Semantic 3D Modeling

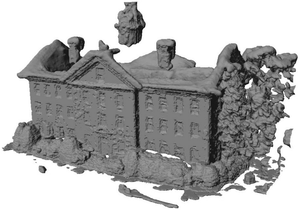
CVPR 2016 Tutorial

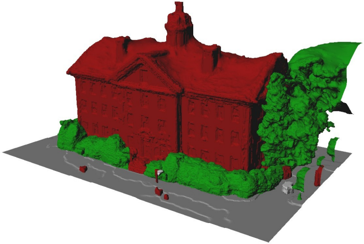
Christian Häne

Computer Vision and Geometry Group, ETHZ

June 2016

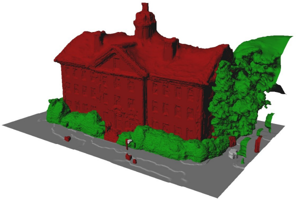
Semantic 3D Modeling

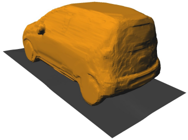


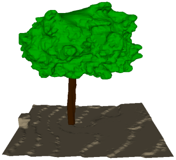


- Semantic 3D Modeling
  - Dense 3D reconstruction
  - Semantic segmentation
- Joint formulation is beneficial
- Richer representation
  - Volume of just the building?
  - Remove vegetation
  - Where are the streets?
  - Are there cars?

Outline







Outline

Motivation

- Input: Set of images

Figure : Some example images

- Goal: Dense 3D reconstruction

Sparse 3D Reconstruction

- Camera locations
- Set of sparse 3D points

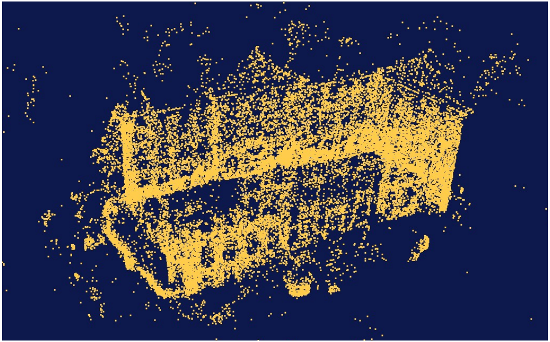


Figure : Sparse point cloud

Dense Stereo Matching

- Depth map for each of the images
- Set of nearby images for matching

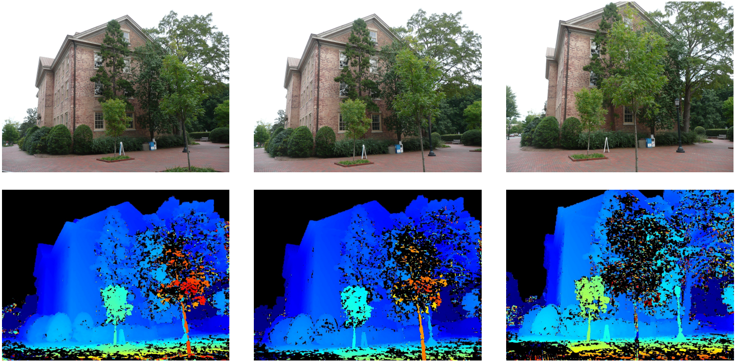
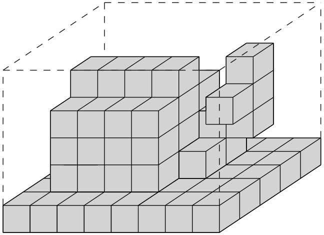
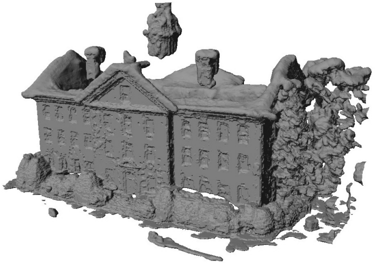


Figure : Some example depth maps (red close, blue far)

Volumetric Depth Map Fusion

- Segmentation of a voxel space Ω
  into free and occupied space: $u(s) \in [0, 1]$
  [Curless and Levoy 1996]
- TV-Flux Fusion [Lempitsky and Boykov 2007, Zach 2008]
- Idea: convex energy
  - Unary term: Computed from depth maps
  - Smoothness term: Penalization of surface area





TV-Flux Fusion: Energy

Energy Minimization

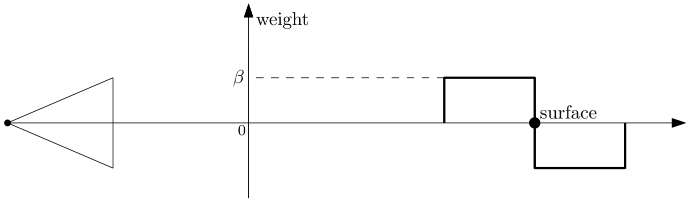
$$E(u) \quad = \int_{\Omega} \left(\rho(s)u(s) + \|(\nabla u)(s)\|_2\right) ds$$

$$\text{s.t.} \;\; u(s) \in [0, 1]$$

- u indicates occupied space
- Total variation (TV) $\|(\nabla u)(s)\|_2$
  - Penalizes surface area [e.g. Chambolle et al. 2010]
- Unary data term: $\rho(s)$
- Optimization: next part of the tutorial

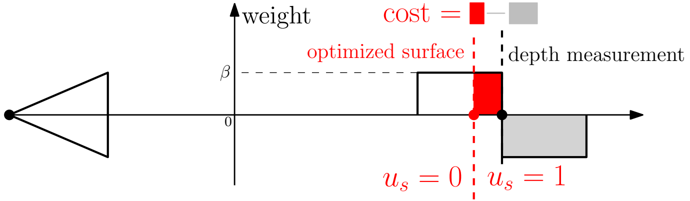
TV-Flux Fusion: Data Term

- Computed from depth maps
- ρ_s: sum of all the weights per voxel



TV-Flux Fusion: Data Term

- Computed from depth maps
- ρ_s: sum of all the weights per voxel



TV-Flux Fusion: Some Results

Figure : Some volumetric fusion results

Anisotropic Minimal Surfaces [Kolev et al. 2010]

- Surface normals estimated from the data used for regularization
- Helps to preserve surface details

Pictures taken from [K. Kolev, T. Pock and D. Cremers ECCV 2010]

- Isotropic smoothness changed to anisotropic version

$$\|(\nabla u)(s)\|_2 \quad \rightarrow \quad \sqrt{(\nabla u)(s)^T D(s)(\nabla u)(s)}$$

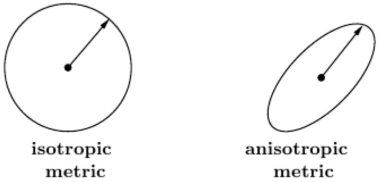


Figure taken from [K. Kolev, T. Pock and D. Cremers ECCV 2010]

Anisotropic Regularization: Results [Kolev et al. 2010]

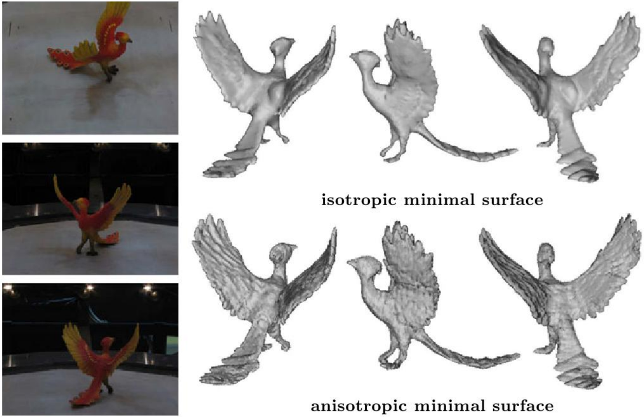


Figure taken from [K. Kolev, T. Pock and D. Cremers ECCV 2010]

