Collaborative User Contributed Dense City Modeling

***Arnold Irschara*[1], *Christopher Zach*[2], *Manfred Klopschitz*[1] and *Horst Bischof*[1]**

[1]*Institute for Computer Graphics and Vision, Graz University of Technology*
[2]*Department of Computer Science, University of North Carolina, Chapel Hill*

Motivation

- **Goal:** Detailed, photorealistic city models
- **Approach:** Wiki principle, interested users contribute photographs
- **Novelty:**
  - Incremental structure-from-motion (SfM): growing forest of reconstructions
  - Global image and 3D model database
  - High quality dense models from unordered sets of images

Method

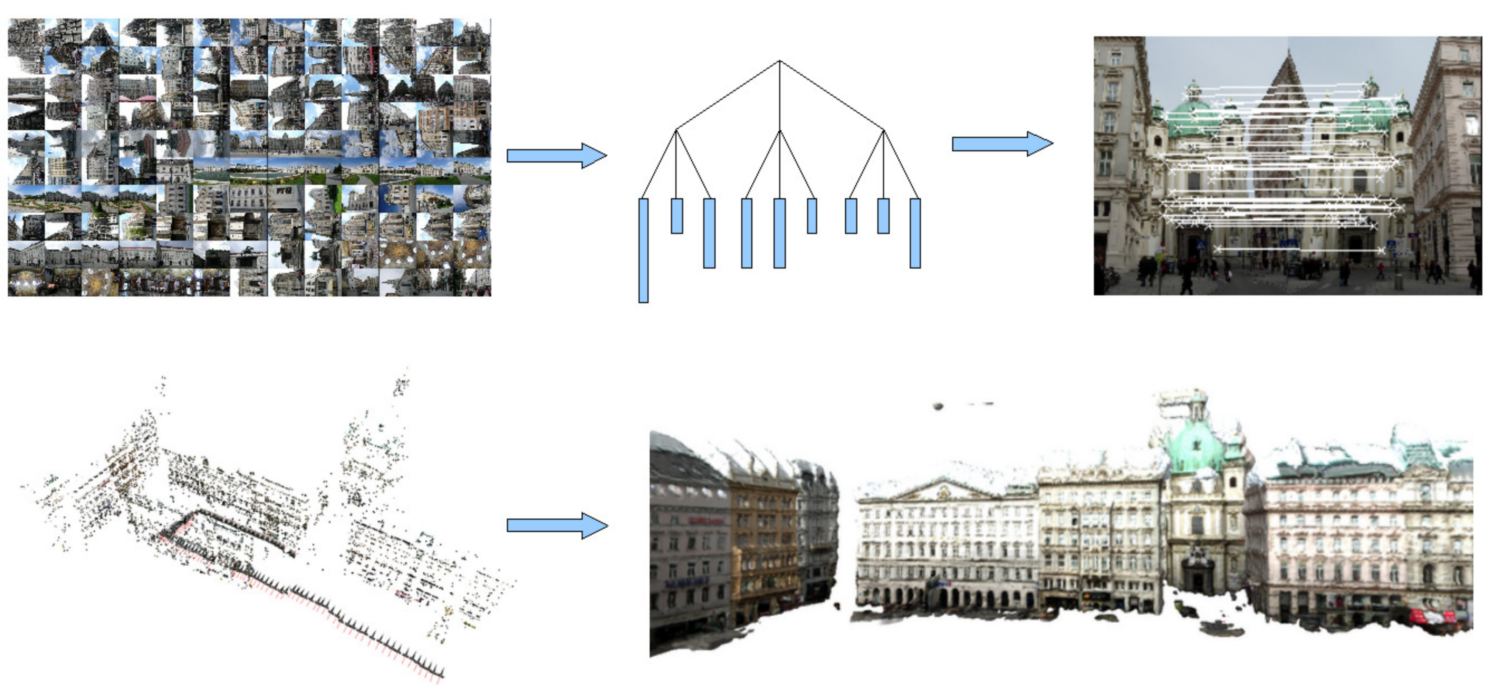

Figure 1: *Overview of the proposed reconstruction pipeline.*

1. **Flexible, simple calibration method**
2. **Reconstruction pipeline**
   - SIFT features (parallelized implementation)
   - Vocabulary tree image retrieval
   - Pair-wise image matching with geometric consistency check
   - Incremental structure-from-motion
     - Relative pose: Five-Point
     - Absolute pose: Three-Point
     - Euclidean bundle adjustment
3. **Dense modeling**
   - Fast GPU Plane-sweep
   - Robust depth map integration:

