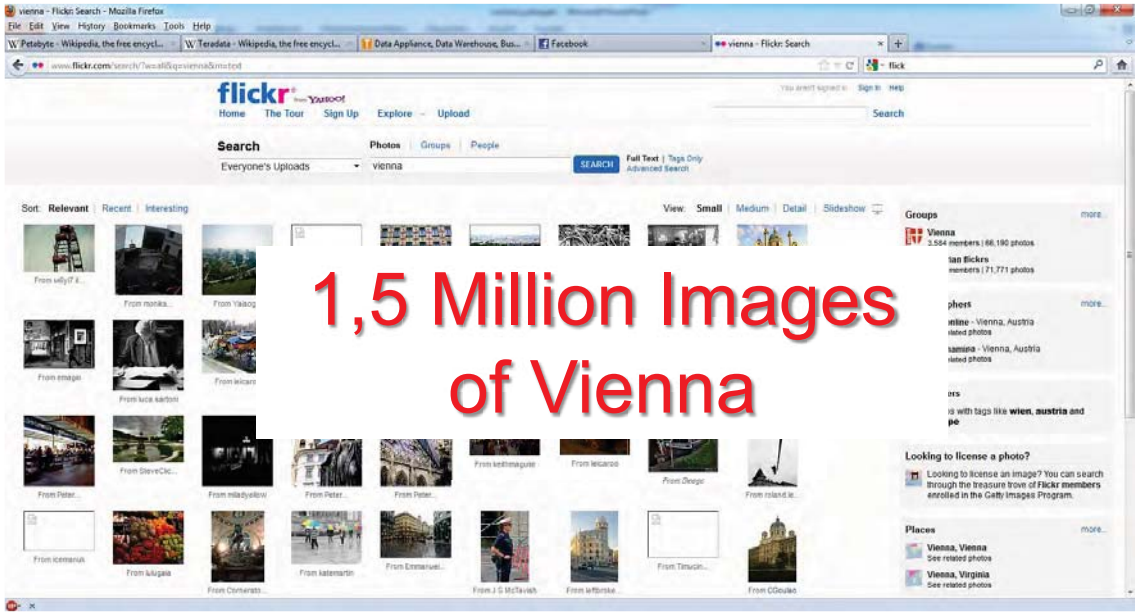
Exploiting redundancy for reliable aerial computer vision

Horst Bischof

Graz University of Technology

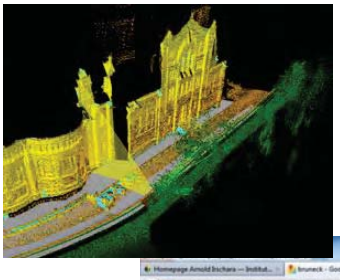
Digital Image Increase

Images Worldwide

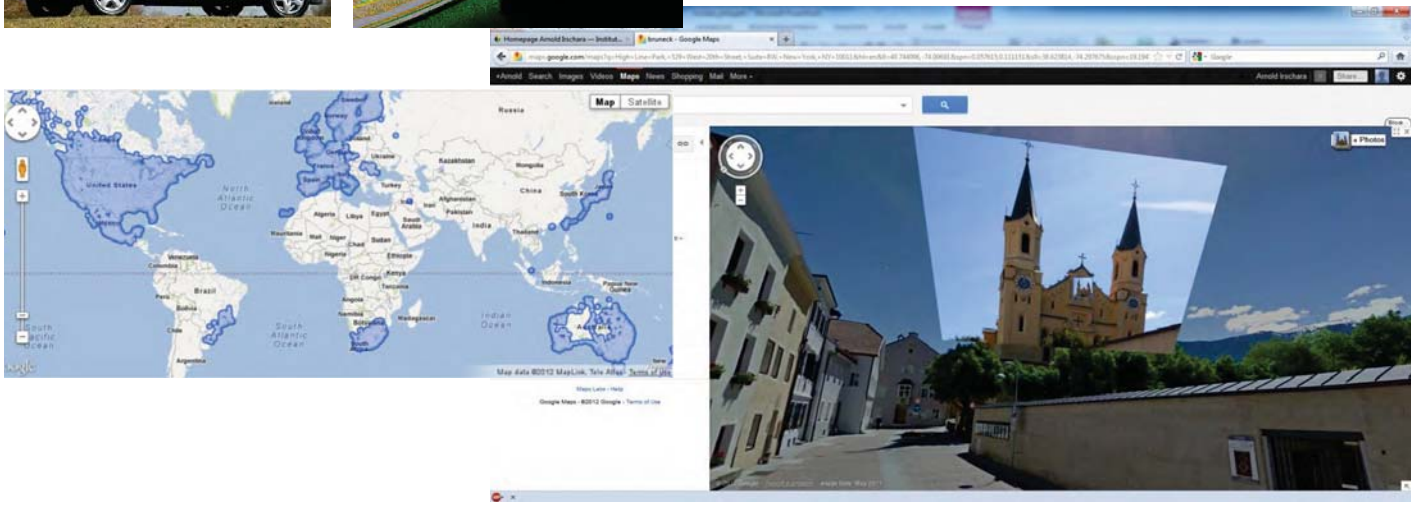


- > 60 billion photos taken in 2011
- > 5 billion photos on flickr
- > 3 billion photos/month on facebook

Terrestrial Image Acquisition



- Google Street View
- ~ 1 billion images (15 Petabyte)
- 3D Geometry
- User Photo Collections



Aerial Photogrammetry

- Clear30 project (Microsoft / DigitalGlobe)
- USA / Europe
- UltraCamG 109 Megapixel (30cm GSD)
- ~150 million (~50 Petabyte)

New Sensor Platforms

- Cost efficient
- Flexible (vertical take off and landing / position hold / GPS waypoint navigation)

[***Towards Fully Automatic Photogrammetric Reconstruction Using Digital Images Taken From UAVs.*** A. Irschara, V.Kaufmann, M. Klopschitz, H. Bischof, and F. Leberl. 100 Years ISPRS 2010.]

