POSE ESTIMATION FOR URBAN ENVIRONMENTS

Daniel Buhrig and Peter Eisert

{daniel.buhrig, peter.eisert}@hhi.fraunhofer.de – Fraunhofer HHI / Humboldt University, Berlin, Germany

Abstract

We propose a robust method for pose estimation in images of urban environments which exploits the architectural properties of men made buildings (i.e. orthogonal structures such as facades with repeating elements). It is targeted e.g. at mobile Augmented Reality (AR) applications, where purely sensor-based estimates of device orientation need to be refined for precise visualization. Images of urban environments contain long and straight lines which allow the determination of the dominant vanishing directions. An intrinsic camera calibration is used to constrain the calculation of three orthogonal vanishing points (VP). The detected VPs define the rotation of the objects and allow a refinement of inaccurate compass data. With support of GPS data and layout of the surroundings, the object rotation in the world coordinate system can be determined.

Keywords: pose estimation, vanishing points.

1 Preprocessing

The upward direction $\boldsymbol{d}_{up}$ in image I often correlates with the up-vector in the real world since users will typically chose to take pictures where vertical lines of world objects stay vertical in the image as well. The focal length of cameras in mobile devices is fixed and the focus in most outdoor applications is close to infinity. This allows the prior determination of the internal camera matrix $\boldsymbol{K}$.

Line Detection. A Hough transformation based method extracts a set L of long straight line segments from I.

Vanishing Point Computation. The vanishing point computation corresponds to the calculation of a common intersection point of a set of lines (supposed to be parallel in 3D). Since measurement noise never allows for a perfect intersection, a maximum likelihood estimate is applied [1]. A set of lines with a single intersection point is fitted to the measured set of lines by minimizing the sum of squared orthogonal distances from the endpoints of the measured lines.

2 Vanishing Points Detection

A set of lines L_v which are within a certain angular deviation of the upward direction $\boldsymbol{d}_{up}$ of I are used to determine the first vanishing point $\boldsymbol{v}_1$. This allows for the calculation of a line $\boldsymbol{l}_h = \omega \boldsymbol{v}_1$ [1] (which can be viewed as the horizon in the image), where $\omega = \boldsymbol{K}^{-T}\boldsymbol{K}^{-1}$ is the image of the absolute conic. The line $\boldsymbol{l}_h$ corresponds to the vanishing line of the scene ground plane. Since the 3D lines of the other two dominant vanishing directions are parallel to the scene ground plane, their imaged 2D lines must intersect $\boldsymbol{l}_h$ at the position of their respective VPs. Thus, the search for $\boldsymbol{v}_2$ and $\boldsymbol{v}_3$ is reduced to a 1D problem.

Point $\boldsymbol{p}_i$ corresponds to the intersection point of $\boldsymbol{l}_i \in L \backslash L_v$ with $\boldsymbol{l}_h$. For each $\boldsymbol{p}$, the set $M(\boldsymbol{p}_i) = \{\boldsymbol{p}_j \mid \|\boldsymbol{p}_i - \boldsymbol{p}_j\| < \alpha\}$ contains all points which are within the neighbourhood defined by the window size $\boldsymbol{\alpha}$. Function $f(\boldsymbol{p}_i) = |M(\boldsymbol{p}_i)|$ represents the number of neighbouring points for the intersection point $\boldsymbol{p}_i$ (see Fig. 1).

The peaks of $f(\boldsymbol{p})$ indicate where the majority of lines intersect $\boldsymbol{l}_h$ at the respective position. The preliminary positions of VPs $\boldsymbol{v}_2$ and $\boldsymbol{v}_3$ are located on $\boldsymbol{l}_h$ where the peaks minimize the deviation from the orthogonality constraint $|\boldsymbol{v}_2^T \, \omega \, \boldsymbol{v}_3|$ [1]. The result is refined by taking the lines which intersect $\boldsymbol{l}_h$ in a close proximity of the preliminary positions and recalculating the common intersection point.

The homogenous representation $\hat{\boldsymbol{v}}_i$ of the three orthogonal VPs $\boldsymbol{v}_i$ define the rotation $\boldsymbol{R} = [\hat{\boldsymbol{v}}_1 \, \hat{\boldsymbol{v}}_2 \, \hat{\boldsymbol{v}}_3]$ of the object. However, the a-priori object orientation in its local coordinate frame is unknown. GPS data and a map representation provide a rough position for the point of view in the world coordinate system relative to the object. This allows the determination of object rotation in the global world coordinate system.

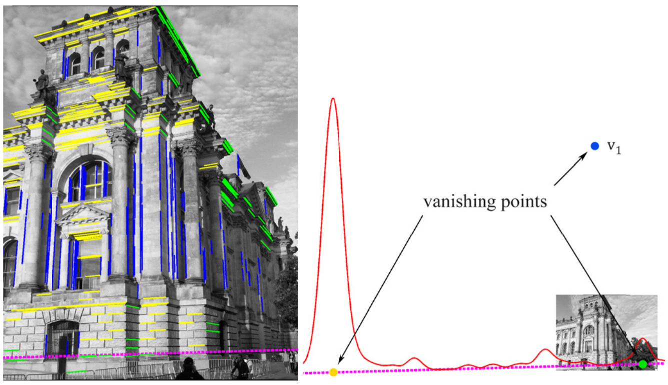

Fig. 1 left: line colors (green, blue, yellow) indicate the respective dominant vanishing direction. The magenta line corresponds to the horizontal line l_h; right: plot of function f representing the number of neighbouring intersection points.

Acknowledgements

This research has been funded by the European 7th Framework Program, under grant VENTURI (FP7-288238).

References

[1] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision,* Cambridge University Press, UK, 2000.