Machine Learning in Graphics and Vision

Prof. Dr.-Ing. Andreas Geiger

Autonomous Vision Group
MPI-IS / University of Tübingen

July 19, 2018

University of Tübingen
MPI for Intelligent Systems
Autonomous Vision Group

Discussion of Lecture Evaluation

Overview

Structured Prediction I

- Graphical Models: Factor Graphs
- Inference: Belief Propagation

Structured Prediction II

- Stereo & Multi-view Reconstruction
- Optical Flow Estimation

Structured Prediction III

- Parameter Estimation
- Deep Structured Models

Disclaimer:

There are many structured prediction problems ...

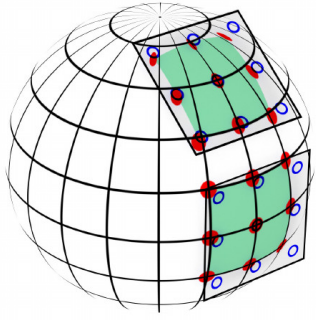
... we can not discuss all of them today

SphereNet: Learning Spherical Representations for Detection

SphereNet: Learning Spherical Representations for Detection

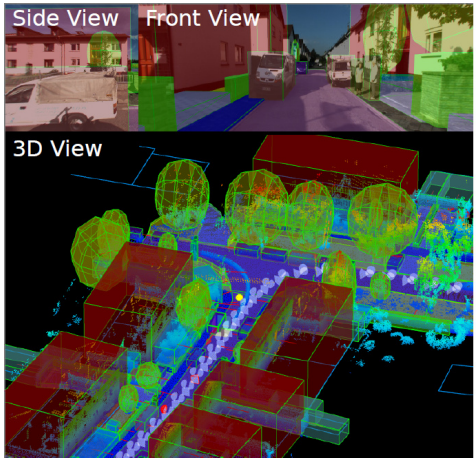
360° Cameras

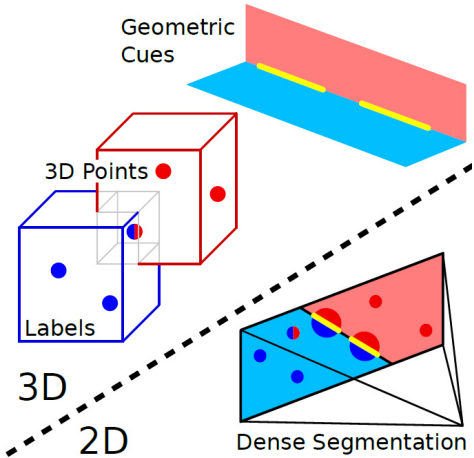
360° Image



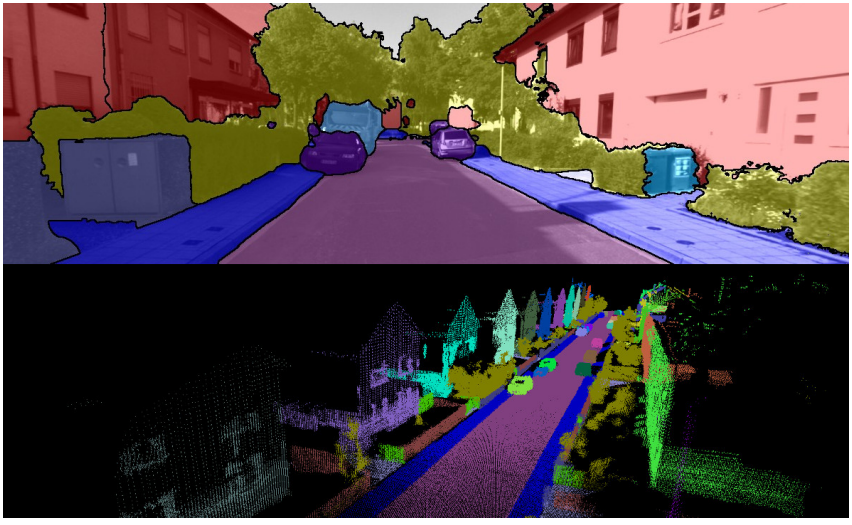
SphereNet Kernel

3D to 2D Semantic Label Transfer



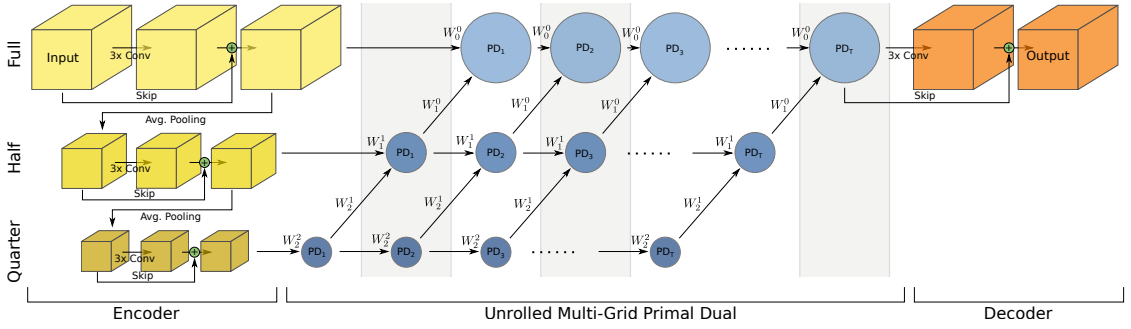


3D to 2D Semantic Label Transfer



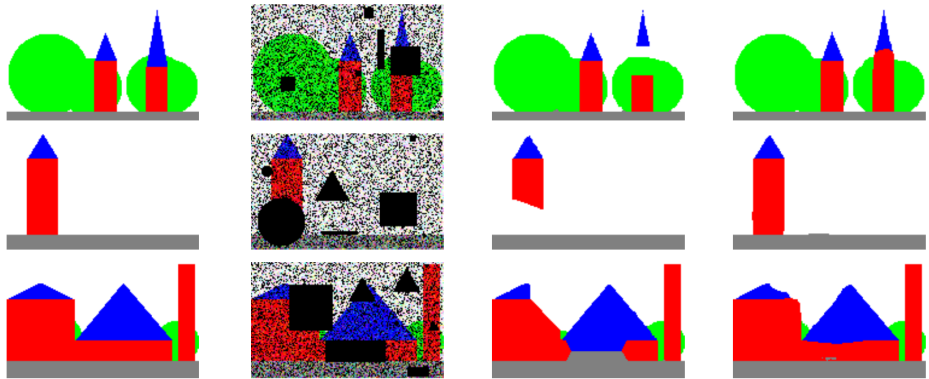
Xie, Kiefel, Sun & Geiger, CVPR 2016

Learning Priors for Semantic 3D Reconstruction

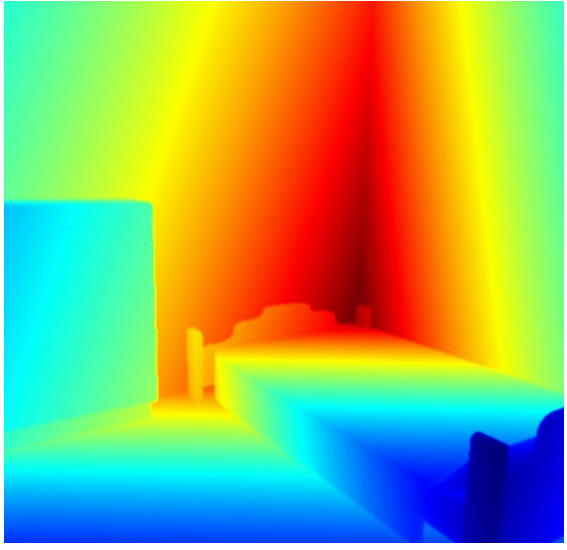


$$\underset{u}{\text{minimize}} \quad \int_{\Omega} \left(\|Wu\|_2 + fu \right) d\mathbf{x} \quad \text{subject to} \quad \forall \mathbf{x} \in \Omega : \sum_{\ell} u_{\ell}(\mathbf{x}) = 1$$

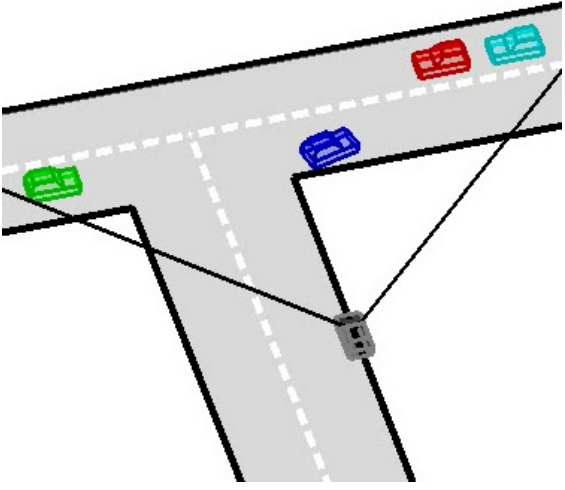
Learning Priors for Semantic 3D Reconstruction



Indoor Scene Understanding

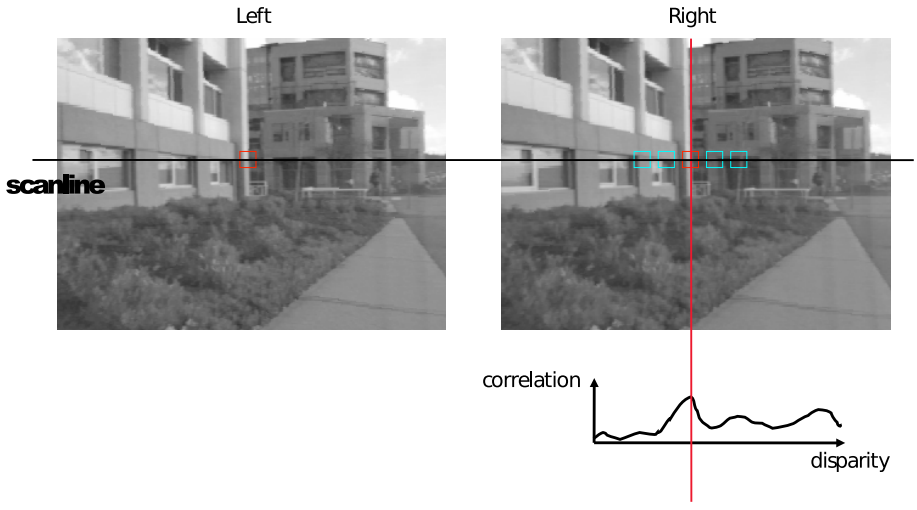


Traffic Scene Understanding

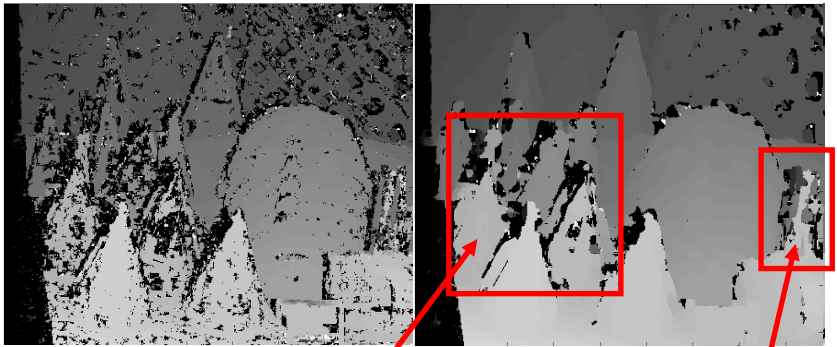


Stereo Reconstruction

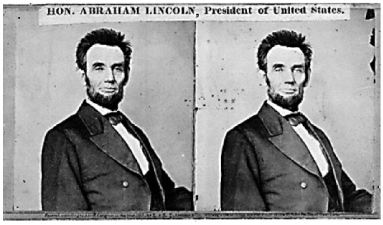
Block Matching



Block Matching – Window Size



Block Matching – Failure Cases



Textureless surfaces

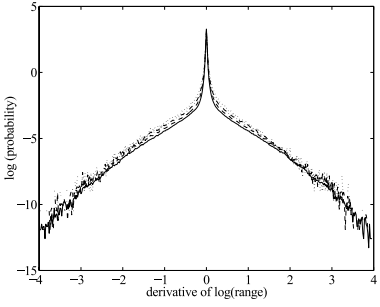
Occlusions, repetition

Non-Lambertian surfaces, specularities

Spatial Regularization

How does the real world look like?