Airborne vs. MAVs

Microsoft
UltraCam$_D$
11500 x 7500

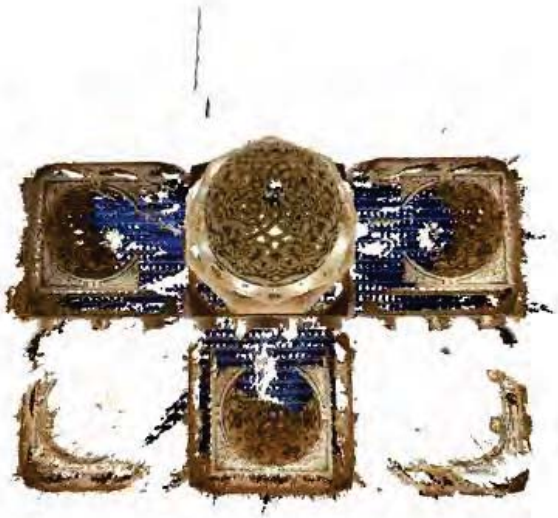
MAV
Pentax A40
4000 x 3000

Applications

Application: Virtual Habitat

[Leberl et al. IEEE Computer 2010]

Application: Architecture and Cultural Heritage



[Zebedin 2010], [Irschara 2010]

Application: Construction Site Monitoring

Application: Mining

Aerial Survey of Erzberg, Austria
May 2, 2012

Aerial Vision Group
Institute for Computer Graphics and Vision
Graz University of Technology, Austria

http://aerial.icg.tugraz.at

Outline

- **3D Reconstruction**
  - Structure from Motion
  - Dense Stereo
  - Redundancy
- **Semantic Classification**
  - Pixel level classification
  - Regularization
  - Fusion
- **Discussion & Outlook**

3D Reconstruction

Structure from Motion

INPUT

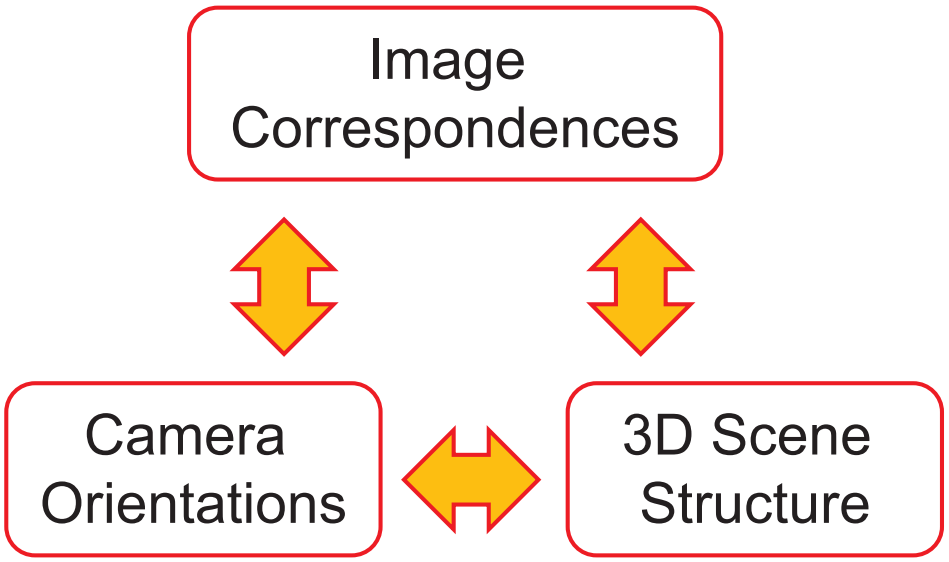
video sequence / unorganized images

OUTPUT

3D model
camera orientations + 3D structure

[D. Nister, “Automatic dense reconstruction from uncalibrated video sequences”. Ph.D., 2001]
[M. Pollefeys et al., “Visual modeling with a handheld camera”. IJCV, 2004]
[N. Snavely et al., “Photo Tourism: Exploring Photo Collections in 3D”. SIGGRAPH, 2006]
[S. Agarwal et al., “Building Rome in a Day”. CVPR, 2009]
[J.-M. Frahm et al., “Building Rome on a Cloudless Day”. ECCV, 2010]

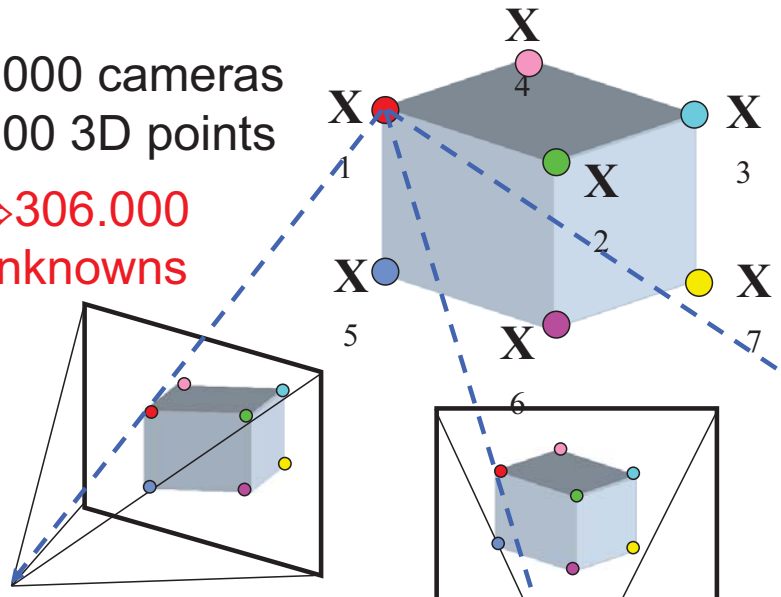
Structure from Motion



Structure from Motion

e.g. 1000 cameras
100.000 3D points

⇒306.000 unknowns



minimize

$g(\mathbf{R},\mathbf{t},[\mathbf{K}],\mathbf{X})$

Camera 1
$\mathbf{R}_1, \mathbf{t}_1, [\mathbf{K}_1]$

Camera 2
$\mathbf{R}_2, \mathbf{t}_2, [\mathbf{K}_2]$

Camera 3
$\mathbf{R}_3, \mathbf{t}_3, [\mathbf{K}_3]$

[N. Snavely CVPR ´09 tutorial]

