VocMatch: Efficient Multiview Correspondence for Structure from Motion

Supplementary Material – Online Resource 1

Michal Havlena and Konrad Schindler

Institute of Geodesy and Photogrammetry, ETH Zürich, Switzerland
{michal.havlena|schindler}@geod.baug.ethz.ch

Sparse 3D Models

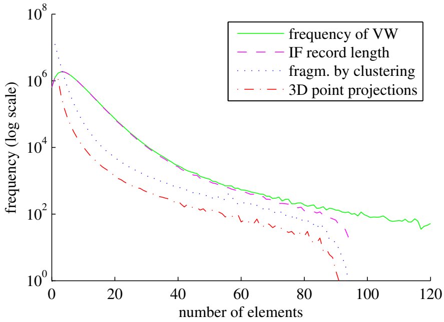
Sample reconstructed 3D models from both datasets are attached in VRML format, to give a more comprehensive impression of the reconstructed point clouds shown in Figures 1 and 5c of the paper. Due to filesize limitations we only include 3D points reconstructed from at least 3 projections (about one third of all 3D points). The models of OLDTOWN (Online Resource 2) and COLOSSEUM (Online Resource 3) therefore look sparser and cleaner than the screenshots of the full models in the paper.

The models can be viewed with any VRML viewer, *e.g.* Octaga Player [3] or FreeWRL [4]. Reconstructed camera poses are available as viewpoints in the models and can in most viewers be traversed using the `PgUp` and `PgDn` keys.

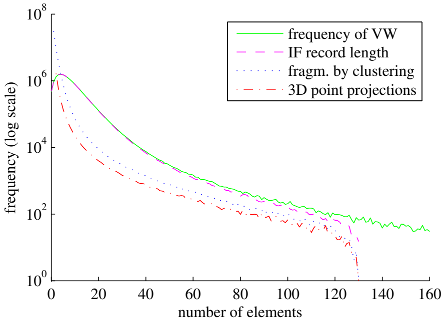
Track Length Statistics

To further evaluate the matching performance of the proposed method, Figure 1 displays histograms of the track lengths after every processing step, for both datasets. Green denotes the number of times different visual words actually occur in the respective dataset; Magenta denotes the lengths of inverted file records after filtering with the uniqueness and rarity constraints; Blue are record lengths after clustering into sub-models and breaking up tracks that span multiple sub-models; Red are the numbers of rays used to reconstruct the 3D points, *i.e.* after filtering out projections identified as outliers during either epipolar geometry estimation or bundle adjustment. Note, the scale of the y-axis is logarithmic.

When comparing the histograms, the gap between green and magenta shows the effect of the uniqueness and rarity conditions. One can clearly see the cut-off at the 1% rarity threshold (94 images for MERGED, 130 for ROME). The gap from magenta to blue shows how many tracks are split because they include images from different clusters, and the gap from blue to red shows how many of the feature points in a track are outliers that do not survive geometric verification. Note that a large fraction of the long tracks, which are especially important for accurate 3D reconstruction, are correct and therefore survive both clustering and geometric verification without being heavily decimated.



(a) MERGED



(b) ROME

Fig. 1. Histograms of track lengths. See text for details

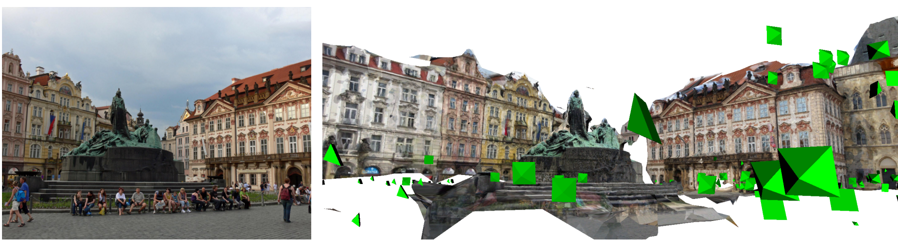
3D Surface Models

To verify that the camera poses obtained with our method are accurate enough for further 3D processing we have also performed dense matching and 3D surface reconstruction, using the method of [2] (authors' original implementation accessible through CMP SfM Web Service [1]). Sample textured 3D surface models are shown in Figure 2.

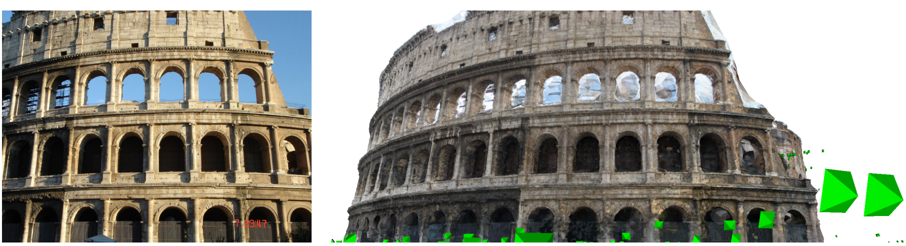
References

1. Heller, J., Havlena, M., Torii, A., Pajdla, T.: CMP SfM web service v1.0. Research Report CTU–CMP–2010–01, CMP Prague (2010)
2. Jancosek, M., Pajdla, T.: Multi-view reconstruction preserving weakly-supported surfaces. In: CVPR (2011)
3. Octaga Visual Solutions: Octaga Player: 3D viewer for interactive content – http://octagavs.com/softwarem/octaga-player (2012)
4. Stewart, J., et al.: FreeWRL X3D/VRML Viewer – http://freewrl.sourceforge.net (2013)

(a) Duomo Milano



(b) Old Town Square Prague



(c) Colosseum

(d) Fontana di Trevi

Fig. 2. 3D surface models, computed with the camera poses estimated from the matches found by the *VocMatch* method. *(Left)* One of the original images from the respective input dataset. *(Right)* Textured 3D surface model rendered from the same viewpoint. Green pyramids denote the poses of other cameras