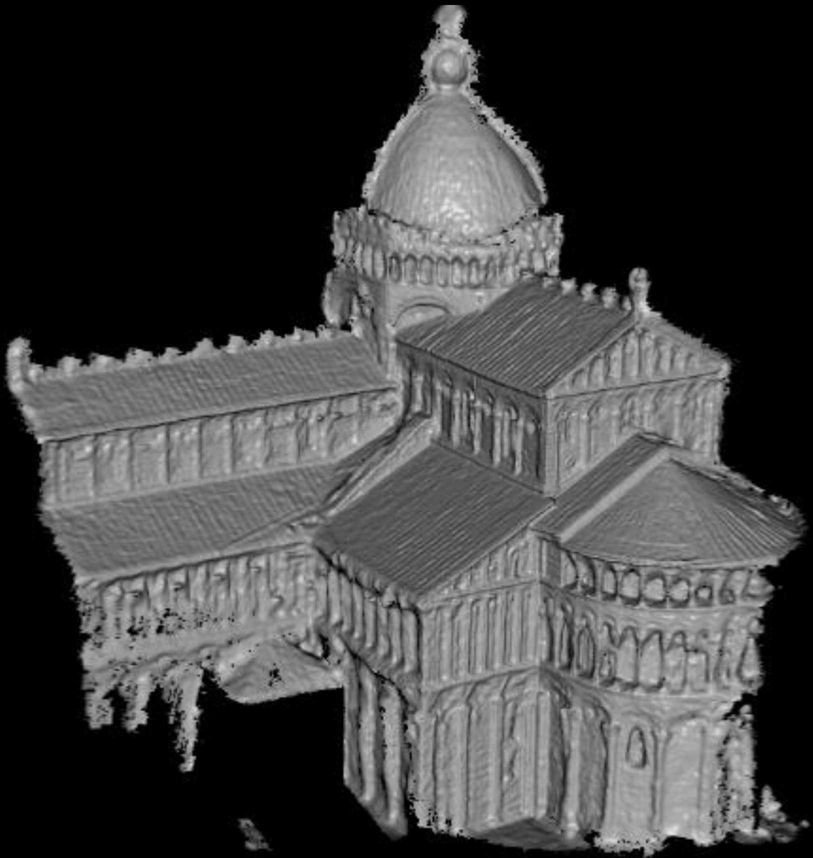


129 *Flickr* images taken by 98 photographers



merged model of Venus de Milo

56 *Flickr* images taken by 8 photographers

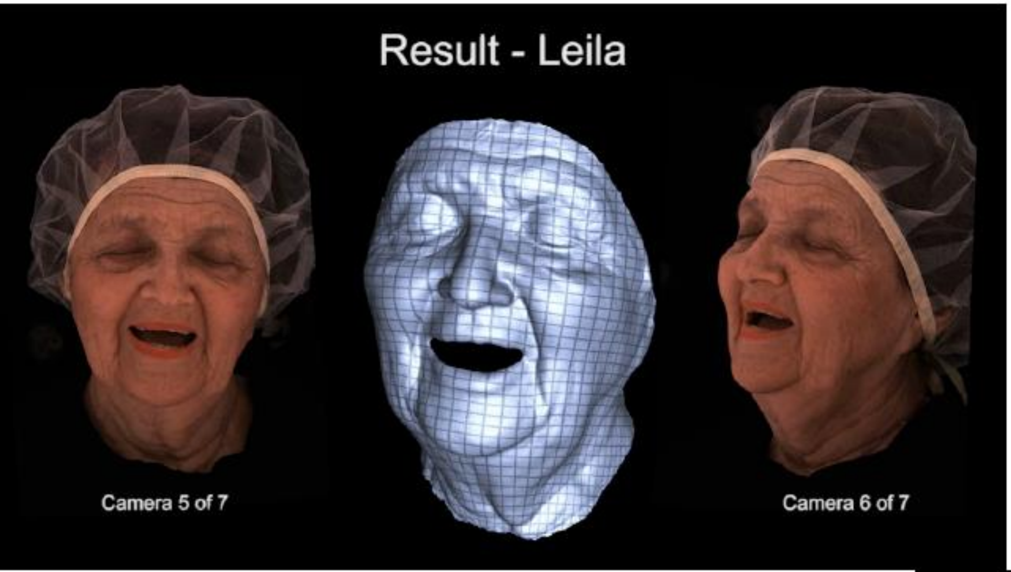


merged model of Pisa Cathedral

Accuracy compared to laser scanned model:
90% of points within *0.25%* of ground truth

Use of Structural Priors

MVS has been so successful



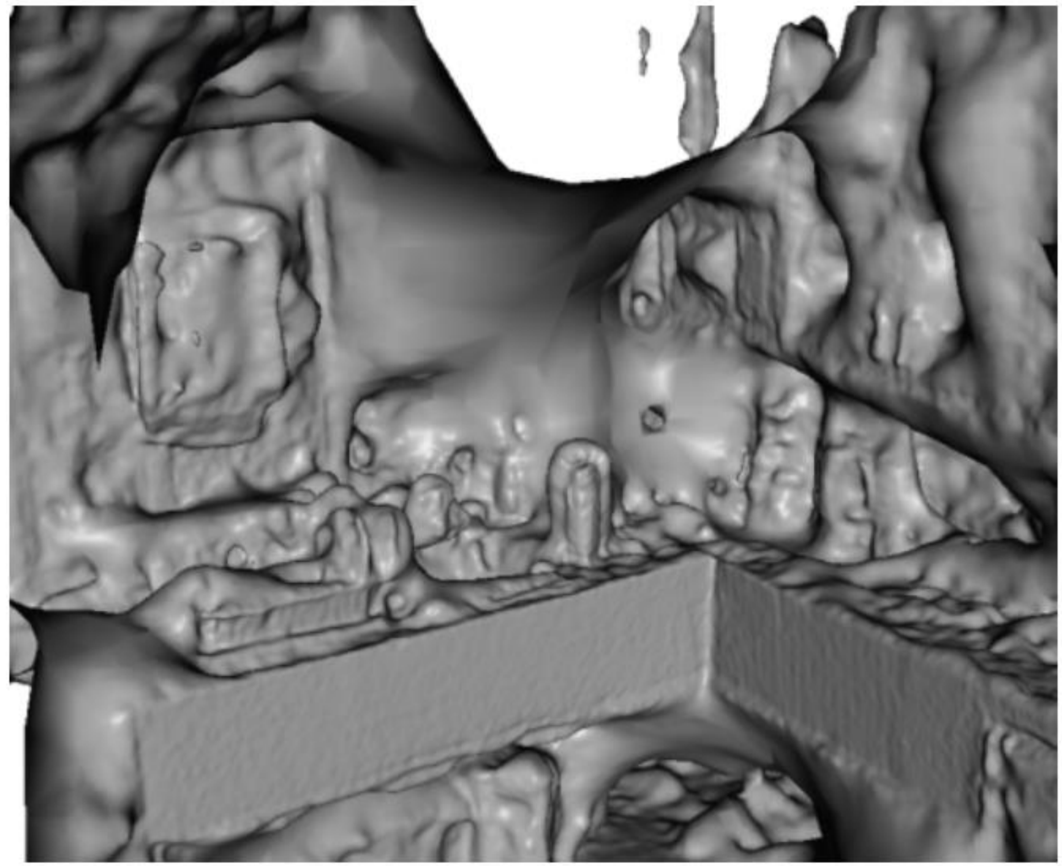
High-quality passive facial performance capture using anchor frames
[T. Beeler, F. Hahn, D. Bradley, B. Bickel,
P. Beardsley, C. Gotsman, M. Gross, 2011]

[Apple maps]

What if no texture...

What if no texture...

Existing Method Fails

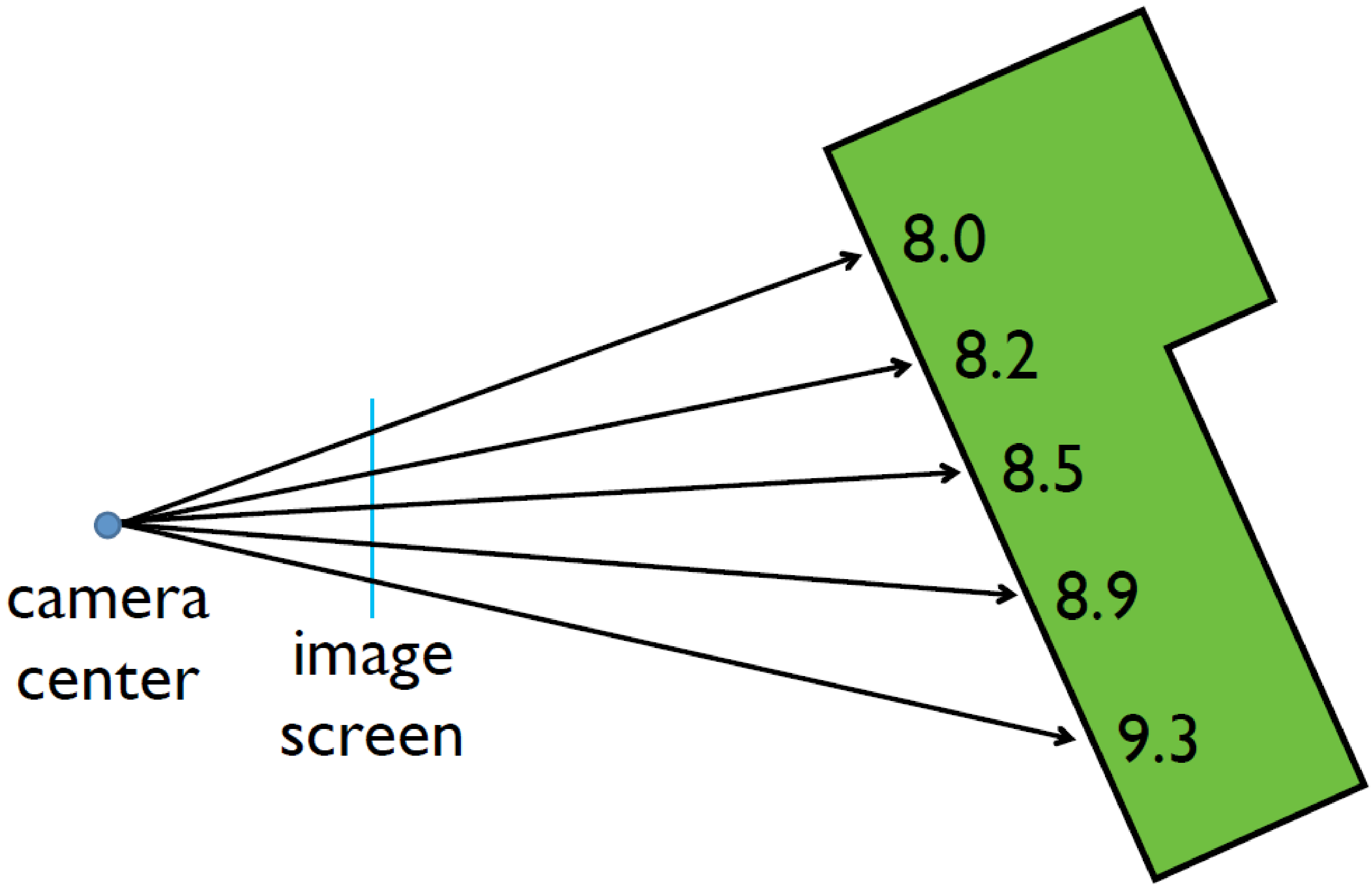


[Furukawa and Ponce, 2007]

Enforce Prior in MVS

- Manhattan-world assumption
  - Planarity
  - Orthogonality

Depthmap Estimation



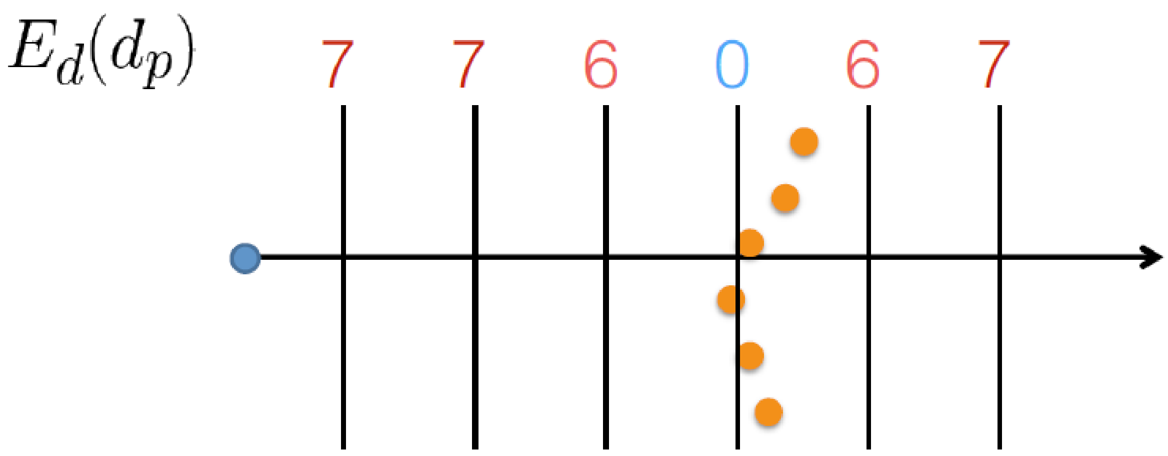
Standard Depthmap Markov Random Field (MRF)

$$E = \sum_p E_d(d_p) + \sum_{\{p,q\} \in \mathbb{N}} E_s(d_p, d_q)$$

A Comparative Study of Energy Minimization Methods for Markov Random Fields with Smoothness-Based Priors [Szeliski et al., PAMI 2008]