Structure from Motion

Images | Pose Prior | Calibration | Pose Prior

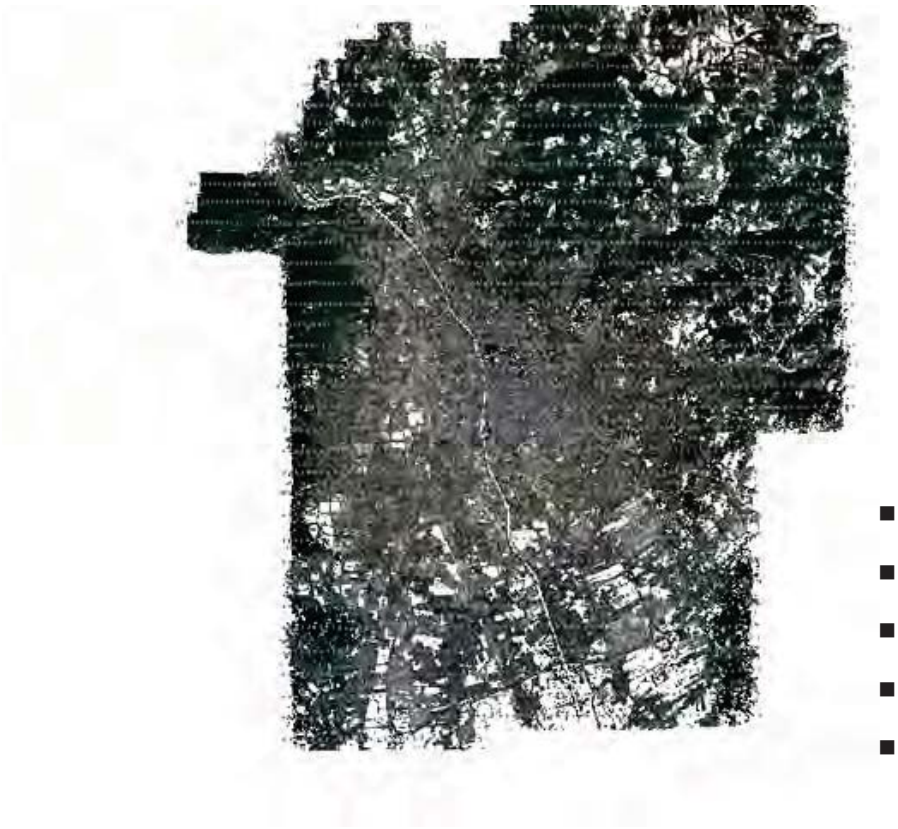
Feature Extraction → Coarse Matching → Detailed Matching → Geometric Verification → Geometric Estimation

Local Descriptors | Image Overlap | Matches | Epipolar Graph | Camera poses 3D points

Compute Unified Device Architecture (CUDA)

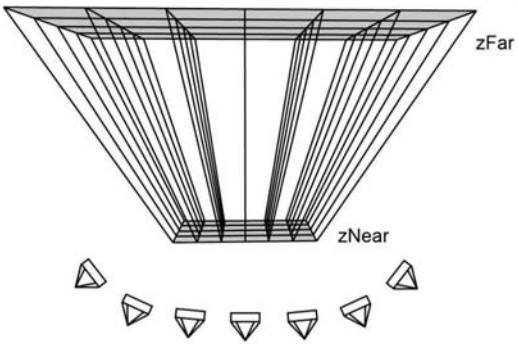
GPU / CPU Speedup 10-20x

Aerial Photogrammetry



- 3000 aerial images
- area ~150 km²
- 80% overlap
- 14,430 x 9,420 pixel
- ~15 cm GSD

Global Depth Map Optimization



Continuous Energy Minimization Approach

$$\min_u \left\{ \int_\Omega |\nabla u(x)| \, dx + \int_\Omega \rho(u(x), x) dx \right\}$$

regularity term — data term

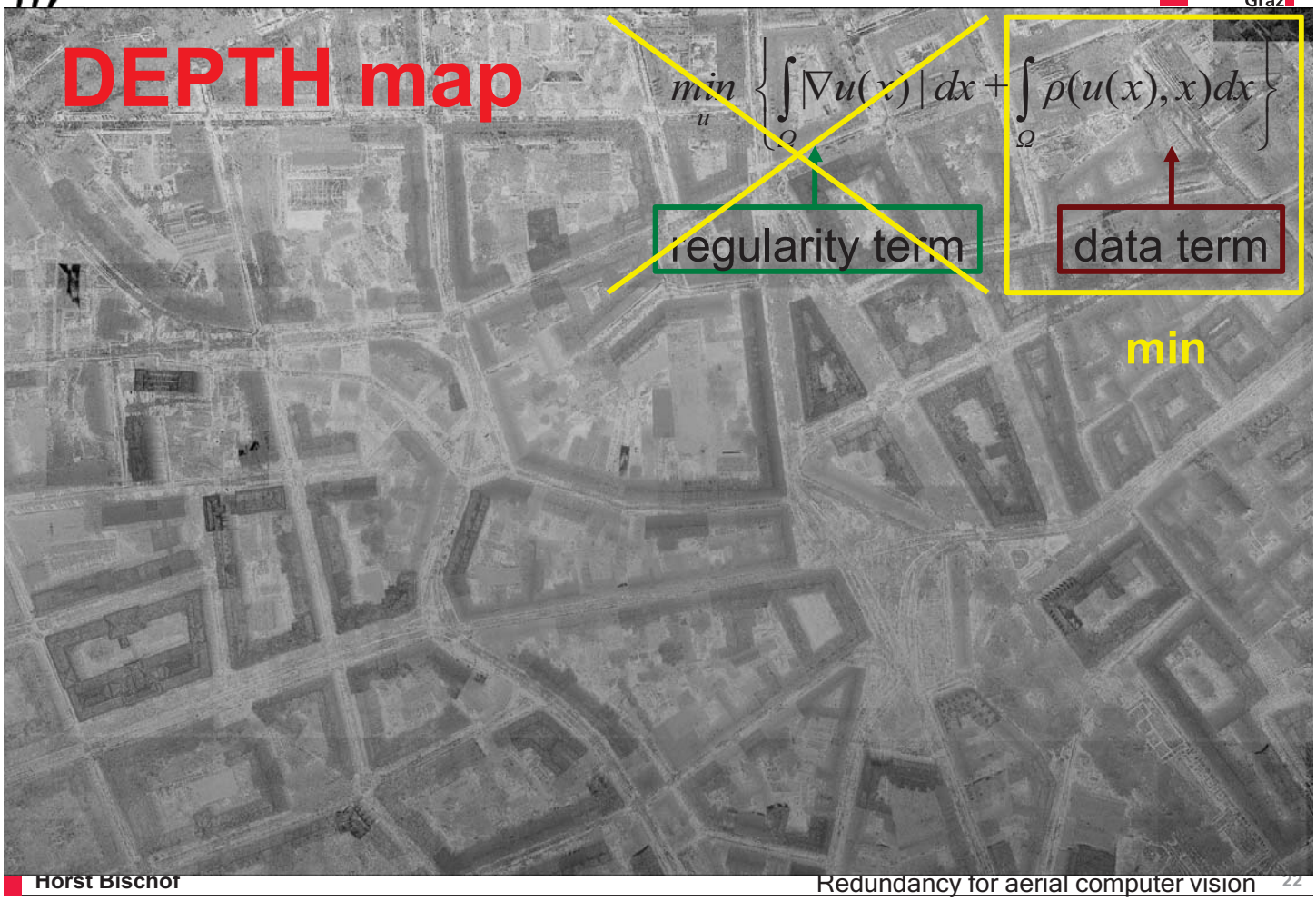
$\rho(u(x), x)$ .. matching costs (NCC)