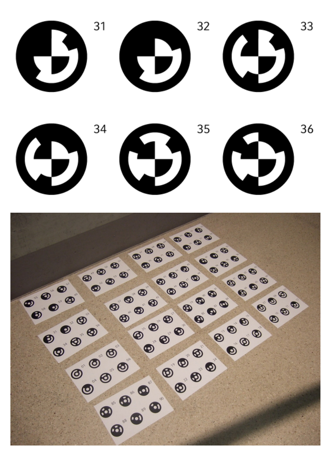

Figure 2: *Calibration targets.*

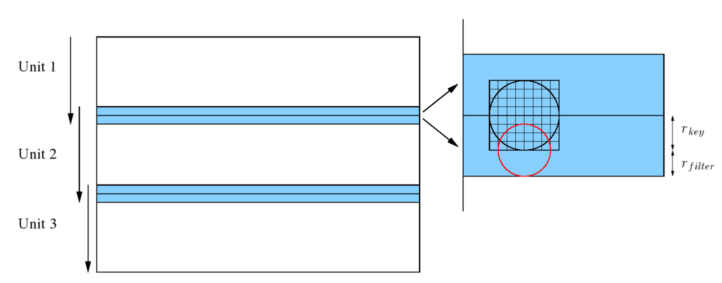

Figure 3: *SIFT parallelization.*

$$E^{TV-L^1}(u) = \int_\Omega \left\{ |\nabla u| + \lambda \sum_i |u - f_i| \right\} d\vec{x}$$

$f_i \ldots$ distance fields from depth maps $\quad |\nabla u| \ldots$ total variation

Results

- Two databases: *Vienna* (2640 images) and *Graz* (7181 images)
- Four different compact digital cameras $2-7$ megapixels; images taken at different days, varying illumination conditions
- Mean reprojection error, calibration: $0.05-0.2$, SfM: $0.3-0.5$ [pixel]
- Processing time $\sim 30s/frame$ (database size 1000), preprocessing (20%), matching (50%), bundle adjustment (30%)

Image Retrieval

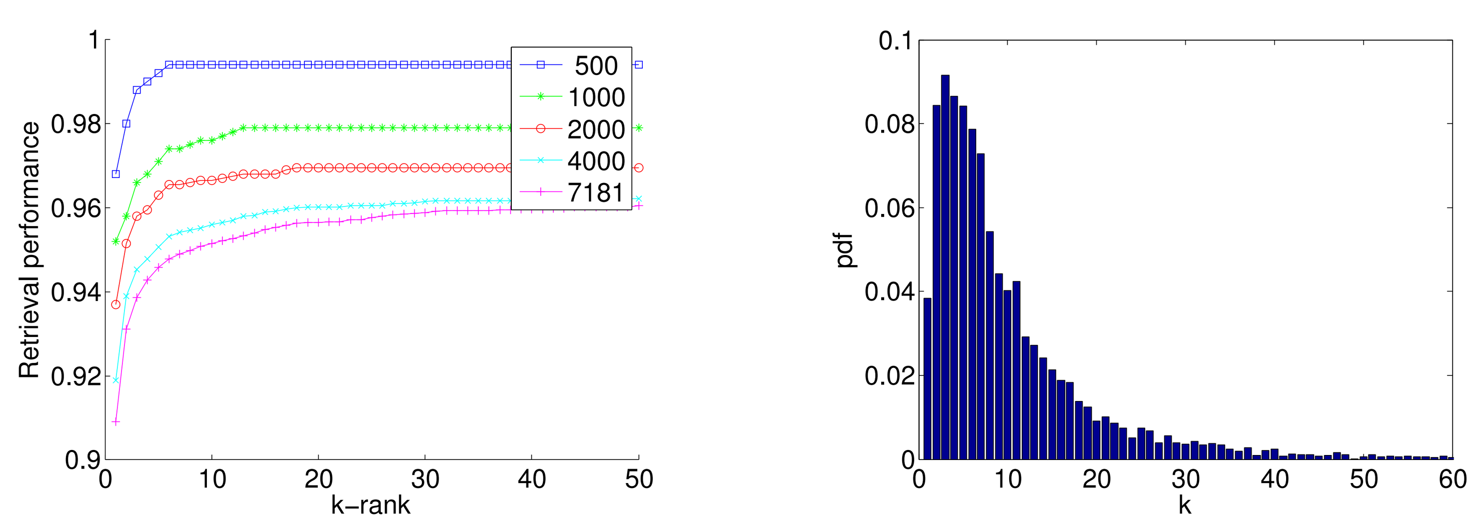

Figure 4: *Left: Image retrieval performance depending on database size. Right: Probability density function of the number of successfully verified epipolar neighbors for a query image in the* Graz *database. On average an image has an overlap with about eight images in the database.*

