Context-based urban terrain reconstruction from UAV-videos for geoinformation applications

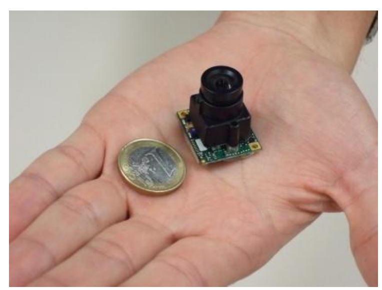
Dimitri Bulatov, Peter Solbrig, Hermann Groß, Peter Wernerus,
Endre Repasi and Christian Heipke

Fraunhofer IOSB

Introduction: Sensors

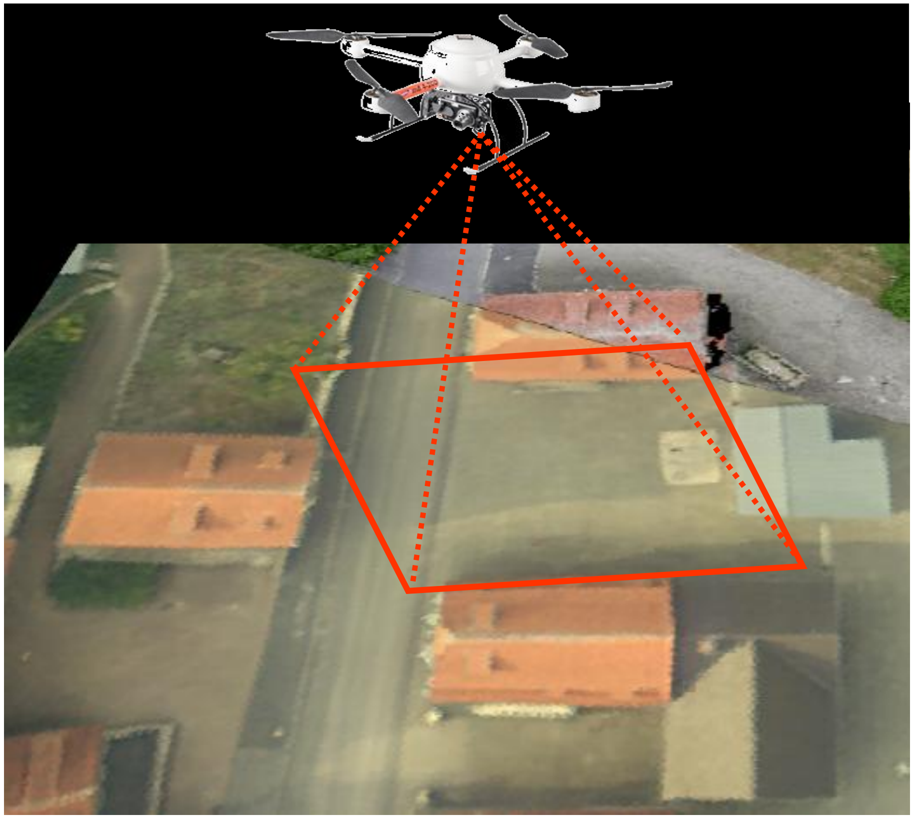
With daylight or infrared camera onboard