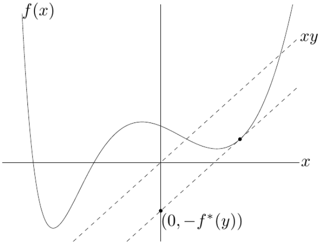
Volumetric 3D Reconstruction Recap

- Pose 3D reconstruction as volumetric segmentation problem
  - Free and occupied space
- Solve optimization problem
- Different data and regularization terms
  - Two examples presented
  - TV-Flux [Zach 2008]
  - Anisotropic regularization [Kolev et al. 2010]
- Many more volumetric reconstruction formulations exist
  - Without regularization:
    - Laser scans [Curless and Levoy 1996]
    - Kinect depth maps [Newcombe et al. 2011]
  - Data adaptive discretization of space [Labatut et al. 2007, Vu et al. 2009]

Outline

Prerequisites [Boyd and Vandenberghe 2004, Browein and Lewis 1999]

- Convex conjugate
  - $F : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$
  - $F^* : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$, convex conjugate of F
  - $F^*(y) = \sup_y \{y^T x - F(x)\}$
  - $F^{**} = F$ iff F is convex and lower semi-continuous (Fenchel-Moreau Theorem)



Boyd and Vandenberghe 2004: “The conjugate function $f^*(y)$ is the maximum gap between the linear function yx and $f(x)$, as shown by the dashed line in the figure.”

Figure from [Boyd and Vanddenberghe 2004]

Prerequisites [Boyd and Vandenberghe 2004, Browein and Lewis 1999]

- Convex conjugate
  - $F : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{\infty\}$
  - $F^* : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{\infty\}$, convex conjugate of F
  - $F^*(y) = \sup_y \{y^T x - F(x)\}$
  - $F^{**} = F$ iff F is convex and lower semi-continuous (Fenchel-Moreau Theorem)
- Indicator function
  - Let C be a nonempty subset of $\mathbb{R}^n$

$$\imath_C(x) = \begin{cases} 0, \text{if } x \in C \\ \infty, \text{if } x \notin C \end{cases}$$

- Proximity operator (generalization of projection)

$$\mathrm{prox}_{\tau F}(y) = \arg\min_x \left\{ \frac{\|x-y\|_2^2}{2\tau} + F(x) \right\}$$

  - Let D be a nonempty closed convex subset of $\mathbb{R}^n$
  - $\mathrm{prox}_{\imath_D}$, projection to D

First-Order Primal-Dual Algorithm I

[Chambolle and Pock 2010, Pock and Chambolle 2011]

- Optimization problems of the form (primal)

$$\min_{z \in \mathbb{R}^n} F(Kz) + G(z)$$

  - $K : \mathbb{R}^n \to \mathbb{R}^m$, linear operator
  - $G : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$, convex function
  - $F : \mathbb{R}^m \to \mathbb{R} \cup \{\infty\}$, convex function

- Primal-dual form

$$\min_{z \in \mathbb{R}^n} \max_{\xi \in \mathbb{R}^m} \langle Kz, \xi \rangle + G(z) - F^*(\xi)$$

  - $\langle \cdot, \cdot \rangle$, inner product
  - F^*, convex conjugate of F

First-Order Primal-Dual Algorithm II

[Chambolle and Pock 2010, Pock and Chambolle 2011]

- Optimization Problem

$$\min_{z \in \mathbb{R}^n} \max_{\xi \in \mathbb{R}^m} \langle Kz, \xi \rangle + G(z) - F^*(\xi))$$

- Initialization

$$(z^0, \xi^0) \in \mathbb{R}^n \times \mathbb{R}^m, \quad \bar{z}^0 = z^0, \quad \tau, \sigma > 0, \quad \theta \in [0, 1]$$

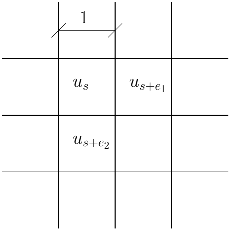
- Iteration

$$\begin{aligned} \xi^{n+1} &= \text{prox}_{\sigma F^*}(\xi^n + \sigma K \bar{z}^n) \\ z^{n+1} &= \text{prox}_{\tau G}(z^n - \tau K^T \xi^{n+1}) \\ \bar{z}^{n+1} &= z^{n+1} + \theta(z^{n+1} - z^n) \end{aligned}$$

- Convergence:
  - Globally convergent if $\theta = 1$ and $\tau\sigma\|K\|^2 < 1$

Application to Volumetric 3D Reconstruction

- Discretize space into regular grid
  Cell length $= 1$
- Gradient: forward differences



$$(\nabla u)_s = \begin{pmatrix} u_{s+e_1} - u_s \\ u_{s+e_2} - u_s \\ \vdots \end{pmatrix} \quad e_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \end{pmatrix} \quad e_2 = \begin{pmatrix} 0 \\ 1 \\ \vdots \end{pmatrix}$$

- Discretized energy

$$E(u) = \sum_{s \in \Omega} \left\{ \|(\nabla u)_s\|_2 + \rho_s u_s + \imath_{[0,1]}(u_s) \right\}$$

- Rewrite as primal-dual saddle point problem $\rightarrow$ primal-dual formulation for TV

$$\|(\nabla u)_s\|_2 = \max_{\|p_s\|_2 \leq 1} \langle (\nabla u)_s, p_s \rangle$$

Final Primal-Dual Saddle Point Problem

- General primal-dual problem

$$\min_{z \in \mathbb{R}^n} \max_{\xi \in \mathbb{R}^m} \langle Kz, \xi \rangle + G(z) - F^*(\xi))$$

- Primal-dual problem for TV-Flux

$$\min_u \max_p \sum_{s \in \Omega} \left\{ \langle (\nabla u)_s, p_s \rangle + \underbrace{\rho_s u_s + \imath_{[0,1]}(u_s)}_{G} - \underbrace{\imath_{\{\leq 1\}}\left(\|p_s\|_2\right)}_{F^*} \right\}$$

- Due to the structure prox operators can be applied point-wise
- $\text{prox}_{\tau G}(u_s) = \text{clamp}_{[0,1]}(u_s - \tau \rho_s)$
- $\text{prox}_{\sigma F^*(p_s)}$: Euclidean projection to unit ball

Example of Update in Code

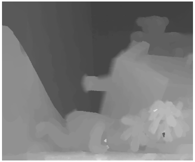
- Update for p variables, p contained in ξ^{n+1}
- $\xi^{n+1} = \text{prox}_{\sigma F^*}(\xi^n + \sigma K \bar{z}^n)$

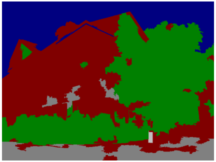
```
#pragma omp parallel for
for (int z = 0; z < depth; z++)     // loop
 for (int y = 0; y < height; y++)   // over
  for (int x = 0; x < width; x++)   // volume
    {
     // xi_n + sigma*K*z_bar
     const int X = x+1 < wdith ? x+1 : x;
     const int Y = y+1 < height ? y+1 : y;
     const int Z = z+1 < depth ? z+1 : z;
     p_w(x,y,z) += sigma*(u_bar(X,y,z) - u_bar(x,y,z));
     p_h(x,y,z) += sigma*(u_bar(x,Y,z) - u_bar(x,y,z));
     p_d(x,y,z) += sigma*(u_bar(x,y,Z) - u_bar(x,y,z));

     // prox_sigma*F^*
     project_to_unit_ball(&p_w(x,y,z), &p_h(x,y,z), &p_d(x,y,z));
    }
```

Outline

Labeling Problems

- Given a set of nodes (pixels, superpixels, voxels)
- Goal: Assign one out of L labels to each node
- Local preference per node plus regularization
  - Energy minimization problem
- Omni present in computer vision
- Most multi-label ($L > 2$) instances NP-hard



Approaches

- Different solution approaches
  - Graph-Cuts
  - Belief propagation
  - Convex relaxation
  - ...
- Christian: Convex relaxation
  - Discrete domain (graphical model)
  - Continuous domain
- Lubor: Graph-Cuts