Markov Random Field (MRF)

$$E = \sum_{p} E_d(d_p) + \sum_{\{p,q\} \in \mathbb{N}} E_s(d_p, d_q)$$

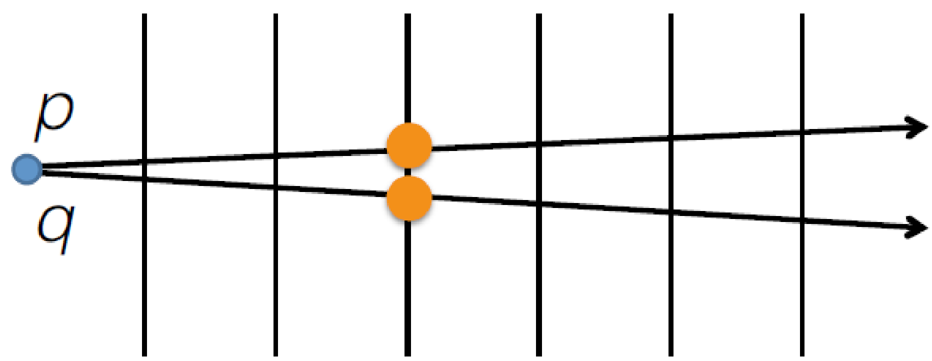


Markov Random Field (MRF)

$$E = \sum_{p} E_d(d_p) + \sum_{\{p,q\}\in\mathbb{N}} E_s(d_p, d_q)$$

$$E_s(d_p, d_q) = |d_p - d_q|$$

0

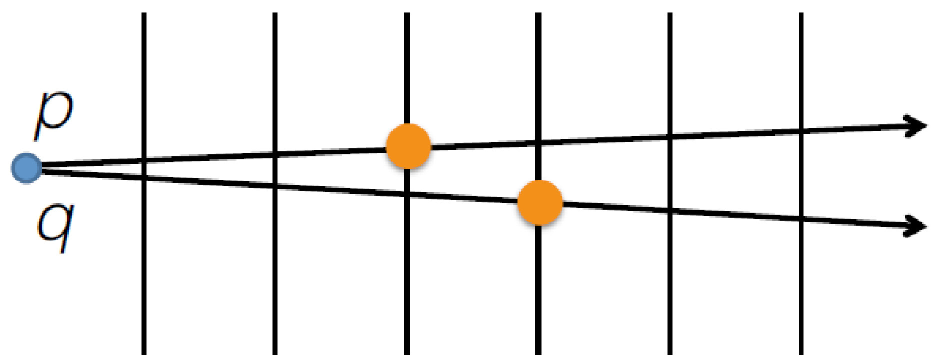


Markov Random Field (MRF)

$$E = \sum_{p} E_d(d_p) + \sum_{\{p,q\}\in\mathbb{N}} E_s(d_p, d_q)$$

$$E_s(d_p, d_q) = |d_p - d_q|$$

1

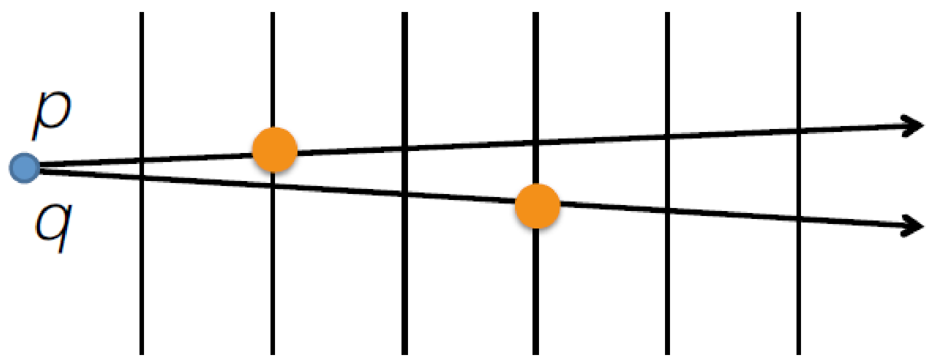


Markov Random Field (MRF)

$$E = \sum_{p} E_d(d_p) + \sum_{\{p,q\} \in \mathbb{N}} E_s(d_p, d_q)$$

$$E_s(d_p, d_q) = |d_p - d_q|$$

2

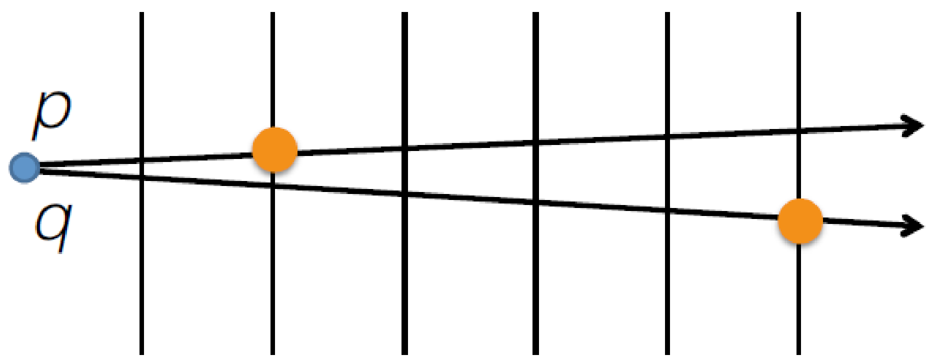


Markov Random Field (MRF)

$$E = \sum_{p} E_d(d_p) + \sum_{\{p,q\} \in \mathbb{N}} E_s(d_p, d_q)$$

$$E_s(d_p, d_q) = |d_p - d_q|$$

4

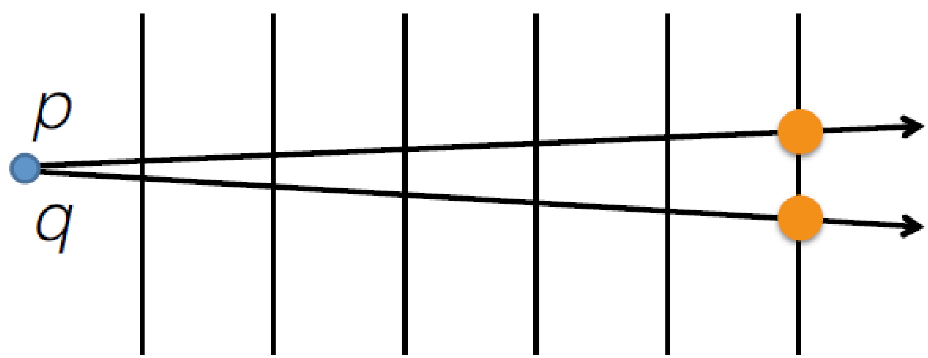


Markov Random Field (MRF)

$$E = \sum_{p} E_d(d_p) + \sum_{\{p,q\} \in \mathbb{N}} E_s(d_p, d_q)$$

$$E_s(d_p, d_q) = |d_p - d_q|$$

0

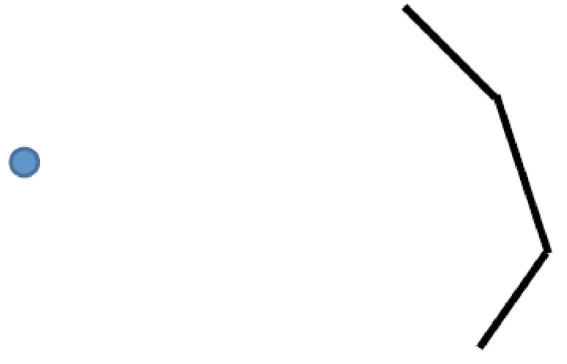


Markov Random Field (MRF)

$$E = \sum_{p} E_d(d_p) + \sum_{\{p,q\}\in\mathbb{N}} E_s(d_p, d_q)$$

Graph-cuts (alpha-expansion)
gives you very good solutions

We want **piecewise planar**



How to enforce piecewise planar

1. Advanced MRF and optimization

    - Global Stereo Reconstruction under Second Order Smoothness Priors [Woodford et al., CVPR 2008] Best Paper Award

2. Integrate with top-down (primitive) approach

    - Manhattan World Stereo [Furukawa et al., CVPR 2009]

    - Piecewise Planar Stereo for Image-based rendering [Sinha et al., ICCV 2009]

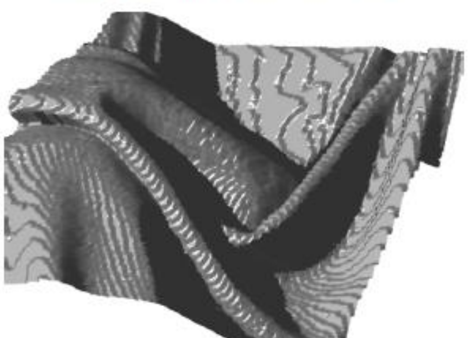
    - Fusion of Feature- and Area-Based Information for Urban Buildings Modeling from Aerial Imagery [Zebedin et al., ECCV 2008]

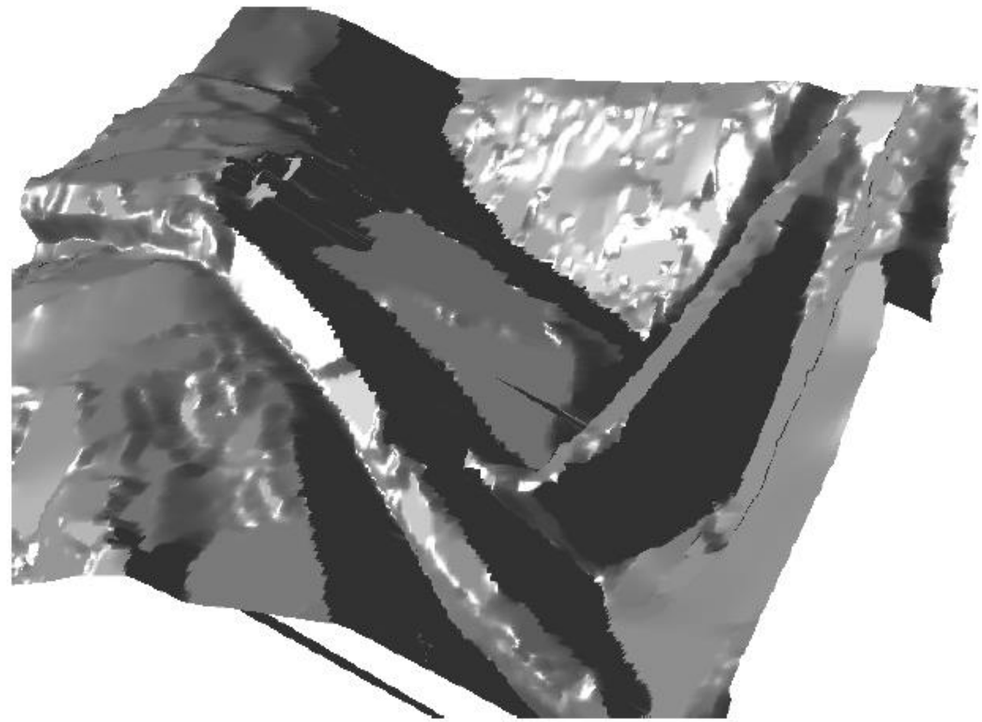
    - Piecewise Planar and Non-Planar Stereo for urban Scene Reconstruction [Gallup et al., CVPR 2010]

Advanced MRF for Depthmap Estimation

Reference image

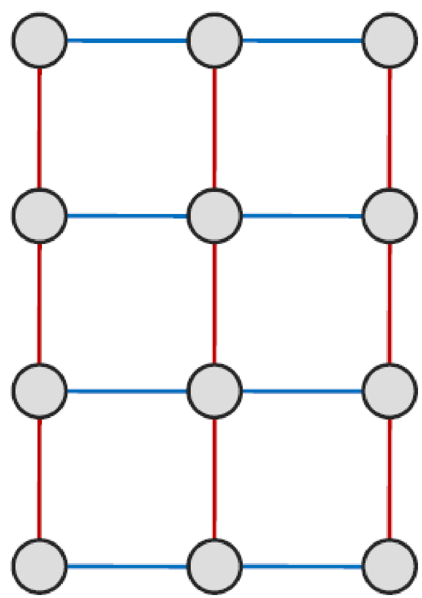
Ground truth





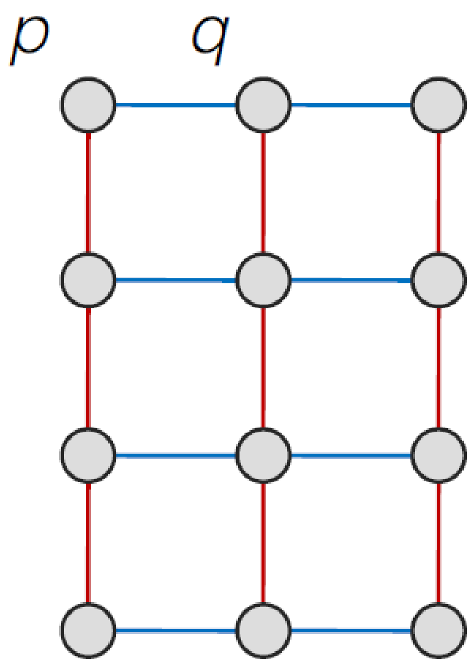
Global Stereo Reconstruction under Second Order Smoothness Priors
[Woodford et al., CVPR 2008] Best Paper Award

Standard MRF



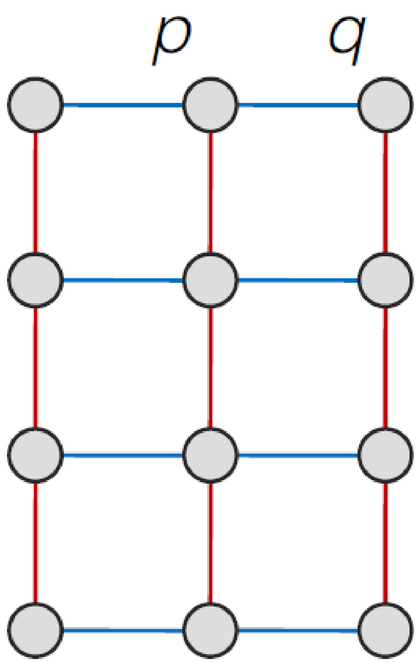
Standard MRF

$$E_s(d_p, d_q) = |d_p - d_q|$$



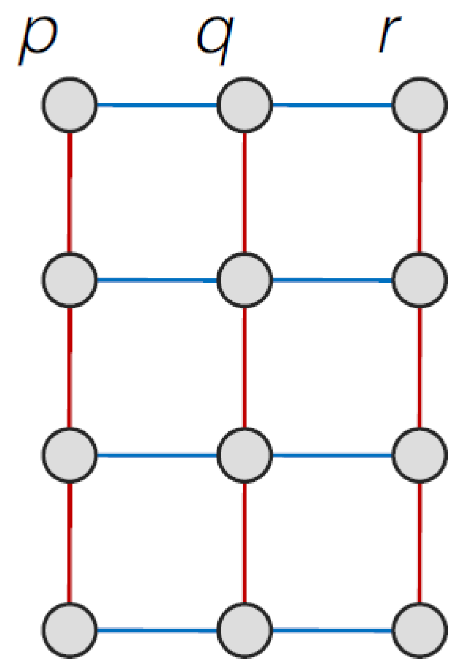
Standard MRF

$$E_s(d_p, d_q) = |d_p - d_q|$$



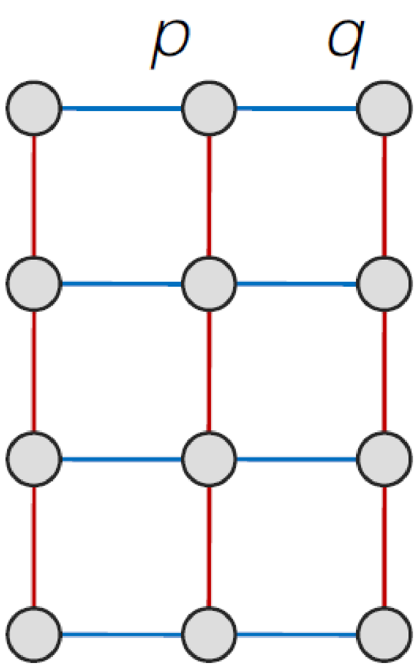
MRF with a triple clique

$$E_s(d_p, d_q, d_r)$$



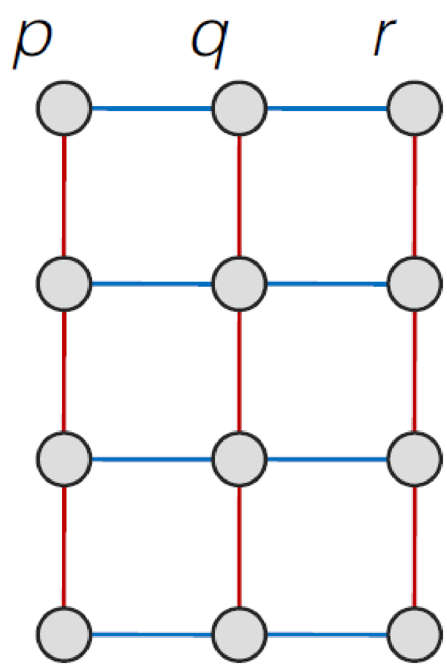
Standard MRF

$$E_s(d_p, d_q) = |d_p - d_q|$$



MRF with a triple clique

$$E_s(d_p, d_q, d_r) = |d_p + d_r - 2d_q|$$



MRF with a triple clique

$$E = \sum_{p} E_d(d_p) + \sum_{\{p,q\}\in\mathbb{N}} E_s(d_p, d_q)$$
$$+ \sum_{\{p,q,r\}\in\mathbb{N}_3} E_s(d_p, d_q, d_r)$$

Optimization becomes a challenge

Standard MRF
(submodular)

- Graph-cuts (alpha-expansion)
- Belief propagation (TRW)

MRF with a triple clique
(non-submodular)

- QPBO
- Belief propagation (TRW?)