Reduced to minimal weight → GPS-data often not available

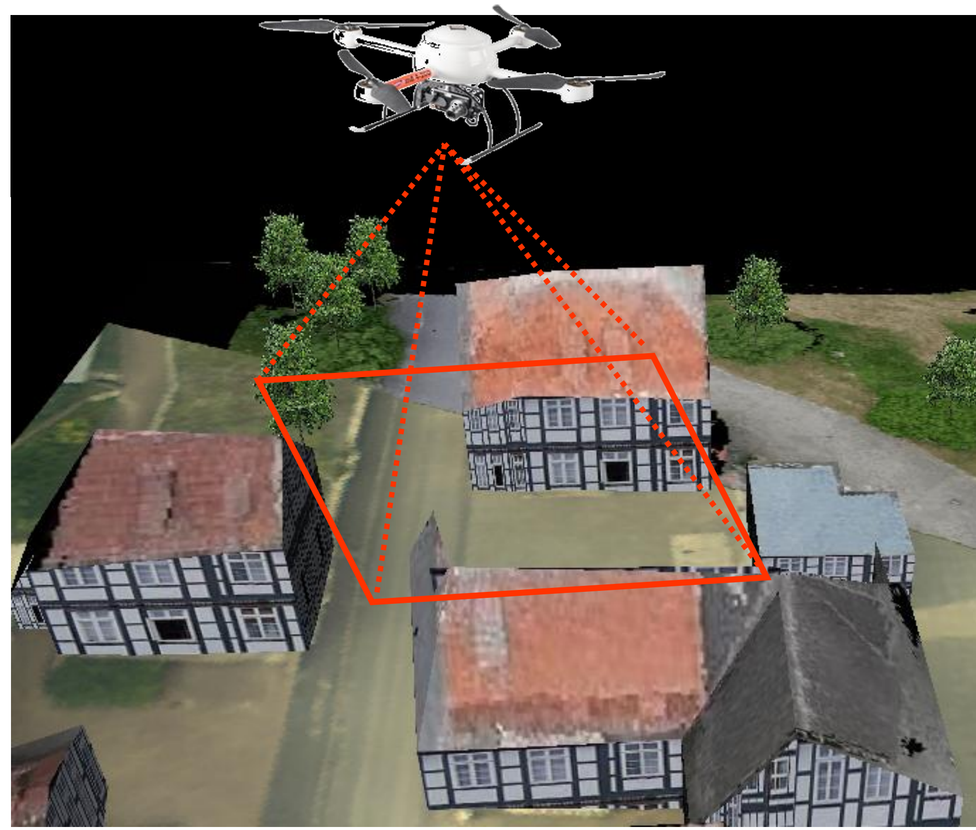


Introduction: Example of a mUAV-video

Introduction: 2D or 3D?



Our previous work (Solbrig et al. 2008)



