Depth Extraction from Video Using Non-parametric Sampling

Kevin Karsch
University of Illinois

Ce Liu
Microsoft Research New England

Sing Bing Kang
Microsoft Research

Problem Statement

Given an image/video, estimate distance from the camera

No parallax necessary Camera motion OK Scene motion OK

Problem Statement

Given an image/video, estimate distance from the camera

No parallax necessary Camera motion OK Scene motion OK

Problem Statement

Given an image/video, estimate distance from the camera

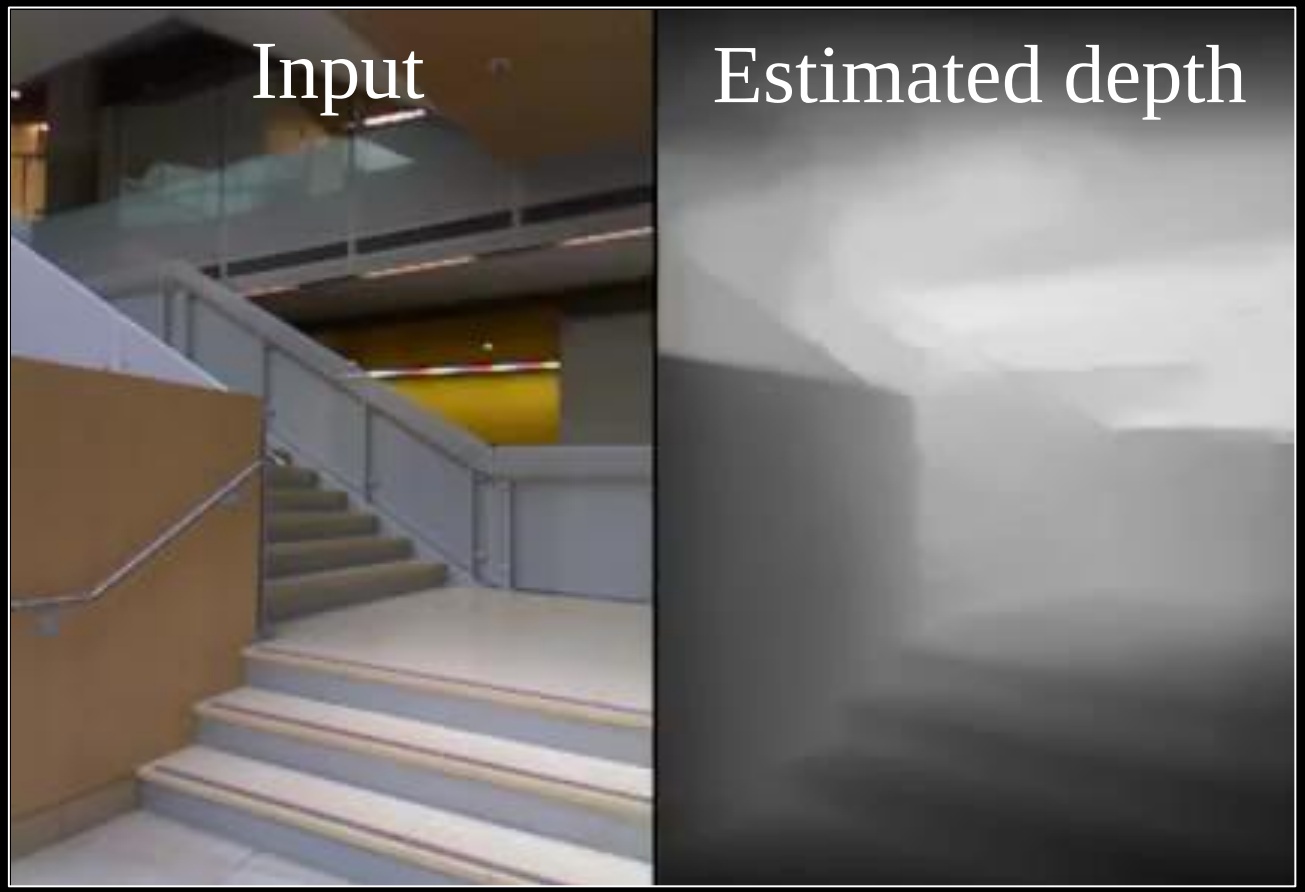
No parallax necessary Camera motion OK Scene motion OK

Input Estimated depth

Problem Statement

Given an image/video, estimate distance from the camera

No parallax necessary Camera motion OK Scene motion OK



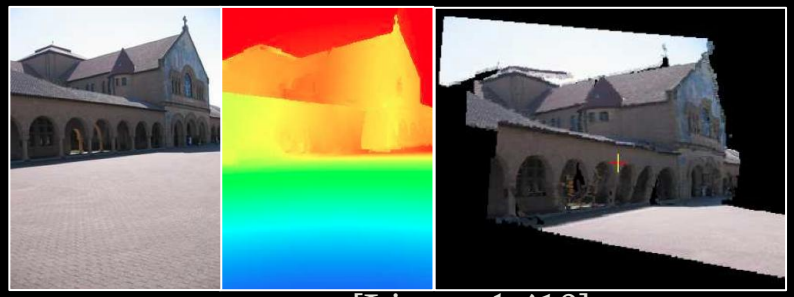
Related Work

[Zhang et al. ’09]