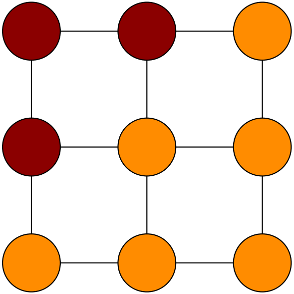
Default Procedure

- Define a (non-convex) energy that solves the problem
- Relax the problem to a convex one
- In case of a continuous formulation, discretize
- Solution of the relaxed problem is approximation to the original one

Discrete Domain

- Describe domain by a graph
  - For images, a node per pixel and edges to the neighbors

- Assign one out of L labels to each node
- θ_s^i: Cost for assigning label i at node s
- θ_{st}^{ij}: Cost for assigning i at s and j at t
- Find assignment that has minimal cost



LP Relaxation I

- LP relaxation by Schlesinger 1976, review Tomas Werner 2007
- Objective
$$\sum_{s,i} \theta_s^i x_s^i + \sum_{s,t} \sum_{i,j} \theta_{st}^{ij} x_{st}^{ij}$$
- θ_s^i and θ_{st}^{ij} cost for assigning a label or a transition
- $x_s^i \in \{0,1\}$ and $x_{st}^{ij} \in \{0,1\}$ exact solution but non-convex problem

Example for position r

$$x_r^{\cdot} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \begin{array}{l} \text{label 0} \\ \text{label 1} \\ \text{label 2} \end{array} \quad \rightarrow \quad \sum_i \theta_r^i x_r^i = \theta_r^1$$

- Indicator variables select data cost / smoothness cost

LP Relaxation II

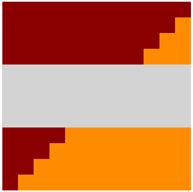
- LP relaxation by Schlesinger 1976, review Tomas Werner 2007

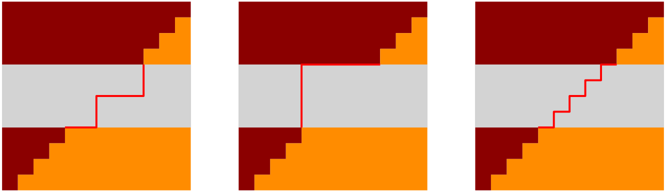
$$
\begin{aligned}
\min_x \quad & \sum_{s,i} \theta_s^i x_s^i + \sum_{s,t} \sum_{i,j} \theta_{st}^{ij} x_{st}^{ij} \\
\text{s.t.} \quad & x_s^i = \sum_j x_{st}^{ij} \qquad x_t^i = \sum_j x_{st}^{ji} \\
& \sum_i x_s^i = 1 \qquad x_s^i \geq 0 \qquad x_{st}^{ij} \geq 0 \qquad \forall s,t,i,j
\end{aligned}
$$

- θ_s^i and θ_{st}^{ij} cost for assigning a label or a transition
- $x_s^i \in \{0,1\}$ and $x_{st}^{ij} \in \{0,1\}$ exact solution but non-convex problem
- Relaxed to linear program $x_s^i \in [0,1]$ and $x_{st}^{ij} \in [0,1]$
- Label assignment through thresholding

Metrication artifacts

- Grid graph based representation
  - Smoothness cost
    - Measured by crossing edges
  - Inpainting example
  - Multiple equally good solutions

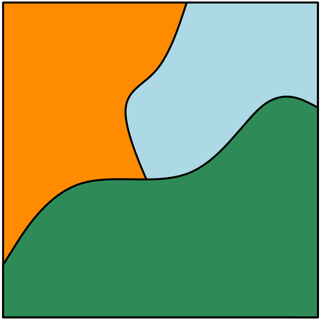




- Penalize true boundary length: Continuous formulation

Continuous Formulation I

- Domain continuous (e.g. image plane), label space discrete
- Domain segmented into areas that have one out of L labels assigned



Continuous Formulation II

- Continuous primal-dual saddle point energy
  [Chambolle et al. 2008, Lellmann and Schnörr 2010]
- Smoothness cost
  - Boundary length between labels i and j weighted by θ^{ij}
  - Needs to form a metric over the label-space

Figure from [Chambolle, Cremers, Pock SIIMS 2012]

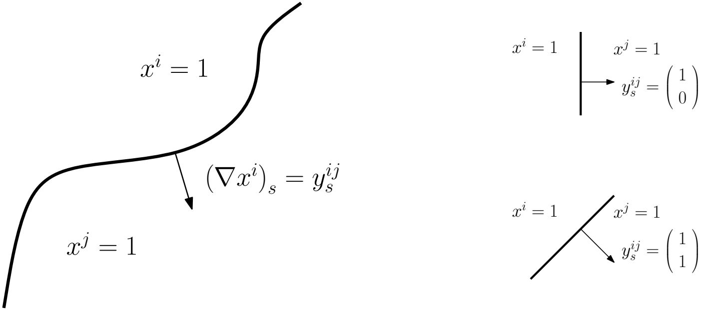
Continuously Inspired Formulation

- Our version: continuously inspired
  [Zach, Häne, Pollefeys, CVPR 2012, TPAMI 2014]
  - Discretized pure primal formulation
  - Extended to non-metric and anisotropic smoothness (more later)

$$
\begin{aligned}
\min_{x,y} \quad & \sum_{s,i} \theta_s^i x_s^i + \sum_s \sum_{i,j:i<j} \theta_s^{ij} \| y_s^{ij} \|_2 \\
\text{s.t.} \quad & \left(\nabla x^i\right)_s = \sum_{j:j<i} y_s^{ji} - \sum_{i:j>i} y_s^{ij} \\
& \sum_i x_s^i = 1 \qquad x_s^i \geq 0 \qquad \forall s, i, j
\end{aligned}
$$

Interpretation of the variables y_s^{ij}

- Defined in the continuous domain
- Consider the following segmentation result (2D example)



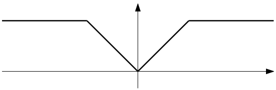
- y_s^{ij} gradient of change from label i to label j
- Important consequence
    - y_s^{ij} aligned with normal direction of the boundary between i and j

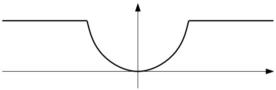
Extensions I

- Original continuous formulation only metric smoothness

$$\theta_s^{ii} = 0 \qquad\qquad \theta_s^{ij} \leq \theta_s^{ik} + \theta_s^{kj}$$

  - Labels have an order
  - Sum of transitions more expensive then direct transition
  - Truncated linear is metric
  - Truncated quadratic is not metric

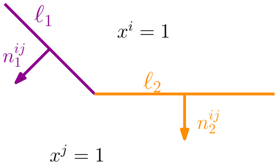




Extensions II

- Anisotropic smoothness sometimes desired
  - Aligning segmentation boundary direction with image edges
  - Well known for binary segmentation [Esedoglu and Osher 2004]
  - Can directly be used in the continuous multi-label formulation [Strekalovskiy and Cremers 2011]
- Our extended formulation:
  - non-metric and anisotropic smoothness

Anisotropic Smoothness Term



- Goal: Penalize boundary length ℓ weighted by its direction n
- Exchange $\theta_s^{ij}\|y_s^{ij}\|_2$ by $\phi_s^{ij}(y_s^{ij})$ [Esedoglu and Osher 2004]
- $\phi^{ij}(\cdot)$: $\mathbb{R}^N \rightarrow \mathbb{R}_0^+$ is a convex positively 1-homogeneous function
  - How do we specify such functions? Next slide
- n_s^{ij} normal of boundary between labels i and j at position s
- Regularizer penalizes boundary length ℓ weighted by $\phi^{ij}(n_s^{ij})$

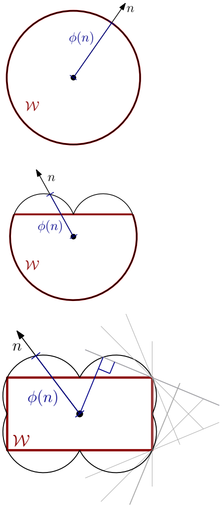
$$\sum_{s\in\Omega}\sum_{k,l:k<l}\phi^{kl}(y_s^{kl}) = \ell_1\phi^{ij}(n_1^{ij}) + \ell_2\phi^{ij}(n_2^{ij})$$

