Descriptor Free Visual Indoor Localization with Line Segments

Branislav Micusik[1], Horst Wildenauer[2]
[1]AIT Austrian Institute of Technology. [2]Zeno Track GmbH.

We present a novel view on the indoor visual localization problem, where we avoid the use of interest points and associated descriptors, which are the basic building blocks of most standard methods. Instead, localization is cast as an alignment problem of the edges of the query image to a 3D model consisting of line segments. The proposed strategy is effective in low-textured indoor environments and in very wide baseline setups as it overcomes the dependency of image descriptors on textures, as well as their limited invariance to view point changes. The basic features of our method, which are prevalent indoors, are line segments. As we will show, they allow for defining an efficient Chamfer distance-based aligning cost, computed through integral contour images, incorporated into a first-best-search strategy. Experiments confirm the effectiveness of the method in terms of both, accuracy and computational complexity.

The visual localization problem stands for estimation of the 3D location of a query image in a given 3D model from visual information only. In recent years the localization problem has been attracting ever-increasing attention in the computer vision community. Yet, most of the proposed methods assume that the 3D model consists of a sparse set of 3D points associated with their image descriptors. On the algorithmic level, the localization problem is mostly a matching challenge, i.e. given 2D point features with their descriptors extracted from a query image, searching for tentative 2D-3D correspondences such that re-sectioning PnP algorithms can be applied.

Standard matching procedures based on comparison of image descriptors across images have a known bottleneck. The problem is rooted in the limited descriptor invariance to certain 3D transformations. Most descriptors are invariant to affine transformations and assume planarity of the supporting image regions from which they are computed. In practice, these relaxations were shown to be feasible and many successful techniques solving the localization problem on large city scale have emerged [3].

Most of the techniques are employed on outdoor scenery, but application in indoor scenarios typically results in a significant performance drop. The reason is that indoor scenes often exhibit a lot of windows, wiry structures, reflections and repetitions, as well as limited texture, see Fig. 1. All this causes standard procedures based on image descriptors to poorly perform indoors.

We introduce a novel proof-of-concept approach to the indoor localization problem, sketched in Fig. 1. Aside from ubiquitous interest point + descriptor methods and representing scenes by 3D point clouds, we build a 3D model consisting of 3D line segments. This allows the matching of the query image to the 3D model to be cast as an alignment problem between two sets of line segments. This strategy avoids a use of explicit, rather vague, line descriptors like in [1, 4]. Instead, our method harnesses geometric information provided by lines, and, despite the simplicity, copes well with extreme baseline changes that break other approaches. We show that evaluating alignment of two line sets in two images can be computed very efficiently with Chamfer matching through the integral contour strategy. Moreover, the proposed alignment cost allows for an effective tree-based search strategy to be employed, resulting in low computation time.

We present results on data taken in an entrance foyer of a modern building with many glass walls, wiry, and repetitive structures, see Fig. 1. Line segments were previously shown to be powerful features for prior 3D modeling of the scene [2], and, as we demonstrate, are as well for localization. We obtained superior performance over point-based localization methodologies on two types of images; those acquired with similar and those with very wide baseline viewpoints compared to the 3D modelling sequence.

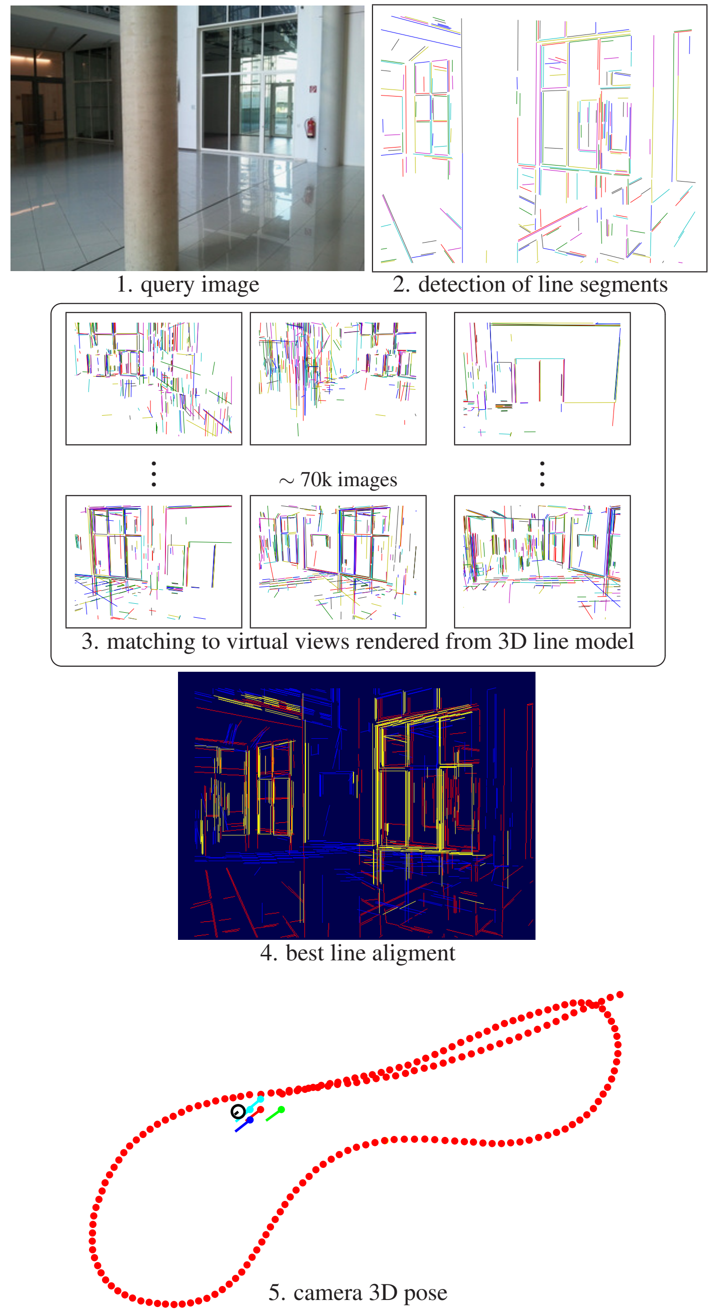

Figure 1: Concept of the localization. Query image (1) is sought for line segments (2) to match to virtual views (3). Best alignment with maximum number of overlapping lines in yellow between the detected in red and rendered blue line segments (4) gives the camera 3D pose (5). The best matches are depicted in red, green, blue and cyan short line with a dot, respectively, the 3D trajectory of a mapping sequence is shown in red, the ground truth pose of the query as a small black circle.

[1] H. Bay, V. Ferrari, and L. Van Gool. Wide-baseline stereo matching with line segments. In *Proc. CVPR*, 2005.

[2] B. Micusik and H. Wildenauer. Structure from motion with line segments under relaxed endpoint constraints. In *Proc. 3DV*, 2014.

[3] T. Sattler, B. Leibe, and L. Kobbelt. Improving image-based localization by active correspondence search. In *Proc. ECCV*, 2012.

[4] B. Verhagen, R. Timofte, and L. Van Gool. Scale-invariant line descriptors for wide baseline matching. In *Proc. Winter Conference on Applications of Computer Vision (WACV)*, 2014.

This is an extended abstract. The full paper is available at the Computer Vision Foundation webpage.