Multiview reconstruction

- Very accurate for videos with moving camera
- May fail for dynamic scenes

[Newcombe and Davidson ’10]
[Furukawa and Ponce ’09]
[Zhang et al. ‘09]
…



[Liu et al. ’10]

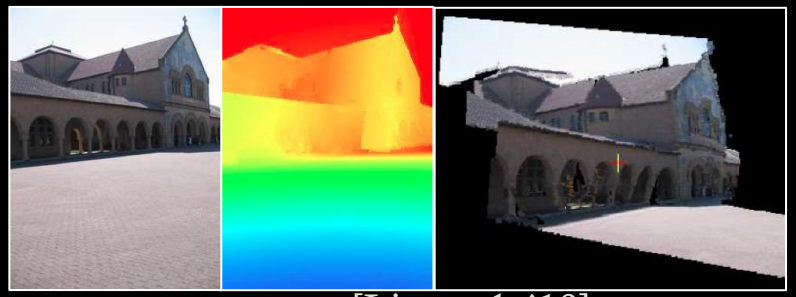
Parametric learning

- Works well for single images
- No literature on extending to video

[Liu et al. ’10]
[Saxena et al. ‘09]
[Hoiem et al. ’05]
…

Related Work

[Zhang et al. ’09]



[Liu et al. ’10]

Multiview reconstruction

- Very accurate for videos with moving camera
- May fail for dynamic scenes

[Newcombe and Davidson ’10]
[Furukawa and Ponce ’09]
[Zhang et al. ‘09]
…

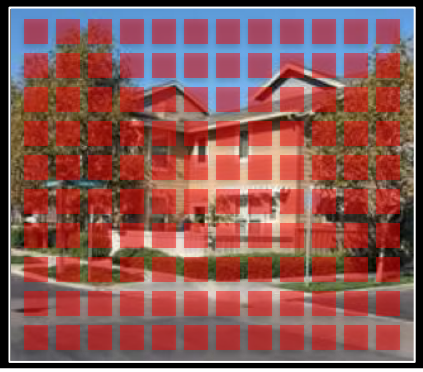
Parametric learning

- Works well for single images
- No literature on extending to video

[Liu et al. ’10]
[Saxena et al. ‘09]
[Hoiem et al. ’05]
…

Training set

Pixel level [Saxena et al. ‘05]



All data

Object level [Lui et al. ‘10]

Tree

Sky

…

Ground

Scene level (ours)