Current work

Overview of the procedure

Step 1: Computation of camera orientation parameters

Step 2: Depth maps extraction

Step 3: Urban terrain reconstruction

Step 4: Registration and geo-referencing

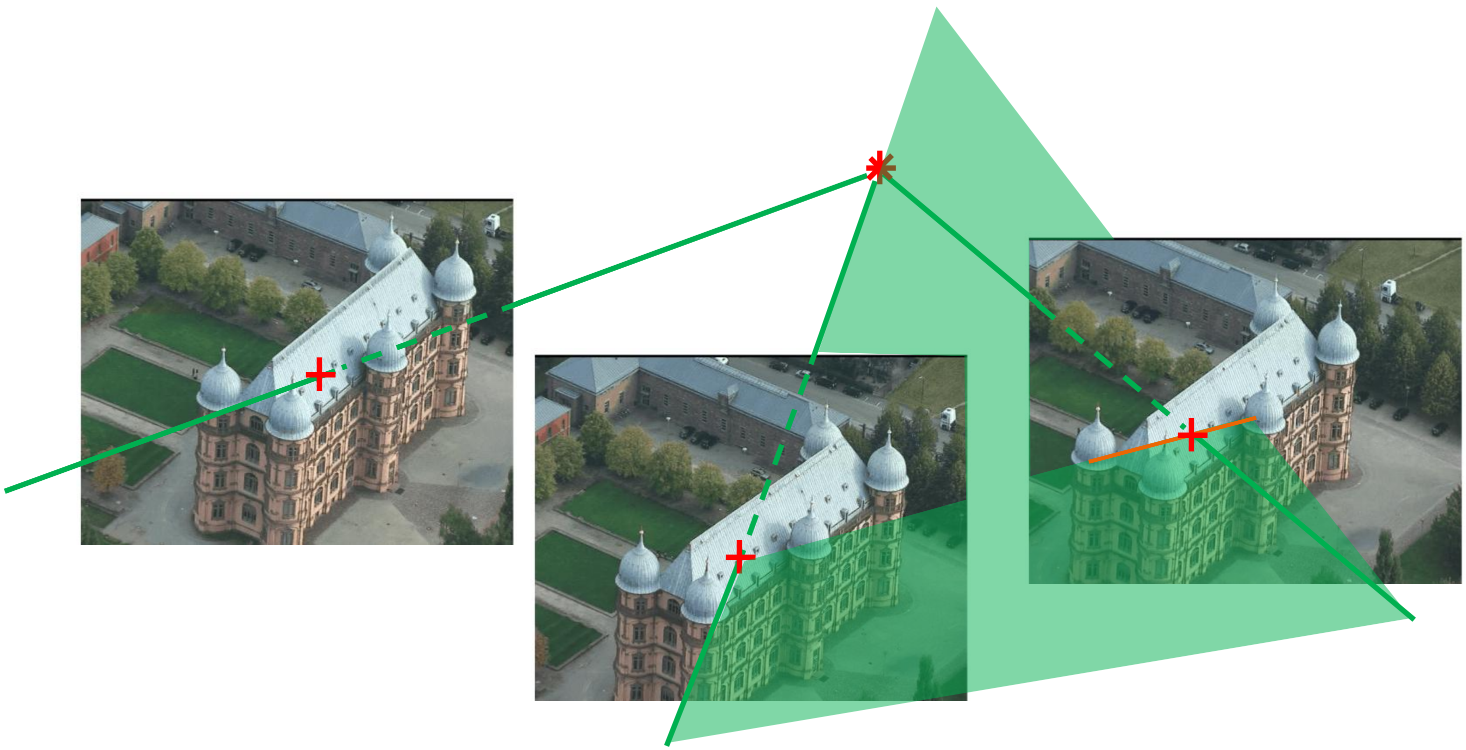
Step 1: Orientation

Computation of camera orientation parameters and a sparse 3D point cloud by means of a structure-from-motion approach (Bulatov 2008)

- Detecting and tracking characteristic points from frame to frame
- Computation of fundamental matrices and 3D points in a projective coordinate system (triangulation)
- Computation of camera matrices in the same coordinate system (camera resection)
- Self-calibration → transformation into an Euclidean coordinate system
- Bundle adjustment

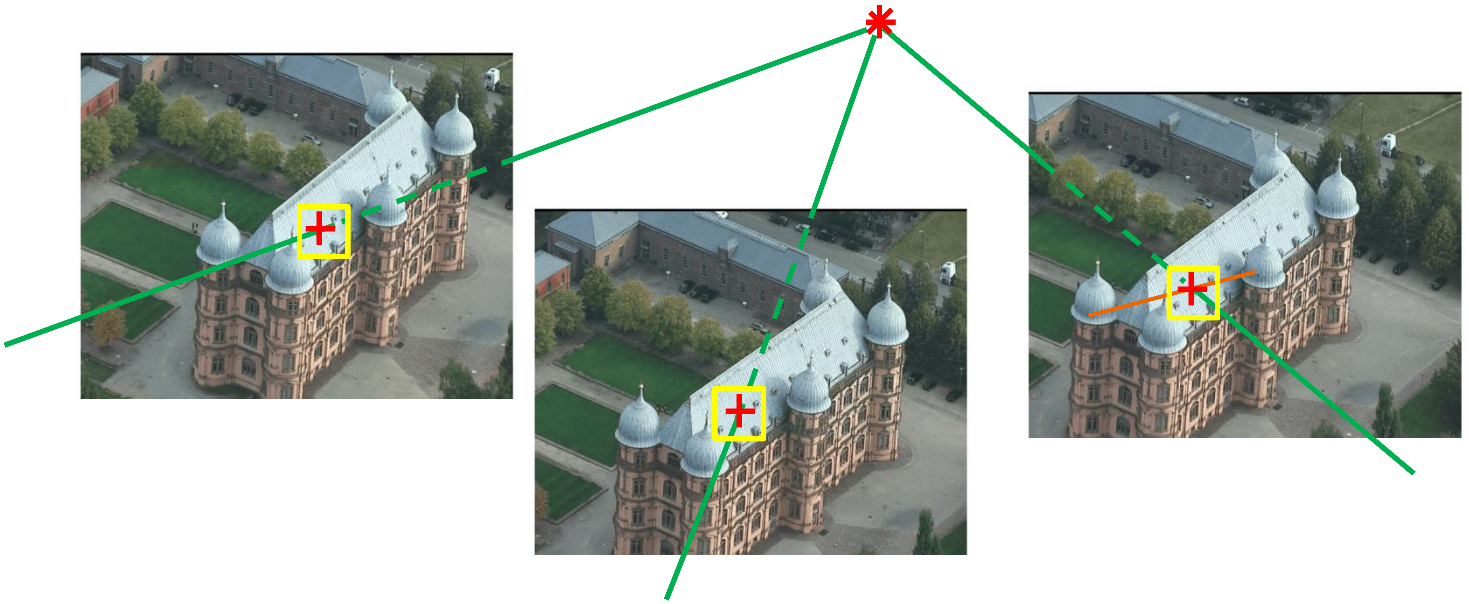
Step 2: Multi-view depth map computation

Given a subsequence of 5-10 images, compute for each pixel of the reference frame the depth value of the corresponding 3D point:



Step 2: Multi-view depth map computation

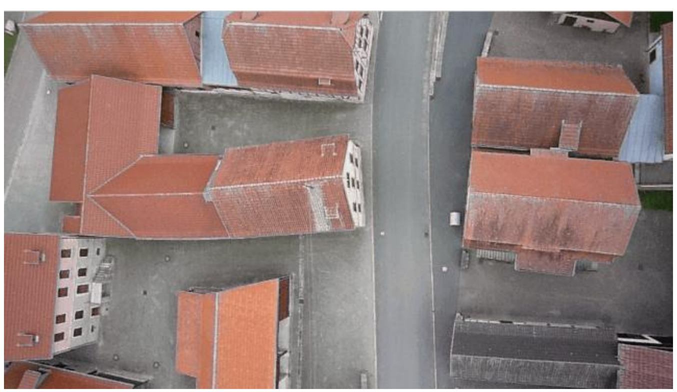
Given a subsequence of 5-10 images, compute for each pixel of the reference frame the depth value of the corresponding 3D point:

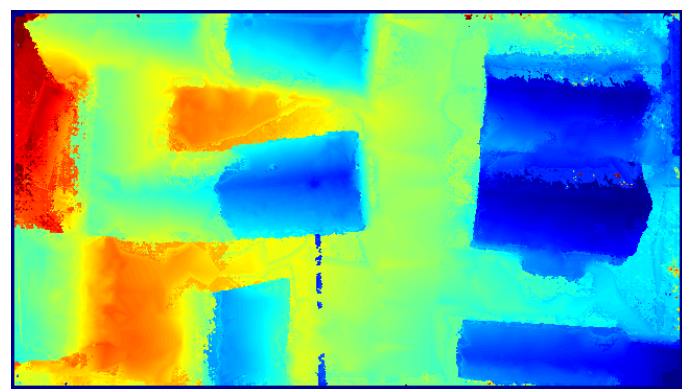


Step 2: Multi-view depth map computation

Our approach (Bulatov et al. 2011):

- Depth values discretization
- Data cost computation (e.g. gray values difference) and aggregation of low cost values
- Triangle-based smoothing (towards the depth map based on the triangular interpolation from already available 3D points)
- Semi-global optimization with 16 smoothness paths (Hirschmüller, 2005)
- Evaluation on triangles

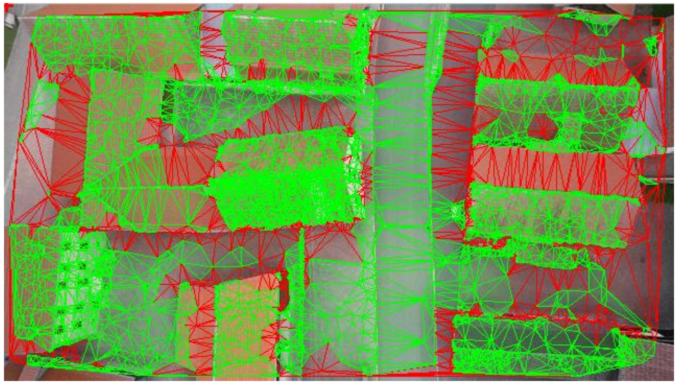


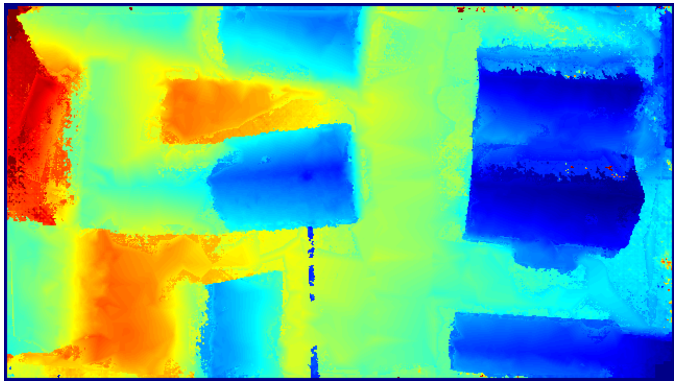


Step 2: Multi-view depth map computation

Our approach (Bulatov et al. 2011):

- Depth values discretization
- Data cost computation (e.g. gray values difference) and aggregation of low cost values
- Triangle-based smoothing (towards the depth map based on the triangular interpolation from already available 3D points)
- Semi-global optimization with 16 smoothness paths (Hirschmüller, 2005)
- Evaluation on triangles

