A Patch Prior for Dense 3D Reconstruction in Man-Made Environments

Christian Häne\textsuperscript{1}, Christopher Zach\textsuperscript{2}, Bernhard Zeisl\textsuperscript{1}, Marc Pollefeys\textsuperscript{1}

\textsuperscript{1} ETH Zürich
\textsuperscript{2} MSR Cambridge

October 14, 2012
Outline

1 Introduction
   • Motivation
   • Related Work

2 Patch prior
   • Piece-wise planar dictionary

3 Applications
   • Stereo matching
   • Depth map fusion
Outline

1 Introduction
   - Motivation
   - Related Work

2 Patch prior
   - Piece-wise planar dictionary

3 Applications
   - Stereo matching
   - Depth map fusion
Motivation

- **Goal**: high quality depth maps in man-made environments
  - **Challenge**: Textureless areas
  - **Weak and ambiguous observations**

- **But**: man-made environments mostly planar
- **Often used prior (Huber) total variation not sufficient**
Related Work: MRF Approach

- 2nd-order smoothness [Woodford et al. 2009]
  - Smoothness enforced on disparity gradients
  - 3rd order MRF
  - Proposal-based optimization

Discrete ground truth  Pairwise MRF  3-clique MRF
Related Work: Sparse Coding for Natural Images

- Natural images have sparse local image statistics
- Image patch is sparse combination of over-complete dictionary patches

\[ I \approx D\alpha \quad \text{with } \alpha \text{ sparse} \]

\[ \alpha^* = \arg \min_{\alpha} \| I - D\alpha \|^2 + \lambda \| \alpha \|_1 \]

\(D\) is a dictionary (hand-crafted or estimated from data)
\(\alpha\) are (sparse) coefficients of dictionary elements

[Elad and Aharon, 2006]
Related Work: Sparse Coding for Depth Images

- Inpainting of sparsely sampled depth maps [Hawe et al. 2011]

- Learn patch dictionary for depth maps [Tošić et al. 2011]
Man-made environments

 Mostly (i.e. piecewise) planar in 3D

Question: Do we really need trained / overcomplete dictionaries to explain depth images in man-made environments?

We believe not: focus on piecewise planar prior
Man-made environments

Mostly (i.e. piecewise) planar in 3D

Question: Do we really need trained / overcomplete dictionaries to explain depth images in man-made environments?

We believe not: focus on piecewise planar prior
Outline

1. Introduction
   - Motivation
   - Related Work

2. Patch prior
   - Piece-wise planar dictionary

3. Applications
   - Stereo matching
   - Depth map fusion
Piece-wise planar dictionaries

- Planar surface in 3D $\implies$ linear gradient in disparity (not depth) images
- Two (horizontal and vertical) dictionaries: only 1D “patches”

- Two elements per dictionary
  - “Slope” patch: generates any slope in 3D
    - Coefficient: $\alpha_{\text{slope}}$
  - “Bias” patch: disparity offset, bias
    - Coefficient: $\alpha_{\text{bias}}$
Regularization of dictionary coefficients

- Goal: *piece-wise* planar patches
  - Penalize spatial change of *slope* coefficients
  - No penalization on *bias* coefficients
- Sparse changes of slope:

Coefficient regularization term

\[ \| \nabla (\alpha^k_p)_{\text{slope}} \| \]

- Link between patch priors and disparities:
  - Penalize local deviation of disparity from planarity
  \[ \implies \text{Reconstruction error} \]
Goal: *piece-wise* planar patches
- Penalize spatial change of *slope* coefficients
- No penalization on *bias* coefficients

Sparse changes of slope:

**Coefficient regularization term**

\[ \| \nabla (\alpha^k_p)_{\text{slope}} \| \]

Link between patch priors and disparities:
- Penalize local deviation of disparity from planarity
  \( \implies \) Reconstruction error
Reconstruction error

\[ \sum_k \| R_p^k u - D_k^k \alpha_p^k \| \]

- \( R_p^k \) extracts a patch
- Reconstructed from dictionary
- Reconstruction error penalized
  - summed over all dictionaries (horizontal and vertical)
Our Energy

**Energy**

\[ E(u, \alpha) = \int \left( \phi_p(u_p) \right) + \eta \sum_k \| R_p^k u - D^k \alpha_p^k \| + \mu \| \nabla (\alpha_p^k)_{\text{slope}} \| \, dp, \]

- Application dependent convex/convexified data term
- Energy convex but non-smooth
- Optimized with proximal methods
- Guaranteed to converge to the global optimum
Outline

1 Introduction
   - Motivation
   - Related Work

2 Patch prior
   - Piece-wise planar dictionary

3 Applications
   - Stereo matching
   - Depth map fusion
Stereo matching

- Rectified stereo image pair $l_0, l_1$
- Recover disparity map $u$

**Data term**

$$\phi_p(u_p) = \lambda|l_1(u_p) - l_0|$$

- Non-convex, first order approximation around $u^0$
  - Pyramidal approach
  - Like optical flow

**Data term, first order approximation**

$$\phi_p(u_p) = \lambda|l_1(u_p^0) + (u_p - u_p^0)\nabla_u l_1 - l_0|$$
Results I

Left input image

Right input image

Total variation

Piece-wise planar
Results for the four urban data sets (available from http://rainsoft.de/software/libelas.html): left input image, depth from stereo using the TV prior, depth from stereo using the piecewise planar prior.
Depth map fusion I

- 25 images, five depth maps
- semi global matching
Depth maps warped to reference (middle)

**Data term**

\[
\phi_p(u_p) = \sum \max\{0, |u_p - u_p^l| - \delta\}
\]

- \(u_p\) reconstructed disparity
- \(u_p^l\) input disparities
- Capped \(L^1\) distance with threshold \(\delta\)
  - Allows deviation within quantization level
Results

Input image
Input depth map
Huber-TV fusion
Piecewise planar

Huber-TV fusion
Piecewise planar fusion
Results II

- Input image
- Input depth map
- Huber-TV fusion
- Piecewise planar
- Huber-TV fusion
- Piecewise planar fusion
Conclusion

- Depth map recovery utilizing patch-based prior
- Inspired by patch based approaches for image processing
- Dictionary for piece-wise planar regularization
  - Only 4 dictionary elements
  - Relatively large neighborhoods possible
  - Without computational drawbacks of higher-order MRFs
- Applied to computational stereo and depth map fusion
Questions

Video

Questions?