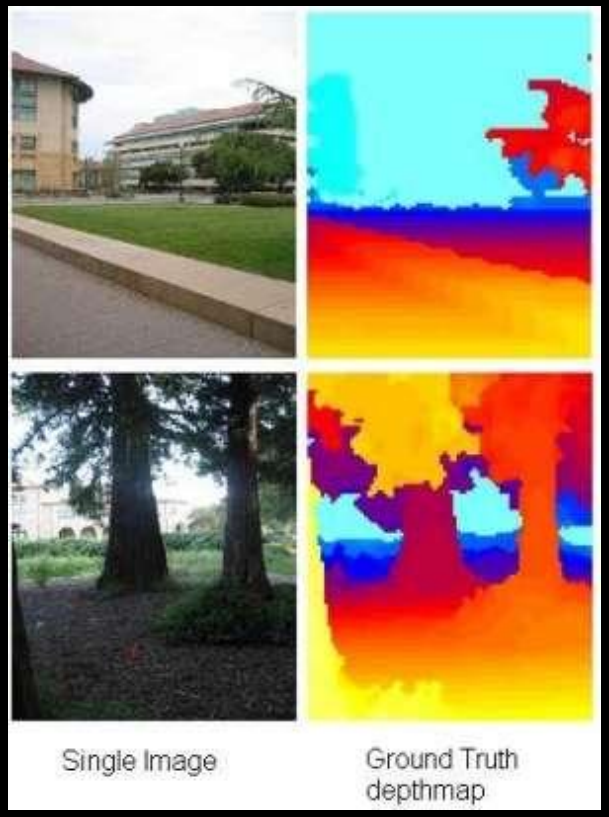
Forest

Office

Street-level

…

RGBD Datasets

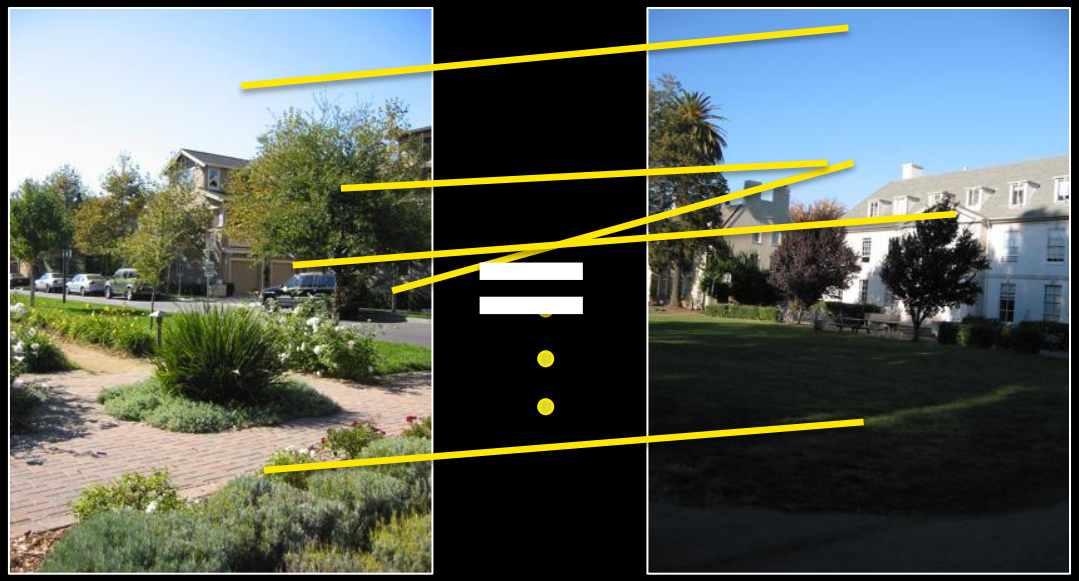


Laser rangefinder
Outdoor scenes
[Saxena et al.]

MSR-V3D
Indoor scenes
(Ours)

SIFT Flow Refresher

- Optical flow using dense SIFT features
  - Larger search window
  - Modified smoothness constraints
- Scenes rearranged so semantics are matched



A B

[Liu et al. ‘08, ‘09]

SIFT Flow Refresher

- Optical flow using dense SIFT features
  - Larger search window
  - Modified smoothness constraints
- Scenes rearranged so semantics are matched

Ψ

Warping operator

A

B

[Liu et al. ‘08, ‘09]

SIFT Flow Refresher

- Optical flow using dense SIFT features
  - Larger search window
  - Modified smoothness constraints
- Scenes rearranged so semantics are matched

Ψ

Warping operator

A

=

B

[Liu et al. ‘08, ‘09]

SIFT Flow Refresher

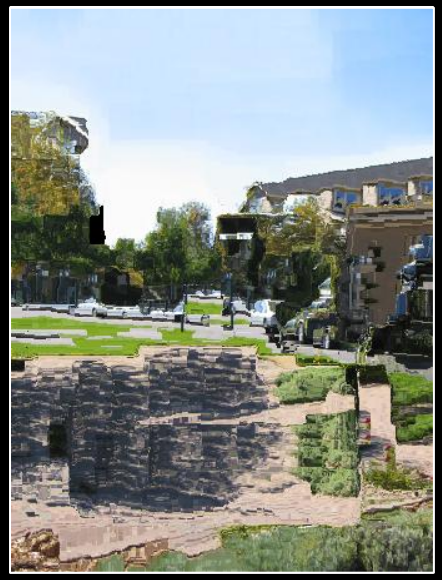
- Optical flow using dense SIFT features
  - Larger search window
  - Modified smoothness constraints
- Scenes rearranged so semantics are matched

Ψ

Warping operator

A

=



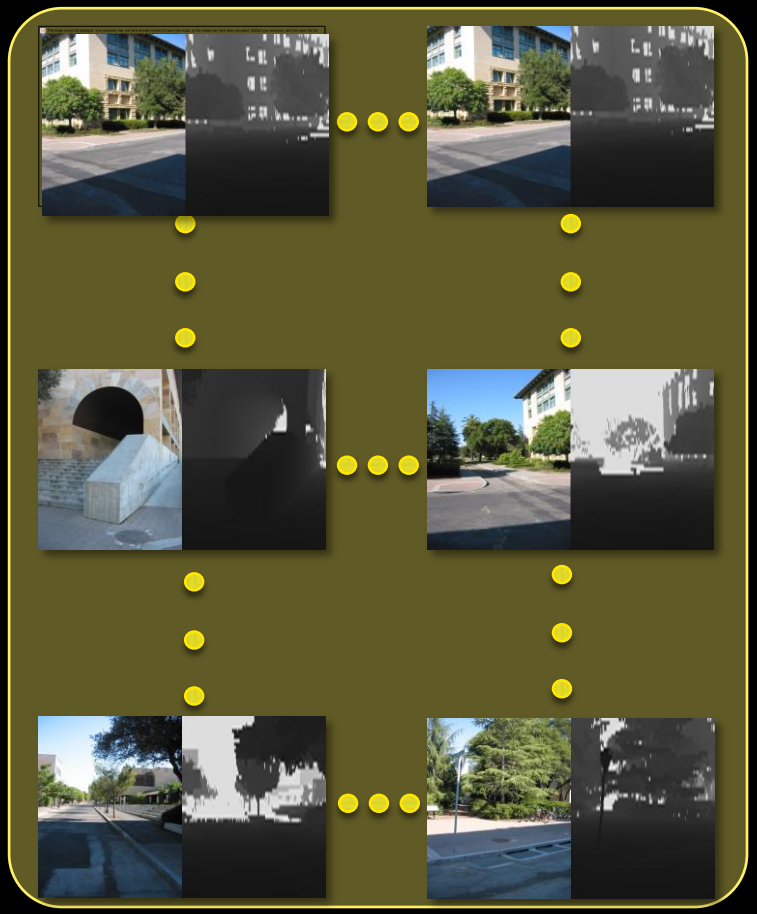
B

[Liu et al. ‘08, ‘09]