Sparse Reconstructions

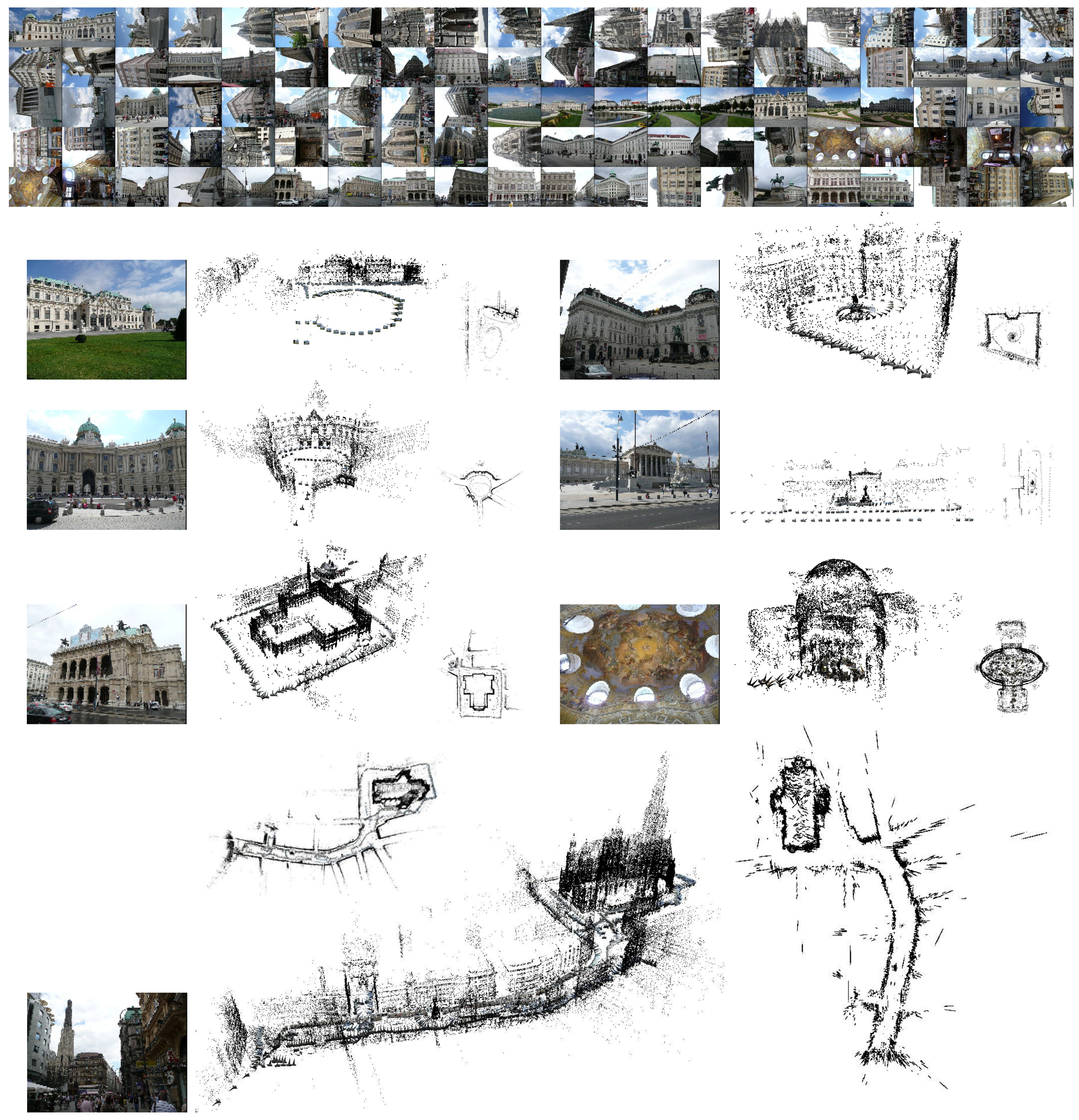

Figure 5: *Seven sparse reconstructions from the* Vienna *database comprising 2640 images (100 shown). Bottom row: St. Stephen's Cathedral reconstruction from 1330 images, 138410 triangulated points. Bottom right: covariances of the 3D points; average uncertainty in depth 25cm, in-plane 4cm.*