- Leverage real-world statistics
- E.g.: Brown range image database [Mumford et al.]



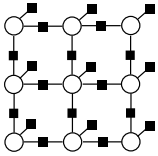
Fast Approximate Energy Minimization via Graph Cuts

[Boykov, Veksler & Zabih, PAMI 1999]

Stereo MRF

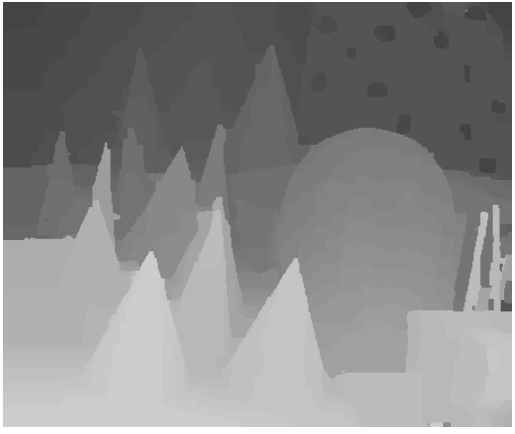
- Specify a loopy Markov Random Field (MRF) on a grid and solve for the whole disparity map $\mathbf{D}$ at once. MAP solution = minimum of energy.

$$p(\mathbf{D}) \propto \exp\left\{ -\sum_i \psi_{data}(d_i) - \lambda \sum_{i \sim j} \psi_{smooth}(d_i, d_j) \right\}$$

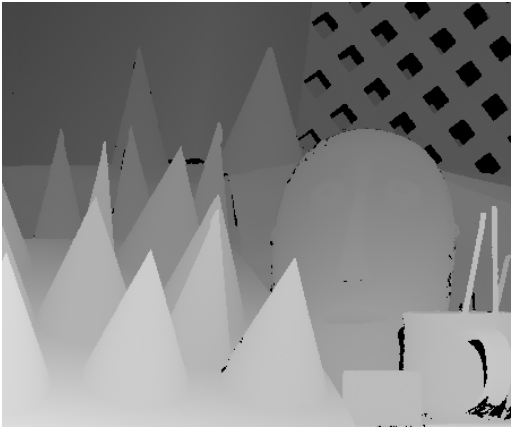


- $i \sim j$: neighboring pixels on a 4-connected grid
- Unary terms: Matching cost $\psi_{data}(d)$
- Pairwise terms: Smoothness between adjacent pixels, e.g.:
  - Potts: $\psi_{smooth}(d, d') = [d \neq d']$
  - Truncated l_1: $\psi_{smooth}(d, d') = \min(|d - d'|, \tau)$
- Solve MRF approximately using BP / graph cuts

Stereo MRF – Results



Inference Results



Ground Truth

Displets: Resolving Stereo Ambiguities using Object Knowledge

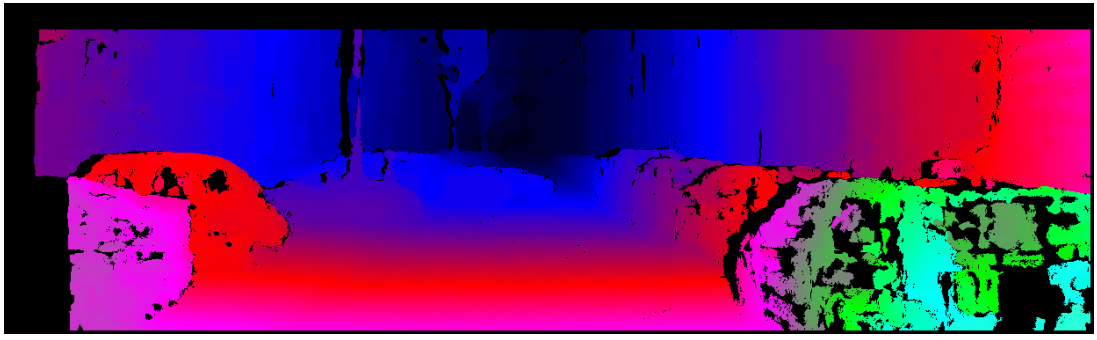
[Güney & Geiger, CVPR 2015]

Integrating Object Knowledge

$$E(D) = \underbrace{\sum_i \psi_i^{\mathcal{A}}(d_i)}_{\text{Appearance}} + \lambda \underbrace{\sum_{i \sim j} \psi_{ij}^{\mathcal{S}}(d_i, d_j)}_{\text{Smoothness}}$$

[Boykov, PAMI 1999]

Integrating Object Knowledge



$$E(D) = \underbrace{\sum_i \psi_i^{\mathcal{A}}(d_i)}_{\text{Appearance}} + \lambda \underbrace{\sum_{i \sim j} \psi_{ij}^{\mathcal{S}}(d_i, d_j)}_{\text{Smoothness}}$$

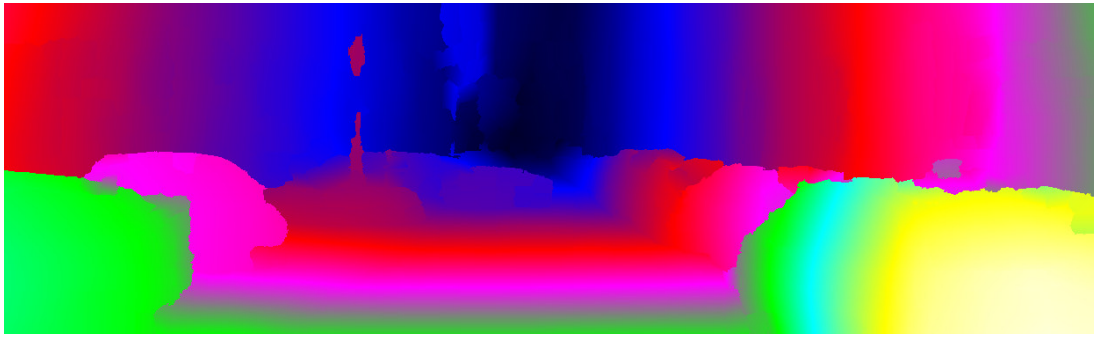
[Boykov, PAMI 1999]

Integrating Object Knowledge

$$E(D,O) = \underbrace{\sum_i \psi_i^{\mathcal{A}}(d_i)}_{\text{Appearance}} + \lambda_{\mathcal{S}} \underbrace{\sum_{i \sim j} \psi_{ij}^{\mathcal{S}}(d_i, d_j)}_{\text{Smoothness}} + \lambda_{\mathcal{O}} \underbrace{\sum_k \psi_k^{\mathcal{O}}(o_k)}_{\text{Object Semantics}} + \lambda_{\mathcal{C}} \underbrace{\sum_k \sum_i \psi_{ki}^{\mathcal{C}}(o_k, d_i)}_{\text{3D Consistency}}$$

[Güney & Geiger, CVPR 2015]

Integrating Object Knowledge



$$E(D,O) = \underbrace{\sum_i \psi_i^{\mathcal{A}}(d_i)}_{\text{Appearance}} + \lambda_{\mathcal{S}} \underbrace{\sum_{i \sim j} \psi_{ij}^{\mathcal{S}}(d_i, d_j)}_{\text{Smoothness}} + \lambda_{\mathcal{O}} \underbrace{\sum_k \psi_k^{\mathcal{O}}(o_k)}_{\text{Object Semantics}} + \lambda_{\mathcal{C}} \underbrace{\sum_k \sum_i \psi_{ki}^{\mathcal{C}}(o_k, d_i)}_{\text{3D Consistency}}$$

[Güney & Geiger, CVPR 2015]

Stereo Reconstruction Summary

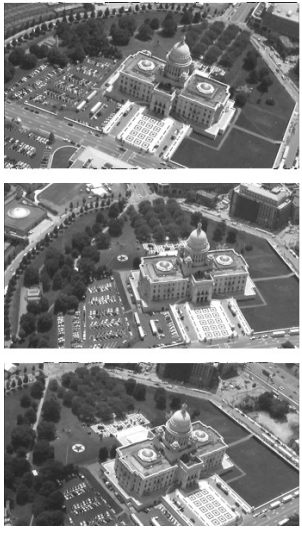
- Block matching suffers from ambiguities
- Choosing window size is problematic (tradeoff)
- Incorporating smoothness constraints can resolve some of the ambiguities and allows for choosing small windows (no bleeding artifacts)
- Can be formulated as inference in a discrete MRF
- MAP solution can be obtained using belief propagation / graph cuts
- Integrating recognition cues can further regularize the problem

Multi-View Reconstruction

Towards Probabilistic Reconstruction

[Ulusoy, Geiger & Black, 3DV 2015]

Volumetric 3D Reconstruction from Multiple Views



Related Work

Online Algorithms:

- [Bonet & Viola, ICCV 1999]
- [Agrawal & Davis, CVPR 2001]
- [Pollard & Mundy, CVPR 2007]

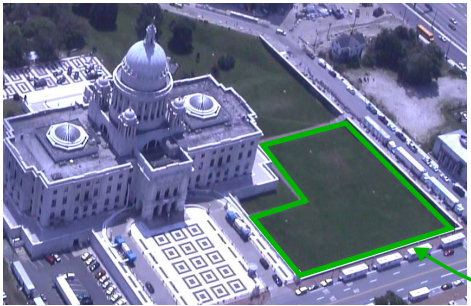
Ray Potentials:

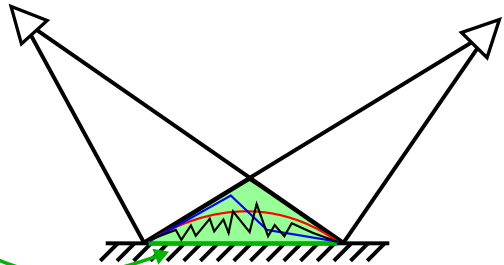
- [Gargallo, Sturm & Pujades, ACCV 2007]
- [Liu & Cooper, PAMI 2014]
- [Savinov, Ladicky, Häne & Pollefeys, CVPR 2015]

Here:

- Joint estimation of voxel occupancy and appearance
- Marginal inference algorithm $\Rightarrow$ Bayes optimal predictions

Reconstruction Ambiguities

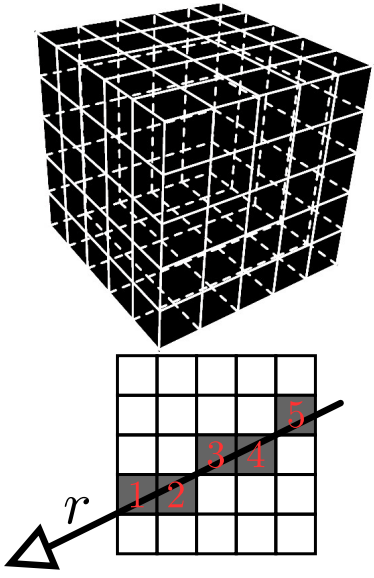




- Image-based 3D reconstruction is a highly **ill-posed** problem
  ⇒ Coping with and exposing **uncertainty** is essential

Can we formulate 3D reconstruction in a probabilistic way?