Algorithm

Input image

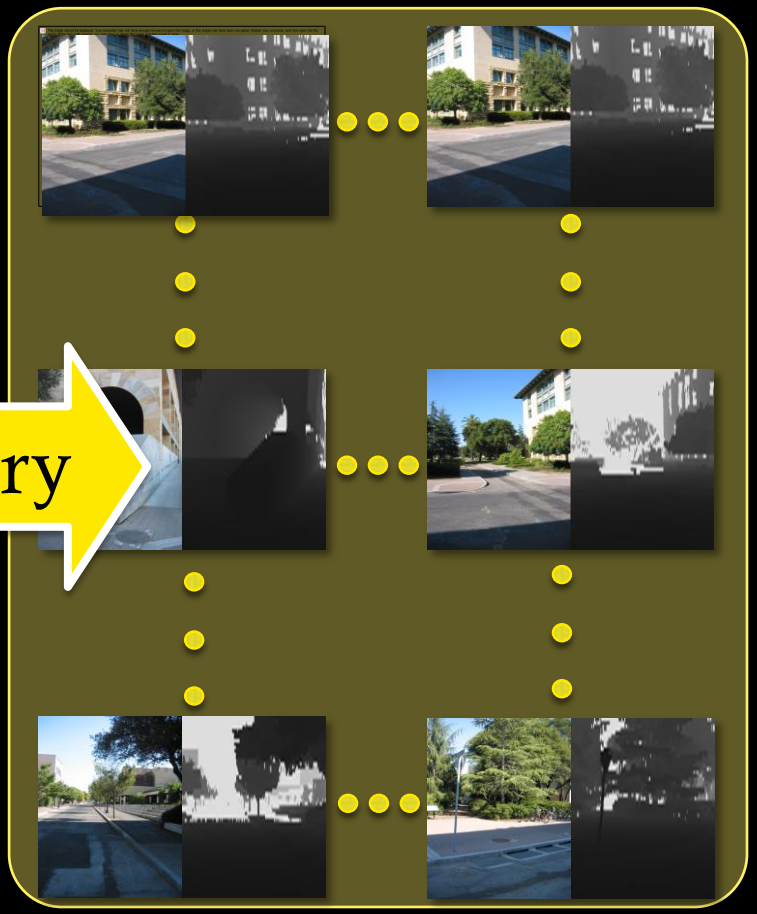
RGBD Database



Algorithm

Input image

RGBD Database



kNN query

Algorithm

Input

Candidates

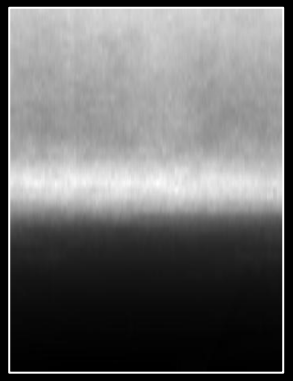
SIFT flow

Warped candidates

Algorithm

Input

Prior



Warped candidates

Depth inference

Inferred depth

Inference

$$\underset{D}{\operatorname{argmin}}\, E(D) = \sum_{i \in \text{pixels}} \left[\sum_{C \in \text{candidates}} w_i \left(|D_i - C_i|_1 + \gamma |\nabla D_i - \nabla C_i|_1 \right) \right] + \underbrace{\alpha s_i |\nabla D_i|_1}_{\text{Spatial smoothness}} + \underbrace{\beta |D_i - \text{prior}_i|_1}_{\text{Match to database mean}}$$

Enforce depth to match candidates

- D : inferred depth
- C : warped candidate depth
- w : depth confidence
- s : image-based weights
- α, β, γ : constant weights

- Both *absolute* and *relative* depth are transferred
- Regularize with smoothness and prior

Inference

$$\underset{D}{\operatorname{argmin}}\, E(D) = \sum_{i \in \text{pixels}} \left[\sum_{C \in \text{candidates}} w_i \left(|D_i - C_i|_1 + \gamma |\nabla D_i - \nabla C_i|_1 \right) \right] + \alpha s_i |\nabla D_i|_1 + \beta |D_i - \text{prior}_i|_1$$

Enforce depth to match candidates

Spatial smoothness

Match to database mean

Not a discrete MRF!