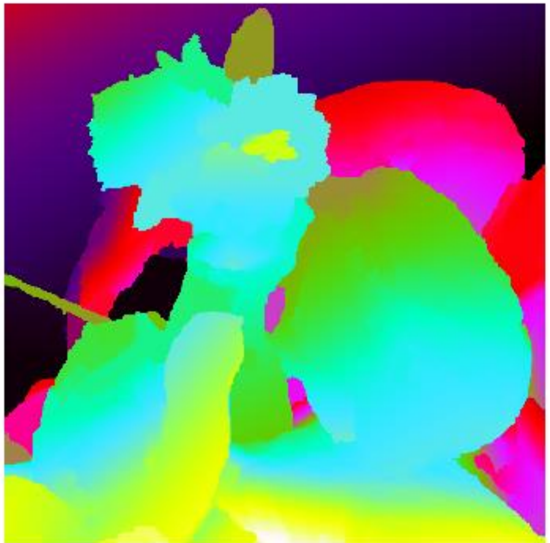
Fusion moves…

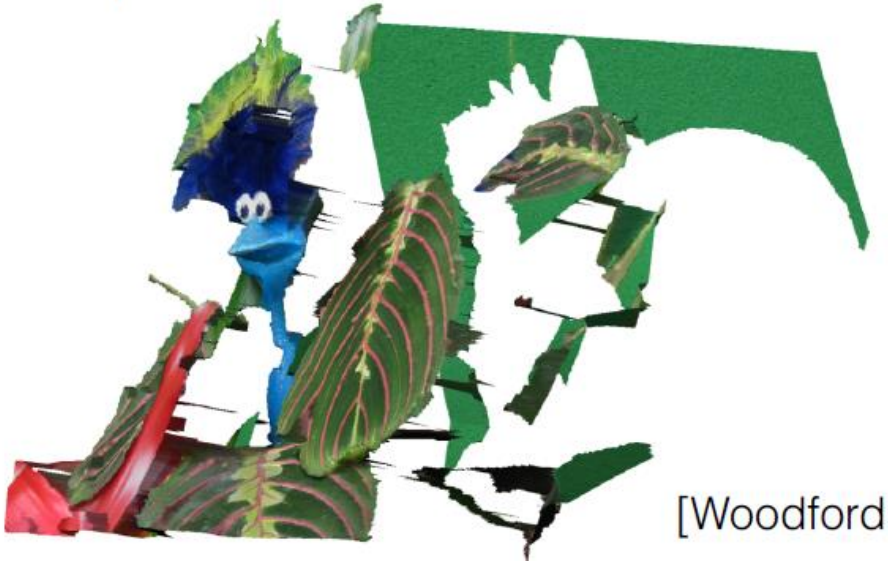
Experimental results

Reference image

Neighboring image

Output depthmap



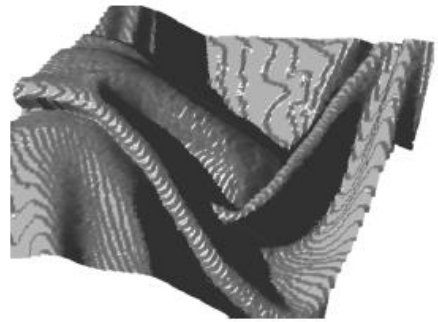


[Woodford et al.]

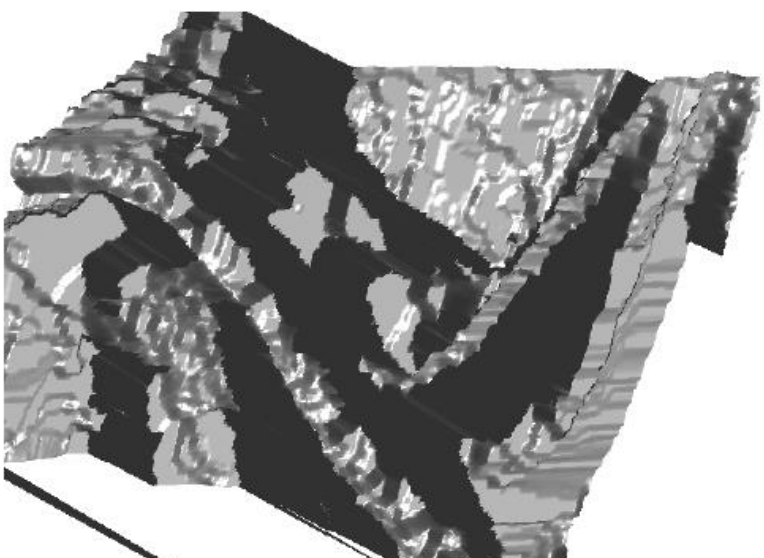
Comparative experiment

Reference image

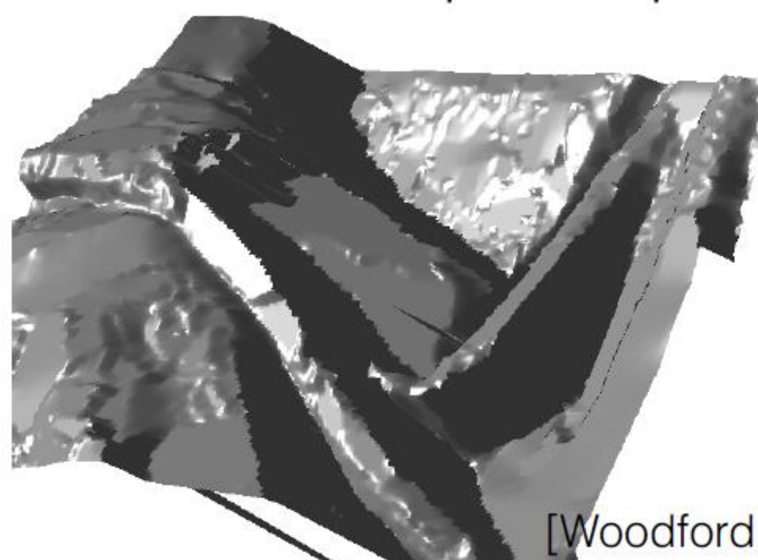
Ground truth



Standard MRF



MRF with a triple clique



[Woodford et al.]

How to enforce piecewise planar

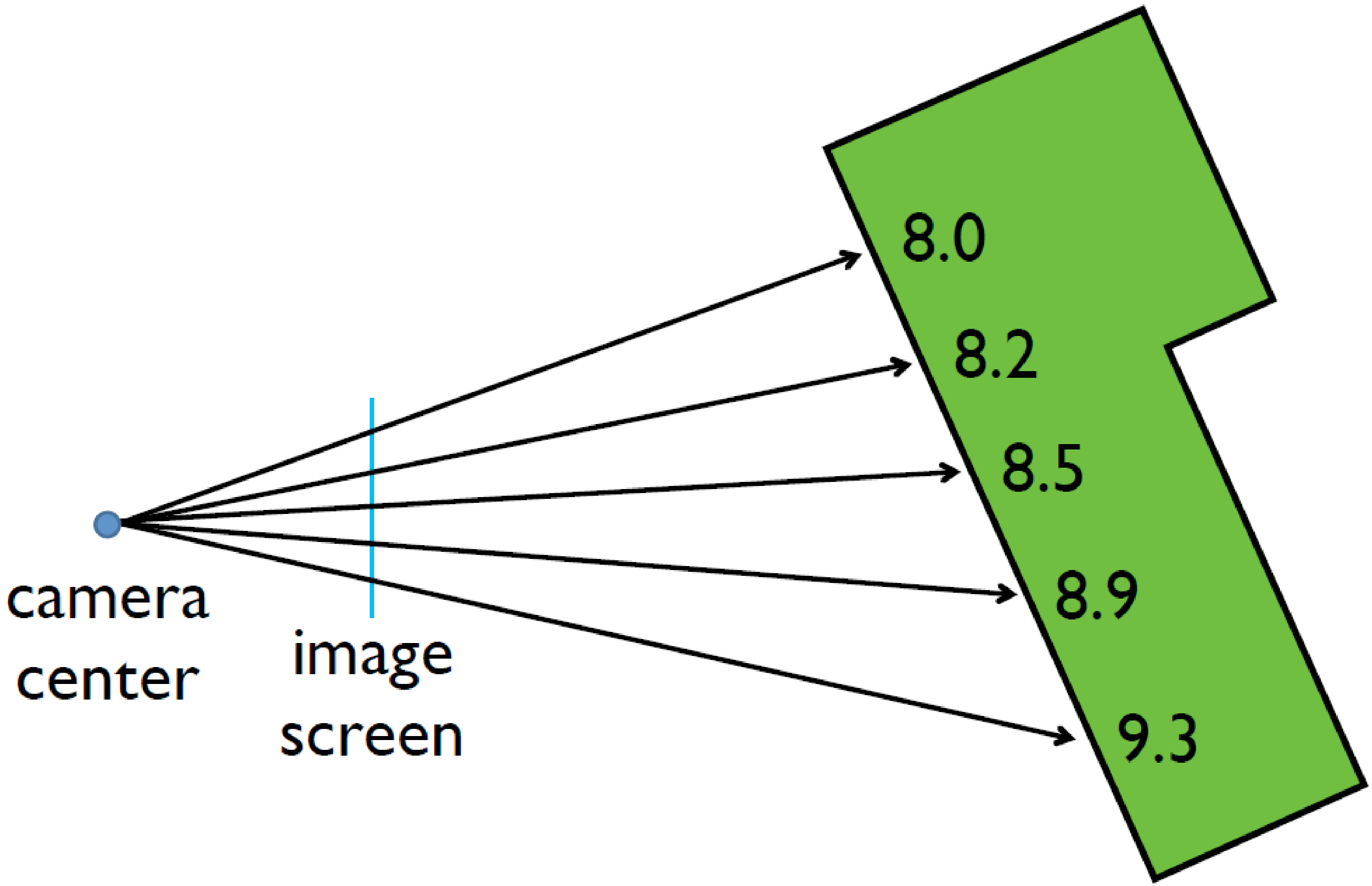
1. Advanced MRF and optimization

   - Global Stereo Reconstruction under Second Order Smoothness Priors [Woodford et al., CVPR 2008] Best Paper Award

2. Integrate with top-down (primitive) approach

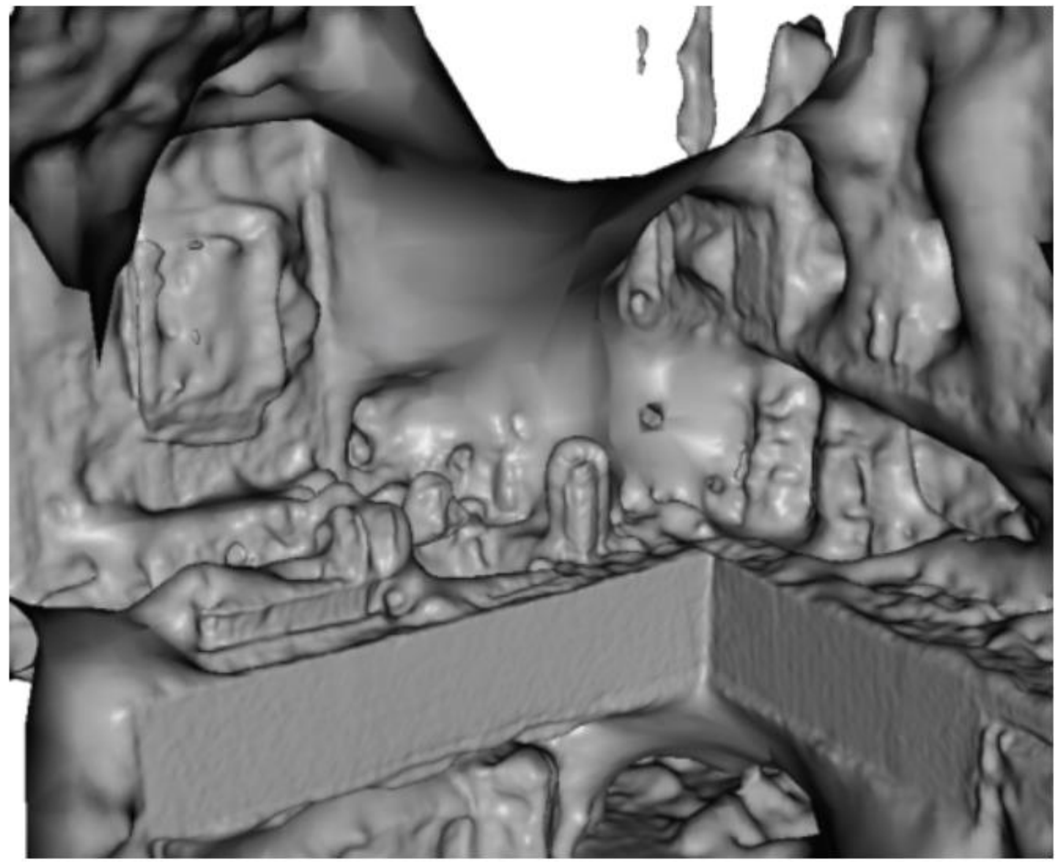
   - Manhattan World Stereo [Furukawa et al., CVPR 2009]
   - Piecewise Planar Stereo for Image-based rendering [Sinha et al., ICCV 2009]
   - Fusion of Feature- and Area-Based Information for Urban Buildings Modeling from Aerial Imagery [Zebedin et al., ECCV 2008]
   - Piecewise Planar and Non-Planar Stereo for urban Scene Reconstruction [Gallup et al., CVPR 2010]