Representation



- **Voxel occupancy:**

$$o_i = \begin{cases} 1 & \text{if voxel } i \text{ is occupied} \\ 0 & \text{if voxel } i \text{ is empty} \end{cases}$$

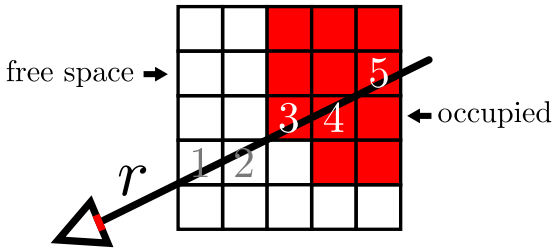
- **Voxel appearance:**

$$a_i \in \mathbb{R}$$

- **Shorthand notation:**

$$\mathbf{o}_r = \{o_1^r, \ldots, o_{N_r}^r\}$$
$$\mathbf{a}_r = \{a_1^r, \ldots, a_{N_r}^r\}$$

Image Formation Process



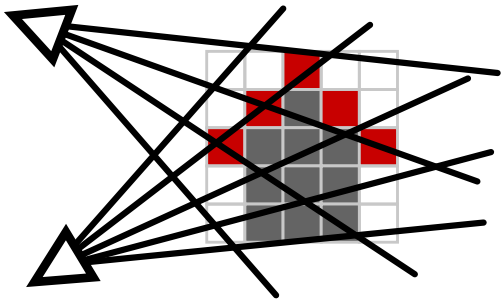
$$I_r = \sum_{i=1}^{N} o_i \prod_{j<i} (1 - o_j)\, a_i$$

- I_r: intensity at pixel r o_i: occupancy of voxel i a_i: appearance of voxel i

Probabilistic Model

Joint Distribution:

$$p(\mathbf{o}, \mathbf{a}) = \frac{1}{Z} \prod_{v \in \mathcal{V}} \underbrace{\varphi_v(o_v)}_{\text{unary}} \prod_{r \in \mathcal{R}} \underbrace{\psi_r(\mathbf{o}_r, \mathbf{a}_r)}_{\text{ray}}$$



Probabilistic Model

Joint Distribution:

$$p(\mathbf{o}, \mathbf{a}) = \frac{1}{Z} \prod_{v \in \mathcal{V}} \underbrace{\varphi_v(o_v)}_{\text{unary}} \prod_{r \in \mathcal{R}} \underbrace{\psi_r(\mathbf{o}_r, \mathbf{a}_r)}_{\text{ray}}$$

Unary Potentials:

$$\varphi_v(o_v) = \gamma^{o_v} (1 - \gamma)^{1 - o_v}$$

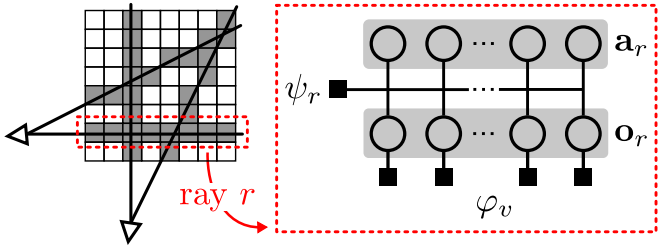
- Most voxels are empty $\Rightarrow \gamma < 0.5$

Probabilistic Model

Joint Distribution:

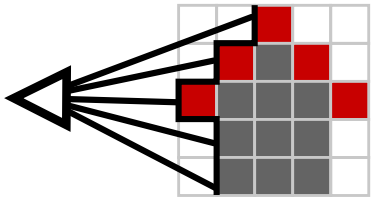
$$p(\mathbf{o}, \mathbf{a}) = \frac{1}{Z} \prod_{v \in \mathcal{V}} \underbrace{\varphi_v(o_v)}_{\text{unary}} \prod_{r \in \mathcal{R}} \underbrace{\psi_r(\mathbf{o}_r, \mathbf{a}_r)}_{\text{ray}}$$

Ray Potentials:

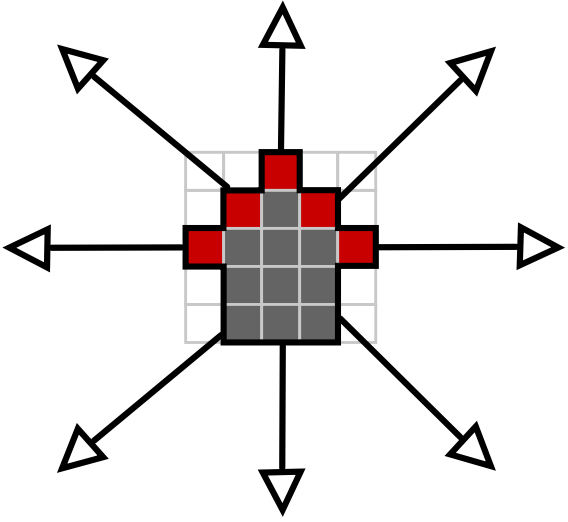


$$\psi_r(\mathbf{o}_r, \mathbf{a}_r) = \sum_{i=1}^{N_r} o_i^r \prod_{j<i} (1 - o_j^r) \underbrace{\mathcal{N}(a_i^r | I_r, \sigma)}_{\text{Gaussian Noise}}$$

3D Reconstruction

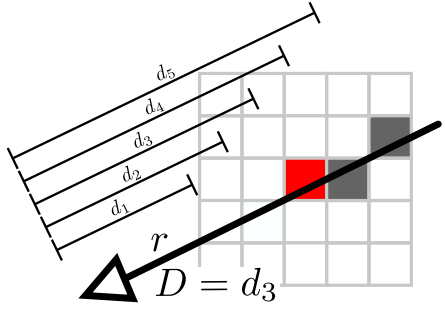


3D Reconstruction



Bayes Optimal Depth Estimation

- Consider a single ray r in space
- Let d_k be the distance from the camera to voxel k along ray r
- Depth $D \in \{d_1, \ldots, d_N\}$: distance to closest occupied voxel



Bayes Optimal Depth Estimation

- Consider a single ray r in space
- Let d_k be the distance from the camera to voxel k along ray r
- Depth $D \in \{d_1, \ldots, d_N\}$: distance to closest occupied voxel
- Optimal depth estimate:

$$
\begin{aligned}
D^* &= \operatorname*{argmin}_{D'} \text{Risk}(D') \\
&= \operatorname*{argmin}_{D'} \mathbb{E}_{p(D)}[\,\Delta(D, D')\,] \\
&= \begin{cases} \text{mean}\,(\,p(D)\,) & \text{if } \Delta(D, D') = (D - D')^2 \\ \text{median}\,(\,p(D)\,) & \text{if } \Delta(D, D') = |D - D'| \end{cases}
\end{aligned}
$$

- Requires depth distribution $p(D)$!

Depth Distribution for Single Ray

Let:

$$o_1 = 0 \quad \cdots \quad o_{k-1} = 0 \quad o_k = 1$$

Then:

$$p(D = d_k) = \underbrace{\sum_{\mathbf{o}_{>k}} \int_{\mathbf{a}} p(\mathbf{o}, \mathbf{a})}_{= p(o_1, \ldots, o_k)}$$

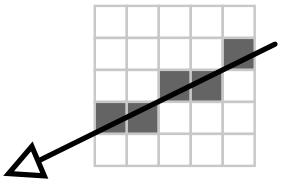
Depth Distribution for Single Ray

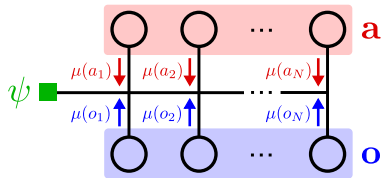
Let:

$$o_1 = 0 \quad \cdots \quad o_{k-1} = 0 \quad o_k = 1$$

Then:

$$p(D = d_k) \propto \sum_{\mathbf{o}_{>k}} \int_{\mathbf{a}} \psi(\mathbf{o}, \mathbf{a}) \prod_i \mu(o_i) \prod_i \mu(a_i)$$





Marginal = Product of Factor & Incoming Messages

Depth Distribution for Single Ray

Let:

$$o_1 = 0 \quad \cdots \quad o_{k-1} = 0 \quad o_k = 1$$

Then:

$$p(D = d_k) \propto \sum_{\mathbf{o}_{>k}} \int_{\mathbf{a}} \psi(\mathbf{o}, \mathbf{a}) \prod_i \mu(o_i) \prod_i \mu(a_i)$$

$$\psi(\mathbf{o}, \mathbf{a}) = \underbrace{\sum_i o_i \prod_{j<i} (1 - o_j) \mathcal{N}(a_i | I, \sigma)}_{= \mathcal{N}(a_k | I, \sigma)}$$

Depth Distribution for Single Ray

Let:

$$o_1 = 0 \quad \cdots \quad o_{k-1} = 0 \quad o_k = 1$$

Then:

$$p(D = d_k) \propto \sum_{\mathbf{o}_{>k}} \int_{\mathbf{a}} \mathcal{N}(a_k | I, \sigma) \prod_i \mu(o_i) \prod_i \mu(a_i)$$

Depth Distribution for Single Ray

Let:

$$o_1 = 0 \quad \cdots \quad o_{k-1} = 0 \quad o_k = 1$$

Then:

$$p(D = d_k) \propto \sum_{\mathbf{o}_{>k}} \int_{\mathbf{a}} \mathcal{N}(a_k | I, \sigma) \prod_{i>k} \mu(o_i) \prod_{i} \mu(a_i) \times \prod_{i \leq k} \mu(o_i)$$

Depth Distribution for Single Ray

Let:

$$o_1 = 0 \quad \cdots \quad o_{k-1} = 0 \quad o_k = 1$$

Then:

$$p(D = d_k) \propto \sum_{\mathbf{o}_{>k}} \int_{\mathbf{a}_{\neq k}} \prod_{i>k} \mu(o_i) \prod_{i \neq k} \mu(a_i)$$

$$\times \prod_{i \leq k} \mu(o_i) \int_{a_k} \mathcal{N}(a_k | I, \sigma) \mu(a_k)$$

Depth Distribution for Single Ray

Let:

$$o_1 = 0 \quad \cdots \quad o_{k-1} = 0 \quad o_k = 1$$

Then:

$$p(D = d_k) \propto \underbrace{\sum_{\mathbf{o}_{>k}} \int_{\mathbf{a}_{\neq k}} \prod_{i>k} \mu(o_i) \prod_{i \neq k} \mu(a_i)}_{=1}$$
$$\times \prod_{i \leq k} \mu(o_i) \int_{a_k} \mathcal{N}(a_k | I, \sigma) \mu(a_k)$$

Depth Distribution for Single Ray

Let:

$$o_1 = 0 \quad \cdots \quad o_{k-1} = 0 \quad o_k = 1$$

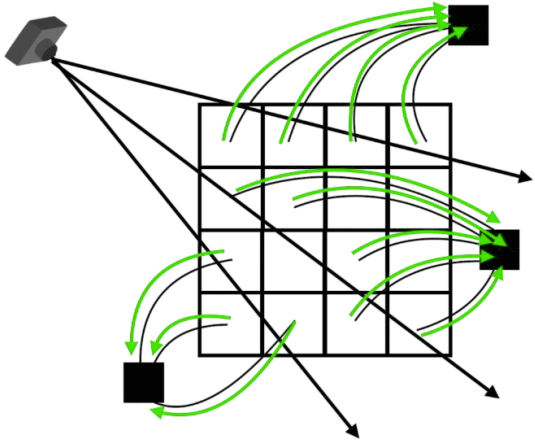
Then:

$$p(D = d_k) \propto \prod_{i \leq k} \mu(o_i) \int_{a_k} \mathcal{N}(a_k | I, \sigma) \mu(a_k)$$

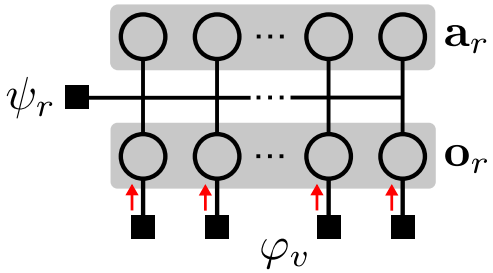
Intuition: Depth $D = d_k \Leftrightarrow$ Voxel k is occupied and visible and explains the observed pixel value.