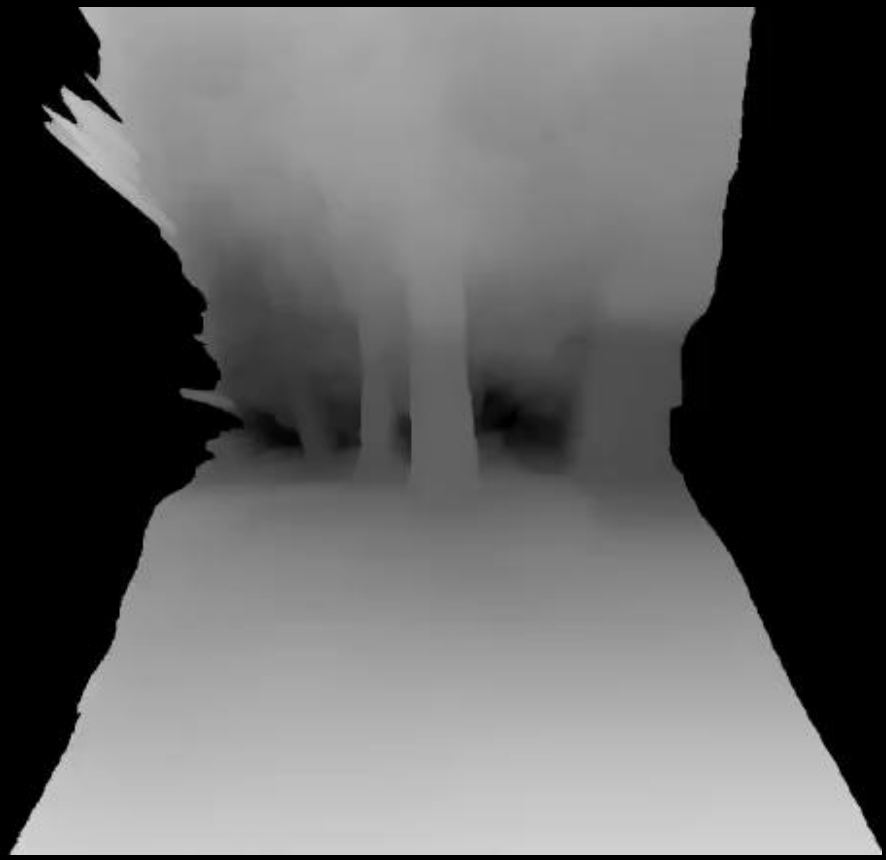
D : inferred depth
C : warped candidate depth
w : depth confidence
s : image-based weights
α, β, γ : constant weights

- Both *absolute* and *relative* depth are transferred
- Regularize with smoothness and prior

Input

True depth

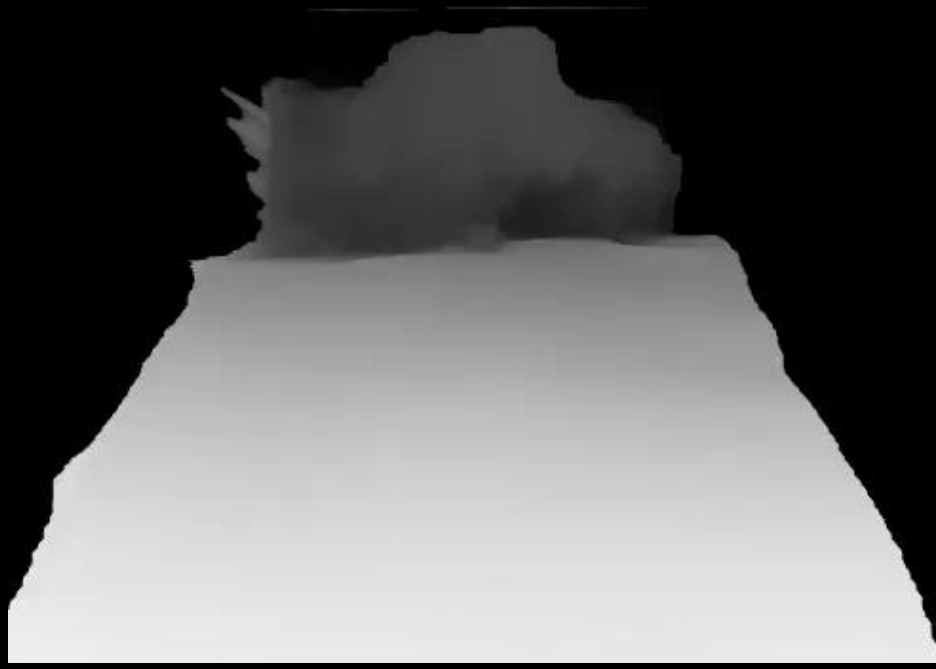
Inferred depth



Input

True depth

Inferred depth



$$\sum_{i \in \text{pixels}} \left[\sum_{C \in \text{candidates}} w_i \left(|D_i - C_i|_1 + \gamma |\nabla D_i - \nabla C_i|_1 \right) \right]$$

$$+ \alpha s_i |\nabla D_i|_1 + \beta |D_i - \text{prior}_i|_1$$

Result *without* relative depth term ($\gamma = 0$)

Result *with* relative depth term ($\gamma > 0$)

$$\sum_{i \in \text{pixels}} \left[\sum_{C \in \text{candidates}} w_i \left(|D_i - C_i|_1 + \gamma |\nabla D_i - \nabla C_i|_1 \right) \right]$$