Standard Depthmap



Planemap

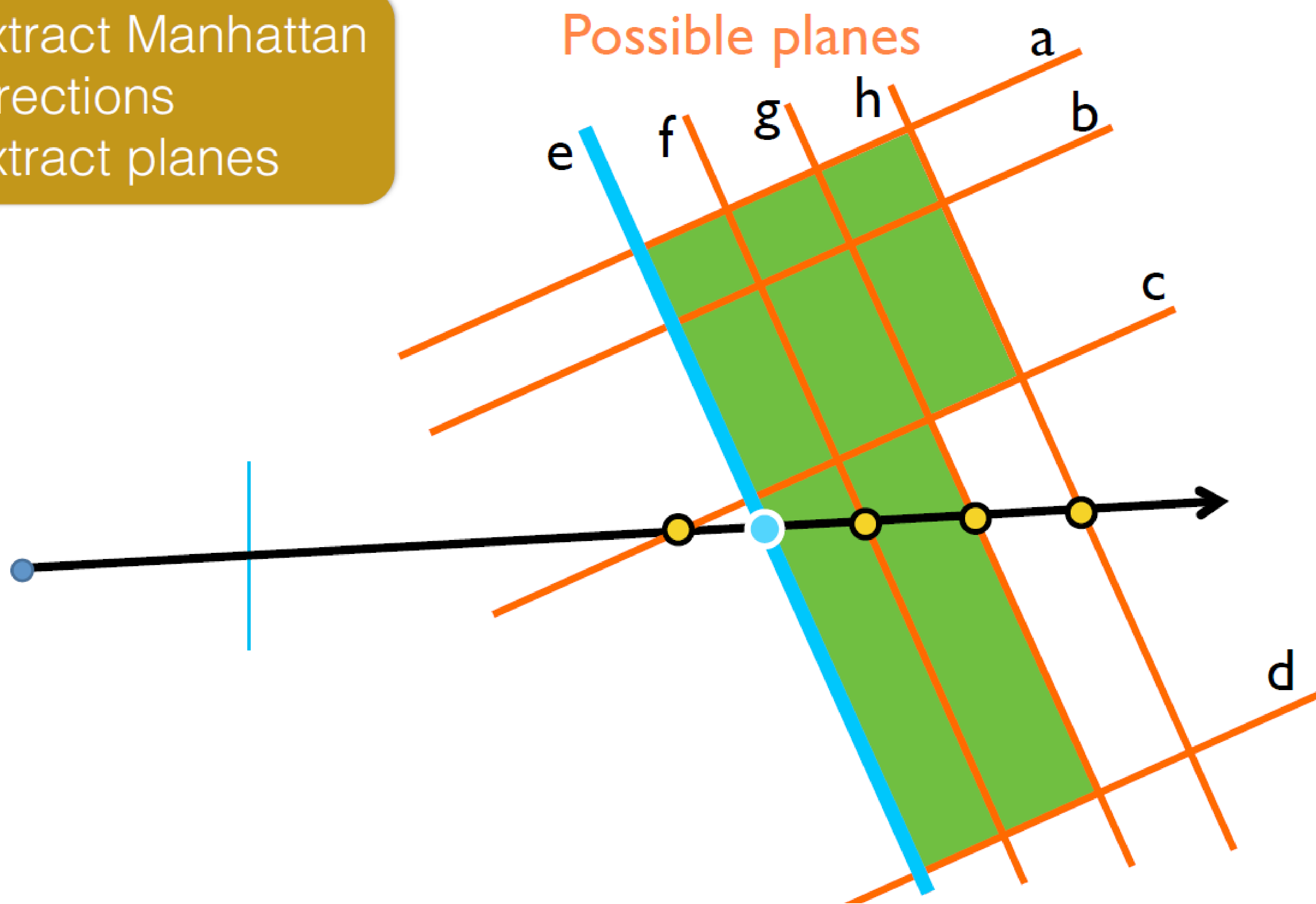
1. Extract Manhattan directions
2. Extract planes



[Furukawa and Ponce, 2007]

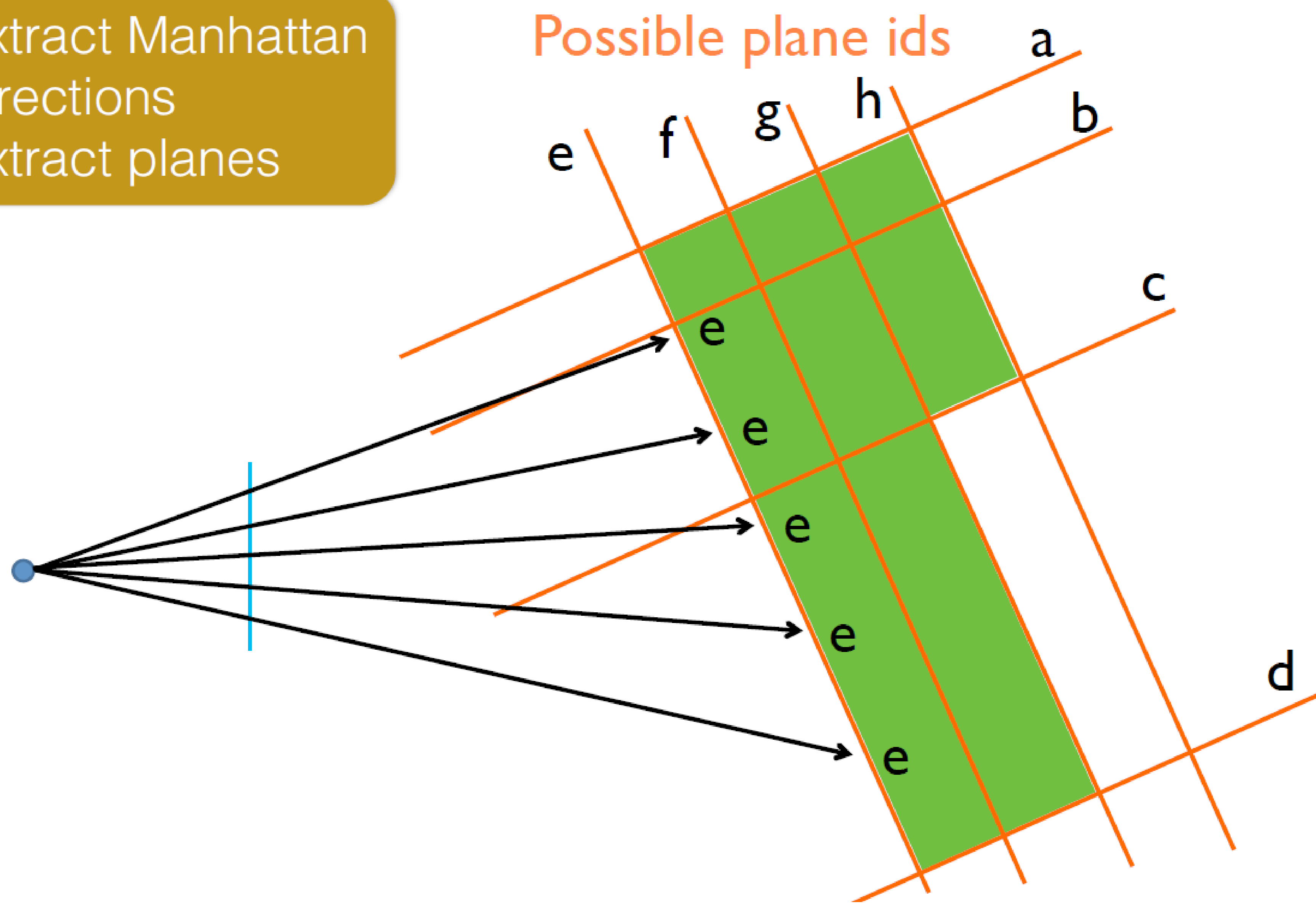
Planemap

1. Extract Manhattan directions
2. Extract planes



Planemap

1. Extract Manhattan directions
2. Extract planes



Depthmap to Planemap

Depthmap (d_p)

$$E = \sum_p E_d(d_p) + \sum_{\{p,q\}\in\mathbb{N}} E_s(d_p, d_q)$$

Planemap (h_p)

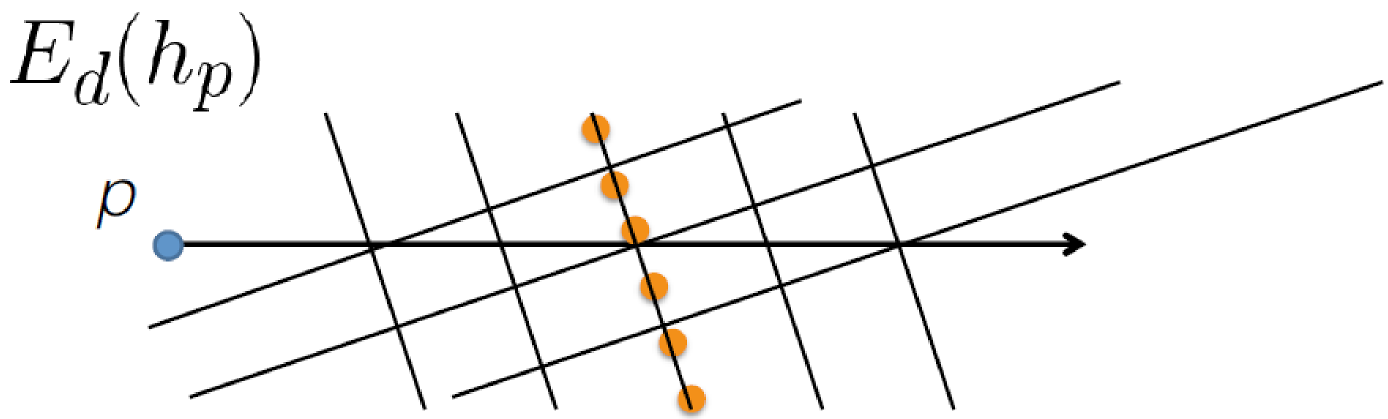
$$E = \sum_p E_d(h_p) + \sum_{\{p,q\}\in\mathbb{N}} E_s(h_p, h_q)$$

Planemap MRF

$$E = \sum_{p} E_d(h_p) + \sum_{\{p,q\} \in \mathbb{N}} E_s(h_p, h_q)$$

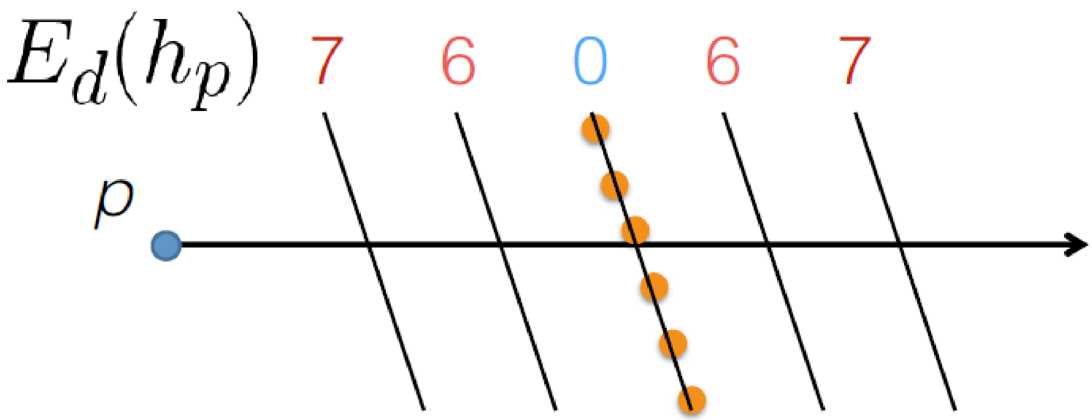
Planemap MRF

$$E = \sum_{p} E_d(h_p) + \sum_{\{p,q\} \in \mathbb{N}} E_s(h_p, h_q)$$



Planemap MRF

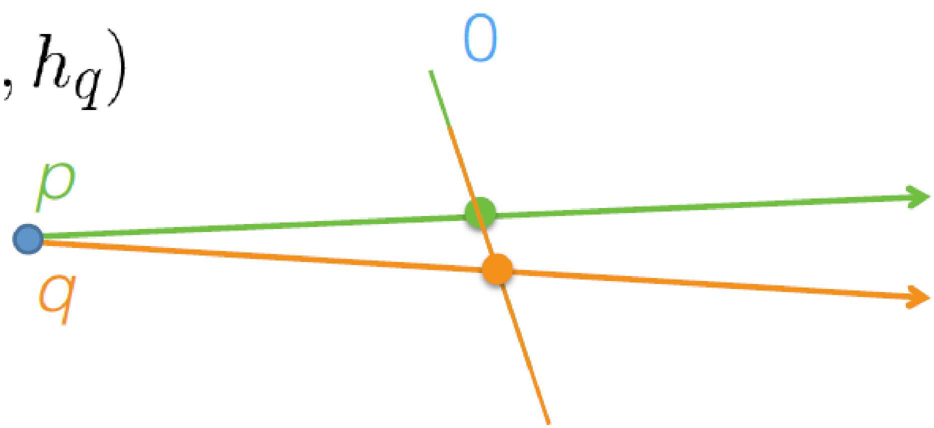
$$E = \sum_{p} E_d(h_p) + \sum_{\{p,q\} \in \mathbb{N}} E_s(h_p, h_q)$$