- How can we obtain $\mu(o_i)$ and $\mu(a_k)$?

Message Passing

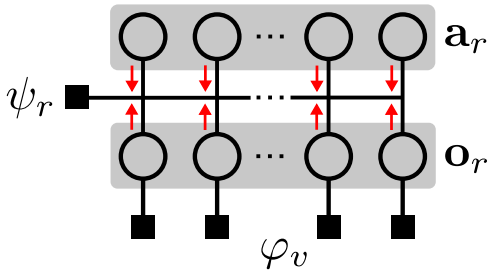


Message Passing



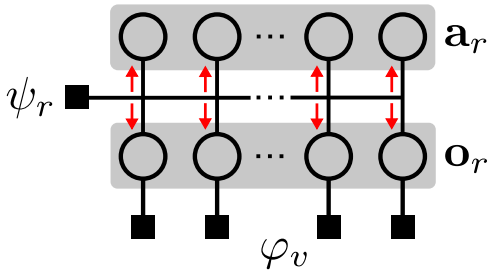
$$\mu_{\varphi_v \to o_v}(o_v) = \varphi_v(o_v)$$

Message Passing



$$\mu_{x \to \psi_r}(x) = \prod_{f \in \mathcal{F} \setminus \psi_r} \mu_{f \to x}(x)$$

Message Passing



$$\mu_{\psi_r \to x}(x) = ?$$

Ray-to-Occupancy Message

$$\mu_{\psi_r \to o_i}(o_i = 1) = \sum_{\mathbf{o}_{j \neq i}} \int_{\mathbf{a}} \psi_r(\mathbf{o}, \mathbf{a}) \prod_{j \neq i} \mu(o_j) \prod_j \mu(a_j)$$

$$\vdots$$

$$= \prod_{k=1}^{i-1} \mu(o_k = 0) \int_{a_i} \nu(a_i)\mu(a_i)$$

$$+ \sum_{j=1}^{i-1} \mu(o_j = 1) \prod_{k=1}^{j-1} \mu(o_k = 0) \int_{a_j} \nu(a_j)\mu(a_j)$$

Photoconsistent voxels behind free voxels are *likely* occupied.

Voxels behind photo-consistent voxels *may be* occupied.

Ray-to-Occupancy Message

$$\mu_{\psi_r \to o_i}(o_i = 1) = \sum_{\mathbf{o}_{j \neq i}} \int_{\mathbf{a}} \psi_r(\mathbf{o}, \mathbf{a}) \prod_{j \neq i} \mu(o_j) \prod_j \mu(a_j)$$

$$\vdots$$

$$= \prod_{k=1}^{i-1} \mu(o_k = 0) \int_{a_i} \nu(a_i)\mu(a_i)$$

$$+ \sum_{j=1}^{i-1} \mu(o_j = 1) \prod_{k=1}^{j-1} \mu(o_k = 0) \int_{a_j} \nu(a_j)\mu(a_j)$$

- Can be computed in linear time (see paper for derivation)
- Similar calculation for $\mu_{\psi_r \to o_i}(o_i = 0)$

Ray-to-Appearance Message

$$\mu_{\psi_r \to a_i}(a_i) = \underbrace{\sum_{j \neq i} \mu(o_j = 1) \prod_{k<j} \mu(o_k = 0) \int_{a_j} \nu(a_j)\mu(a_j)}_{\text{Constant}}$$

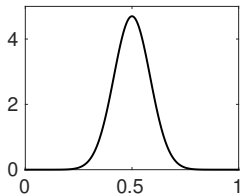
$$+ \underbrace{\mu(o_i = 1) \prod_{k<i} \mu(o_k = 0)}_{\text{Weight}} \times \underbrace{\nu(a_i)}_{\text{Gaussian}}$$

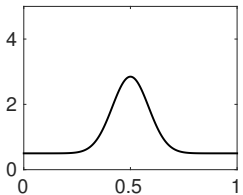
- Can also be computed in linear time (see paper for derivation)
- Constant: measures how well voxels $j \neq i$ explain evidence
- Weight: measures how likely voxel i is the first occupied voxel
- Gaussian: measures photoconsistency of a_i wrt. the image

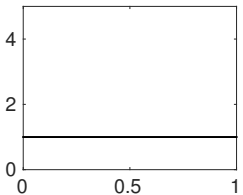
Ray-to-Appearance Message

$$\mu_{\psi_r \to a_i}(a_i) = \underbrace{\sum_{j \neq i} \mu(o_j = 1) \prod_{k<j} \mu(o_k = 0) \int_{a_j} \nu(a_j)\mu(a_j)}_{\text{Constant}}$$

$$+ \underbrace{\mu(o_i = 1) \prod_{k<i} \mu(o_k = 0)}_{\text{Weight}} \times \underbrace{\nu(a_i)}_{\text{Gaussian}}$$







Ray-to-Appearance Message

$$\mu_{\psi_r \to a_i}(a_i) = \underbrace{\sum_{j \neq i} \mu(o_j = 1) \prod_{k<j} \mu(o_k = 0) \int_{a_j} \nu(a_j)\mu(a_j)}_{\text{Constant}}$$

$$+ \underbrace{\mu(o_i = 1) \prod_{k<i} \mu(o_k = 0)}_{\text{Weight}} \times \underbrace{\nu(a_i)}_{\text{Gaussian}}$$

- **Integrals:** Monte Carlo approximation
- **Messages:** Mixture-of-Gaussian representation:

$$\mu_{a_v \to \psi_r}(a_v) = \prod_{r' \in \mathcal{R}_v \setminus r} \mu_{\psi_{r'} \to a_v}(a_v)$$

Inference

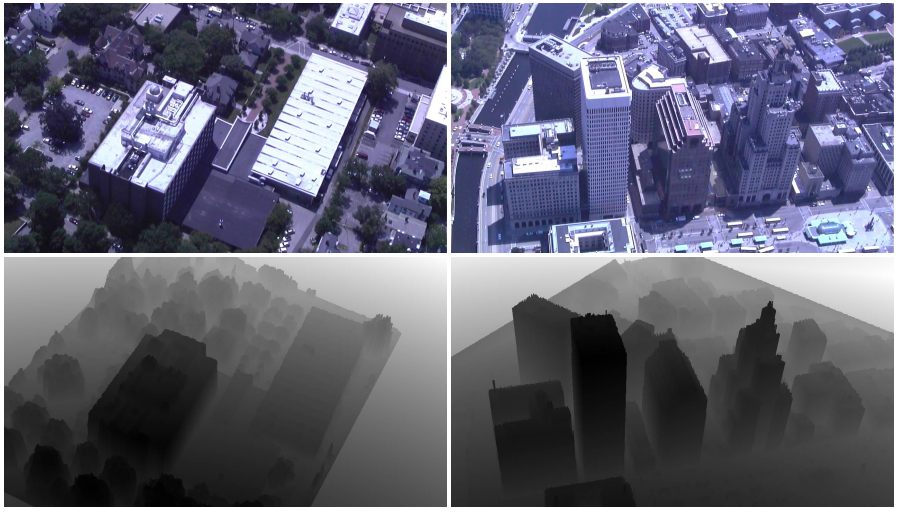
Challenges:

1. MRF comprises discrete and continuous variables
2. Ray potentials are high-order
3. Each pixel defines a factor

Our Approach:

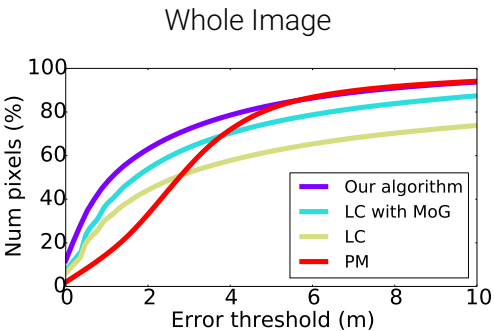
1. Approximate continuous belief propagation
   $\Rightarrow$ Update MoG's via importance sampling
2. Messages can be calculated in linear time
   $\Rightarrow$ Exact (but technical) derivation of messages
3. Octree implementation & GPGPU parallelization

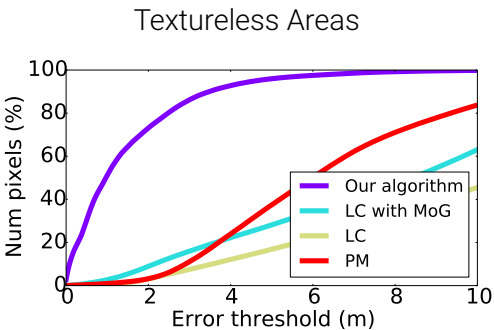
Experimental Results



[Restrepo *et al*., ISPRS 2014]

Quantitative Results

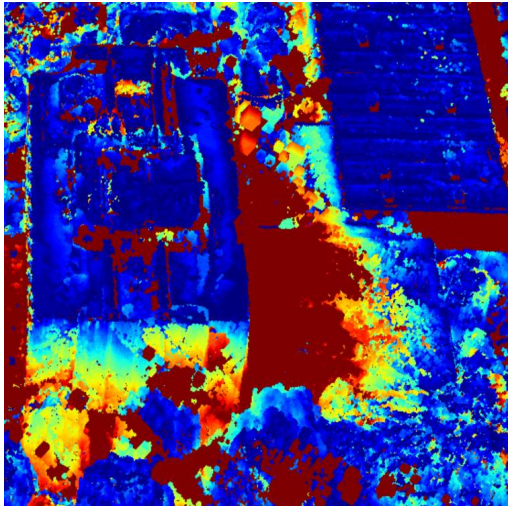




Qualitative Results

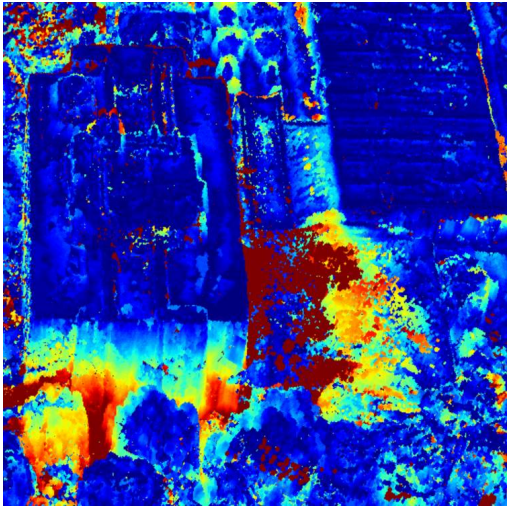
Input Image

Qualitative Results



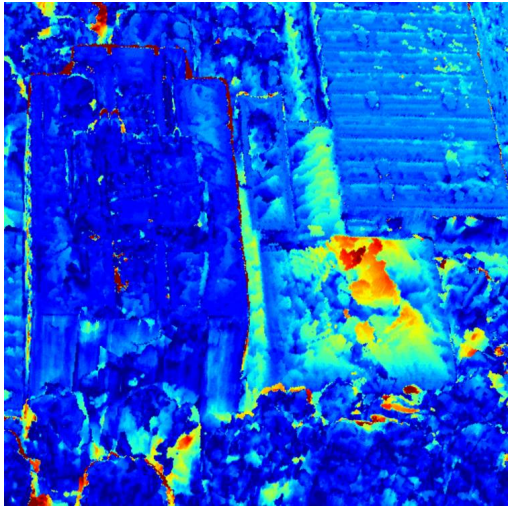
Error Map (Liu & Cooper)