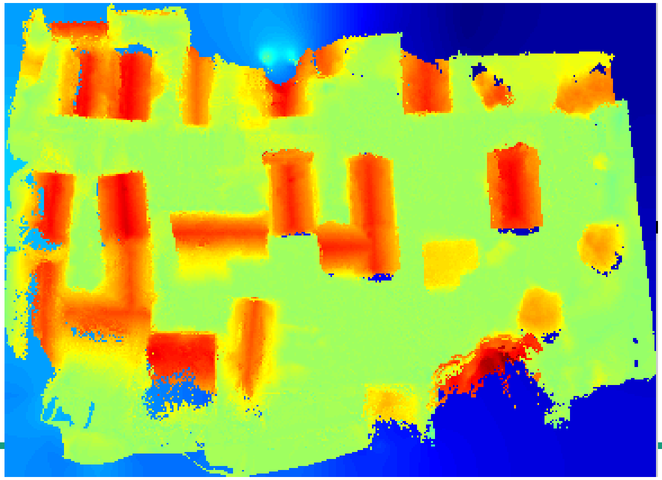




Step 3: Urban terrain modeling

Texture image rendering, DTM and DSM generation

- DSM generation with the median value of the *z*-coordinate in each cell of a rasterized image $z(x, y)$, the same with the texture image
- DTM is obtained by setting ground cells and then minimization of a suitable functional (our current choice: Neumann differential equation with border condition)
- Cells that satisfy the condition DSM – DTM > threshold are supposed to belong either to buildings or to vegetation



Step 3: Urban terrain modeling

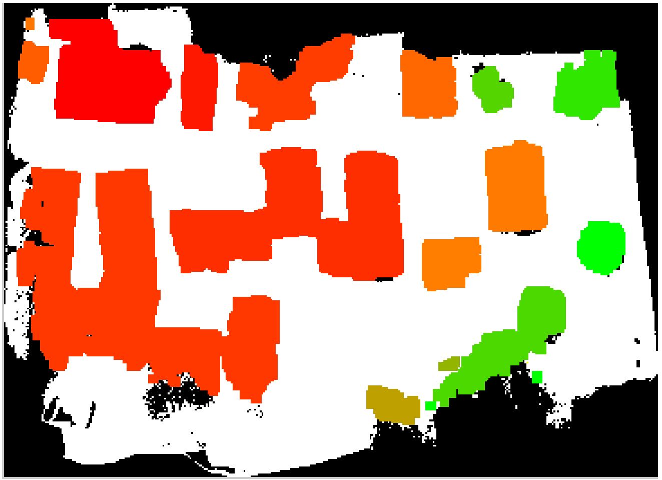
Identification of vegetation

- Straight lines computation (Burns et al. 1986) in reference images
- Each region R is assigned *lineness*-measure $lm(R) = {}^{1}\!/\!\sqrt{\text{area}(R)} \sum_{l \in \underline{R}} \text{length}(l)$, so that isolated trees are given low lineness measures
- Identification of tree-like pixels by means of color distribution