Planemap MRF

$$E = \sum_{p} E_d(h_p) + \sum_{\{p,q\} \in \mathbb{N}} E_s(h_p, h_q)$$

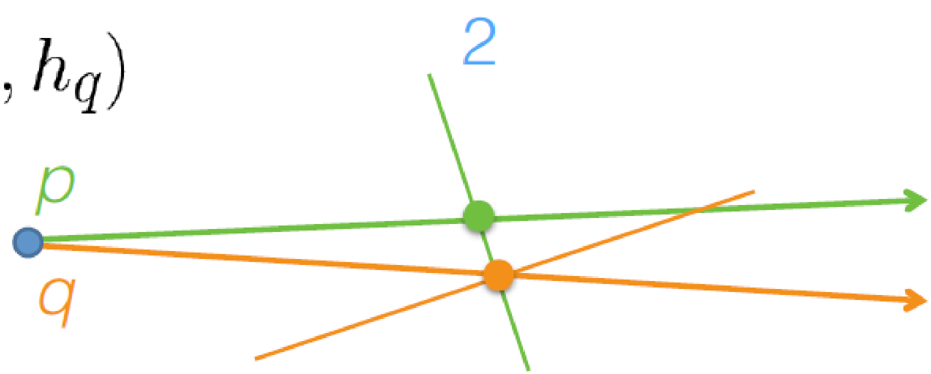
$E_s(h_p, h_q)$



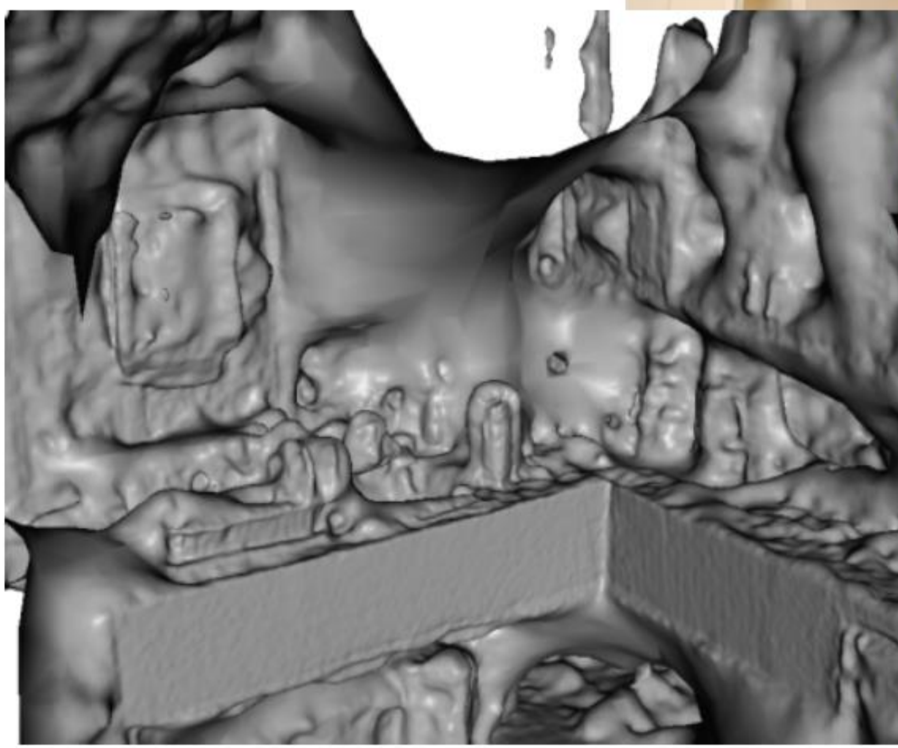
Planemap MRF

$$E = \sum_{p} E_d(h_p) + \sum_{\{p,q\}\in\mathbb{N}} E_s(h_p, h_q)$$

$E_s(h_p, h_q)$



Comparison

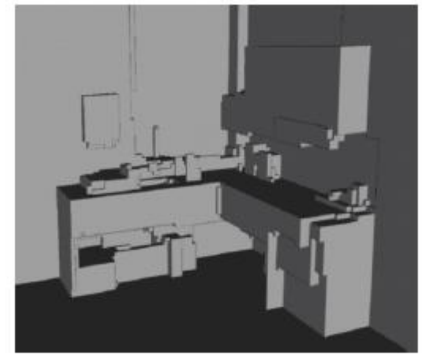


Standard method

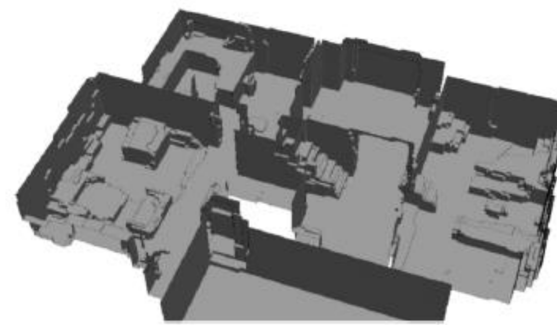
Manhattan Planemap

Reconstruction Results

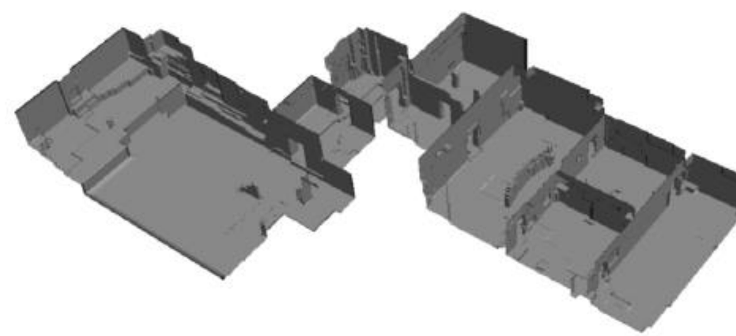
Kitchen - 22 images



house - 148 images



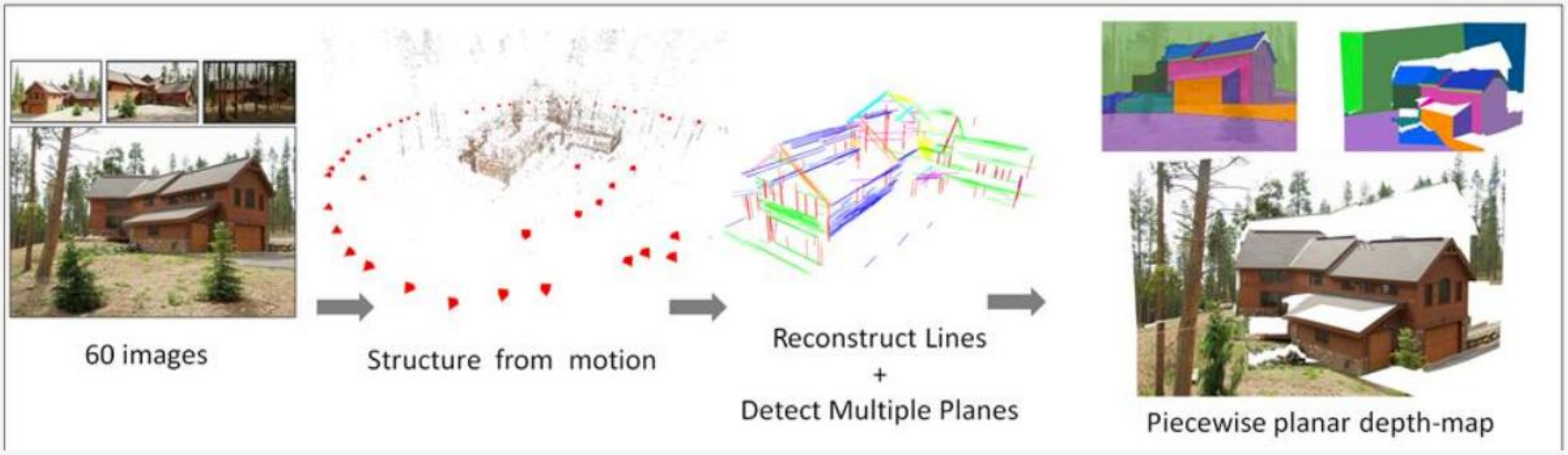
gallery - 492 images



How to enforce piecewise planar

1. Advanced MRF and optimization
    - Global Stereo Reconstruction under Second Order Smoothness Priors [Woodford et al., CVPR 2008] Best Paper Award
2. Integrate with top-down (primitive) approach
    - Manhattan World Stereo [Furukawa et al., CVPR 2009]
    - Piecewise Planar Stereo for Image-based rendering [Sinha et al., ICCV 2009]
    - Fusion of Feature- and Area-Based Information for Urban Buildings Modeling from Aerial Imagery [Zebedin et al., ECCV 2008]
    - Piecewise Planar and Non-Planar Stereo for urban Scene Reconstruction [Gallup et al., CVPR 2010]

Relaxing Mahnattan

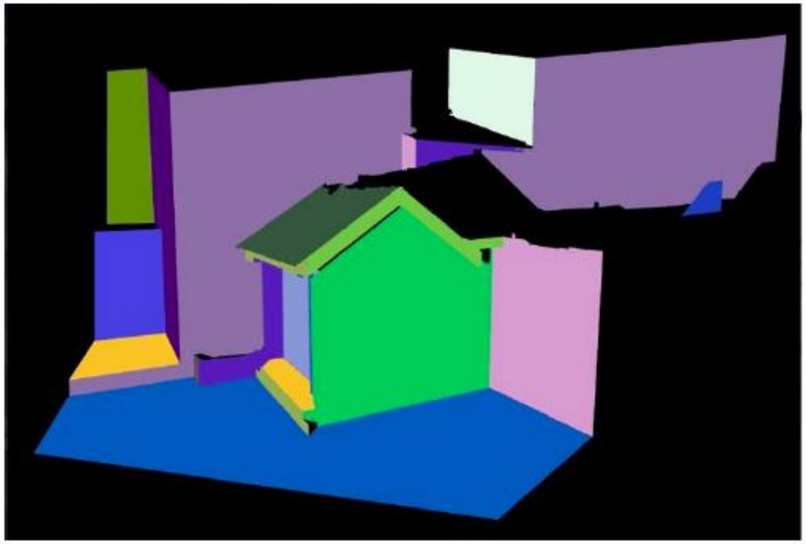


- Use sparse lines + sparse points to detect planes
- MRF + Graph-cuts

Piecewise Planar Stereo for Image-based rendering [Sinha et al., ICCV 2009]

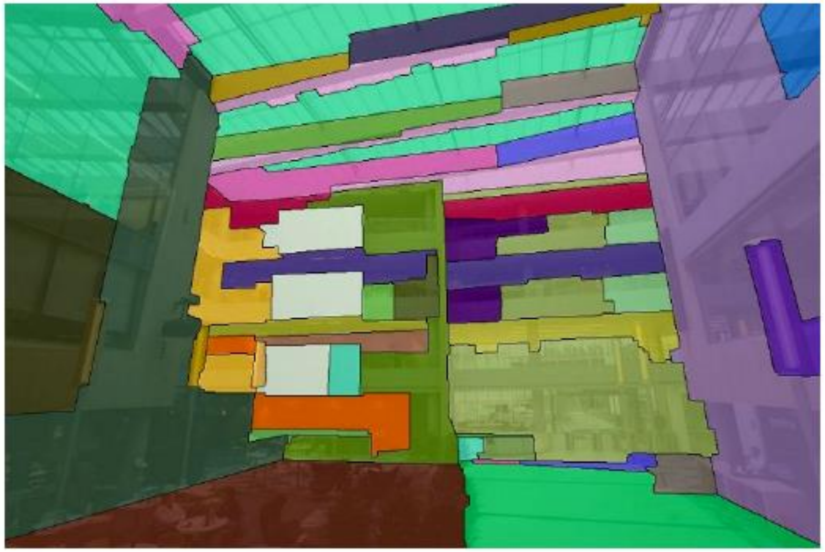
Relaxing Manhattan

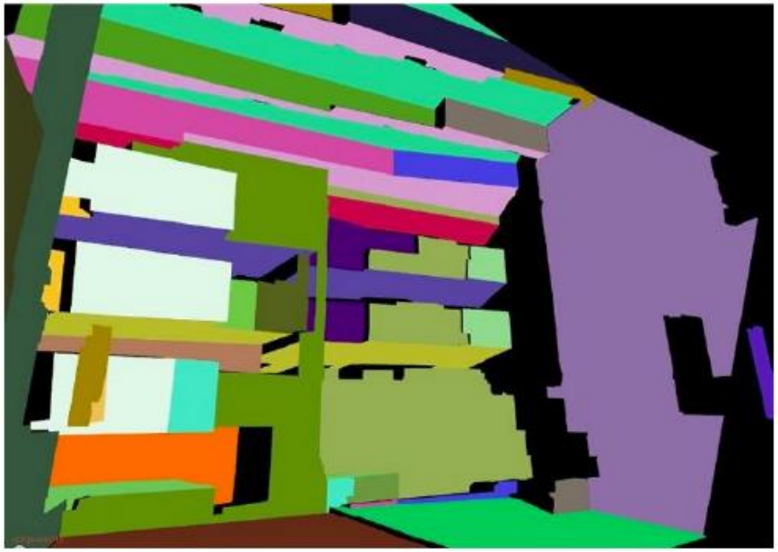
[Sinha et al.]



Relaxing Manhattan

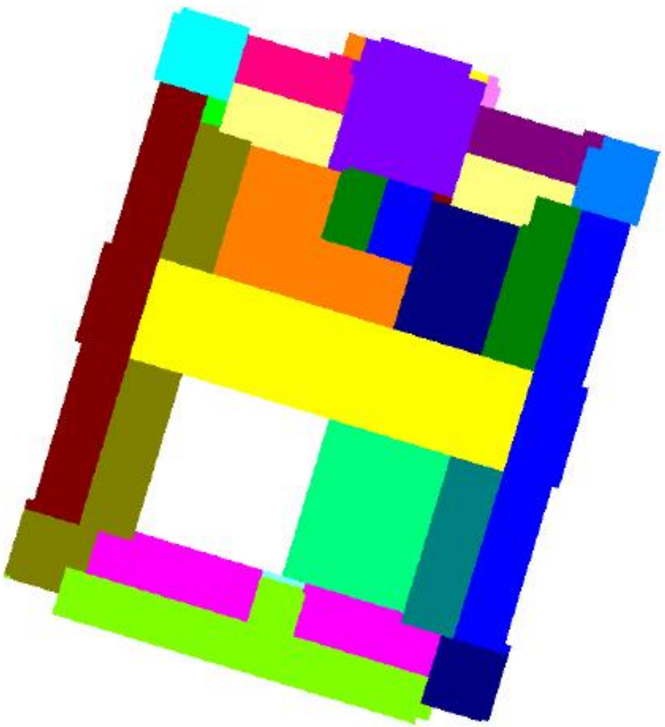
[Sinha et al.]