Wulff Shape

- Specifying a function $\phi(\cdot)$ directly can be hard
- Solution: Wulff shape $\mathcal{W}_\phi$ [Esedoglu and Osher 2004]
  - Convex shape
  - Originates form crystallography [Wulff 1901]
- All possible $\phi(\cdot)$ can be specified by

$$\phi(y) = \max_{\mu \in \mathcal{W}_\phi} \mu^T y$$

- Defining $\phi(\cdot)$ through $\mathcal{W}_\phi$ often easier
- Wulff shapes can be learned from training data



Non-Metric

- LP-relaxation does allow for arbitrary smoothness
- Continuously inspired formulation allows only for metrics
- Where is the difference?
  - LP relaxation contains x_{st}^{ij} variables that have to be non-negative
  - Continuously inspired formulation contains y_s^{ij} that are in $[-1, 1]^N$
  - And hence, no non-negative x_s^{ii}
- Fixed by introducing non-negative pseudo-marginals
  - Split positive and negative part of y_s^{ij} into individual variables
  - $x_s^{ij} := \max\{0, y_s^{ij}\}$ and $x_s^{ji} := -\min\{0, y_s^{ij}\}$
  - $y_s^{ij} = x_s^{ij} - x_s^{ji}$ still present in the formulation
  - x_s^{ii} and non-negativity constraint added
  - This allows non-metric smoothness for the discretized case

Final Convex Multi-Label Formulation

- The final formulation allows for non-metric and anisotropic smoothness at the same time

$$
\begin{aligned}
\min_x \quad & \sum_{s,i} \theta_s^i x_s^i + \sum_s \sum_{i,j:i<j} \phi_s^{ij}\left(x_s^{ij} - x_s^{ji}\right) \\
\text{s.t.} \quad & x_s^i = \sum_j \left(x_s^{ij}\right)_k \qquad x_s^i = \sum_j \left(x_{s-e_k}^{ji}\right)_k \\
& \sum_i x_s^i = 1 \qquad x_s^i \geq 0 \qquad x_s^{ij} \geq 0 \qquad \forall s, i, j, k
\end{aligned}
$$

- e_k: k-th canonical basis vector

Summary

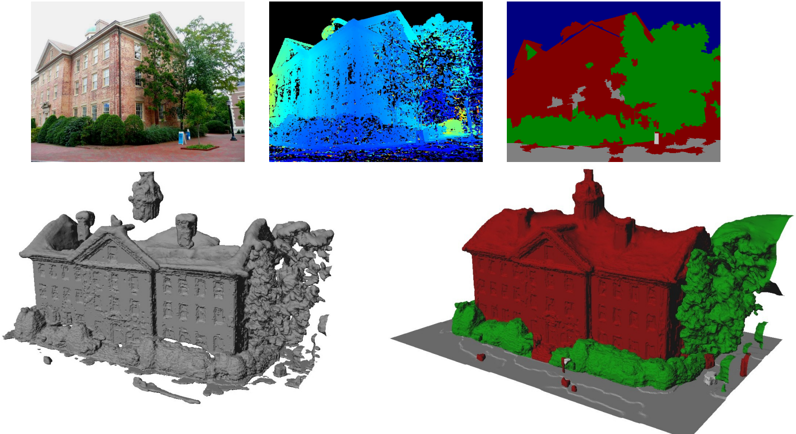
- LP-Relaxation
  - Arbitrary smoothness cost
  - Metrication artifacts
- Continuous formulation
  - Penalizes boundary length
  - Only metric smoothness
- Extensions (continuously inspired formulation)
  - Anisotropic and non-metric smoothness

Outline

C. Häne, C. Zach, A. Cohen, R. Angst, M. Pollefeys, **Joint 3D Scene Reconstruction and Class Segmentation**, CVPR 2013

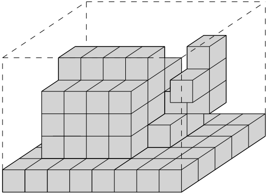
Idea

- Two challenging problems
  - Image segmentation
  - Dense 3D modeling
- Object category influences desired surface smoothness
- Optimize both jointly

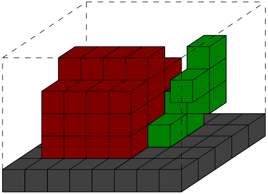


Formulation

- Baseline Method: Volumetric depth map fusion
  - Segmentation of a voxel space into free and occupied space: $u_s \in [0, 1]$
- Our joint fusion
  - Labeling of a voxelspace into L labels: $x_s^i \in [0, 1]$ and $\sum_i x_s^i = 1$
  - We use: free space, building, ground, vegetation, clutter
- Convex Energy
  - Unary term
    - Connects image appearance and depth maps
  - Smoothness term
    - Dependent on surface orientation and involved labels



Baseline



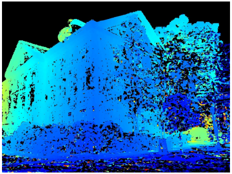
Proposed

Joint Fusion: Inference Overview

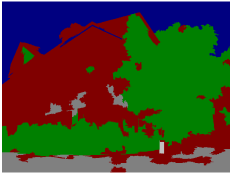
- Input

  - Camera poses
  - Vertical direction [Cohen et al. 2012]

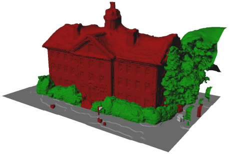
- Joint Fusion



+



→



Convex Continuous Multi-Label Formulation

Energy Minimization

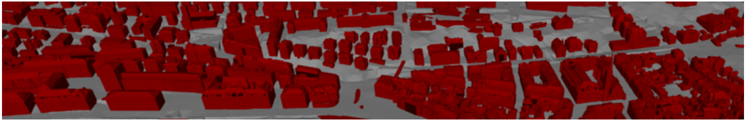
$$E(x,y) \quad = \sum_{s\in\Omega}\left(\sum_i \rho_s^i x_s^i + \sum_{i,j:i<j}\phi^{ij}(y_s^{ij})\right)$$

- Subject to marginalization and normalization constraints
- $x_s^i \in [0,1]$: indicating whether label i is chosen at voxel s
- $y_s^{ij} \in [-1,1]^3$: label transition gradients
- ρ_s^i: unary term, cost for label i at location s
- ϕ^{ij}: anisotropic surface area penalization between label i and j
- Optimized using primal-dual algorithm [Chambolle and Pock 2011]

Input to Energy

- For unary term: image based classifier + depth maps

- For smoothness term: geometric priors (anisotropic smoothness)



Appearance Likelihoods

- Image based semantic classifier:
  - Boosted decision tree classifier
    [STAIR Vision Library, Gould et al. 2010]
  - Context-based classifier
    [ALE Library, Ladicky et al. ICCV 2009]
- 5 classes: sky, building, ground, vegetation, clutter
- Negative log likelihoods $\sigma_{\text{class } i}$, per image pixel

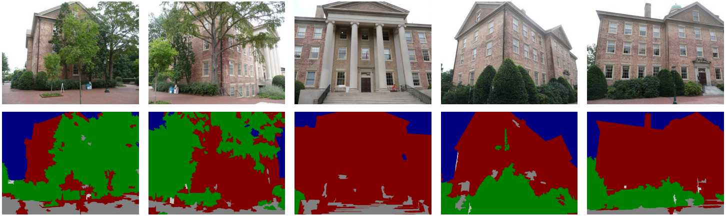
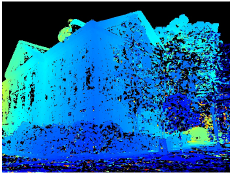
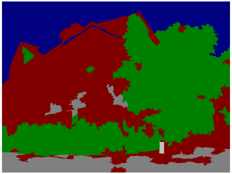


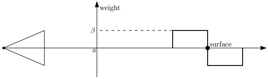
Figure : Best cost labels, STAIR Vision Library