Qualitative Results



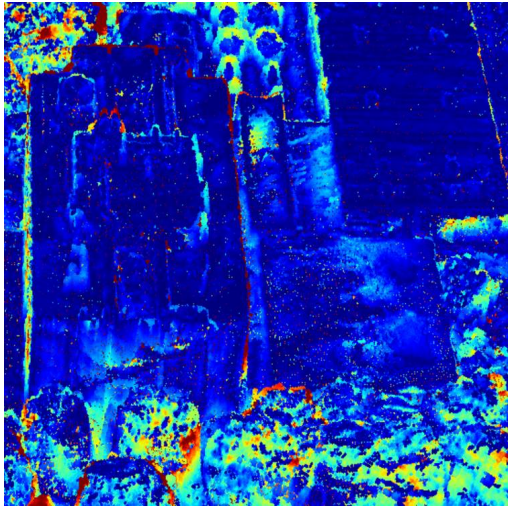
Error Map (Liu & Cooper with MoG)

Qualitative Results



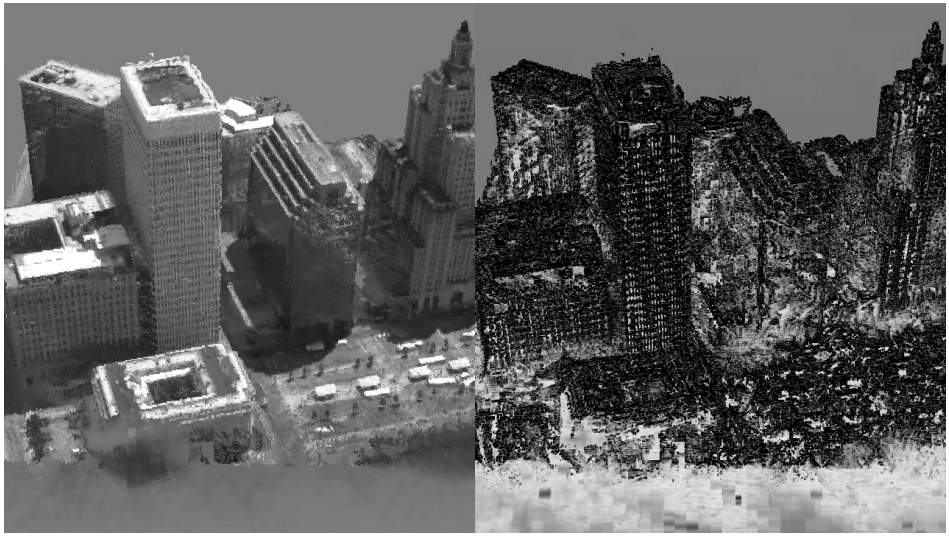
Error Map (Pollard & Mundy)

Qualitative Results

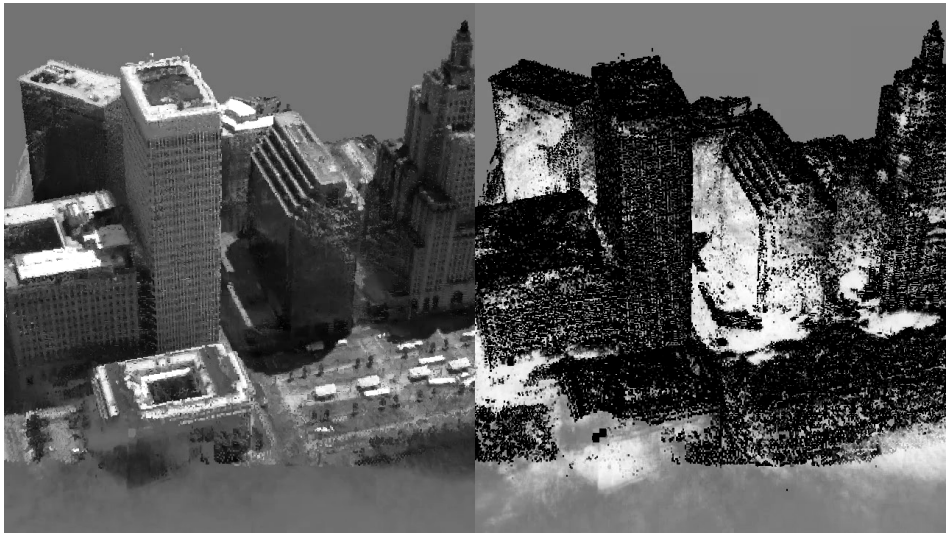


Error Map (Our Approach)

Qualitative Results: Pollard & Mundy

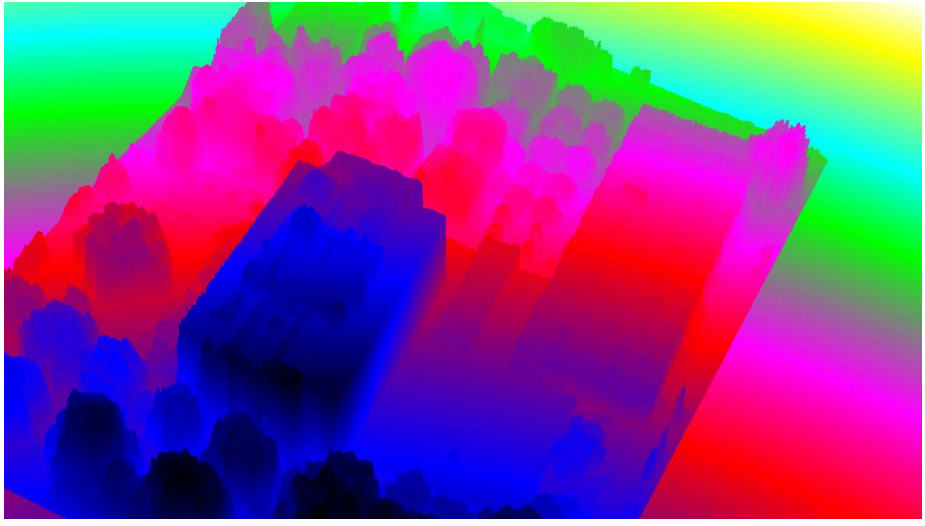


Qualitative Results: Our Results



Qualitative Results: Our Results

Code Available on GitHub

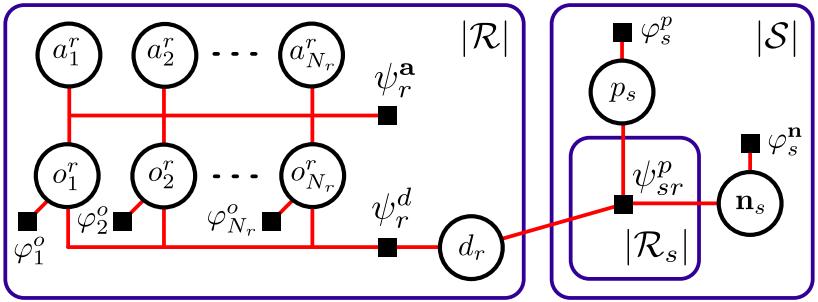


www.github.com/aliosmanulusoy

Patches, Planes and Probabilities: A non-local Prior for 3D Reconstruction

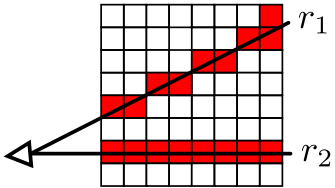
[Ulusoy, Black & Geiger, CVPR 2016]

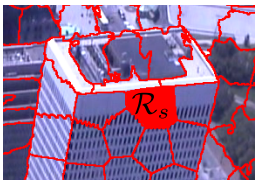
Probabilistic Model



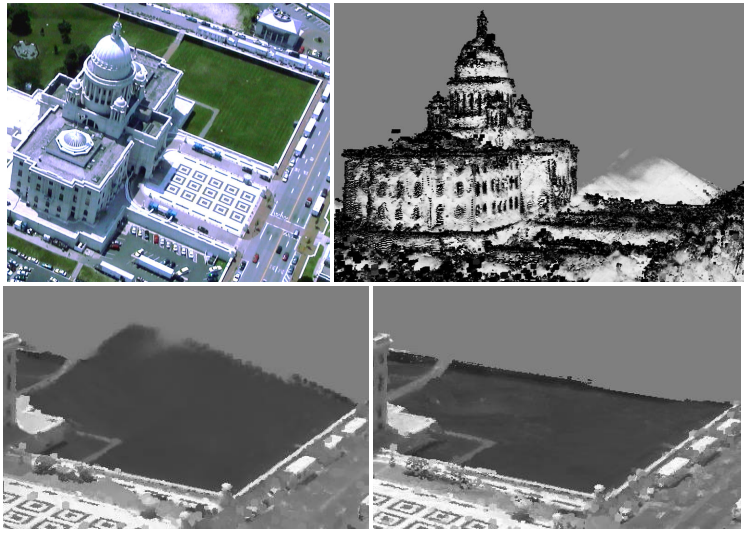
Ray Potentials & Occupancy Potentials

Patch Prior





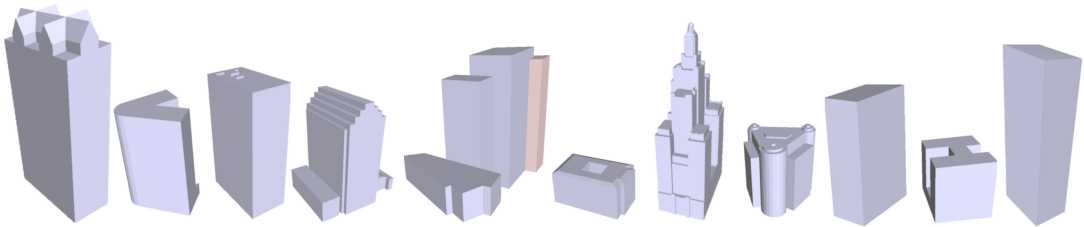
Results



Semantic Multi-view Stereo: Jointly Estimating Objects and Voxels

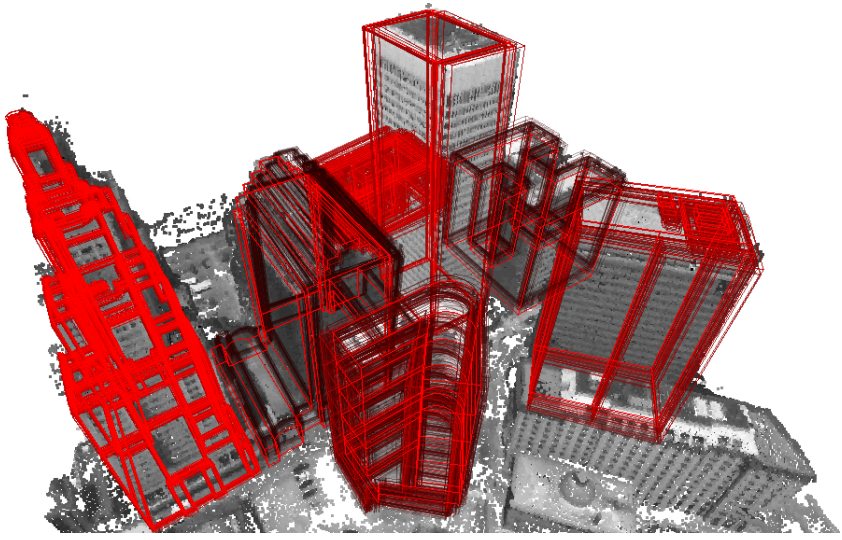
[Ulusoy, Black & Geiger, CVPR 2017]