Enable Curved Surfaces

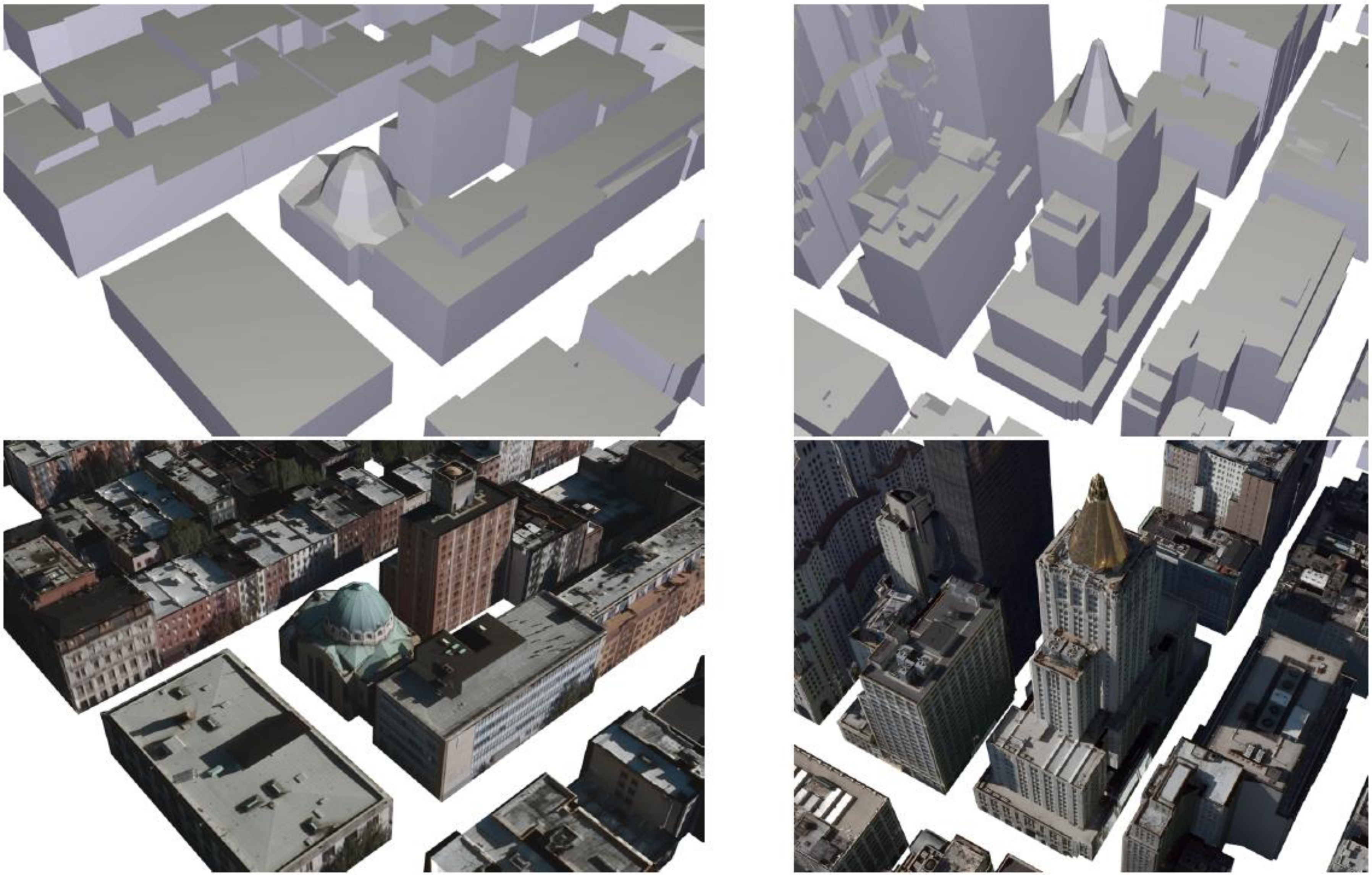
- Building reconstruction from a top down view



Fusion of Feature- and Area-Based Information for Urban Buildings Modeling from Aerial Imagery [Zebedin et al., ECCV 2008]

Enable Curved Surfaces

[Zebedin et al.]



How to enforce piecewise planar

1. Advanced MRF and optimization
   - Global Stereo Reconstruction under Second Order Smoothness Priors [Woodford et al., CVPR 2008] Best Paper Award

2. Integrate with top-down (primitive) approach
   - Manhattan World Stereo [Furukawa et al., CVPR 2009]
   - Piecewise Planar Stereo for Image-based rendering [Sinha et al., ICCV 2009]
   - Fusion of Feature- and Area-Based Information for Urban Buildings Modeling from Aerial Imagery [Zebedin et al., ECCV 2008]
   - Piecewise Planar and Non-Planar Stereo for urban Scene Reconstruction [Gallup et al., CVPR 2010]

3D Reconstruction from Video

ETH

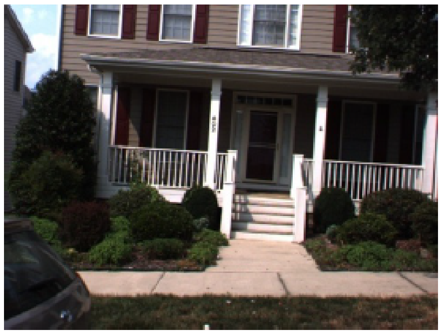
Idea

Video Frames

Enforce planarity where it looks like a building

First step: Planar and Non-Planar Stereo

Video



Real-Time Stereo

Labels = $\{\pi_1, \cdots, \pi_N, \pi_\infty, \textit{non-plane}, \textit{discard}\}$

planes non-plane discard

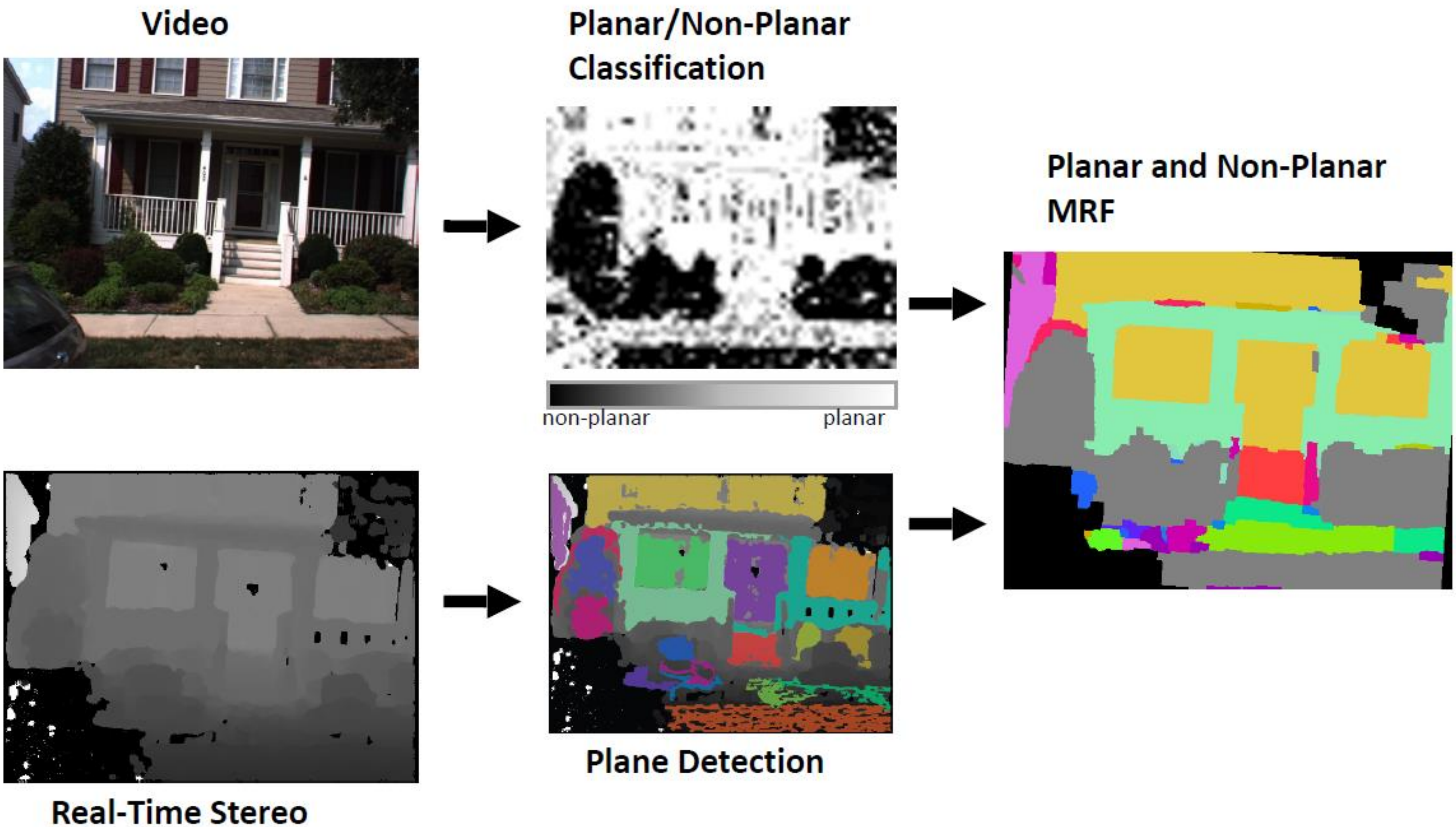
Plane Detection

Planemap (w/ non-planar))

Second step: Appearance Prior

Video Frames

Algorithm


Video
Planar/Non-Planar Classification
non-planar
planar
Planar and Non-Planar MRF
Plane Detection
Real-Time Stereo

Large-Scale Multi-View Stereo

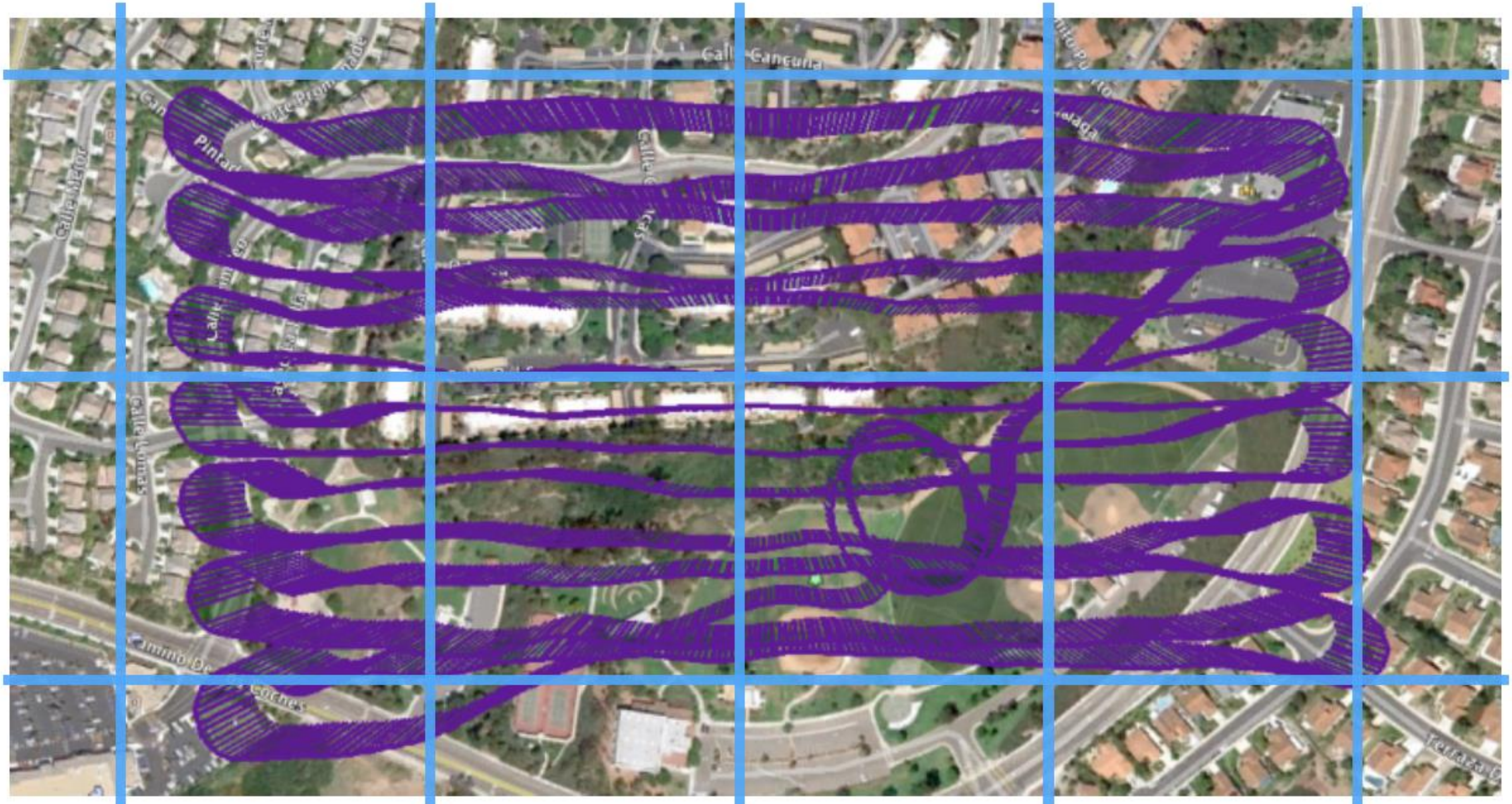
Input images

[diydrones.com]

Input images

[diydrones.com]

Divide and Reconstruct

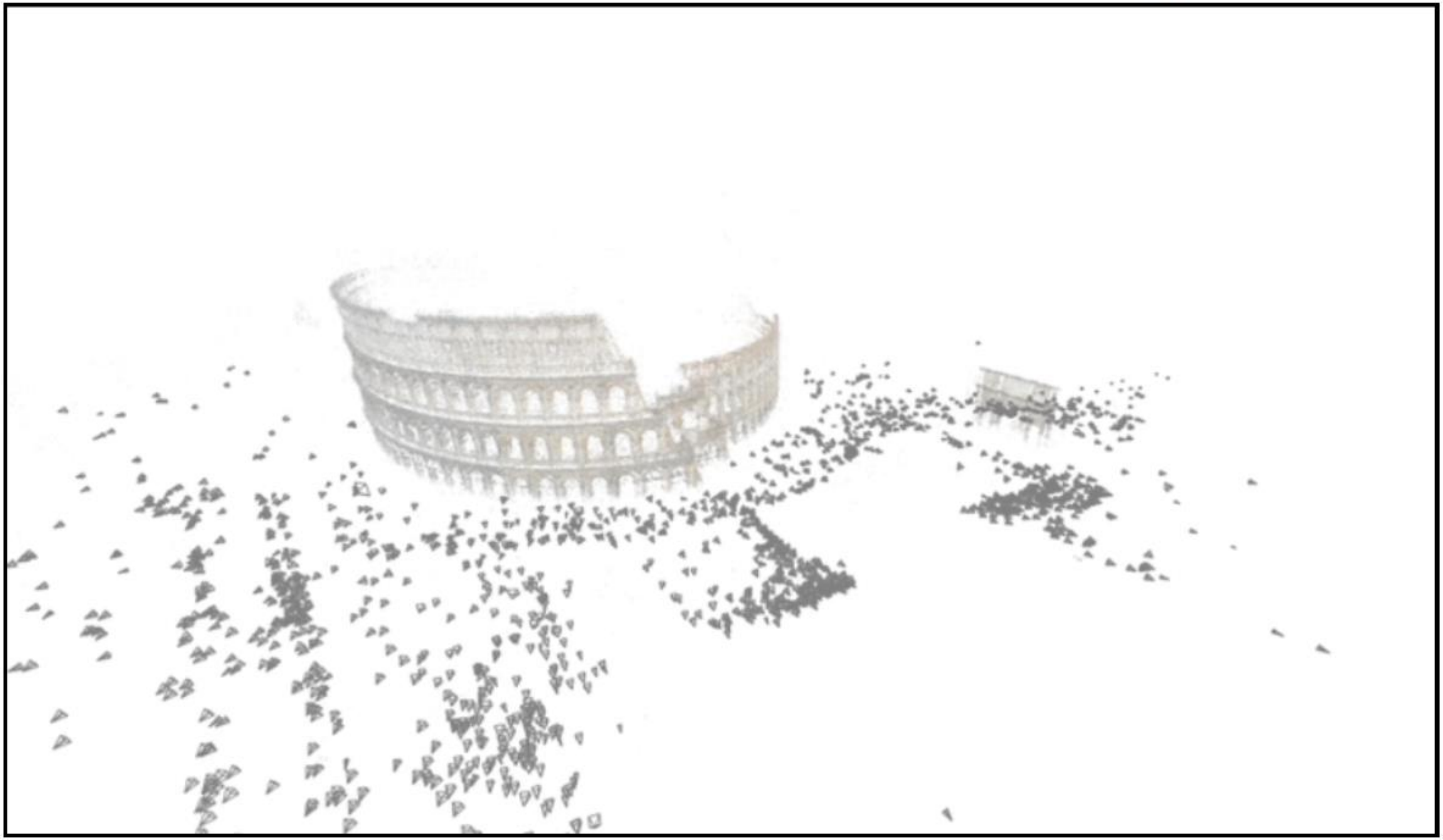


[diydrones.com]