Unary Term I





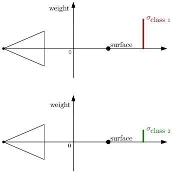
- Computed from depth maps and class likelihoods



Volumetric fusion

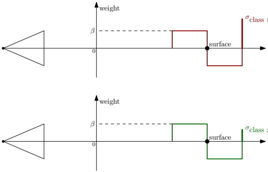
(depth only)

+



Class likelihoods

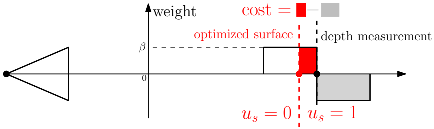
=



Joint fusion (solid classes)

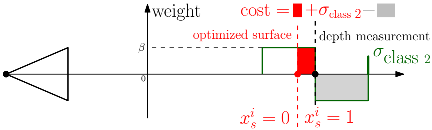
Unary Term II

- Cost induced for a single ray



Volumetric fusion

(depth only)



Joint fusion

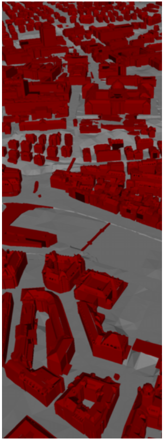
(depth and class likelihoods)

  - Cost for free-space indirect through solid classes
  - High likelihood for sky but no depth $\rightarrow$ free space preference along ray

- Approximates true ray likelihoods faithfully for the important cases
  - Assuming only one transition occurs in the band around the measured depth

Penalization According to Normal Orientation

- Vertical ground free space transition penalized differently then e.g. building free space transition



Convex Smoothness Term

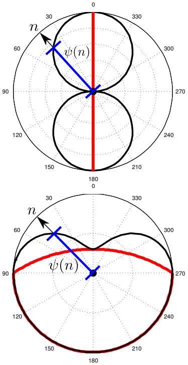
$$\psi^{ij}(\cdot;\theta^{ij}) : \mathbb{S}^2 \to \mathbb{R}_0^+$$

- Direction dependent cost of normal
  - Parameterized according to parameters θ^{ij}
  - Cadastral 3D city model as training data
  - Maximum-likelihood estimation (MLE)

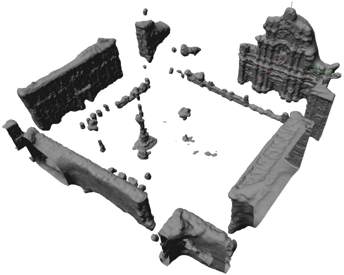
Smoothness term ϕ^{ij}

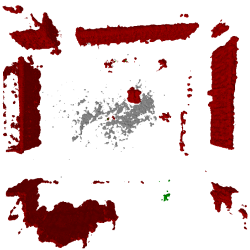
- Split into isotropic and anisotropic part
- Isotropic part: relative frequency of transitions
  - Independent of normal n
  - Determined as constant C^{ij}
- Anisotropic part:
  Selected from a collection of Wulff shapes $W_{\psi^{ij}}$
  - Line Segment
  - Half-sphere plus spherical cap
- Parameters of Wulff shape trained via MLE
  - Which shape
  - Shape parameters: grid search

$$\phi^{ij}(n) = \psi^{ij}(n) + C^{ij}$$

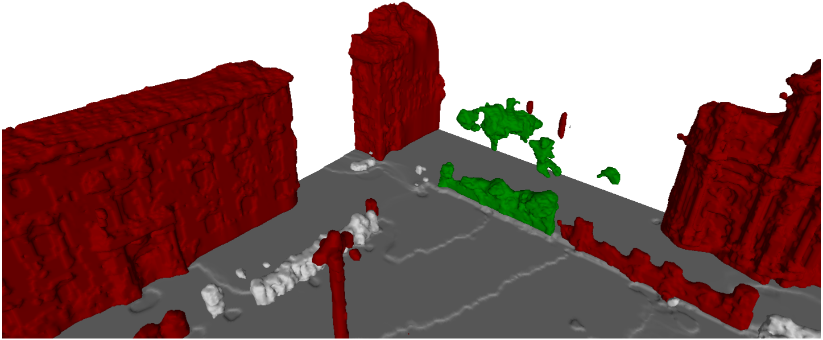


Evolution During Optimization



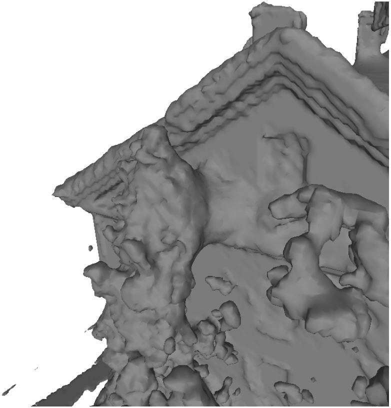


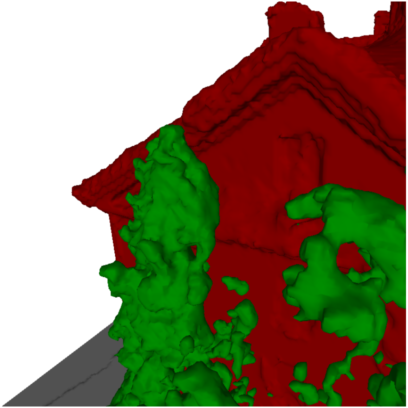
Weakly Observed Surfaces I



- Ground and buildings correctly reconstructed

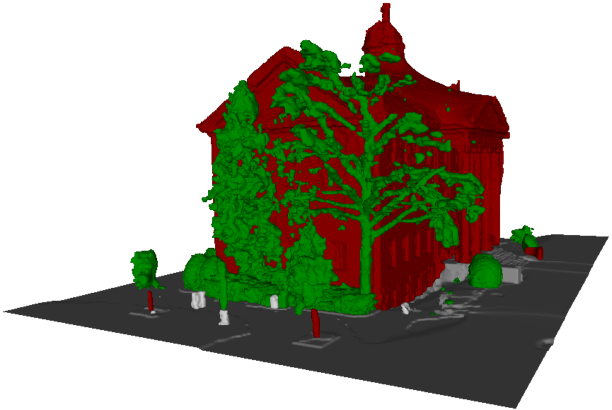
Weakly Observed Surfaces II



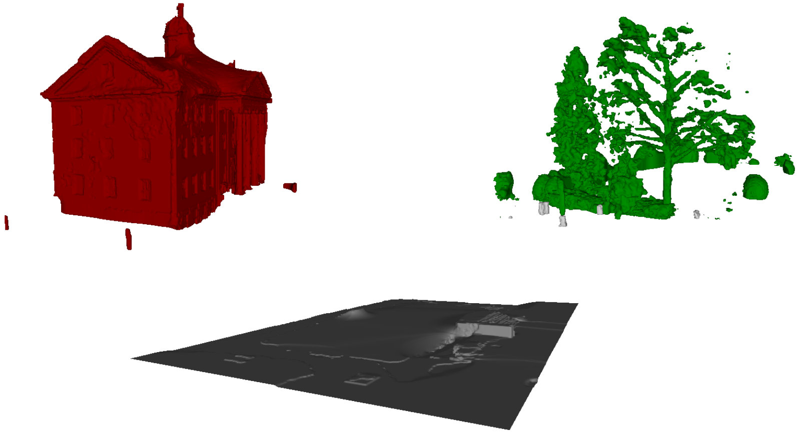


- Vegetation correctly separated from the building

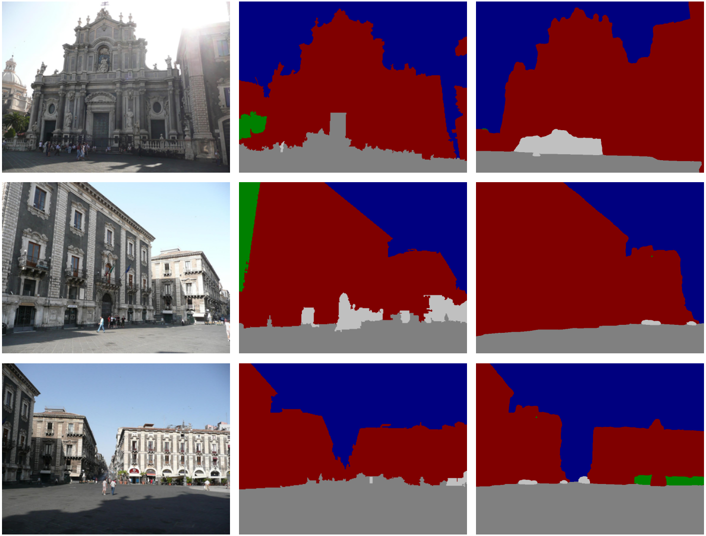
Unobserved Surfaces



Unobserved Surfaces



Advantages to Image Based Segmentation I



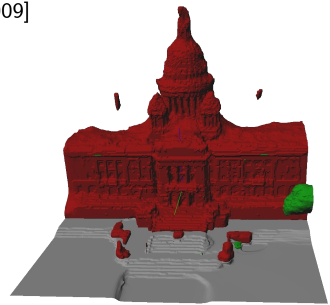
Input Image [Ladicky ICCV 2009] Joint Fusion