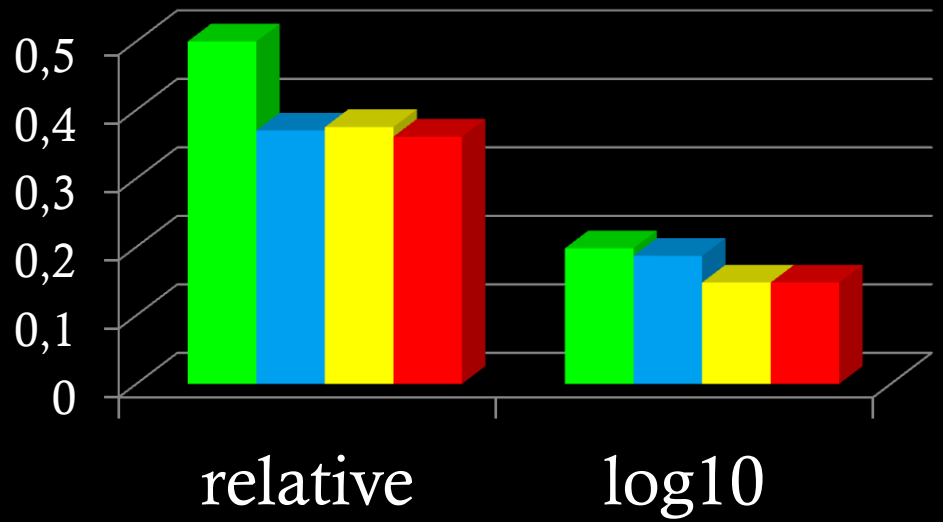
$$+ \alpha s_i |\nabla D_i|_1 + \beta |D_i - \text{prior}_i|_1$$

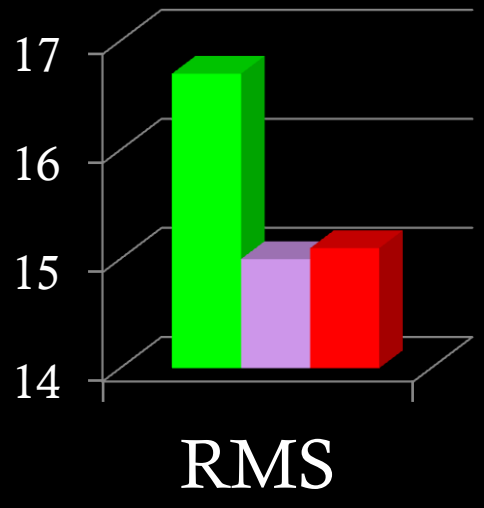
Result *without* relative depth term ($\gamma = 0$)

Result *with* relative depth term ($\gamma > 0$)

Evaluation: Make3D Dataset

Method
Depth MRF [Saxena et al. '05]
Make3D [Saxena et al. '09]
θ-MRF [Li et al. '11]
Semantic Labels [Liu et al. '10]
Depth Transfer (ours)

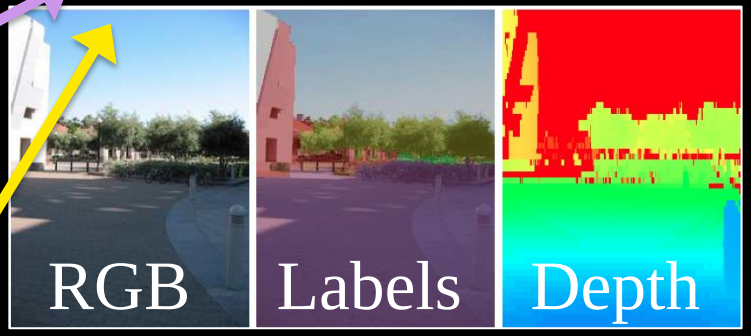


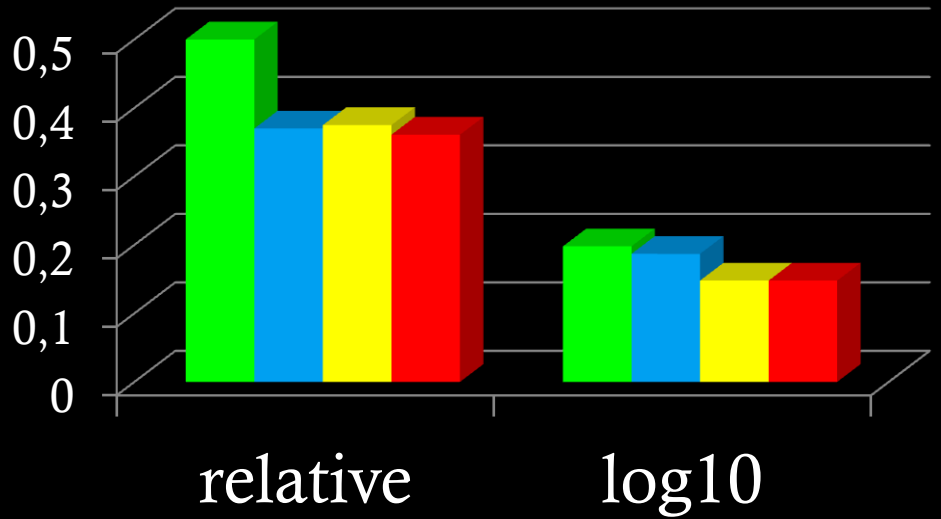


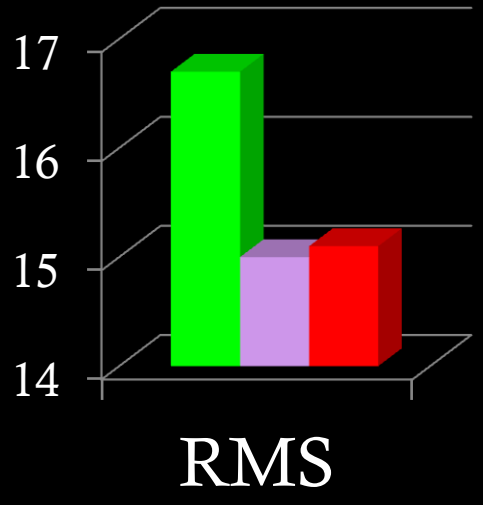
Evaluation: Make3D Dataset

Method
Depth MRF [Saxena et al. '05]
Make3D [Saxena et al. '09]
θ-MRF [Li et al. '11]
Semantic Labels [Liu et al. '10]
Depth Transfer (ours)