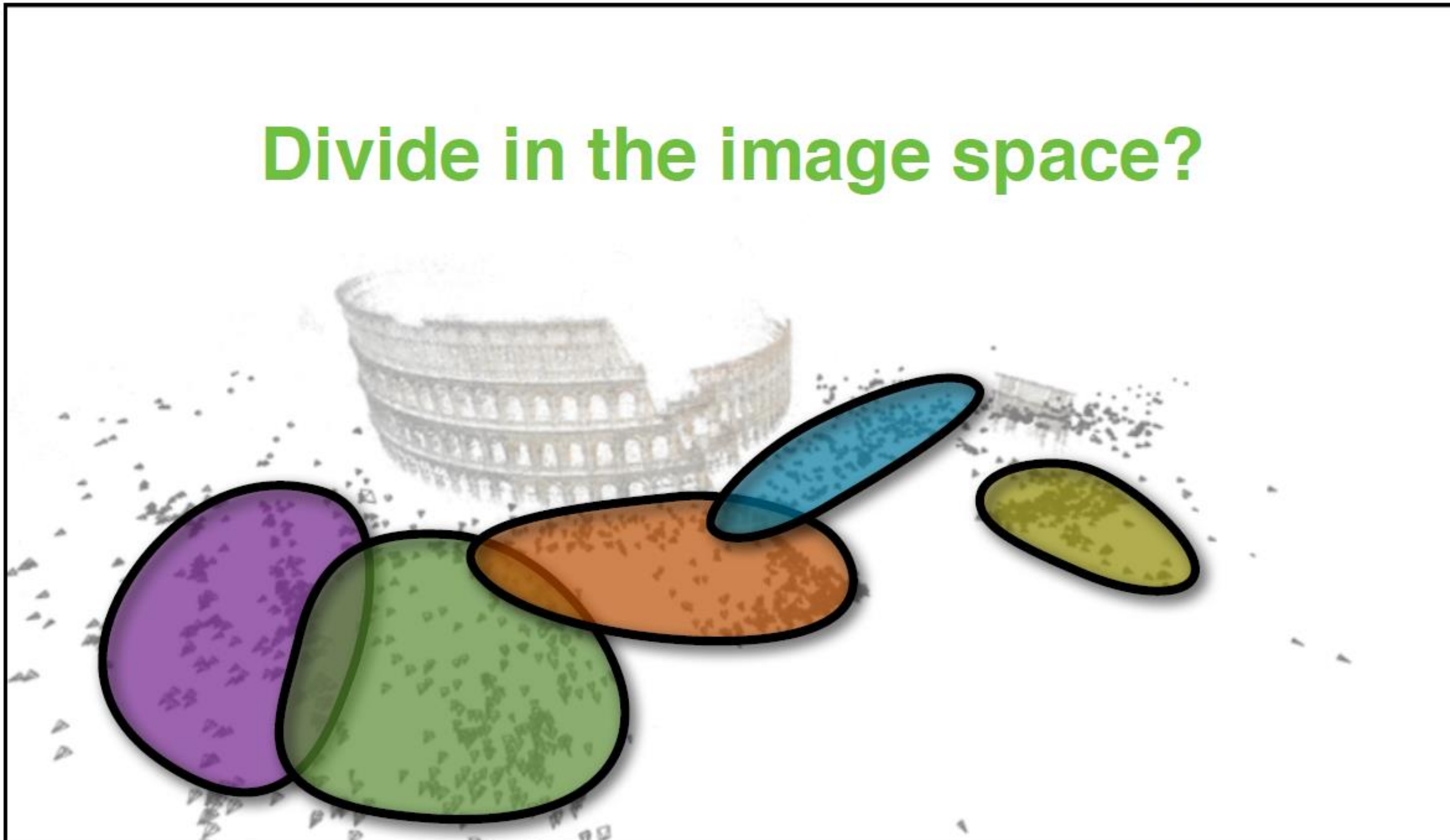
Large-Scale MVS for Unorganized Photos (Internet Photos)

"*Building Rome in a Day*" [Agarwal, Snavely, et al., 2009]

Divide is a challenge



Divide is a challenge



Divide is a challenge

Divide in the model space?

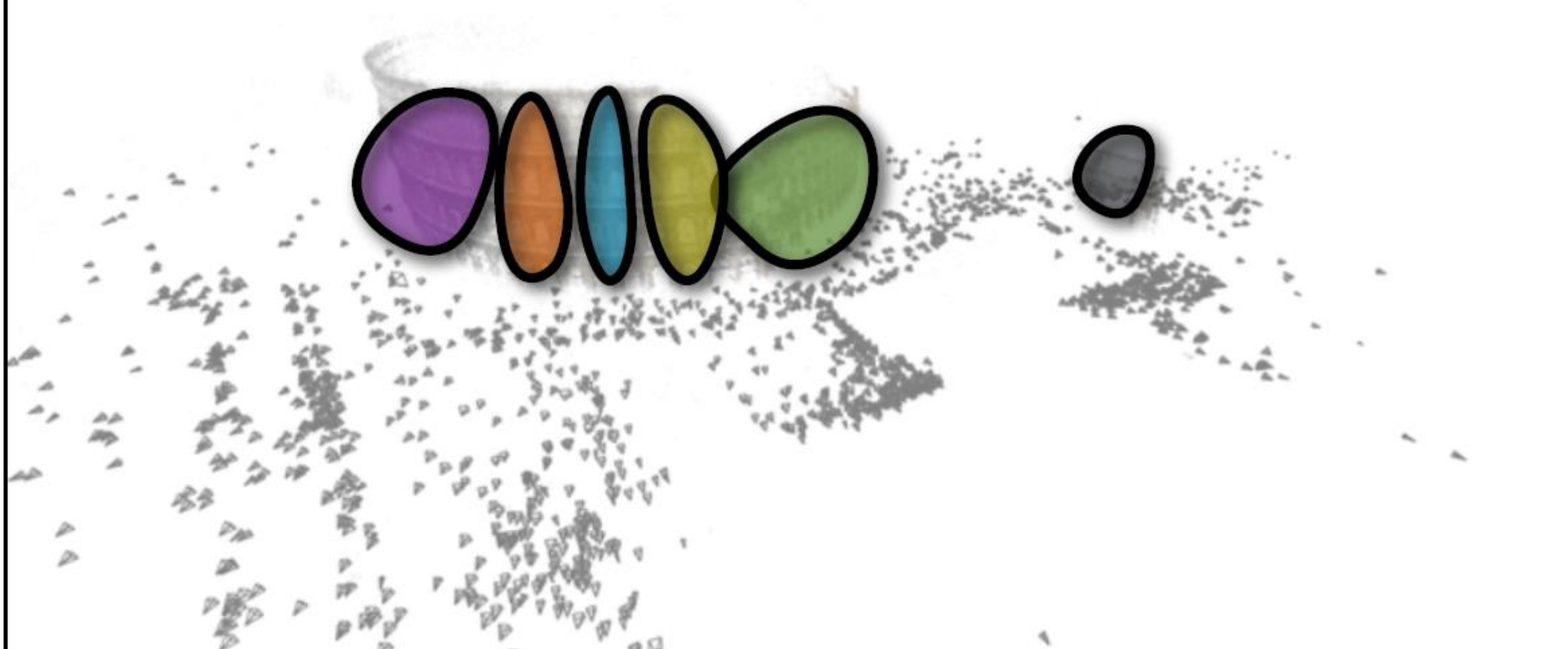
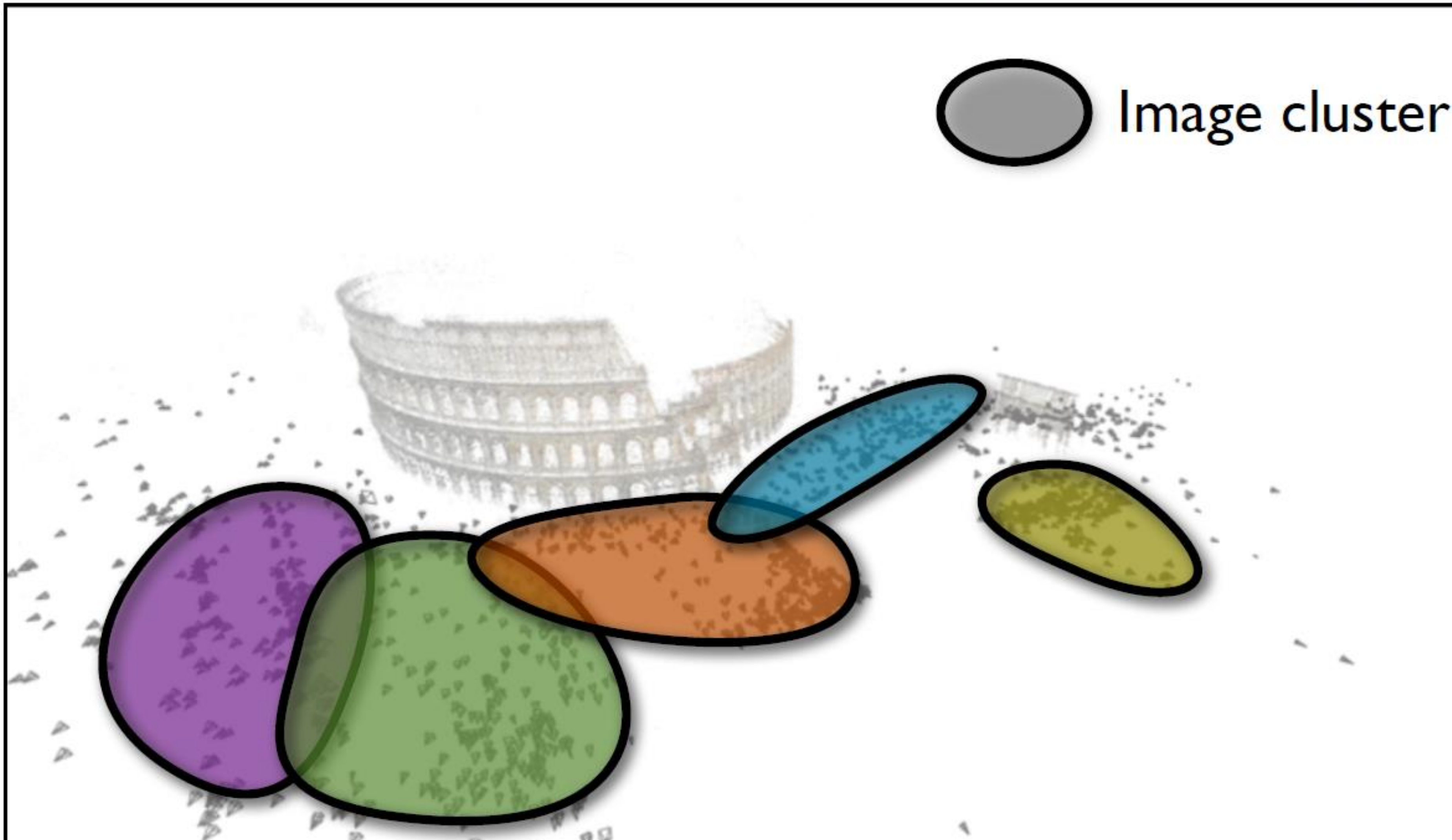
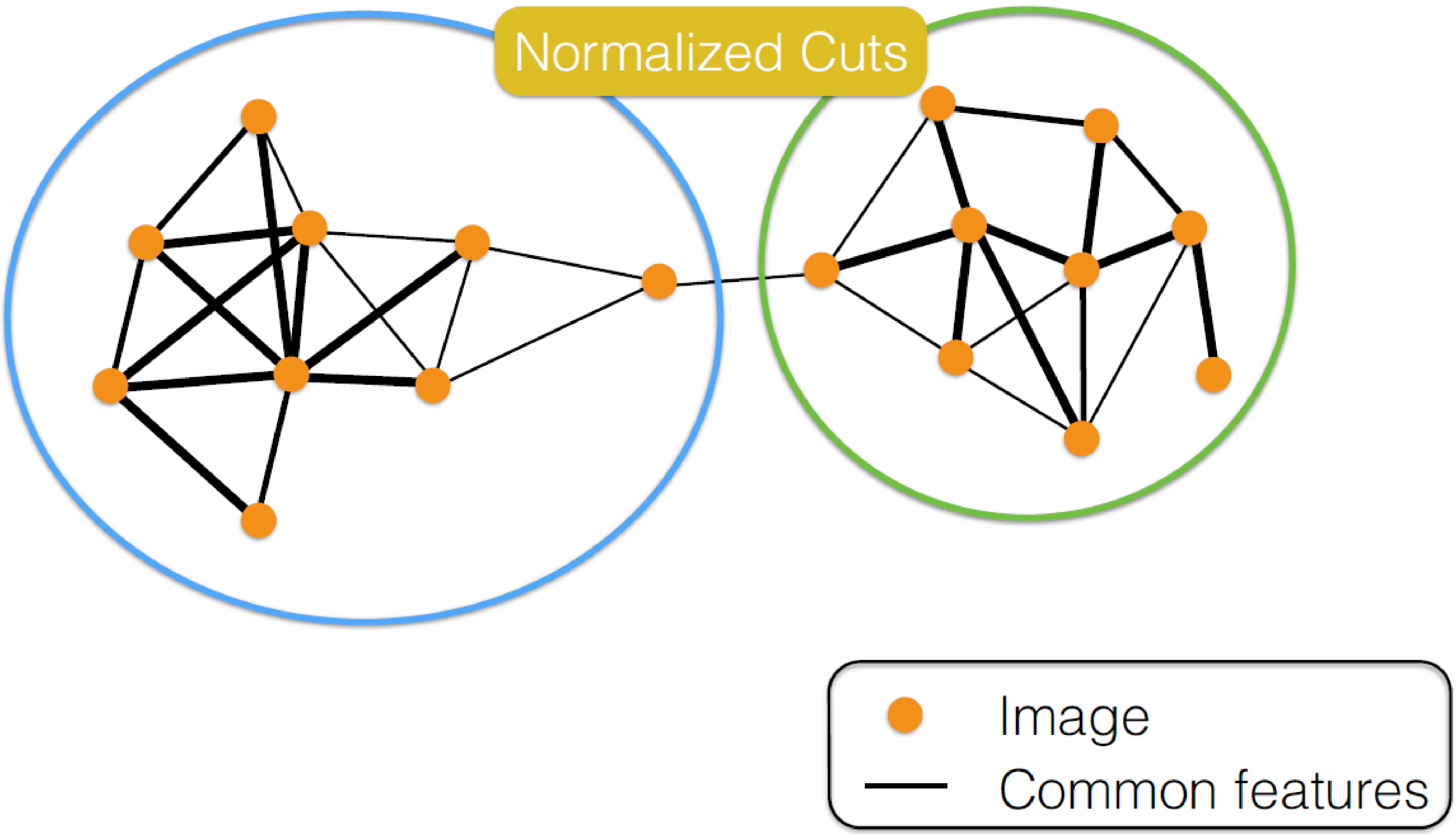