Advantages to Image Based Segmentation II

Images

[Ladicky et al. ICCV 2009]

- Single image segmentations not consistent
    - Ambiguous cases eg. ground or building
- Joint fusion disambiguates



Joint Fusion

Strength of the Relaxation

- Relaxation: integral solution not guaranteed
- Experiments show relaxation is strong
  - Similar relaxations experimentally shown to be strong [Chambolle et al. 2008]

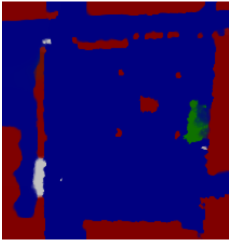


Figure : Slice through volume shows that relaxation is strong

Comparison to Volumetric Fusion

Conclusion: Joint Reconstruction and Class Segmentation

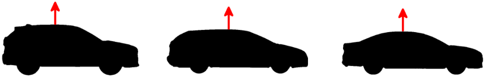
- Framework to jointly optimize for reconstruction and segmentation
  - To our knowledge: first volumetric approach
- Geometry and class segmentation coupled tightly
  - Previous joint methods on depth maps use height only [Ladicky et al. 2012]
  - We use priors on the normal direction
- Convex energy can be optimized globally
- Relaxation strong
- Geometric priors trained from cadastral 3D model
- Joint formulation improves best cost labeling and geometry

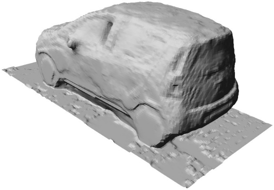
Outline

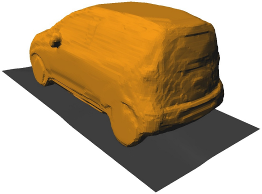
C. Häne, N. Savinov, M. Pollefeys, **Class Specific 3D Object Shape Priors Using Surface Normals**, CVPR 2014

Idea

- Some object classes hard to reconstruct
  - Lack of texture
  - Transparency
  - Reflection
- Solution: shape prior
  - Shapes within object class similar
  - Local distribution of surface normals
  - Roof of car always close to horizontal







Formulation

- Baseline Method: Volumetric depth map fusion
  - Segmentation of a voxel space into free and occupied space
- Shape prior formulation
  - Voxel space **aligned** with object of **known class**
  - 3 labels
    free space, ground, object
- Unary term
  - Computed from depth maps, local preference for solid class
- Smoothness term
  - Dependent on surface orientation, position and involved labels

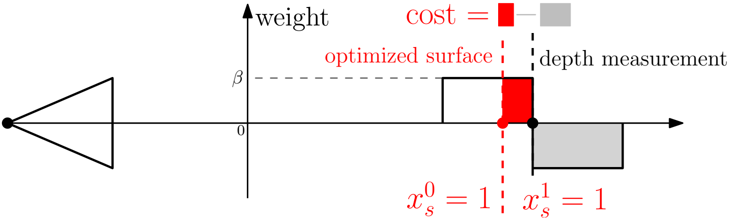
Energy Formulation

Minimization

$$E(x,y) \quad = \sum_{s\in\Omega} \left(\sum_i \rho_s^i x_s^i + \sum_{i,j:i<j} \phi_{s}^{ij}(y_s^{ij}) \right)$$

- Subject to marginalization and normalization constraints
- $x_s^i \in [0,1]$: indicating whether label i is chosen at voxel s
- $y_s^{ij} \in [-1,1]^3$: label transition gradients
- ρ_s^i: unary term, cost for label i at location s
- ϕ_{s}^{ij}: anisotropic surface area penalization between label i and j

Unary data term



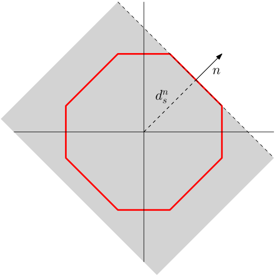
- Both solid classes use same data cost
- Label chosen solely based on the shape prior

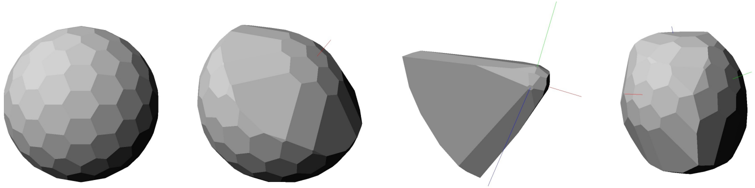
Smoothness Term (Shape Prior Training)

- Training data aligned
- Per voxel s
  - Collect normals of all training samples
  - Generate histogram over normal directions $\rightarrow P(n_s)$
  - Bins predefined (regular sampling of sphere)
  - Define anisotropic smoothness ϕ_s that reflects the distribution (next slide)
- Transition taken into account
  - free space $\leftrightarrow$ object
  - ground $\leftrightarrow$ object

Discrete Wulff Shape

- Parameterization for W_{ϕ_s}:
  Intersection of half spaces
  - n half space normal
  - d_s^n distance of half-space boundary to origin
- It holds $\phi_s(n) = d_s^n$ [Esedoglu and Osher 2004]
- We set $d_s^n \leftarrow -\log(P_s(n))$

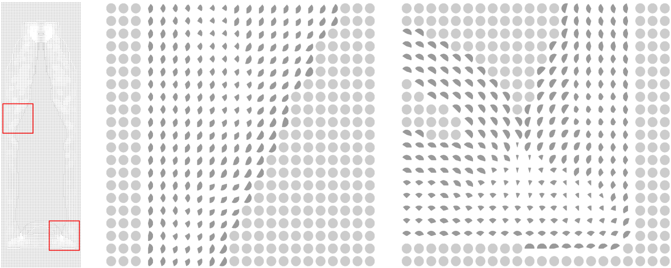




Visualization of the shape prior

- Training data

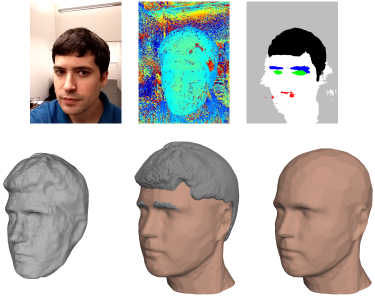
- Vertical cut through the shape prior



Results

Semantic Head Reconstruction

- Low resolution images captured on mobile phone with rolling shutter
- Image based classifier necessary
- Manual alignment not possible