$\int_\Omega |\nabla u(x)| \, dx$ .. Total Variation (TV) regularization:

➢ enforces spatial smoothness / discontinuity preserving

[T. Pock et al., "A convex formulation of continuous multi-label problems". ECCV, 2008]


KEY view


DEPTH map
$\min_u \left\{ \int_\Omega |\nabla u(x)|\, dx + \int_\Omega \rho(u(x), x) dx \right\}$
regularity term
data term
min

DEPTH map

$$\min_u \left\{ \int_\Omega |\nabla u(x)|\, dx + \int_\Omega \rho(u(x), x) dx \right\}$$

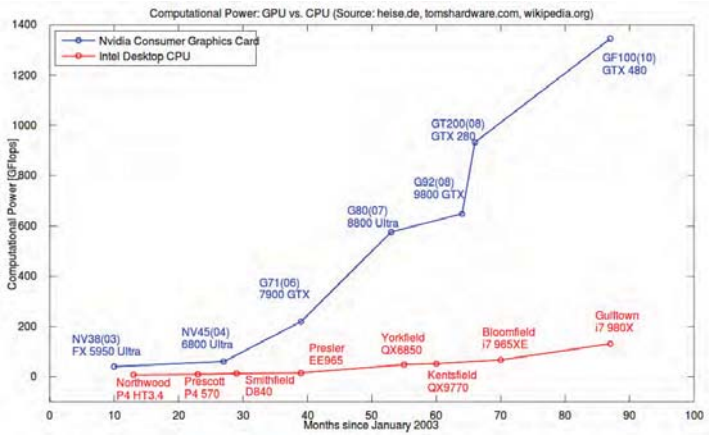
regularity term

data term

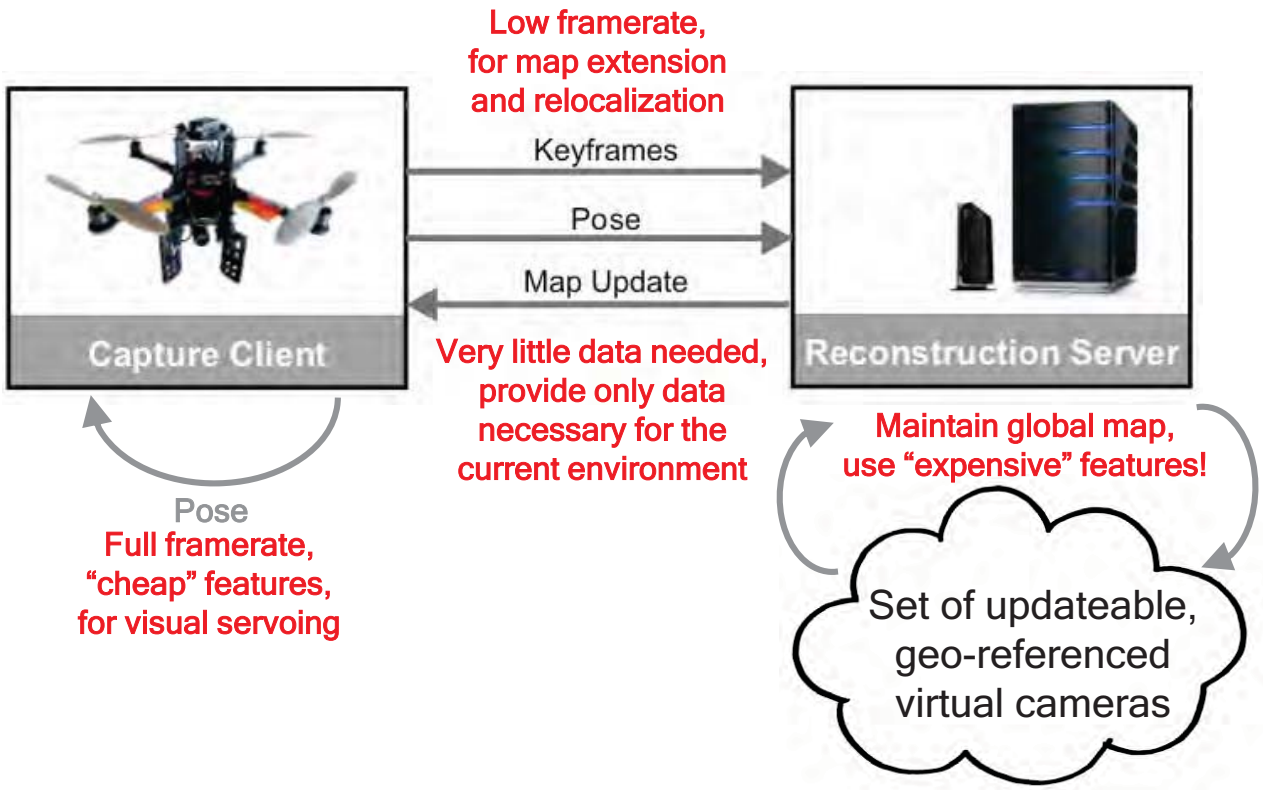
min

Primal-Dual Optimization

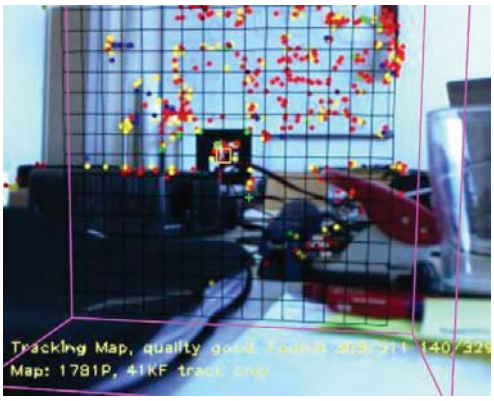
- The resulting convex optimization problems are solved via a first-order primal-dual algorithm [Chambolle, Pock 2010]
- Features a provable optimal convergence for convex problems with known structure
- Can be efficiently parallelized on massively parallel hardware such as GPUs