3D Shape Priors

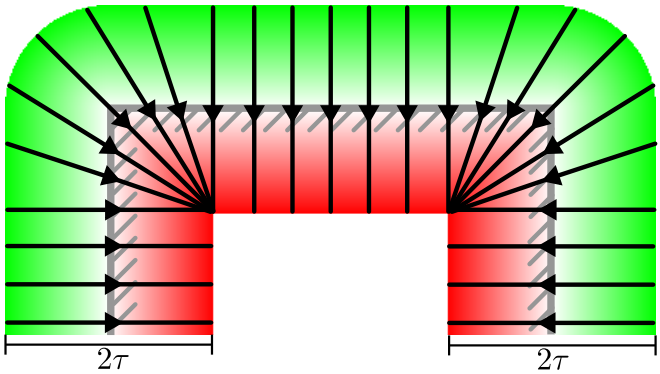


- For many scenes exclusive prior knowledge is available!
  - GPS tags can be used to retrieve 3D Warehouse models
  - 3D models of IKEA furniture for indoor scenes
- Challenges:
  - Often only coarse and inaccurate models
  - Unknown orientation and approximate location
  - Occlusions and object size

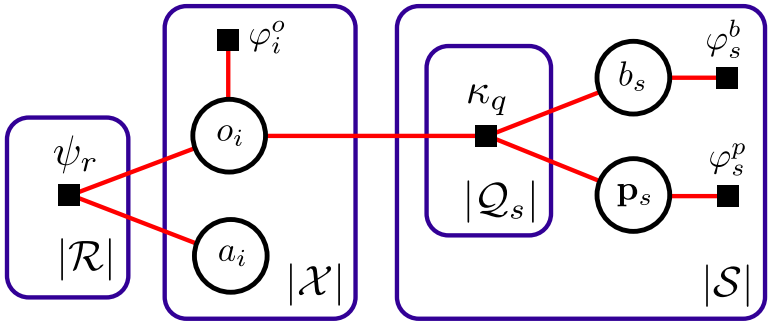
Probabilistic Model Fitting and 3D Reconstruction



Raylet Potentials



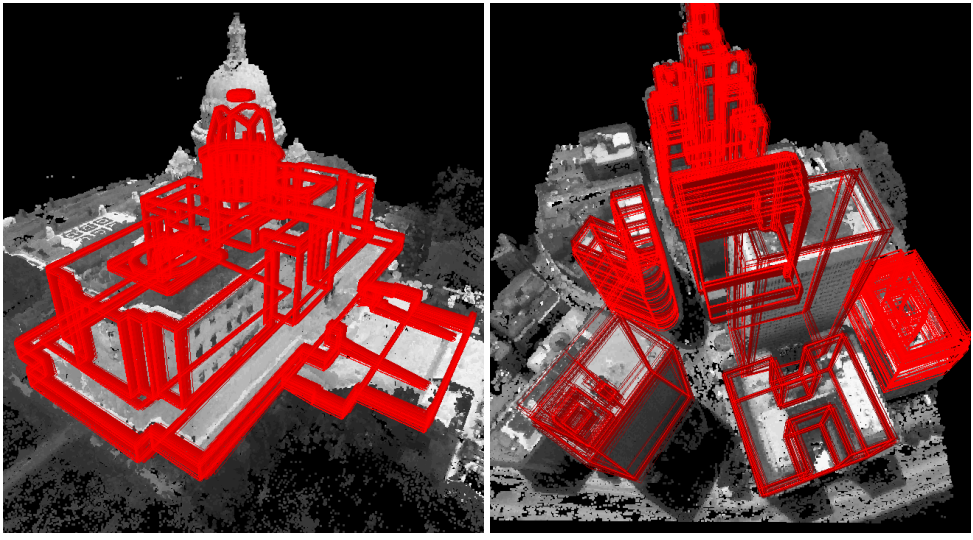
Probabilistic Model



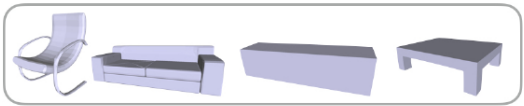
o_i/a_i: occupancy/appearance $b_s/\mathbf{p}_s$: shape model presence/pose

$\mathcal{R}$: #rays $\mathcal{X}$: #voxels $\mathcal{Q}_s$: #raylets $\mathcal{S}$: #shape models

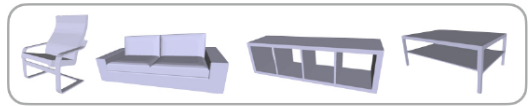
Results: Outdoor



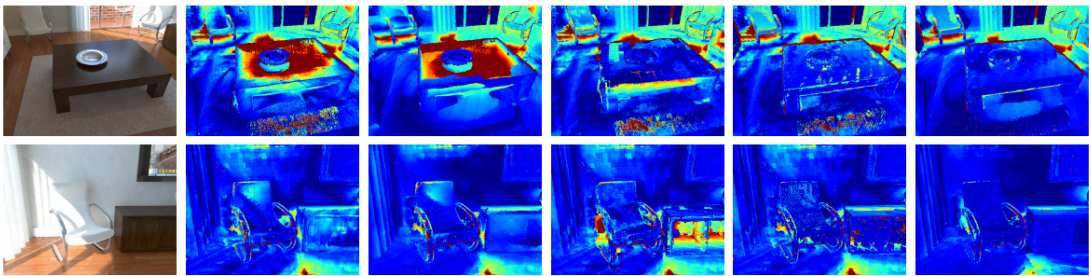
Results: Indoor



(a) Correct shape models for LIVINGROOM.



(b) Approximate shape models from IKEA [23].



(c) Ref image (d) No prior (e) Planarity prior (f) IKEA prior (g) Correct prior (h) Object+Planarity

Multi-View Reconstruction Summary

Pros:

- Probabilistic formulation is tractable as ray factors decompose
- Non-local constraints via joint inference in 2D and 3D
- CAD priors can help disambiguate textureless regions
- Using octrees reconstruction up to 1024^3 voxels possible

Cons:

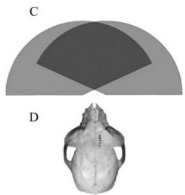
- Only approximate inference possible (highly loopy)
- Relatively slow: several minutes per scene on a GPU
- Appearance term too simplistic and not robust
- Resolution limited to discretization into voxels (as opposed to meshes)

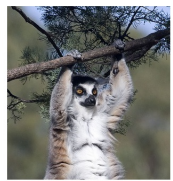
Optical Flow

Stereo vs. Optical Flow

Stereo

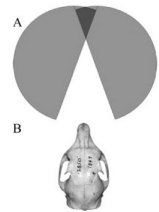
- 2 images at same time
- Only camera motion
- 1D estimation problem
- Monkeys



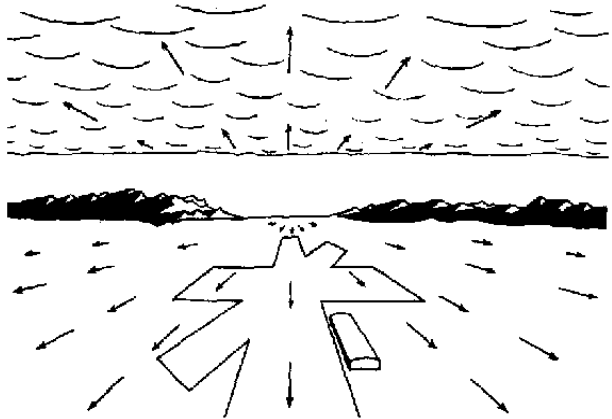


Optical Flow

- 2 images at 2 time steps
- Camera and object motion
- 2D estimation problem
- Squirrels

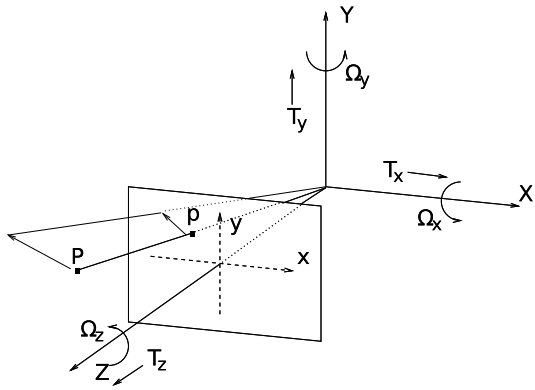


Optical Flow



[J. J. Gibson, 1950: The Ecological Approach to Visual Perception]

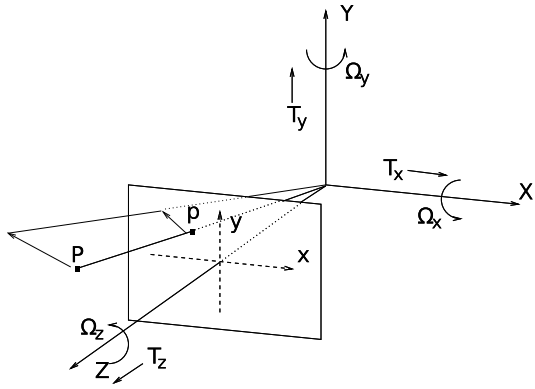
Optical Flow



Motion field:

- 2D motion field representing the **projection of the 3D motion** of points in the scene onto the image plane
- Can be the result of camera motion or object motion (or both)!

Optical Flow



Optical flow:

- 2D velocity field describing the **apparent motion** in the image (*i.e.*, the displacement of pixels looking “similar”)
- Optical flow $\neq$ motion field! Why?

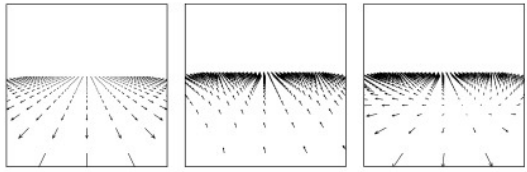
Thought Experiment

- Lambertian ball
  rotating in 3D
- What does the 2D
  motion field look like?
- What does the 2D
  optical flow field look like?

Thought Experiment

- Stationary specular ball
  moving light source
- What does the 2D
  motion field look like?
- What does the 2D
  optical flow field look like?

Optical Flow Field



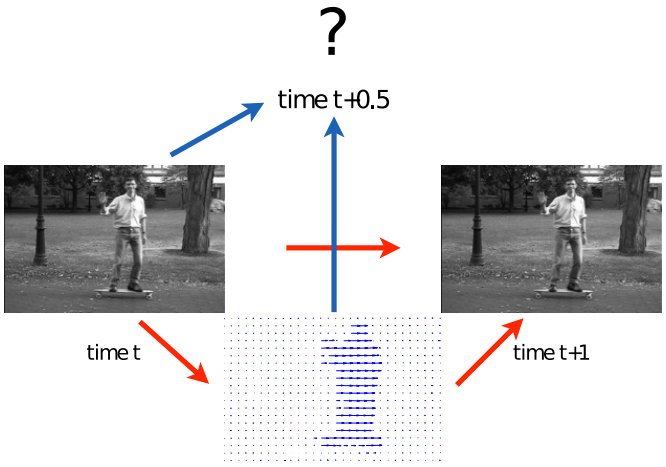
Optical flow fields tell us something (maybe ambiguous) about:

- The **3D structure** of the world
- The **motion of objects** in the viewing area
- The **motion of the observer** (if any)

In contrast to stereo:

- No epipolar geometry $\Rightarrow$ 2D estimation problem!