F. Maninchedda, C. Häne, B. Jacquet, A. Delaunoy, M. Pollefeys, **Semantic Reconstruction of Heads**, submitted

Energy Formulation with Non-Convex Alignment

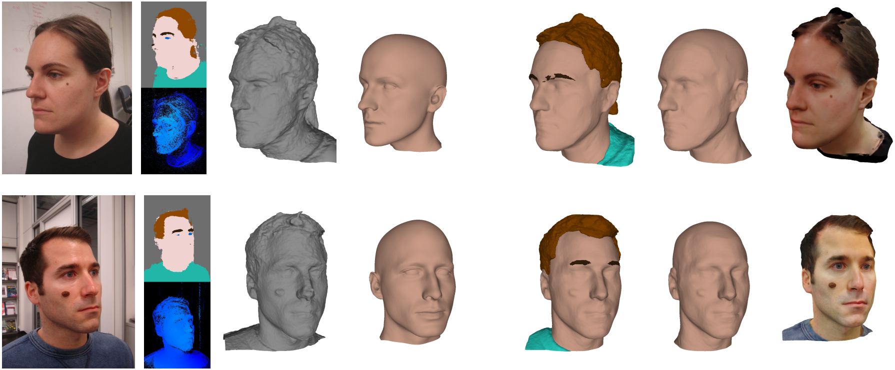
Minimization

$$E(x, y) = \sum_{s \in \Omega} \left(\sum_{i} \rho_s^i(\mathcal{T}) x_s^i + \frac{1}{\alpha^2} \sum_{i,j:i<j} \phi_s^{ij}(\mathcal{T}, y_s^{ij}) \right)$$

- Subject to marginalization and normalization constraints
- $x_s^i \in [0, 1]$: indicating whether label i is chosen at voxel s
- $y_s^{ij} \in [-1, 1]^3$: label transition gradients
- ρ_s^i: unary term, cost for label i at location s
- ϕ_s^{ij}: anisotropic surface area penalization between label i and j
- $\mathcal{T}$: similarity transform with scaling factor α
- Alternating optimization (reconstruction, alignment)

Overview of the Method

Comparisons



Input TV-Flux Statistical Model Our Semantic Reconstruction

Conclusion: Class Specific 3D Object Shape Priors

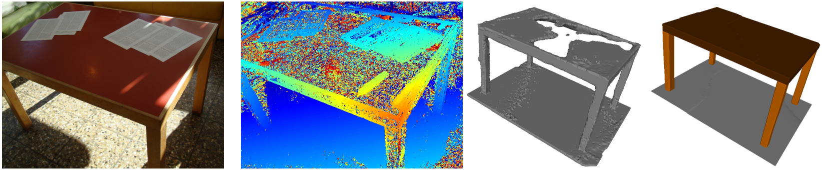
- Shape prior based on surface normals
- Allows for shape variation
- Preserves expected surface details
- Recovers hidden surfaces
  - Ground underneath the car
  - Skin underneath hair
- Support points are directly inferred by the optimization
- Shortcomings
  - Alignment between shape prior and input data
  - Different Wulff shape at each voxel

Outline

R. Karimi, C. Häne, M. Pollefeys, **Segment Based 3D Object Shape Priors**, CVPR 2015

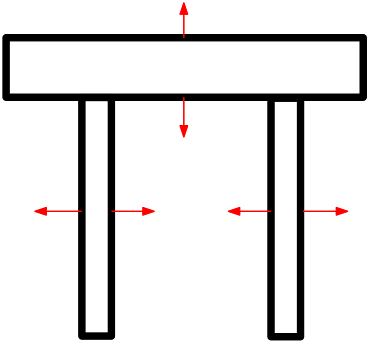
Idea

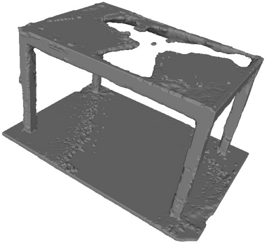
- Idea: Split object into convex segments
- Global (spatially homogeneous) smoothness terms
- Alignment only in terms of rotation

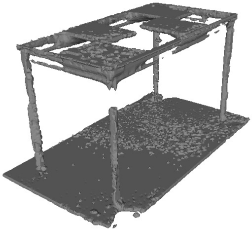


Splitting

- Why splitting?
  - Table top mostly horizontal
  - Table legs mostly vertical
  - Single anisotropic smoothness not distinctive
  - Individual segments very distinctive







Energy Formulation

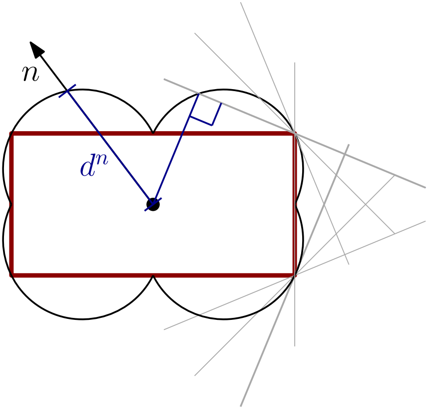
Minimization

$$E(x, y) \quad = \sum_{s \in \Omega} \left(\sum_i \rho_s^i x_s^i + \sum_{i,j:i<j} \phi^{ij}(y_s^{ij}) \right)$$

- Subject to marginalization and normalization constraints
- $x_s^i \in [0, 1]$: indicating whether label i is chosen at voxel s
- $y_s^{ij} \in [-1, 1]^3$: label transition gradients
- ρ_s^i: unary term, cost for label i at location s
- ϕ^{ij}: anisotropic surface area penalization between label i and j
  - For this formulation spatially homogeneous

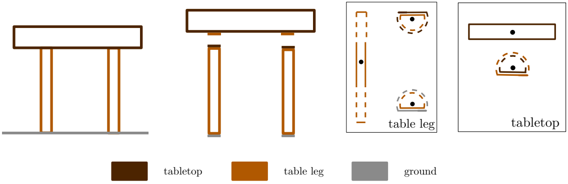
Smoothness Term

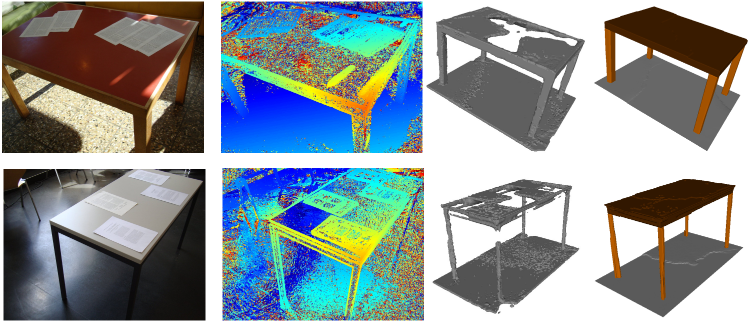
- Reconstructed shape has lowest energy for given volume, if it has the same shape as the Wulff shape



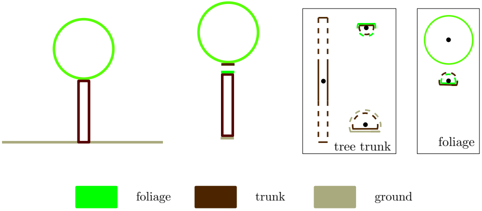
- Multiple transitions per segment
  - E.g. table top connects to leg and free space

Results: Table

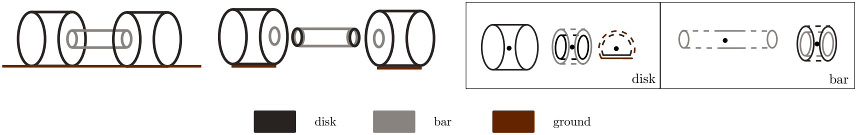


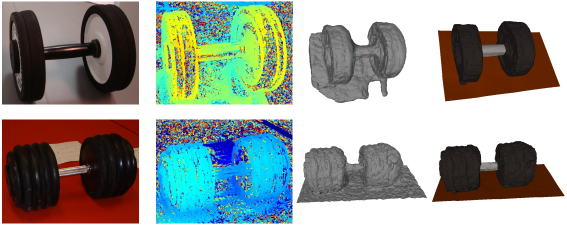


Results: Tree



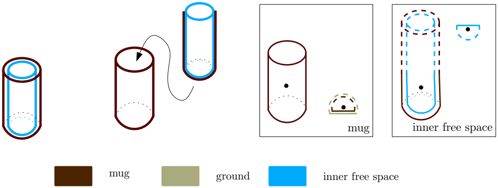
Results: Dumbbell

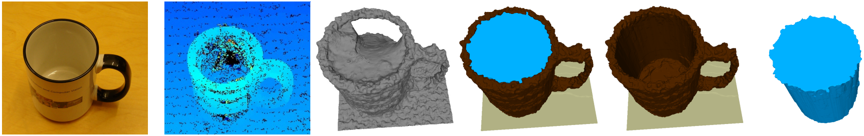




Results: Mug

- Convex segment of free space inside the mug





Video

Conclusion: Segment Based 3D Object Shape Priors

- Through splitting object into convex segments
  - Global (spatially homogeneous) smoothness term
  - Alignment only in terms of rotation
- Non-convex objects
  - Splitting free space

