Joint Vanishing Point Extraction and Tracking

Till Kroeger[1], Dengxin Dai[1], Luc Van Gool[1,2]
[1]Computer Vision Laboratory, D-ITET, ETH Zurich. [2]VISICS, ESAT/PSI, KU Leuven

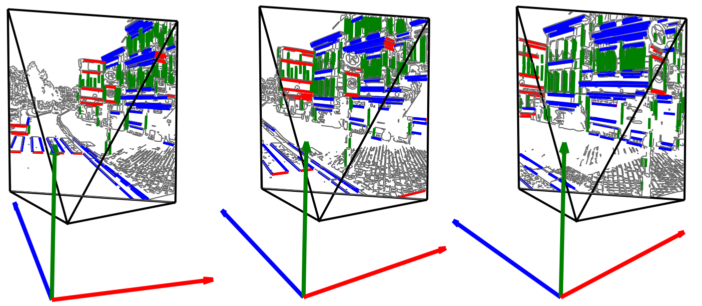

Figure 1: Tracked vanishing directions are shown together with associated imaged line segments in three frames of a sequence.

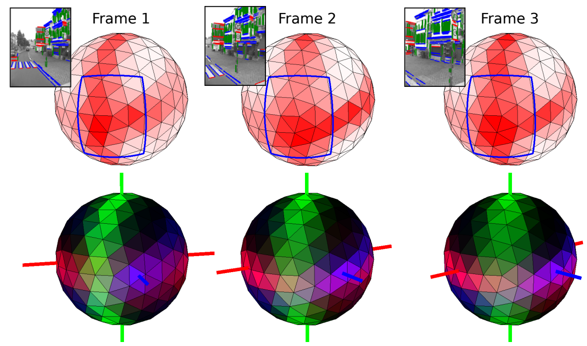

Figure 2: **Top:** Probabilistic spherical occupancy grid for VP location over a sequence of three frames. For visualization purposes probabilistic evidence in each bin is aggregated from all line segments. Line-VP associations are modeled as free variable in the proposed method. **Bottom:** Resulting VP tracks and color-coded association of line evidence to each VP. For visualization purposes a low discretization of 80 bins (instead of 5120) is used.

Introduction. A vanishing point (VP) is the point of convergence of a set of parallel lines in the imaged scene under a projective transformation. Man-made structures often consist of geometric primitives, such as multiple sets of parallel or orthogonal planes and lines in the scene. Because of this, the detection of VPs provides strong cues for the extraction of knowledge about the unknown 3D world structure. Detected VPs have been used as a low-level input to many computer vision tasks, such as 3D reconstruction, autonomous navigation, camera calibration and pose estimation.

Many applications, which take video sequences or unordered image sets as input, require VP estimates in every frame and VP identities across views or frames. Usually, when this is needed, the camera pose is assumed to be known for every frame [1, 5], thereby rendering the VP association across images simple, or separate steps for VP detection and tracking (particle filters [6], greedy assignment [4]) are used. Since pose knowledge can only be obtained through expensive odometry or external motion measurements, it will often not be available. Separate VP detection and tracking often results in missed detections or loss and re-initialization of VP tracks due to weak line support in some frames.

Method. We propose a method for the simultaneous VP extraction over all images of (calibrated) monocular image sequences with unknown camera poses. We borrow from recent advances in multi-target tracking [2, 7] and model the problem as a variant of a network-flow tracking problem. We discretize the set of possible VPs by creating a *probabilistic spherical occupancy grid*. For a short sequence such a grid is shown in Fig. 2.

The proposed discretization can now be used in a directed acyclic graph, similar to flow networks in multi-target tracking-by-detection. Using such graphs for probabilistic multi-target tracking, object detections or probable object locations are linked across time through pairwise object transition arcs. The best (i.e. most probable) set of object trajectories is extracted using Linear Programing.

Linear Program. Given I_t line segments in each of T frames the optimal VP tracks through J possible discretized VP locations are given by λ^*, which minimizes the LP objective function

$$f(\lambda) = \sum_t^T \sum_j^J \Bigg[\left(\sum_i^{I_t} \lambda_l(j,i,t) \cdot S_l(j,i,t) \right) + \left(\sum_{j'}^{J} \lambda_t(j,t,j') \cdot C_t(j,t,j') \right) + \lambda_b(j,t) \cdot C_u(j,t) + \lambda_s(j,t) \cdot C_s(j,t) + \lambda_e(j,t) \cdot C_e(j,t) \Bigg],$$

where $\lambda = [\lambda_l, \lambda_t, \lambda_b, \lambda_s, \lambda_e]$ are binary variables, indicating active ([l]ine, [t]ransition, VP [b]in, [s]tart, [e]xit) arcs. We set start, ending and VP bin unary costs uniformly to $C_s = C_r = C_u = -\log(1/J)$. The transition cost C_t between two VP bins j and j' increases with enclosed angle. The line-VP consistency scores S_l indicates how consistent line segment i fits to VP j and increases with enclosed angle. Costs are strictly positive and inhibit VP track creation. Scores can be negative and support VP track creation.

The LP is solved subject to the following constraints:

C1. Flow conservation: VP bins are maximally traversed by one track.

C2. Line-VP association: Only active VP bins have line segments associated to it. A line can at most be linked to one VP.

C3. Non-Maximum Suppression: For an active VP bin, we suppress other neighboring active VP bins.

C4. Angle preservation: For two active VP tracks, we require constancy of the enclosed angle.

C5. Orthogonality (optional): We can optionally enforce that all tracked VPs have to be mutually orthogonal at all times.

C1 and C2 are essential to enforce the graph structure. C3 is only needed for strong noise in line endpoints, and C4 for horizontal VPs near/at infinity. If a Manhattan world is assumed, C5 can be included.

Experiments. We evaluated our approach for three scenarios: joint VP detection and tracking on a new street-view dataset (48 sequences, total of 14K frames) for joint VP detection and tracking, VP detection and tracking when camera poses are known, and single-frame orthogonal VP detection on the York Urban Dataset (YUD) [3].

Acknowledgments: This work was supported by the European Research Council project *VarCity (#273940)*.

[1] Matthew E. Antone and Seth Teller. Automatic Recovery of Relative Camera Rotations for Urban Scenes. In *CVPR*, 2000.

[2] Jérôme Berclaz, François Fleuret, Engin Türetken, and Pascal Fua. Multiple Object Tracking using K-Shortest Paths Optimization. *PAMI*, 2011.

[3] Patrick Denis, James H. Elder, and Francisco J. Estrada. Efficient Edge-Based Methods for Estimating Manhattan Frames in Urban Imagery. In *ECCV*, 2008.

[4] Wael Elloumi, Sylvie Treuillet, and Rémy Leconge. Tracking Orthogonal Vanishing Points in Video Sequences for a Reliable Camera Orientation in Manhattan World. In *CISP*, 2012.

[5] Michael Hornáček and Stefan Maierhofer. Extracting Vanishing Points across Multiple Views. In *CVPR*, 2011.

[6] Peyman Moghadam and Jun Feng Dong. Road Direction Detection Based on Vanishing-Point Tracking. *IROS*, 2012.

[7] Li Zhang, Yuan Li, and Ramakant Nevatia. Global Data Association for Multi-Object Tracking Using Network Flows. In *CVPR*, 2008.