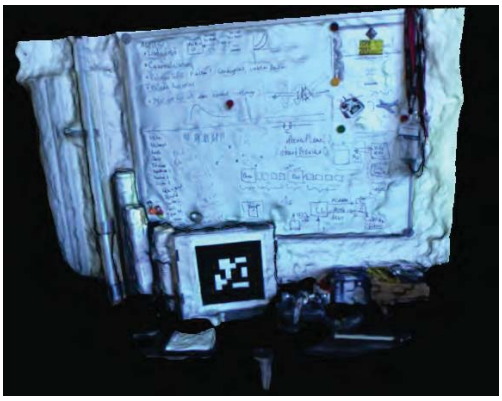


Distributed Visual SLAM

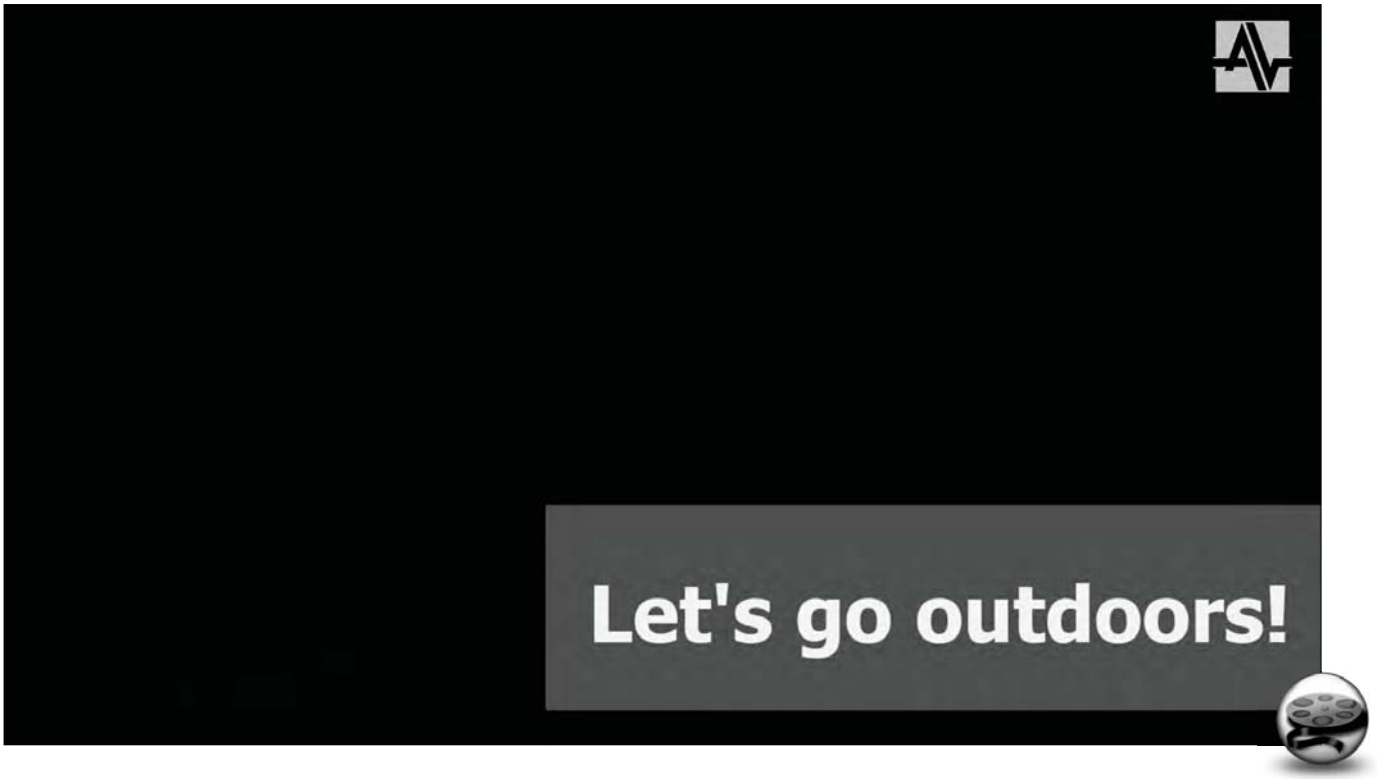


Dense Reconstruction On-the-Fly



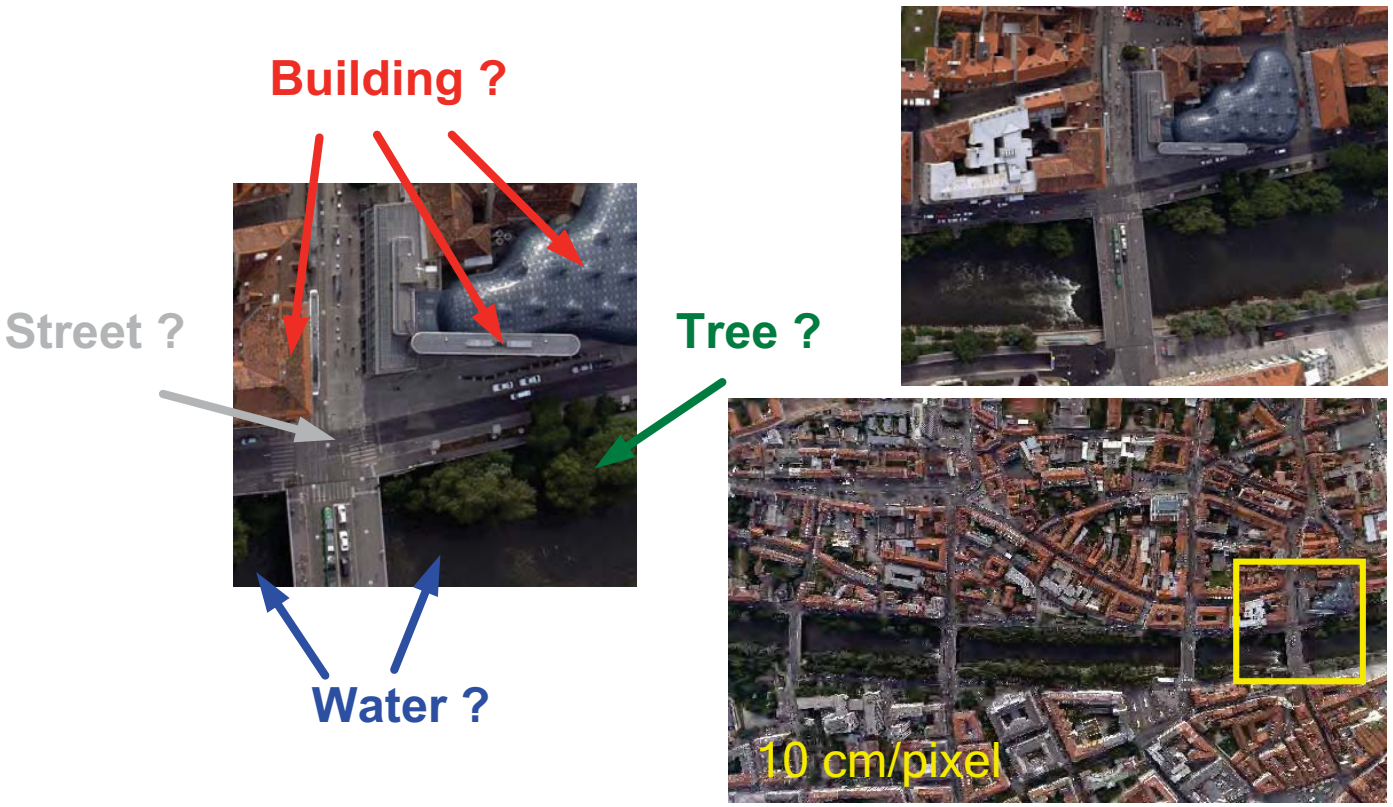


Dense Reconstruction On-the-Fly

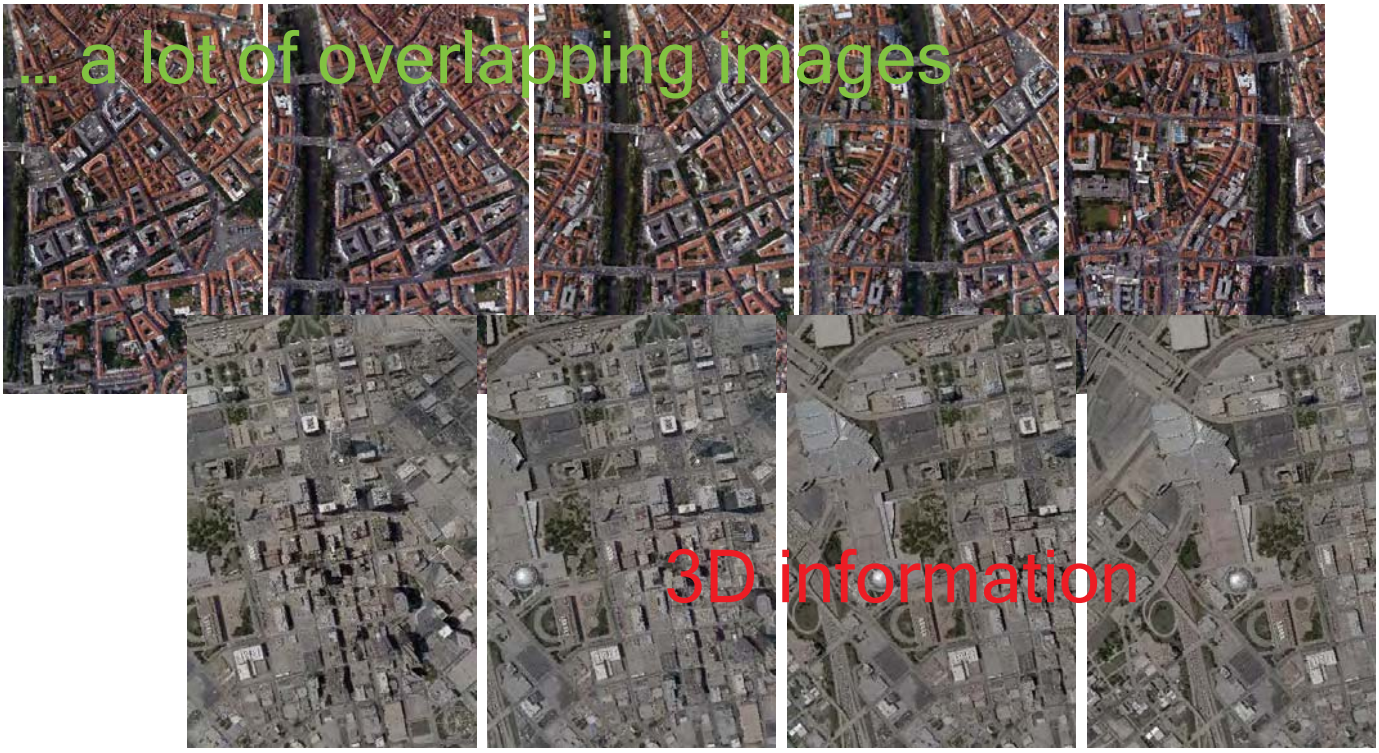


Semantic Segmentation

Motivation – Semantic Interpretation



Motivation



Fusion - Model

- TV-L^1 denoising model [Nikolova 04]

$$\min_u \left\{ \|\nabla u\|_1 + \lambda \sum_{i,j \in \Omega} |u_{i,j} - f_{i,j}| \right\}$$