Outline

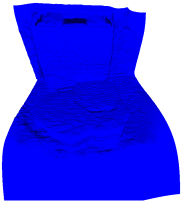
B. Jacquet, C. Häne, R. Angst, M. Pollefeys, **Multi-Body Depthmap Fusion with Non-Intersection Constraints**, ECCV 2014

Problem Setting

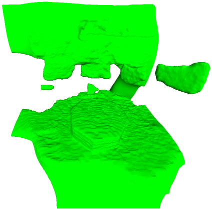
- Rigid objects move around
- Camera poses with respect to both rigid objects can be computed
- Goal: Volumetric reconstruction of both rigid objects


Volumetric 3D Reconstruction

- Volumetric Reconstruction for each object independently



Background



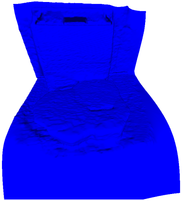
Box

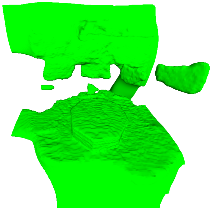
- Problem: Objects do not separate
- Solution: Enforce that occupied voxels do not intersect

$$\sum_{o \in \mathcal{O}, t \in \mathcal{T}} x^{o}_{t_s} \leq 1$$

Result Box

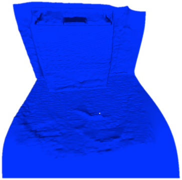
Without constraints

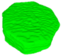




Background

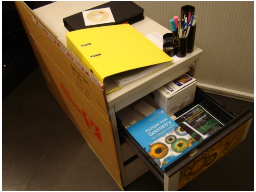
Box

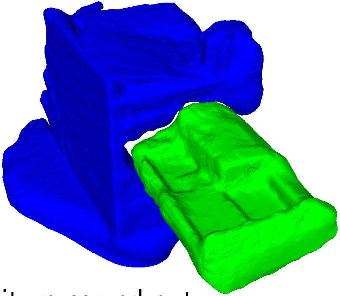




With constraints

Result Drawer

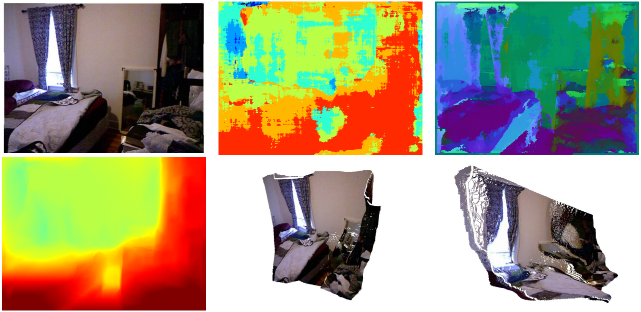
- Interior of the furniture carved out

Outline

C. Häne, L. Ladicky, M. Pollefeys, **Direction Matters: Depth Estimation with a Surface Normal Classifier**, CVPR 2015

Idea

- Idea: Combine single view depth and normal estimates
- Leads to more faithful single view depth estimates

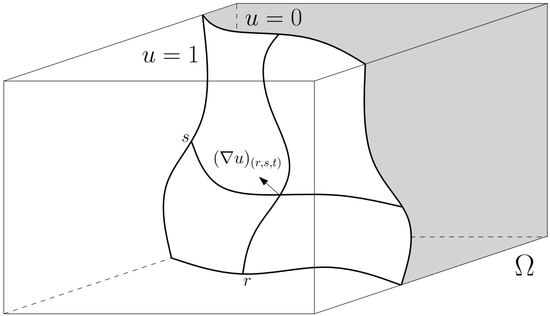


(top row) input to method, (bottom row) results

Formulation

- Task: Each pixel gets a depth label assigned
- 2D multi-label problem lifted to 3D two-label problem [Roy and Cox, 98]
  - Instead of assigning label $\ell_{(r,s)}$ to pixel (r,s)
  - Segmenting volume Ω into before/behind assigned depth label

$$u_{(r,s,t)} = \begin{cases} 0 & \text{if } \ell_{(r,s)} < t \\ 1 & \text{else,} \end{cases}$$



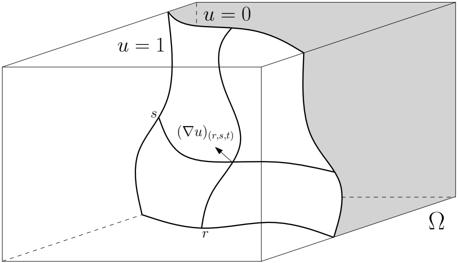
Convex Energy

Energy Minimization

$$E(u) = \sum_{r,s,t} \left\{ \rho_{(r,s,t)} \left| (\nabla_t u)_{(r,s,t)} \right| + \phi_{(r,s,t)} \left((\nabla u)_{(r,s,t)} \right) \right\}$$

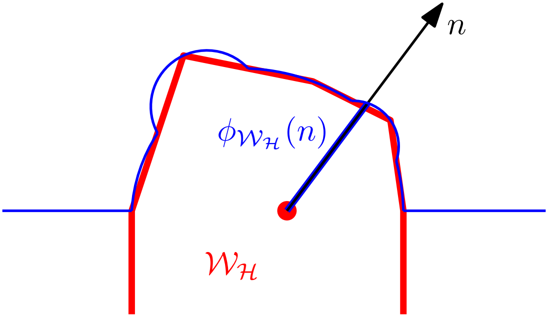
$$\text{s.t. } u_{(r,s,0)} = 0 \quad u_{(r,s,L)} = 1 \quad \forall (r,s)$$

- $\phi_{(r,s,t)}$: Anisotropic surface regularizer, convex pos. 1-homogeneous
- $\rho_{(r,s,t)}$: Cost for assigning label t to pixel (r,s)



Regularization Term

- Input: Classifier scores for a discrete set of normal directions
- Goal: Penalizing surface directions according to the classifier scores
- Discrete Wulff shape



Results

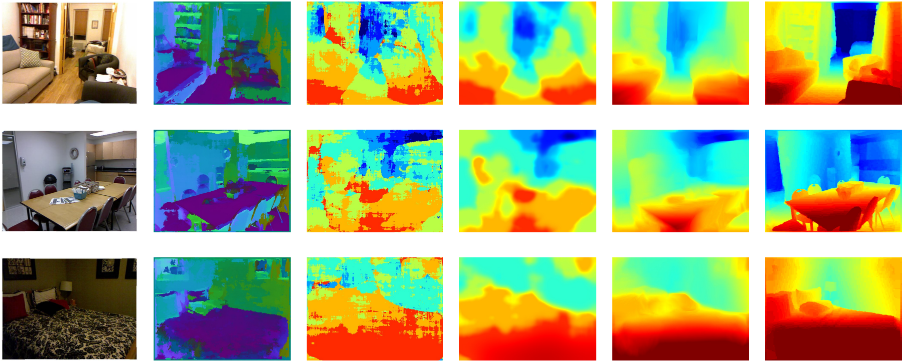


image | normals [Ladicky ECCV'14] | depth [Ladicky CVPR'14] | w/o normals | proposed | ground truth

- Conclusion: Normal directions contribute valuable information

Conclusion

- Reasoning about semantics and geometry jointly
  - Improves results
  - Shape prior formulations
  - Leads to more complete understanding of the scene

Questions ?