Applications: Video Interpolation / Frame Rate Adaption

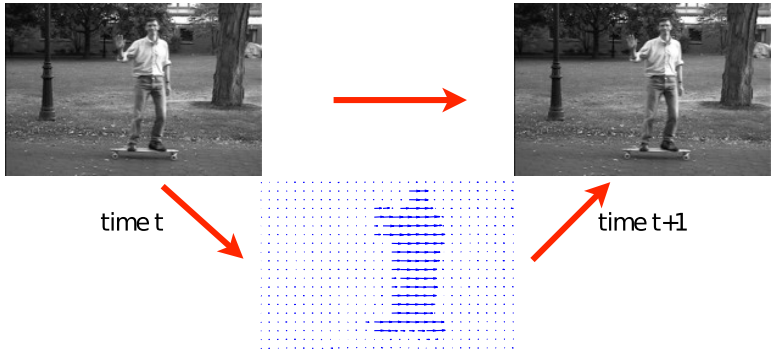


- If we know the image motion we can compute images at intermediate frames

Applications: Video Interpolation / Frame Rate Adaption

- If we know the image motion we can compute images at intermediate frames

Applications: Video Compression

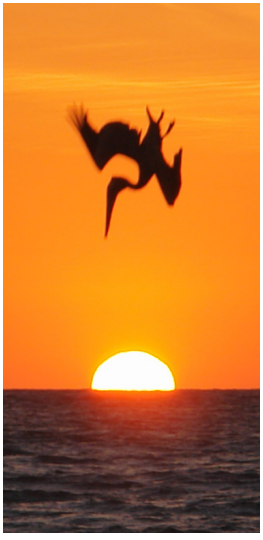


- To compress an image sequence, we can predict new frames using the optical flow field and only store how to “fix” the prediction
- Flow fields are smooth, thus easier to compress/store than images!

The Northern Gannet

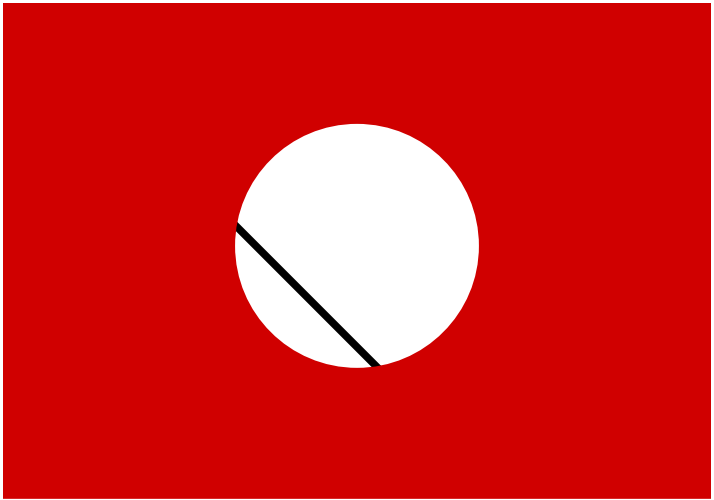
The Northern Gannet

The Northern Gannet



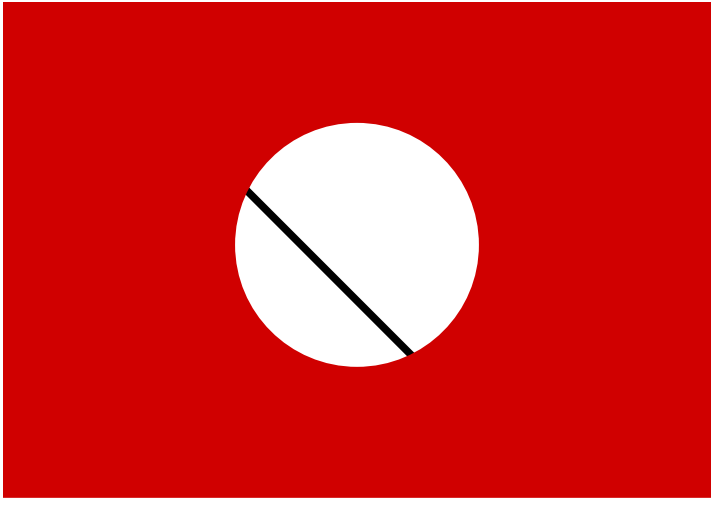
Aperture Problem

In which direction does the line move?



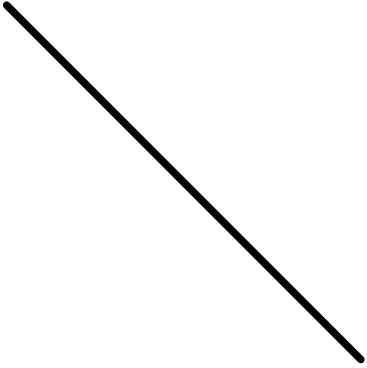
Aperture Problem

In which direction does the line move?



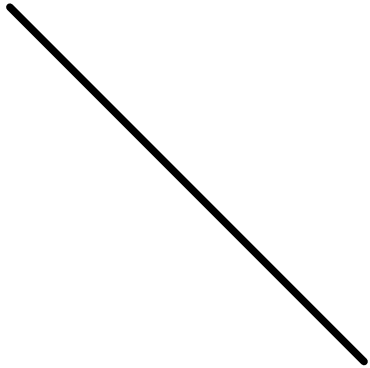
Aperture Problem

Now the full picture …



Aperture Problem

Now the full picture …



Aperture Problem

- Barber Pole: What is the motion field? What is the optic flow field?

Determining Optical Flow

[Horn & Schunck, Artificial Intelligence 1981]

Horn-Schunck Optical Flow

Horn-Schunck Model:

- Consider the image I as a function of continuous variables x, y, t
- Consider $u(x, y)$ and $v(x, y)$ as continuous flow fields
- Goal: Minimizing the following energy functional

$$
\begin{aligned}
E(u, v) &= \iint \underbrace{\left(I(x + u(x, y), y + v(x, y), t + 1) - I(x, y, t)\right)^2}_{\text{quadratic penalty for brightness change}} \\
&+ \lambda \cdot \underbrace{\left(\|\nabla u(x, y)\|^2 + \|\nabla v(x, y)\|^2\right)}_{\text{quadratic penalty for flow change}} \, dx \, dy
\end{aligned}
$$

with regularization parameter λ and $\nabla = \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}\right)$.

Horn-Schunck Optical Flow

$$\begin{aligned} E(u,v) &= \iint \left(I(x+u(x,y),y+v(x,y),t+1) - I(x,y,t)\right)^2 \\ &+ \lambda \cdot \left(\|\nabla u(x,y)\|^2 + \|\nabla v(x,y)\|^2\right) dx\, dy \end{aligned}$$

- Minimizing this directly is a hard problem because the energy is highly non-convex and has many local optima
- Solution: linearize the brightness constancy assumption

$$\begin{aligned} E(u,v) &= \iint \left(I_x(x,y,t)u(x,y) + I_y(x,y,t)v(x,y) + I_t(x,y,t)\right)^2 \\ &+ \lambda \cdot \left(\|\nabla u(x,y)\|^2 + \|\nabla v(x,y)\|^2\right) dx\, dy \end{aligned}$$

Horn-Schunck Optical Flow

$$
\begin{aligned}
E(u,v) &= \iint \left(I_x(x,y,t)u(x,y) + I_y(x,y,t)v(x,y) + I_t(x,y,t)\right)^2 \\
&+ \lambda \cdot \left(\|\nabla u(x,y)\|^2 + \|\nabla v(x,y)\|^2\right) dx\, dy
\end{aligned}
$$

- Imposes quadratic penalty on optical flow constraint and gradient of flow field
- This energy is convex and thus has a unique optimum
- The flow can be estimated by discretizing it spatially
  and performing gradient descent on the discretized objective:

$$
\begin{aligned}
E(\mathbf{U},\mathbf{V}) &= \sum_{x,y} \left(I_x(x,y)\, u_{x,y} + I_y(x,y)\, v_{x,y} + I_t(x,y)\right)^2 \\
&+ \lambda \cdot \big((u_{x,y} - u_{x+1,y})^2 + (u_{x,y} - u_{x,y+1})^2 + \\
&\qquad (v_{x,y} - v_{x+1,y})^2 + (v_{x,y} - v_{x,y+1})^2\big)
\end{aligned}
$$

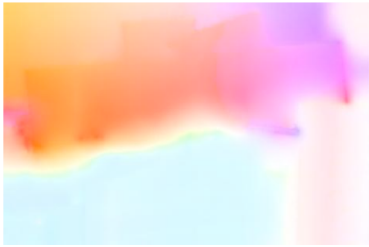
Horn-Schunck Optical Flow

Discretized Objective:

$$
\begin{aligned}
E(\mathbf{U},\mathbf{V}) &= \sum_{x,y} \left(I_x(x,y)\,u_{x,y} + I_y(x,y)\,v_{x,y} + I_t(x,y)\right)^2 \\
&+ \lambda \cdot \big((u_{x,y} - u_{x+1,y})^2 + (u_{x,y} - u_{x,y+1})^2 + \\
&\qquad (v_{x,y} - v_{x+1,y})^2 + (v_{x,y} - v_{x,y+1})^2\big)
\end{aligned}
$$

- Differentiate wrt. $\mathbf{U}, \mathbf{V}$ and set the gradient to 0
- Results in a huge (but sparse) linear system
- Can be solved using standard techniques (*e.g.*, Gauss-Seidel, SOR)
- What would happen for $\lambda = 0$?
- However: Linearization works only for small motions!
  - Coarse-to-fine estimation & warping

Results of Horn & Schunck



Horn & Schunck Optical Flow

- Our HS results are quite a bit better than the results using LK
- However, the flow is very smooth, *i.e.*, to overcome ambituities we need to set λ to a high value which oversmooths flow discontinuities
- Why?
- We use a quadratic penalty for penalizing changes in the flow

A Framework for Robust Estimation of Optical Flow

[Black & Anandan, ICCV 1993]

Probabilistic Interpretation

Optimization problem can be interpreted as MAP inference in MRF:

$$p(x) = \frac{1}{Z} \exp\{-E(x)\}$$

Gibbs energy: $E(\mathbf{U}, \mathbf{V}) = \sum_{x,y} \left(I_x(x,y)\, u_{x,y} + I_y(x,y)\, v_{x,y} + I_t(x,y)\right)^2$

$$+ \quad \lambda \cdot \big((u_{x,y} - u_{x+1,y})^2 + (u_{x,y} - u_{x,y+1})^2 + (v_{x,y} - v_{x+1,y})^2 + (v_{x,y} - v_{x,y+1})^2\big)$$

Gibbs distribution (prior and likelihood):

$$p(\mathbf{U}, \mathbf{V}) \propto \prod_{x,y} \exp\left\{-\left(I_x(x,y)\, u_{x,y} + I_y(x,y)\, v_{x,y} + I_t(x,y)\right)^2\right\}$$

$$\times \quad \exp\left\{-\lambda(u_{x,y} - u_{x+1,y})^2\right\} \times \exp\left\{-\lambda(u_{x,y} - u_{x,y+1})^2\right\}$$

$$\times \quad \exp\left\{-\lambda(v_{x,y} - v_{x+1,y})^2\right\} \times \exp\left\{-\lambda(v_{x,y} - v_{x,y+1})^2\right\}$$

