Incremental Line-based 3D Reconstruction using Geometric Constraints

Manuel Hofer
hofer@icg.tugraz.at
Andreas Wendel
wendel@icg.tugraz.at
Horst Bischof
bischof@icg.tugraz.at

Institute for Computer Graphics and Vision
Graz University of Technology
Austria

Generating accurate 3D models of man-made objects and urban scenery from an image sequence is a challenging task. Traditional Structure-from-Motion (SfM) approaches often fail because of the high amount of untextured objects and wiry structures present. At the very least, these objects are poorly represented in the resulting point clouds. Since most man-made objects can be approximated by line segments, line-based 3D reconstruction techniques can be used as an alternative. While common appearance-based approaches usually deliver accurate results for a wide range of urban scenery, they cannot be directly applied to wiry structures. Since the resulting matching scores are based on the surroundings of a line segment, explicit matching would fail for these structures due to the changing background. To overcome such limitations, methods which do not rely on appearance-based line segment matching can be applied [1, 3]. Those algorithms, which assume known cameras, are usually based on generating a large set of possible 3D line segment hypotheses, gradient based scoring, and spatial clustering. These are very time consuming steps, which have to be computed offline after all cameras are available and oriented correctly.

We propose a novel line-based 3D modelling approach, which extends the principles presented in [1, 3] by incremental hypotheses clustering and geometric verification steps, without the need of time consuming scoring in the image space. We demonstrate how fusing this approach with an incremental point-based SfM [2] leads to an online 3D reconstruction method, which is able to cover wiry- and repeated structures, as well as solid objects.

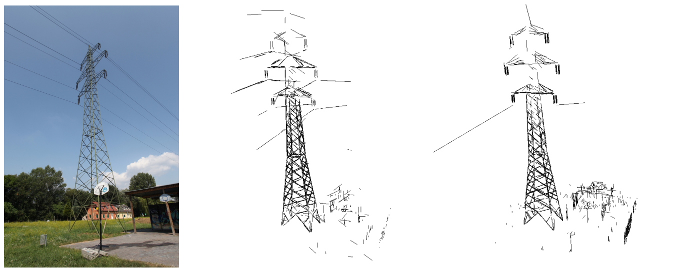
To perform incremental SfM we need to have an initial geometry involving at least two views. Therefore, given two images I_1 and I_2 and their respective sets of 2D line segments, we create an initial set of 3D line segment hypotheses H by computing all possible line segment matches between the two images. To limit the number of potential matches, we exploit epipolar constraints using the corresponding cameras C_1 and C_2. For each putative match we compute a 3D line segment K_h by triangulating the corresponding 2D line segments from I_1 and I_2. Each match results in a new hypothesis $h \in H$. Each hypothesis has a score $s(h)$ and a corresponding camera $C^*(h)$ defined as

$$s(h) = 1 - min\left\{ \left| \left\langle \frac{\overrightarrow{K_h}}{\|\overrightarrow{K_h}\|}, \frac{\overrightarrow{C_i}}{\|\overrightarrow{C_i}\|} \right\rangle \right| \right\}, \quad C^*(h) = \underset{C_i}{\operatorname{argmax}}(s(h)) \qquad (1)$$

where $\overrightarrow{K_h}$ is the directional vector of K_h, and $\overrightarrow{C_i}$ denotes the camera ray of camera C_i, and $\langle \cdot, \cdot \rangle$ is the inner product.

To perform incremental hypothesis merging for further incoming images, we need to define a spatial grouping radius r_{space}, which we derive from the image space dynamically to be scale invariant. Therefore, we define a maximum uncertainty σ in the image space. To bring this value to 3D space, we first compute a specific grouping radius $r_{space}(h)$ for each hypothesis $h \in H$. Therefore, we project the 3D line segment K_h back into the two supporting images. We then shift the resulting 2D line segments in the same orthogonal direction by σ, and triangulate them to obtain a shifted 3D line segment $\hat{K_h}$. The radius $r_{space}(h)$ is defined as the maximum distance between $\hat{K_h}$, and the infinite line passing through K_h. To be robust against imprecise triangulation, we compute a characteristic grouping radius $r_{space}(C_i)$ for each view, by using the median of all referenced radii. This allows us to adapt the system to severe viewpoint changes.

When a new image I_i is available we integrate it into our existing reconstruction, based on the current set of previously computed images. Therefore we have to define a set of neighboring views $N(I_i)$ for I_i. As above, we compute all possible matches for each line segment l, with the segments in $N(I_i)$. For each possible correspondence we try to add l to an existing hypothesis. We create a triangulated 3D line segment and compute the spatial distances to all candidate hypotheses. If we find a hypothesis h for which the spatial distance is below r_{space}, we consider l to be a part of h, and re-compute the score $s(h)$ and update the corresponding view $C^*(h)$. The estimated 3D line segment K_h has to be adapted as well, incorporating the newly triangulated line segment. For each possible match, for which we cannot find an existing hypothesis to be added to, we create a new hypothesis in the same way as during initialization.



(a) (b) (c)

Figure 1: (a) An example image from the *Pylon* sequence by [1] (106 images). (b) Reconstruction obtained by [1] (67 minutes, lines only). (c) Our reconstruction, $M = 10$, $\lambda = 4$ (9 minutes, including SfM).

After all line segments have been matched, we compute the current inlier set. Therefore, we sort the hypotheses $h \in H$ descending by the number of supporting line segments. If two hypotheses have the same number of line segments, we order them according to their reprojection error. We compute the current inlier set by iterating over the sorted hypotheses set. For h to be an inlier the following criteria have to be fulfilled: the number of supporting line segments has to be at least λ, and the score $s(h)$ has to be higher than 0.5. If this holds, the hypothesis is valid and all other hypotheses related to any of the segments referenced by h, are skipped during the iteration. If the validity criteria are not satisfied, the hypothesis is considered to be an outlier. This incremental grouping procedure prevents evaluation of a very large set of hypotheses at the end of the algorithm. However, using our proposed method might as well produce a huge hypotheses set for large image sequences. Hence, we need to remove unpromising hypotheses from time to time. To achieve this, we evaluate the number of supporting line segments in a hypothesis h compared to the number of views, which have been matched with the corresponding view $C^*(h)$. If there are less than λ line segments that agree on h, and $C^*(h)$ has been matched with at least $2 \cdot \lambda$ views, then hypothesis h is permanently removed from the hypothesis set.

Figure 1 shows results for a wiry structure, using the *Pylon* sequence from our previous work [1]. As we can see, our new approach has even less outliers due to the improved scoring approach and the automatic grouping radius selection. Additionally we manage to reconstruct the scene significantly faster, even though we also perform pose estimation, while in [1] the cameras are assumed to be known beforehand. For a more extensive evaluation about the parameters and additional testcases, we kindly refer to the full paper.

[1] M. Hofer, A. Wendel, and H. Bischof. Line-based 3D reconstruction of wiry objects. *Computer Vision Winter Workshop*, 2013.

[2] C. Hoppe, M. Klopschitz, M. Rumpler, A. Wendel, S. Kluckner, H. Bischof, and G. Reitmayr. Online feedback for structure-from-motion image acquisition. *British Machine Vision Conference*, 2012.

[3] A. Jain, C. Kurz, T. Thormaehlen, and H. Seidel. Exploiting global connectivity constraints for reconstruction of 3D line segments from images. *Conference on Computer Vision and Pattern Recognition*, 2010.