Step 3: Urban terrain modeling

Identification of vegetation

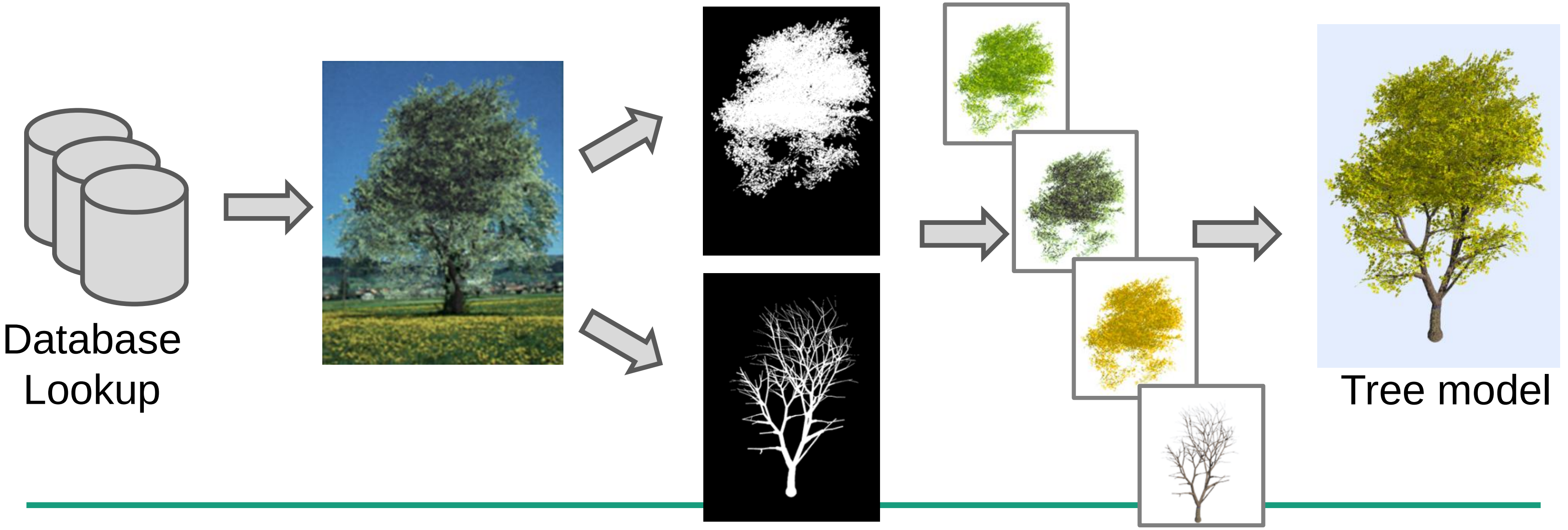
- Straight lines computation (Burns et al. 1986) in reference images
- Each region *R* is assigned *lineness*-measure $lm(R) = \frac{1}{\sqrt{\text{area}(R)}} \sum_{l \in \underline{R}} \text{length}(l)$, so that isolated trees are given low lineness measures
- Identification of tree-like pixels by means of color distribution



Step 3: Urban terrain modeling

Modeling Trees

- Default: Creating an artificial image where all pixels that do not belong to the tree are completely transparent, then superposition of 2-3 such images
- Database lookup (search for an appropriate model), model representation, and model adaption (e.g. seasonal changes) are optionally possible
- Height of the tree is given by the DSM

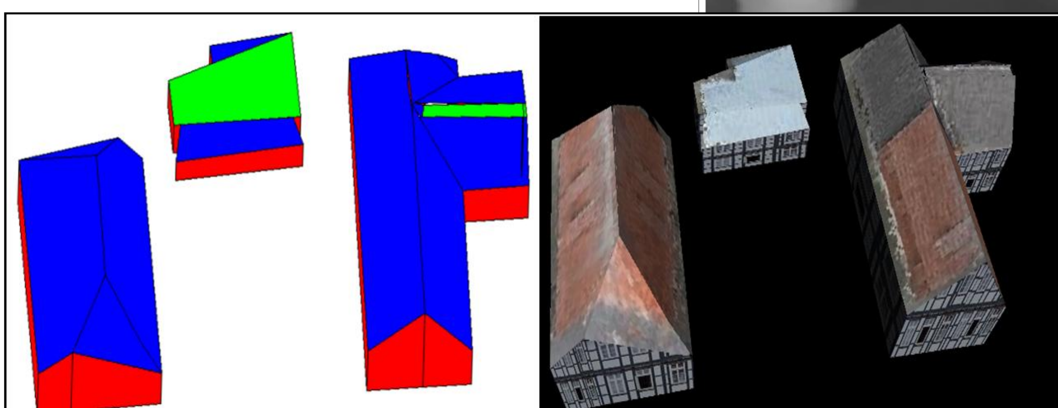


Fraunhofer IOSB

Step 3: Urban terrain modeling

Modeling buildings (Groß et al. 2005)