- Robust against outliers
- Preserves sharp edges

$\Omega = \{(i,j) : 1 \leq i \leq W, 1 \leq j \leq H\}$

$u \in \mathbb{R}^{WH}$ ... sought solution

$f \in \mathbb{R}^{WH}$ ... observation

λ ... controls data fidelity

Fusion - Model

- Extension to multiple observations [Zach07]
- $$\min_u \left\{ \|\nabla u\|_1 + \lambda \sum_{k=1}^{K} \sum_{i,j \in \Omega} |u_{i,j} - f_{i,j}^k| \right\}$$

Some observations

$f^k \in \mathbb{R}^{WH} \ldots$ k-th observation

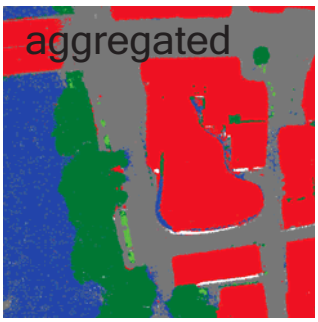
$K \ldots$ number of observations

Fusion – Semantic Interpretation

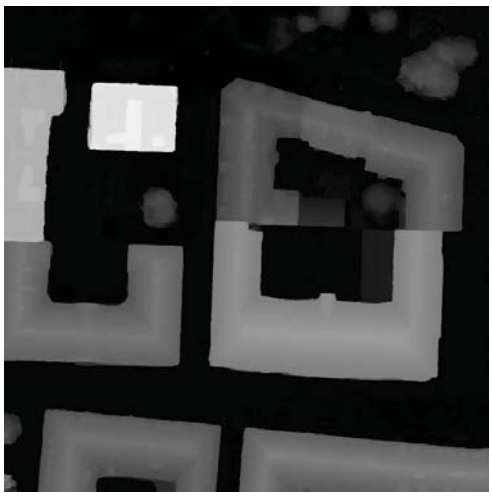
- Refinement of initial class assignments
- Continuous formulation of the Potts model [Pock09]

$$\min_{E_i,\, i=1,\ldots,K} \left\{ \frac{1}{2} \sum_{i=1}^{K} \mathrm{Per}(E_i; \Omega) + \lambda \sum_{i=1}^{K} \int_{E_i} f_i(x)\, dx \right\}$$

$$\bigcup_{i=1}^{N} E_i = \Omega \qquad E_i \cap E_j = \emptyset \; \forall i \neq j$$