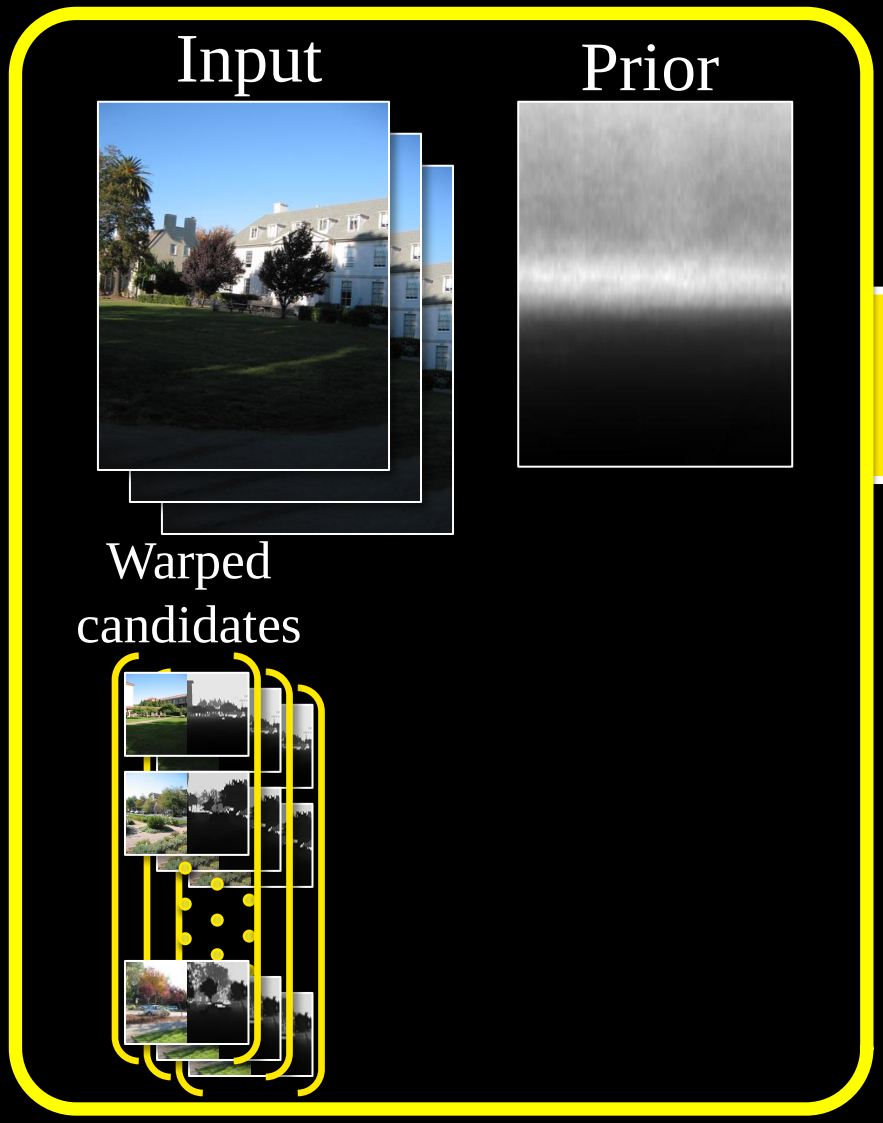
RGBD+labeled data







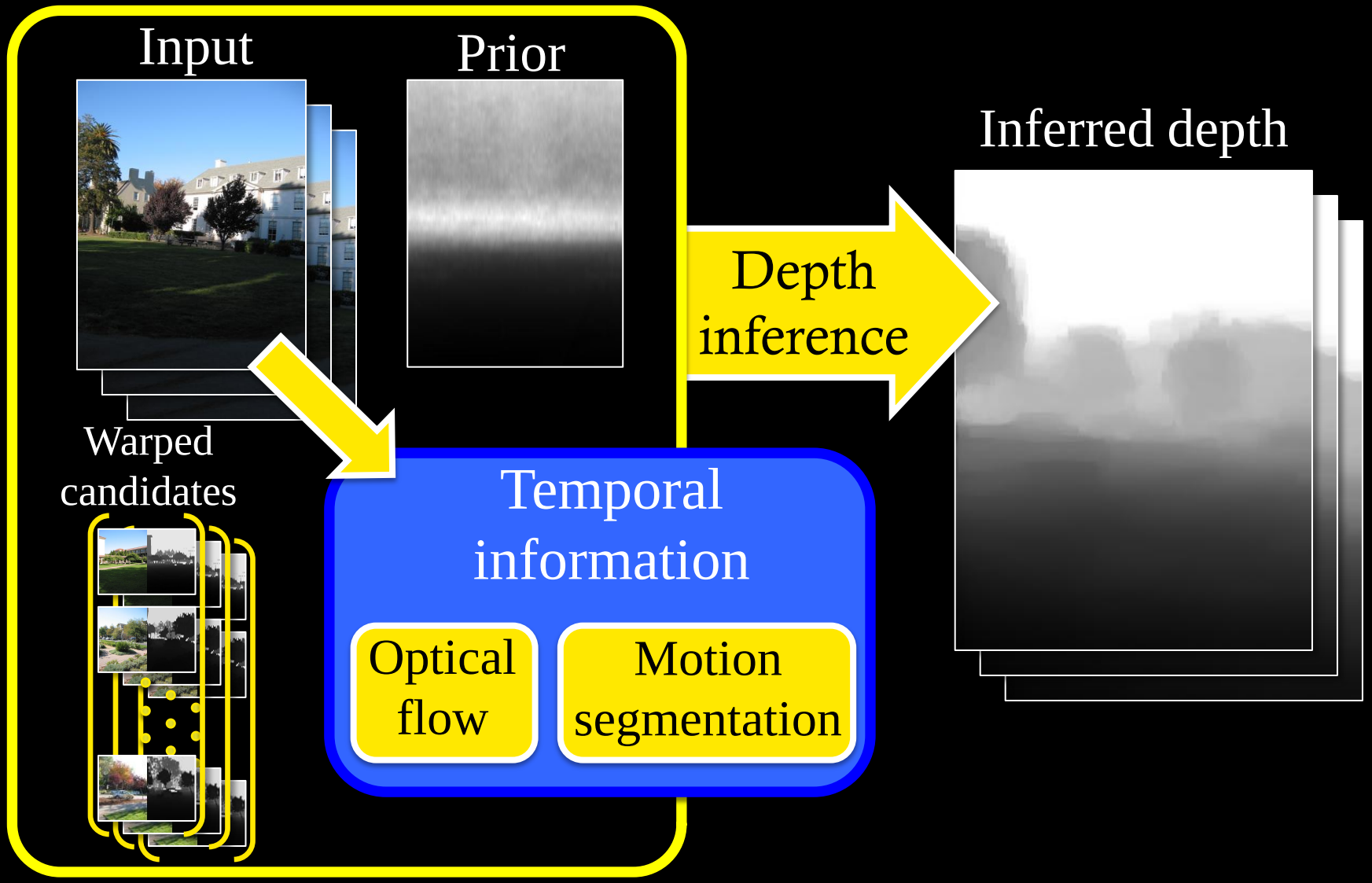
Video Extension



Depth inference

Inferred depth

Video Extension



Video Inference

m : binary motion mask
$\mathcal{M}$: hypothesized depth of motion mask
ζ, η : constant weights

$$\underset{D}{\operatorname{argmin}}\, E_{\text{video}}(D) =$$

$$\underbrace{E(D)}_{\text{Single image objective}} + \sum_{i \in \text{pixels}} \zeta\, t_i |\nabla_{flow} D_i|_1 + \eta\, m_i |D_i - \mathcal{M}_i|_1$$

- Depth changes are gradual frame-to-frame
- Moving objects are usually on the ground

Video Inference

m : binary motion mask
$\mathcal{M}$: hypothesized depth of motion mask
ζ, η : constant weights

$$\underset{D}{\operatorname{argmin}}\, E_{\text{video}}(D) =$$

$$\underbrace{E(D)}_{\text{Single image objective}} + \sum_{i \in \text{pixels}} \underbrace{\zeta\, t_i |\nabla_{flow} D_i|_1}_{\text{Smooth along direction of optical flow}} + \eta\, m_i |D_i - \mathcal{M}_i|_1$$

- Depth changes are gradual frame-to-frame
- Moving objects are usually on the ground