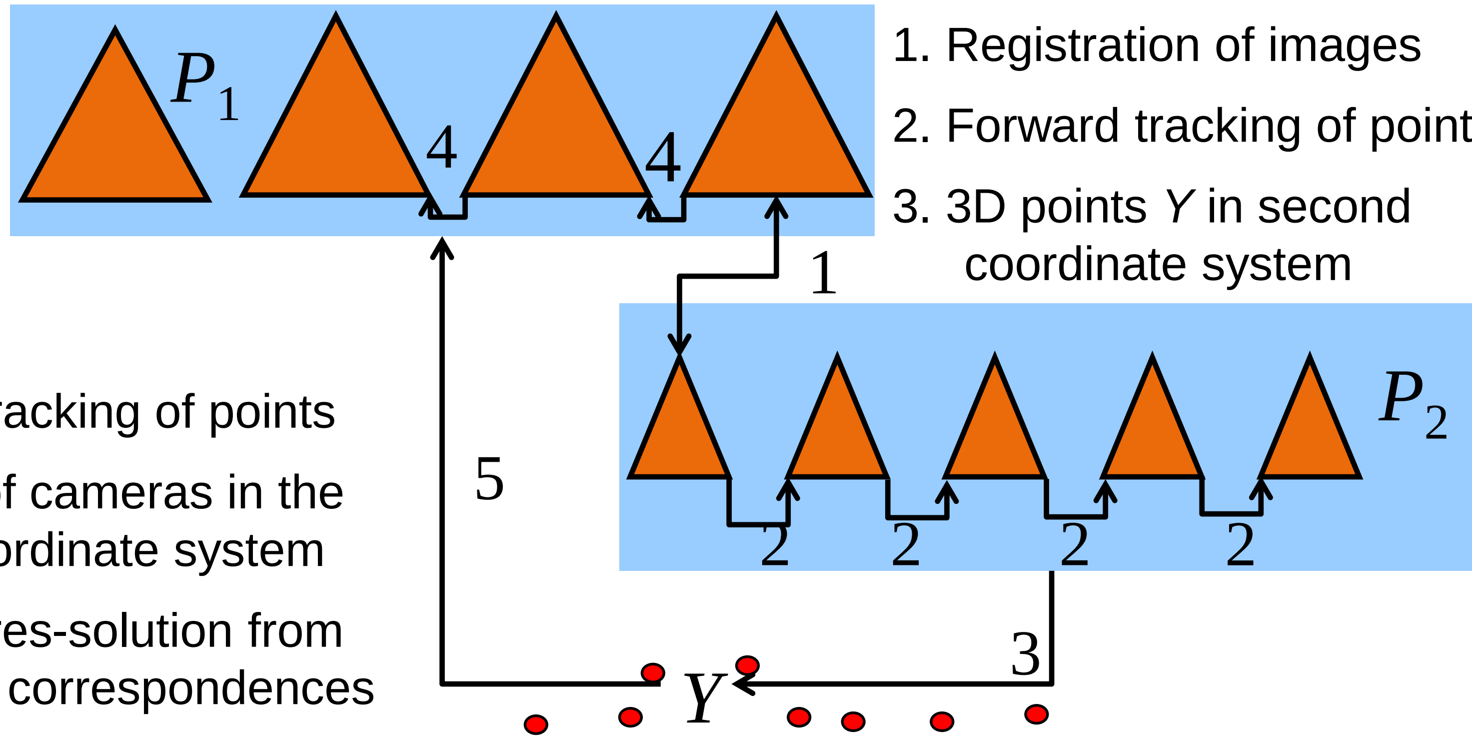
- Description of building outlines with rectangular polygons
- Roof detail analysis by means of normal vector computation and clustering, morphological operations and polygonization of roof surfaces
- Texturing



Step 4: Registration and geo-referencing

Registration of two different UAV-flights

- In absence of any internal navigation data, it is possible to obtain the spatial transformation connecting two Euclidean reconstructions from a pair of images covering the same region of the terrain.