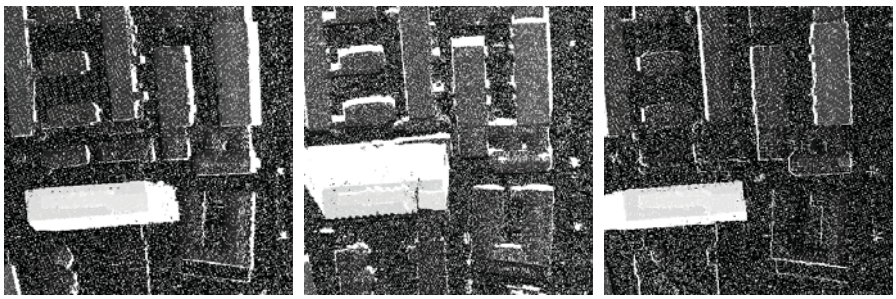


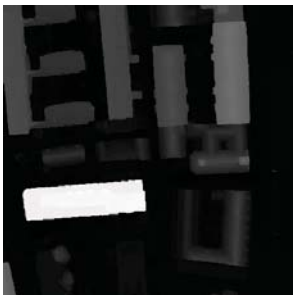
Fusion - Color and Height

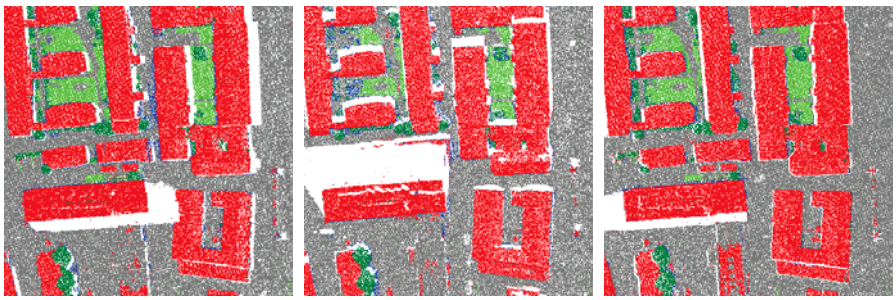


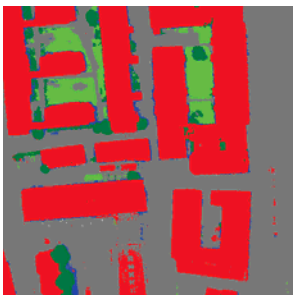
Color: Wavelet-Fusion

Height: TV-Fusion



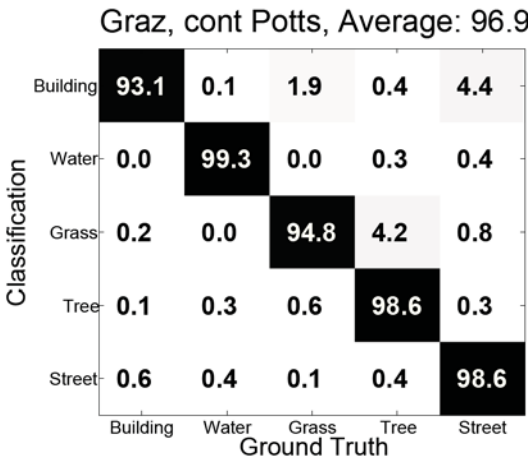


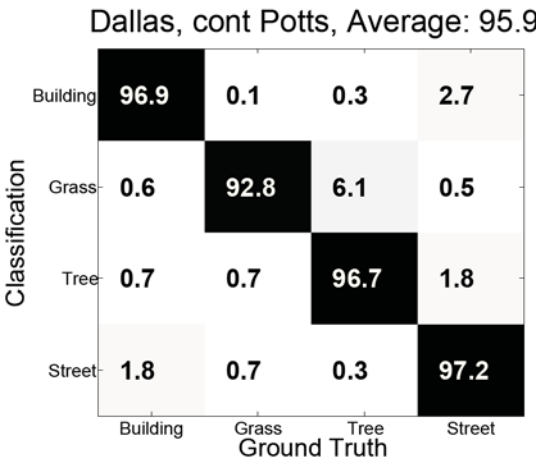




Some observations in ortho-view

Evaluation – Redundant Interpretation

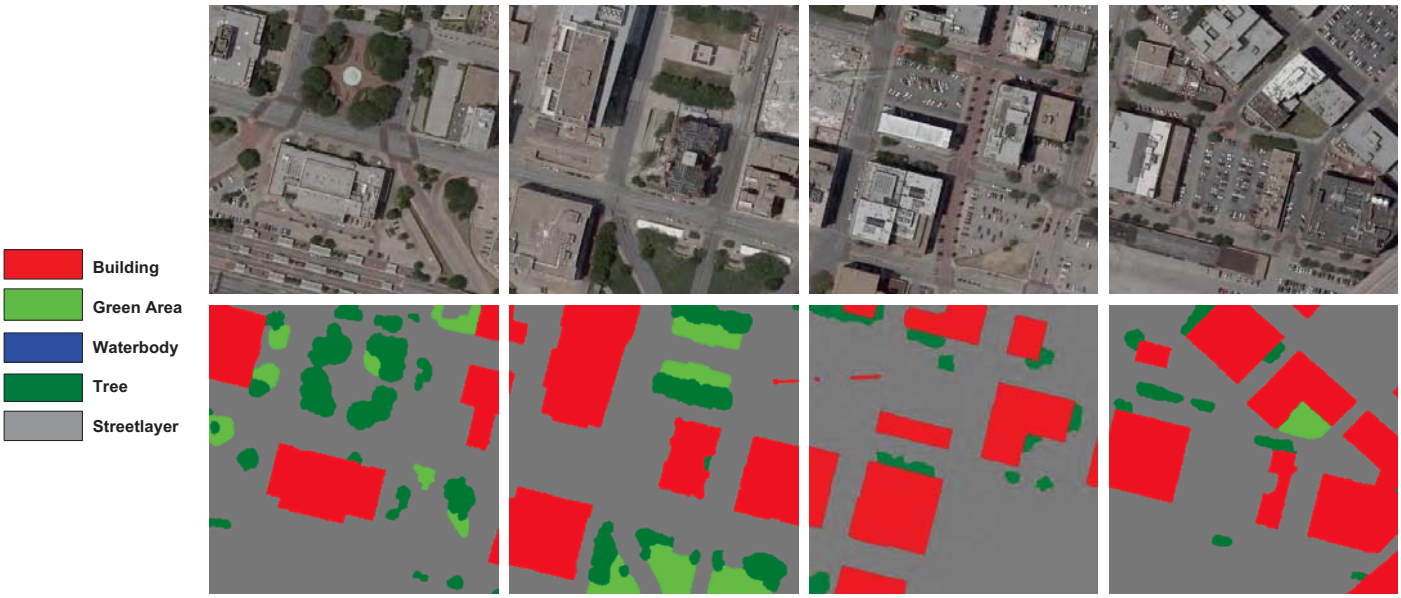




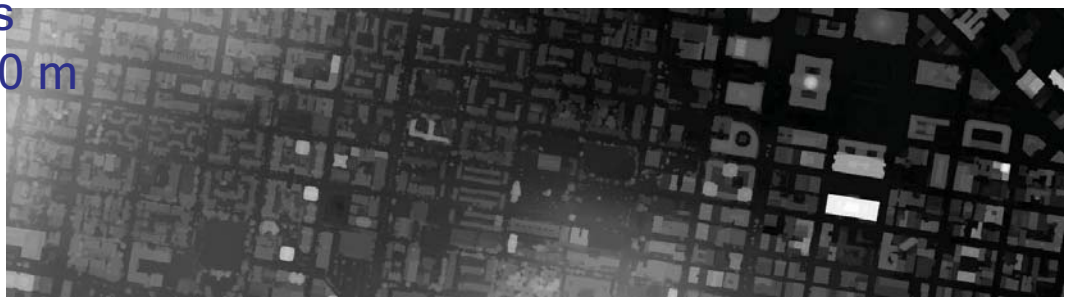
Remember evaluation on single images:
Graz 89.5%, Dallas 92.5%

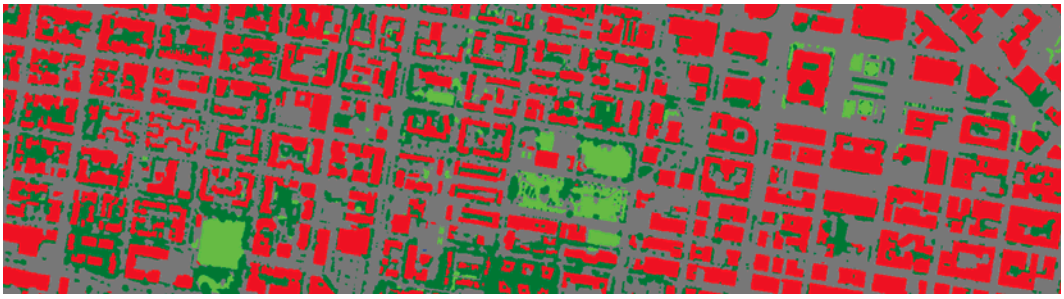
Large-Scale Results

Dallas, 4 tiles, each 240 x 240 m, 15 cm



SF
10 x 3 tiles
2500 x 700 m
15 cm





Building
Green Area
Waterbody
Tree
Streetlayer

Large-scale Results

Graz, 7 km^2, 155 images, 20 x 10 tiles, 8 cm

Building
Green Area
Waterbody
Tree
Streetlayer

Conclusions & Future Work

- Aerial based 3D reconstruction of cities yield many interesting applications
- Appearance and 3D information help

- Redundancy helps: Key to automatization

- Robust computation of a **semantic 3D model**

- Future: Use of MAVs
  - Flexible
  - Cheap

Videos/Code/Papers see

www.gpu4vision.org

aerial.icg.tugraz.at

Acknowledgments

Funding provided by:

- FWF under Project Higher Order Variatonal Methods for Computer Vision
- FIT-IT Program funded by BMVIT under Project VMGPU, PEGASUS, CONSTRUCT, HOLISTIC
- NVidea