Real-time Joint Estimation of Camera Orientation and Vanishing Points

Jeong-Kyun Lee, Kuk-Jin Yoon
Computer Vision Laboratory, Gwangju Institute of Science and Technology.

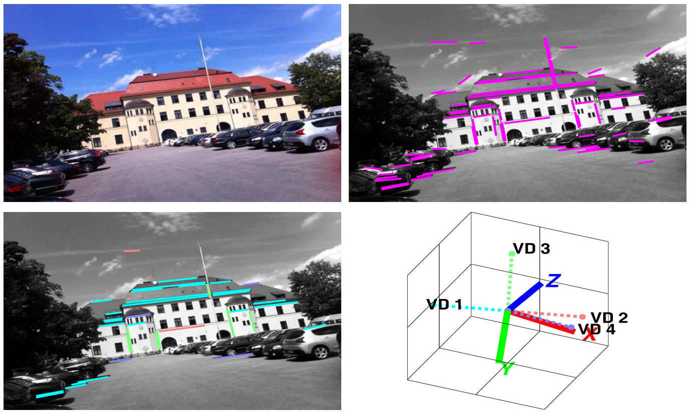

Figure 1: In a general scene (top left), given the line segments detected in the image sequence (top right), the proposed method clusters parallel lines (bottom left) and jointly estimates the camera orientation and vanishing points (bottom right).

A widely-used approach for estimating camera orientation in sequential images is to use point at infinity, *i.e.*, vanishing points (VPs), because a VP is a translation-invariant feature and therefore the rotation estimation can be more accurate by using VPs. The rotation estimation from VPs can be formulated easily by finding VP correspondences with the minimum distance and then solving a least squares problem for the geometric attribute, which is expressed as

$$\begin{bmatrix} \mathbf{d}' \\ 0 \end{bmatrix} = \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0}_{1\times 3} & 1 \end{bmatrix} \begin{bmatrix} \mathbf{d} \\ 0 \end{bmatrix} = \begin{bmatrix} \mathbf{R}\mathbf{d} \\ 0 \end{bmatrix}, \tag{1}$$

where a vanishing direction (VD) $\mathbf{d}$, back-projection of a VP, is transformed into $\mathbf{d}'$ by a 4×4 matrix representing rotational and translational transformations $\mathbf{R}$ and $\mathbf{t}$ in Euclidean 3D space. This method does not work in practical applications since VP correspondences are unreliable due to spurious or noisy line segments. In recent years, the literature [1, 2, 4] has considered to enforce an orthogonal constraint between VPs, called the Manhattan world constraint. This allows us to quickly find the reliable VP correspondences and thus improves the accuracy and robustness of the rotation estimation even if noisy line segments are extracted. Nevertheless, these methods still have significant limitations: 1) finding VP correspondences is based on the Manhattan World constraint, and 2) the rotation estimate is prone to be inaccurate and unstable in the presence of many noisy or spurious parallel line segments.

In this paper, we propose a novel method that jointly estimates camera orientation and VPs as shown in Fig. 1. The proposed method has several advantages: no demand for the Manhattan world assumption, robust and accurate rotation estimation against noisy line segments, and real-time processing. In addition, a feature management technique is also proposed to remove false positives of clustered parallel line segments or classify newly detected lines, and thus the proposed method can be performed well in practice.

The proposed method for the joint estimation of camera orientation and VPs is based on nonlinear Bayesian filtering, the extended Kalman filter (EKF) in this work. The state vector $\mathbf{x}$ is defined by $\mathbf{x} = \left[\mathbf{x}_v^{\mathrm{T}},\ \mathbf{y}_1^{\mathrm{T}},\ \mathbf{y}_2^{\mathrm{T}},\ \cdots\right]^{\mathrm{T}}$, where $\mathbf{x}_v$ is a camera state vector and $\mathbf{y}_i$ a VD vector. We consider the constant angular velocity model as a system model of the Bayesian filtering since we aim only at estimating the camera orientation in an image sequence. A measurement model is designed from two geometric properties: one is (1) and another that the VD of a line is orthogonal to the normal vector of the plane composed of the line and the center of projection as shown in Fig. 2. The measurement model h_{ij} for the i-th VD of the state vector and the j-th line feature is defined by

$$h_{ij} = \mathbf{d}_i^{\mathrm{T}} \mathbf{R}_{WC} \mathbf{n}_{ij}, \tag{2}$$

where $\mathbf{d}_i$ is the i-th unit VD vector, $\mathbf{R}_{WC}$ the rotational matrix for the camera orientation, and $\mathbf{n}_{ij}$ the unit normal vector of the plane defined by the j-th line feature parallel to the i-th VD in the camera coordinates. A value of the measurement model h_{ij} itself is considered as a measurement residual, which is used to update the current state. The measurements are acquired from several line segment tracking methods that can be selected depending on whether the performance should be accurate or fast. The methods are detailed in the paper. In addition, some outliers are rejected by using an outlier rejection technique in similar manner of [3] and spurious line segments are deleted by the propose feature management technique. Since the techniques consider the temporal consistency and the geometric correlation between a line, a VP, and camera orientation, the performance of the proposed method can be more accurate and robust than that of the existing methods that match unreliable and inconsistent VPs extracted from noisy or spurious line segments.

In the experiments, we demonstrate the superiority of the proposed method through an extensive evaluation using synthetic and real datasets and comparison with other state-of-the-art methods. Consequently, the proposed method provides a highly accurate and robust performance in the real-world scenes.

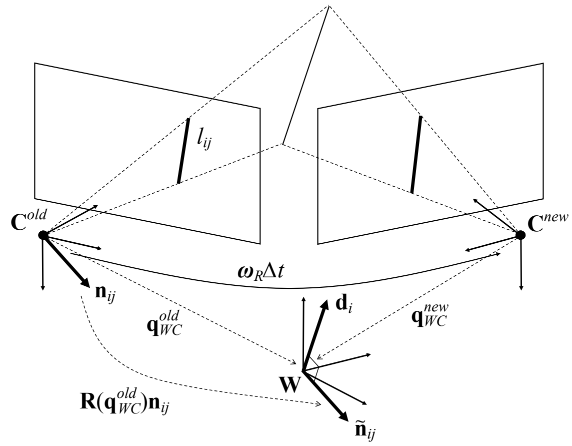

Figure 2: When the camera moves from $\mathbf{C}^{old}$ to $\mathbf{C}^{new}$, the camera orientation $\mathbf{q}_{WC}^{old}$ is changed to $\mathbf{q}_{WC}^{new}$ according to the rotation $\omega_C \Delta t$. A normal $\mathbf{n}_{ij}$ corresponding to a line l_{ij} is orthogonal to a vanishing direction $\mathbf{d}_i$ of the line l_{ij}.

[1] J.-C. Bazin, C. Demonceaux, P. Vasseur, and I. Kweon. Rotation estimation and vanishing point extraction by omnidirectional vision in urban environment. *IJRR*, 31(1):63–81, 2012.

[2] J.-C. Bazin, Y. Seo, C. Demonceaux, P. Vasseur, K. Ikeuchi, I. Kweon, and M. Pollefeys. Globally optimal line clustering and vanishing point estimation in manhattan world. In *Proc. CVPR*, 2012.

[3] Javier Civera, Oscar G Grasa, Andrew J Davison, and JMM Montiel. 1-point ransac for ekf-based structure from motion. In *IROS*, 2009.

[4] W. Elloumi, S. Treuillet, and R. Leconge. Real-time camera orientation estimation based on vanishing point tracking under manhattan world assumption. *JRTIP*, pages 1–16, 2014.

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