Dense Models

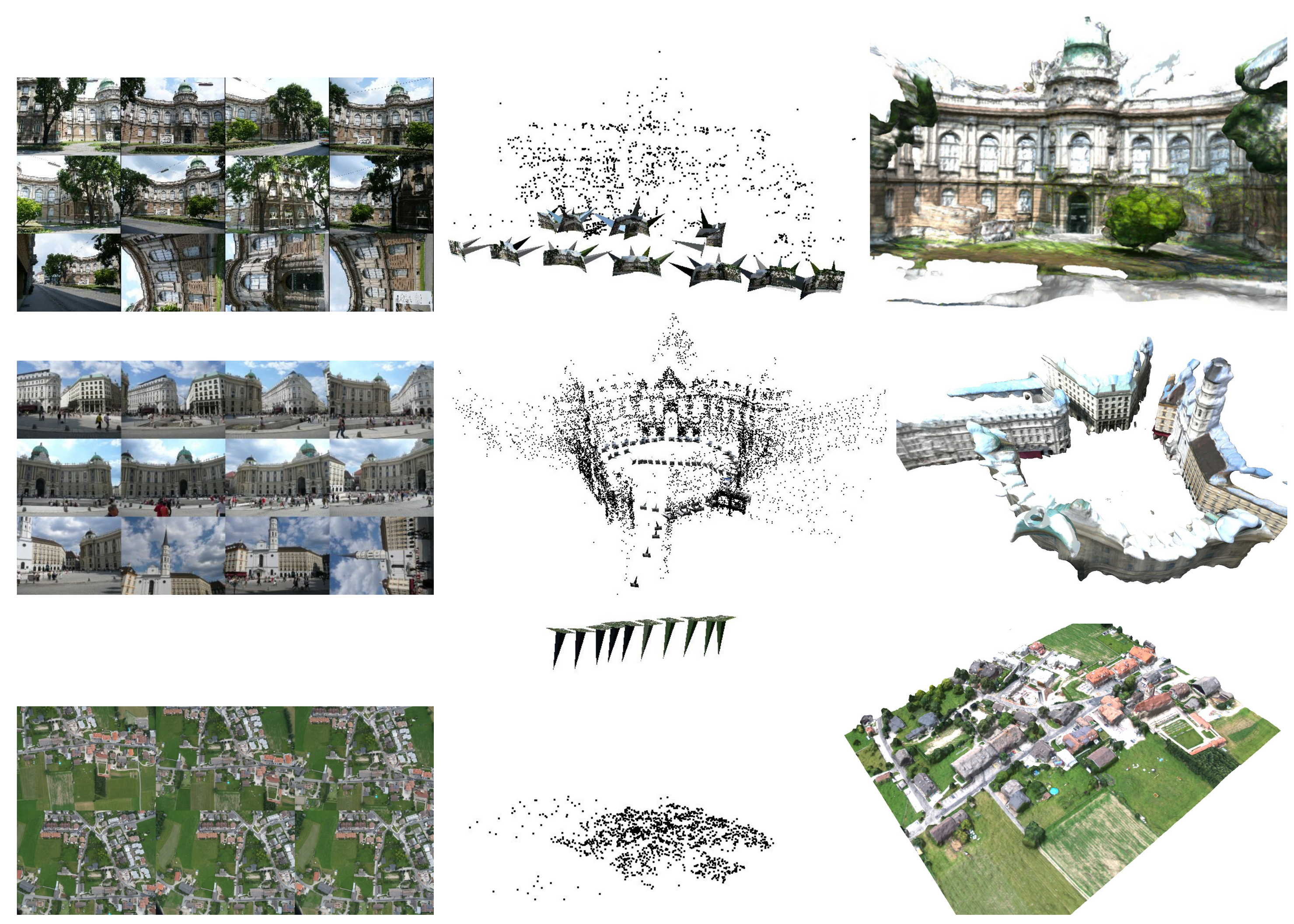

Figure 6: *From left to right: some sample images, sparse reconstructions and final dense models of different scenes: Alte Galerie at Landhausmuseum Joanneum (*Graz*, 49 images), Michaeler square (*Vienna*, 110 images) and a reconstruction from ten aerial images.*

References

[1] A. Irschara, C. Zach, and H. Bischof. Towards wiki-based dense city modeling. In *Workshop on Virtual Representations and Modeling of Large-scale environments (VRML)*, 2007.

[2] C. Zach, T. Pock, and H. Bischof. A globally optimal algorithm for robust TV-L^1 range image integration. In *IEEE International Conference on Computer Vision (ICCV)*, 2007.

[3] C. Zach, M. Sormann, and K. Karner. Scanline optimization for stereo on graphics hardware. In *International Symposium on 3D Data Processing, Visualization and Transmission (3DPVT)*, 2006.

Acknowledgements: This work is partly funded by the Vienna Science and Technology Fund (WWTF), the EU Integrated Project "IPCity" FP6-2004-IST-4-27571 and the Virtual Earth Academic Research Collaboration funded by Microsoft.