Robust Regularization

- Quadratic penalties translate to Gaussian distributions for prior and likelihood:

$$\begin{aligned} p(\mathbf{U},\mathbf{V}) \quad &\propto \quad \prod_{x,y} \exp\left\{-\left(I_x(x,y)\,u_{x,y} + I_y(x,y)\,v_{x,y} + I_t(x,y)\right)^2\right\} \\ &\times \quad \exp\left\{-\lambda(u_{x,y} - u_{x+1,y})^2\right\} \times \exp\left\{-\lambda(u_{x,y} - u_{x,y+1})^2\right\} \\ &\times \quad \exp\left\{-\lambda(v_{x,y} - v_{x+1,y})^2\right\} \times \exp\left\{-\lambda(v_{x,y} - v_{x,y+1})^2\right\} \end{aligned}$$

- Both assumptions are invalid in practice (occlusions/discontinuities)
- Formulation with robust data term and smoothness penalties:

$$\begin{aligned} p(\mathbf{U},\mathbf{V}) \quad &\propto \quad \prod_{x,y} \exp\left\{-\rho_D\left(I_x(x,y)\,u_{x,y} + I_y(x,y)\,v_{x,y} + I_t(x,y)\right)\right\} \\ &\times \quad \exp\left\{-\lambda\,\rho_S(u_{x,y} - u_{x+1,y})\right\} \times \exp\left\{-\lambda\,\rho_S(u_{x,y} - u_{x,y+1})\right\} \\ &\times \quad \exp\left\{-\lambda\,\rho_S(v_{x,y} - v_{x+1,y})\right\} \times \exp\left\{-\lambda\,\rho_S(v_{x,y} - v_{x,y+1})\right\} \end{aligned}$$

Robust Regularization

Formulation with robust data term and smoothness penalties:

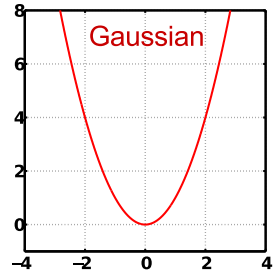
$$\begin{aligned} p(\mathbf{U},\mathbf{V}) \quad &\propto \quad \prod_{x,y} \exp\left\{-\rho_D\left(I_x(x,y)\,u_{x,y} + I_y(x,y)\,v_{x,y} + I_t(x,y)\right)\right\} \\ &\times \quad \exp\left\{-\lambda\,\rho_S(u_{x,y} - u_{x+1,y})\right\} \times \exp\left\{-\lambda\,\rho_S(u_{x,y} - u_{x,y+1})\right\} \\ &\times \quad \exp\left\{-\lambda\,\rho_S(v_{x,y} - v_{x+1,y})\right\} \times \exp\left\{-\lambda\,\rho_S(v_{x,y} - v_{x,y+1})\right\} \end{aligned}$$

- But how to choose $\rho_D(\cdot)$ and $\rho_S(\cdot)$?
- We want a prior that allows for discontinuities in the optical flow field and a likelihood that allows for occlusions in the photoconsistency term
- Thus, we need something more heavy-tailed than a Gaussian distribution,*e.g.*, a Student-t distribution (Lorentzian penalty):

$$p(x) \propto \left(1 + \frac{x^2}{2\sigma^2}\right)^{-\alpha} \quad \Rightarrow \quad \rho(x) = ? - \log\left(p(x)\right) = \alpha \log\left(1 + \frac{x^2}{2\sigma^2}\right)$$

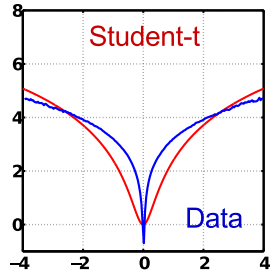
Robust Regularization

- Student-t distribution / Lorentzian penalty:



$$p(x) \propto \left(1 + \frac{x^2}{2\sigma^2}\right)^{-\alpha}$$

negative
log-density
(i.e. energy)



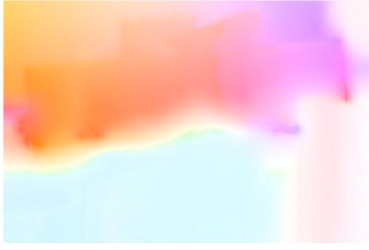
$$\rho(x) = \alpha \log\left(1 + \frac{x^2}{2\sigma^2}\right)$$

- Has been proposed in [Black-Anandan, ICCV 1993]
- How to estimate the parameters? Learn them from data!

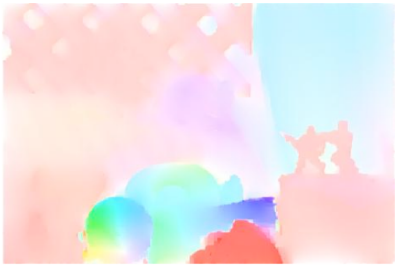
Robust Regularization

- How to obtain optical flow ground truth data?
- There is no sensor which can measure optical flow directly
- Synthesize optical flow fields using
  - A set of natural geometries (*e.g.*, Brown range database)
  - A set of natural camera motions

Result of [Horn-Schunck, AI 1981]



Results of [Black-Anandan, ICCV 1993]



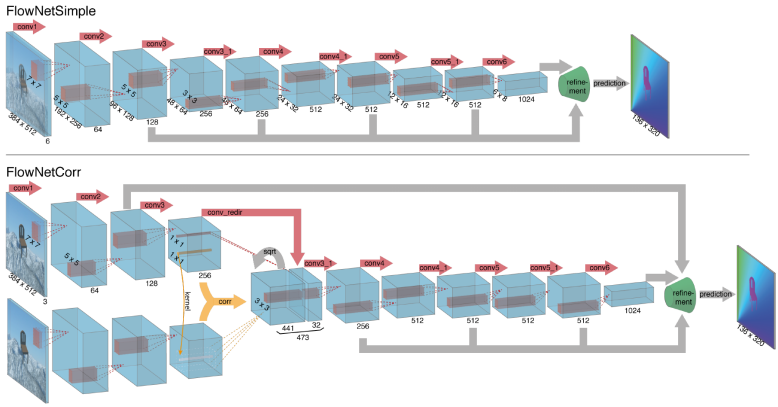
Results of [Sun *et al.*, CVPR 2010]

Deep Learning for Optical Flow Estimation

FlowNet: Learning Optical Flow with Convolutional Networks

[Dosovitskiy *et al.*, ICCV 2015]

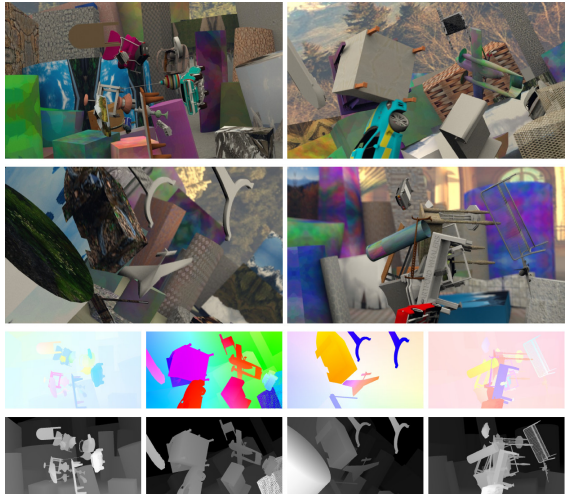
FlowNet



[Dosovitskiy *et al*., ICCV 2015]

- Encoder: Convolution with stride, decoder: Upconvolution, skip-connections
- Multi-scale loss (EPE in pixels), curriculum learning, synth. training data

FlowNet – Synthetic Datasets



Flying Things

FlowNet – Synthetic Datasets

Monkaa

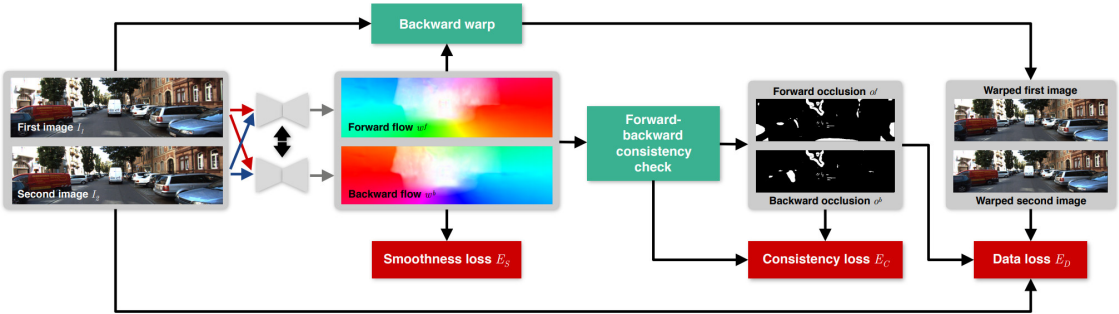
FlowNet – Synthetic Datasets

MPI Sintel

UnFlow: Unsupervised Learning of Optical Flow with a Bidirectional Census Loss

[Meister, Hur & Roth, AAAI 2018]

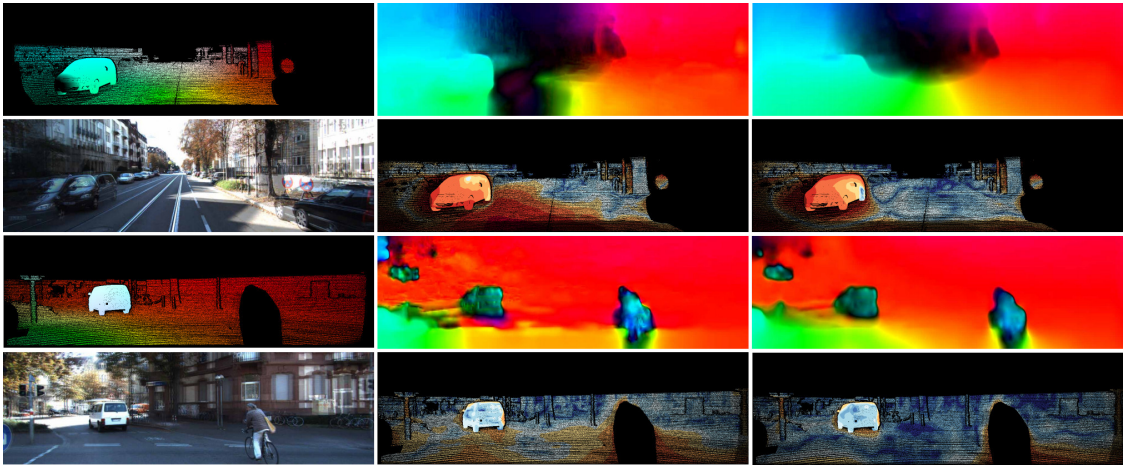
UnFlow: Unsupervised Optical Flow



[Meister *et al.*, AAAI 2018]

- Learn optical flow without supervision by warping images into the other frame
- Recycles photometric/smoothness terms as loss functions for NN training

UnFlow: Unsupervised Optical Flow – Results



Optical Flow Summary

- Classical OF approaches state-of-the-art until 2016
- DL based methods on par or better since 2017
- But require
  - Big models
  - Enormous amount of (synthetic) training data
  - GPU compute time
  - Sophisticated curriculum learning schedules
- Classical methods still best on related scene flow task
- Top performing DL methods borrow many elements from classical methods (*e.g.*, warping, cost volume, coarse-to-fine estimation, loss functions)

Questions?