Video Inference

m : binary motion mask
$\mathcal{M}$: hypothesized depth of motion mask
ζ, η : constant weights

$$\underset{D}{\operatorname{argmin}}\, E_{\text{video}}(D) =$$

$$\underbrace{E(D)}_{\text{Single image objective}} + \sum_{i \in \text{pixels}} \underbrace{\zeta\, t_i |\nabla_{flow} D_i|_1}_{\text{Smooth along direction of optical flow}} + \underbrace{\eta\, m_i |D_i - \mathcal{M}_i|_1}_{\text{Coerce moving objects to be “grounded”}}$$

- Depth changes are gradual frame-to-frame
- Moving objects are usually on the ground
  - Motion mask = threshold flow-weighted, relative pixel differences
  - Ce Liu’s optical flow http://people.csail.mit.edu/celiu/OpticalFlow

Input

Inferred depth

Results

MSR-V3D evaluation

Input Kinect* Ours Input Kinect* Ours

*Naïve hole filling applied to Kinect data (for visualization only)

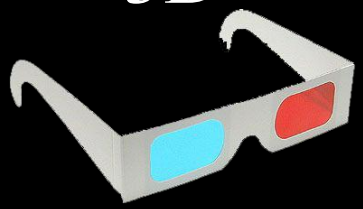
Limitations

Application: 2D-to-3D

Input

Depth

Anaglyph
“3D”



Thanks!

More results, code and dataset available at:
http://kevinkarsch.com/depthtransfer

Our 2D-to-3D

Youtube 2D-to-3D