Image-clustering based on a model

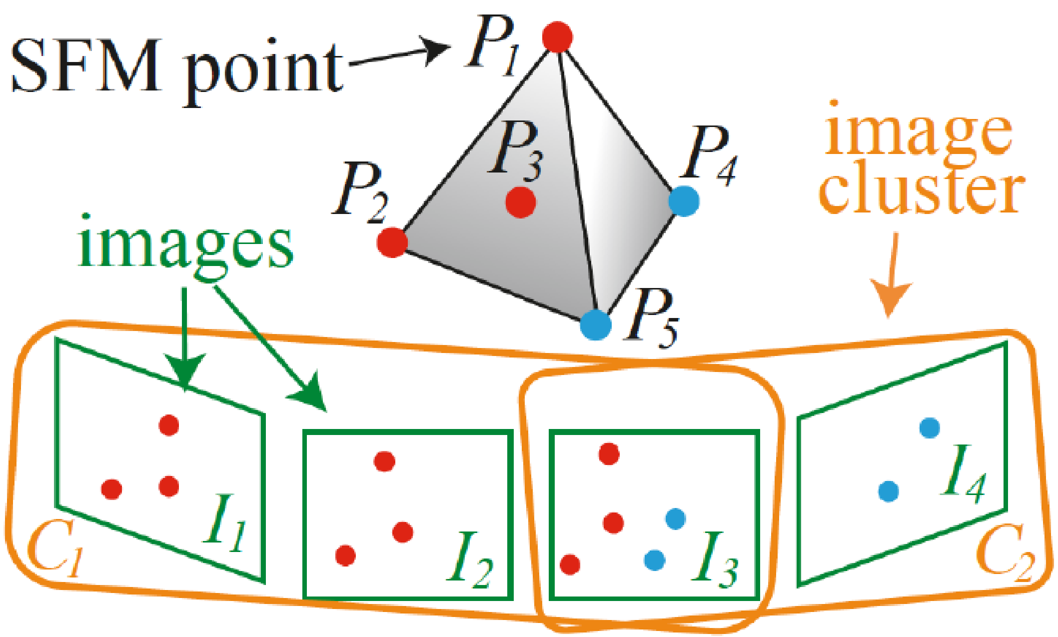
"Towards Internet-scale Multi-View Stereo" [furukawa et al., 2010]



One way to look at the clustering problem



Formulation



SFM points
$\{P_1, P_2, \ldots\}$

Images
$\{I_1, I_2, \ldots\}$

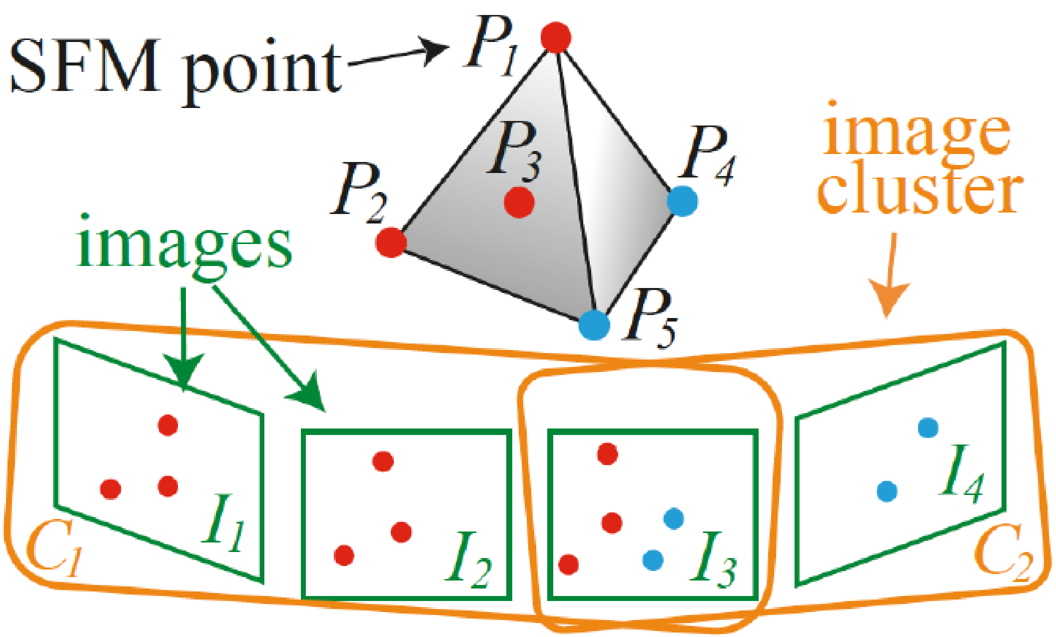
Image clusters
$\{C_1, C_2, \ldots\}$

P1 is reconstructed well in *C1*
based on (image resolutions/distribution)

→ "*P1* is covered by *C1*."

"*P4* is covered by *C2*."

Formulation



SFM points
$\{P_1, P_2, \ldots\}$

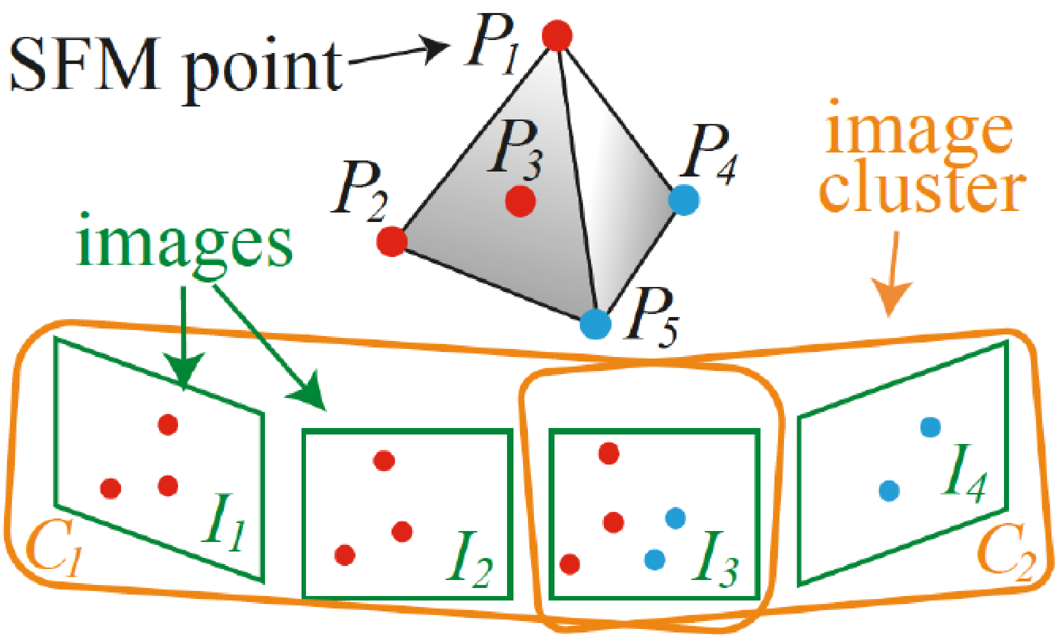
Images
$\{I_1, I_2, \ldots\}$

Image clusters
$\{C_1, C_2, \ldots\}$

We want to

1. remove redundant images
2. keep each cluster small (memory limit)
3. “cover” many points

Formulation



SFM points
$\{P_1, P_2, \ldots\}$

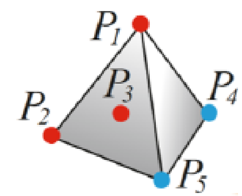
Images
$\{I_1, I_2, \ldots\}$

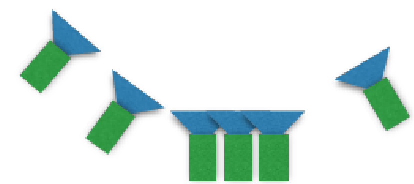
Image clusters
$\{C_1, C_2, \ldots\}$

Minimize $\sum_k |C_k|$ subject to (compactness)

- $\forall k \;\; |C_k| \leq \alpha,$ (size)
- $\forall i \;\; \frac{\{\text{\# of covered points in } I_i\}}{\{\text{\# of points in } I_i\}} \geq \delta.$ (coverage)

Image Clustering Algorithm



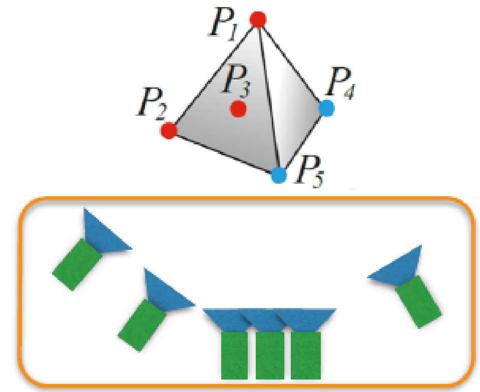


$$\text{Minimize} \sum_k |C_k| \text{ subject to} \qquad \text{(compactness)}$$

- $\forall k \ \ |C_k| \leq \alpha,$ (size)
- $\forall i \ \ \dfrac{\{\text{\# of covered points in } I_i\}}{\{\text{\# of points in } I_i\}} \geq \delta.$ (coverage)

Image Clustering Algorithm

1. Start with all the images in a cluster



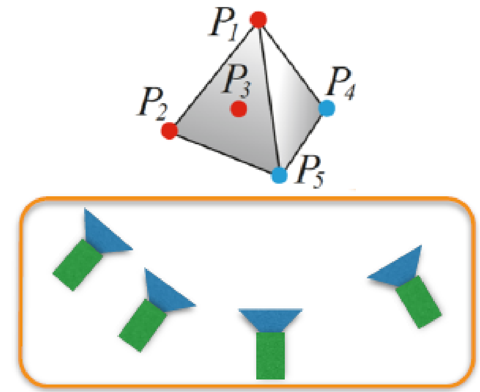
C1

$$\text{Minimize} \sum_k |C_k| \text{ subject to} \qquad \text{(compactness)}$$

- $\cancel{\forall k \ |C_k| \leq \alpha,}$ (size)
- $\forall i \ \dfrac{\{\text{\# of covered points in } I_i\}}{\{\text{\# of points in } I_i\}} \geq \delta.$ (coverage)

Image Clustering Algorithm

1. Start with all the images in a cluster
2. Optimize *compactness* while keeping *coverage*



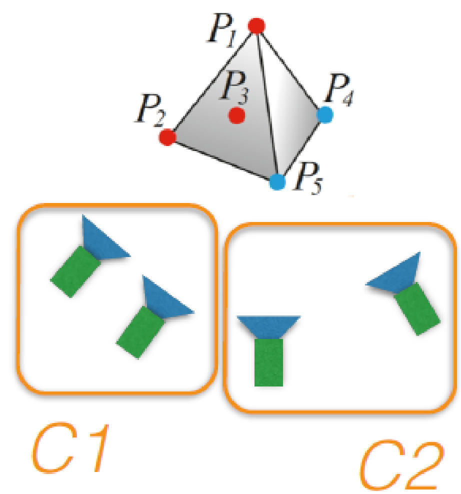
C1

$$\text{Minimize} \quad \sum_k |C_k| \quad \text{subject to} \qquad \text{(compactness)}$$

- $\cancel{\forall k \quad |C_k| \leq \alpha,}$ (size)
- $\forall i \quad \frac{\{\text{\# of covered points in } I_i\}}{\{\text{\# of points in } I_i\}} \geq \delta.$ (coverage)

Image Clustering Algorithm

1. Start with all the images in a cluster
2. Optimize *compactness* while keeping *coverage*
3. If *size* constraint is broken, split a cluster

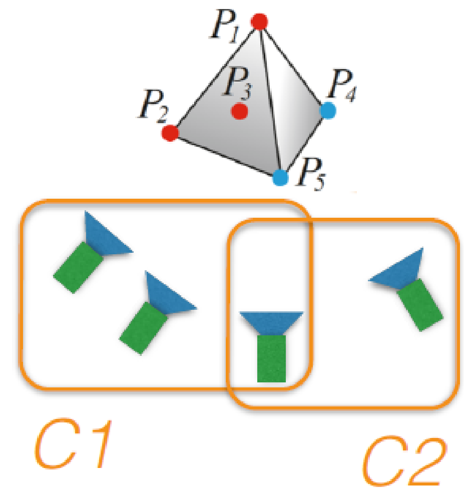


$$\text{Minimize } \sum_k |C_k| \text{ subject to} \qquad \text{(compactness)}$$

- $\forall k \ \ |C_k| \leq \alpha,$ (size)
- $\forall i \ \ \cancel{\frac{\{\text{\# of covered points in } I_i\}}{\{\text{\# of points in } I_i\}}} \geq \delta.$ (coverage)

Image Clustering Algorithm

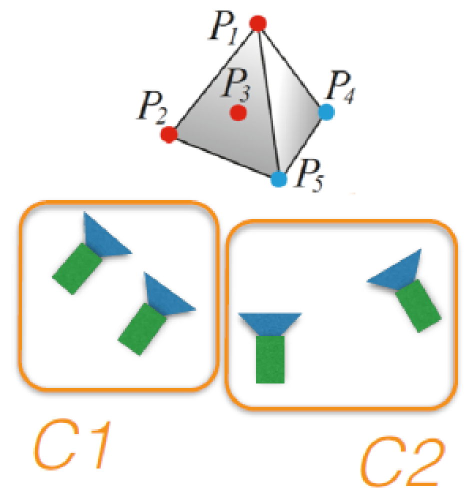
1. Start with all the images in a cluster
2. Remove redundant images while keeping *coverage*
3. If *size* constraint is broken, split a cluster
4. Add an image to each cluster to satisfy *coverage*



Minimize $\sum_k |C_k|$ subject to (compactness)

- $\cancel{\forall k \ |C_k| \leq \alpha,}$ (size)
- $\forall i \ \frac{\{\text{\# of covered points in } I_i\}}{\{\text{\# of points in } I_i\}} \geq \delta.$ (coverage)

Image Clustering Algorithm



1. Start with all the images in a cluster
2. Remove redundant images while keeping *coverage*
3. If *size* constraint is broken, split a cluster
4. Add an image to each cluster to satisfy *coverage*
5. Repeat (3,4) until *size* and *coverage* are satisfied

$$\text{Minimize} \sum_k |C_k| \text{ subject to} \qquad \text{(compactness)}$$

- $\cancel{\forall k \ \ |C_k| \leq \alpha,}$ **?** $\qquad$ (size)
- $\forall i \ \ \dfrac{\{\text{\# of covered points in } I_i\}}{\{\text{\# of points in } I_i\}} \geq \delta.$ $\qquad$ (coverage)

St. Peter's Basilica

- 1275 images
- 4 clusters
- 6M 3D points