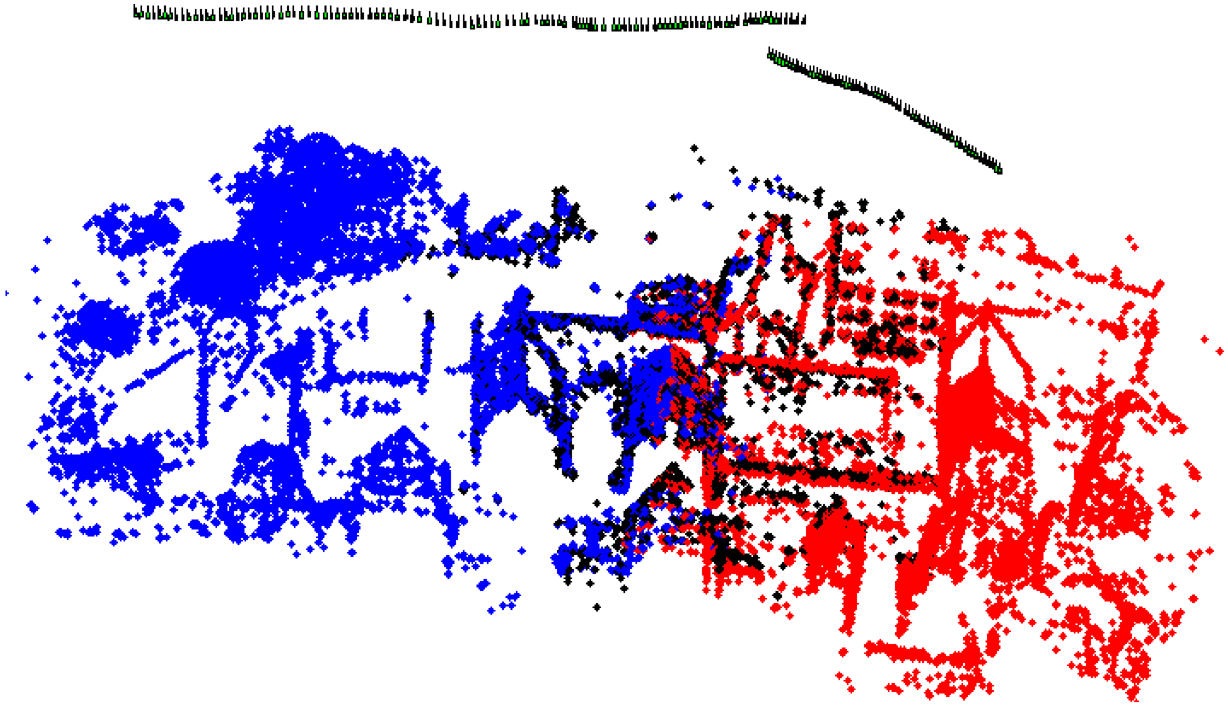


1. Registration of images
2. Forward tracking of points
3. 3D points *Y* in second coordinate system
4. Backward tracking of points
5. Resection of cameras in the second coordinate system
6. Least-squares-solution from ≥2 camera correspondences

Step 4: Registration and geo-referencing

Registration of two different UAV-flights

- In absence of any internal navigation data, it is possible to obtain the spatial transformation connecting two Euclidean reconstructions from a pair of images covering the same region of the terrain.



Step 4: Registration and geo-referencing

Automatic geo-referencing of the synthetic color image

- In case of Nadir videos, a 2D homography is a suitable registration tool with an orthophoto
- Because of different radiometry, a descriptor-based approach must be applied. Our choice: (Shechtman and Irani 2007)

Result

Outlook

- Quantitative evaluation of several reconstruction procedures
- Applications of 3D urban terrain reconstruction (e. g. change detection)

Conclusions

- A robust procedure for geo-referenced urban terrain modeling from UAV-based Nadir videos has been developed at Fraunhofer IOSB
- It is possible to model DTMs, buildings and vegetation. These components are easily exchangeable depending on season, day-time etc. This makes our tool suitable for AR and VR applications
- The modular structure of the procedure allows modifying parts of the programs without dismissing the entire concept
- Integration of new sources of information (GIS-data, short-wave infrared, LIDAR point clouds etc.) is suggestible
- Geo-referencing and parameter choice still require user's interaction, but these draw-backs can be corrected in the future

References

Bulatov, D., 2008. Towards Euclidean reconstruction from video sequences, In: *Int. Conf. Computer Vision Theory and Applications,* (2), pp. 476-483.

Bulatov, D., Wernerus, P., Heipke, C., 2011. Multi-view dense matching supported by triangular meshes, *ISPRS Journal of Photogrammetry and Remote Sensing*, accepted for publication.

Burns, J.B., Hanson, A.R., Riseman, E.M. 1986. Extracting straight lines. *Transactions on Pattern Analysis and Machine Intelligence*, 8(4), pp. 425-455.

Gross, H., Thönnessen, U., v. Hansen, W., 2005. 3D-Modeling of urban structures, In: *Joint Workshop of ISPRS/DAGM Object Extraction for 3D City Models, Road Databases, and Traffic Monitoring CMRT05*, In: *Int. Arch. of Photogrammetry and Remote Sensing*, 36, Part 3/W24, pp. 137-142.

Hirschmüller, H., 2008. Stereo processing by semi-global matching and mutual information, In: *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(2), pp. 328-341.

Shechtman, E., Irani, M., 2007. Matching local self-similarities across images and videos, In: *IEEE Conf. on Computer Vision and Pattern Recognition*, Minneapolis, USA.

Solbrig, P., Bulatov, D., Meidow, J., Wernerus, P., Thönnessen, U., 2008. Online annotation of airborne surveillance and reconnaissance videos, In: *11th International Conf. on Information Fusion,* Köln, Germany, pp. 1131-1138.