

# Egomotion Estimation Using Binocular Spatiotemporal Oriented Energy

Hao Zhong  
zhhao31@gmail.com  
Richard P. Wildes  
wildes@cse.yorku.ca

Department of Electrical Engineering  
and Computer Science  
York University  
Toronto, Ontario, Canada

Camera egomotion estimation is concerned with the recovery of a camera's motion (e.g., instantaneous translation and rotation) as it moves through its environment. It has been demonstrated to be of both theoretical and practical interest and therefore has been the subject of considerable research. In the light of previous work on egomotion estimation, the present paper makes three contributions. First, a novel algorithm for egomotion estimation based on binocularly matched orientation measurements in visual spacetime,  $(x, y, t)$ , is presented. Basing the estimation on spacetime orientation measurements makes it possible to recover egomotion without the need to establish temporal correspondences or convert disparity into 3D world coordinates. Second, the algorithm is based on a formal analysis that relates binocularly matched orientations in visual spacetime to camera egomotion. It appears that this relationship has not been presented previously. Third, the estimation algorithm has been realized in software and evaluated quantitatively on a novel laboratory dataset with groundtruth as well as qualitatively on both indoor and outdoor real-world datasets. Performance is evaluated relative to comparable alternative algorithms. As part of the evaluation, a new binocular video dataset is introduced that includes groundtruth egomotion and is available to the community. Figure 1 overviews the proposed approach.

**Technical approach.** Binocularly matched, local measurements of spatiotemporal oriented energy (SOE) serve as the data on which the developed approach to egomotion estimation operates. SOEs provide an integrated way to capture spatial appearance and temporal characteristics of an image sequence [3]; therefore, they have the potential to support recovery of egomotion via consideration of temporal dynamics of spatial information as a function of egomotion. For present purposes, local SOE measurements are recovered separately in the left and right streams of an input binocular video via convolution with a bank of Gaussian second derivative filters and their Hilbert transforms, which are combined in quadrature to yield energy measurements. To establish correspondences between points in the left and right image sequences any reliable algorithm for establishing binocular correspondence could be applied on a framewise basis to the original image sequences. Here, since SOEs are available and previously have been shown useful for stereo video matching [3], that matching approach is applied to establish the needed left-right correspondences.

Given matched pointwise SOEs in the left and right image streams, egomotion estimation is based on an analysis that parameterizes binocularly corresponding orientations in visual spacetime,  $(x, y, t)$ , in terms of camera egomotion. For these purposes, camera egomotion is defined in terms of instantaneous six degrees-of-freedom translation,  $\mathbf{T}$ , plus rotation,  $\Omega$ , as the camera traverses an otherwise rigid 3D environment. By appealing to brightness constancy, it is further shown that orientations in visual spacetime that yield locally minimal energy are indicative of global egomotion. Correspondingly, global egomotion can be estimated by steering local SOE measurements to the direction that yields minimal energy as parameterized by  $\mathbf{T}$  and  $\Omega$ . Due to the nonlinear dependence of the resulting objective function on  $\mathbf{T}$  and  $\Omega$ , Gauss-Newton refinement is employed to obtain the solution. To facilitate efficient processing with large capture range, estimation is performed within a coarse-to-fine refinement scheme.

**Empirical evaluation.** The proposed approach to egomotion estimation has been realized in C++ for execution on standard PCs. On a 3.4GHz processor with 16GB RAM, the execution rate is  $\approx 84$  milliseconds/frame for  $512 \times 384$  images, beyond the time required for SOE filtering and stereo matching. Significantly, previous research has shown that both SOE filtering and stereo matching can be done in real-time, e.g., [3]. Thus, the overall approach has potential for real-time applications.

The implementation has been evaluated on three datasets. The first

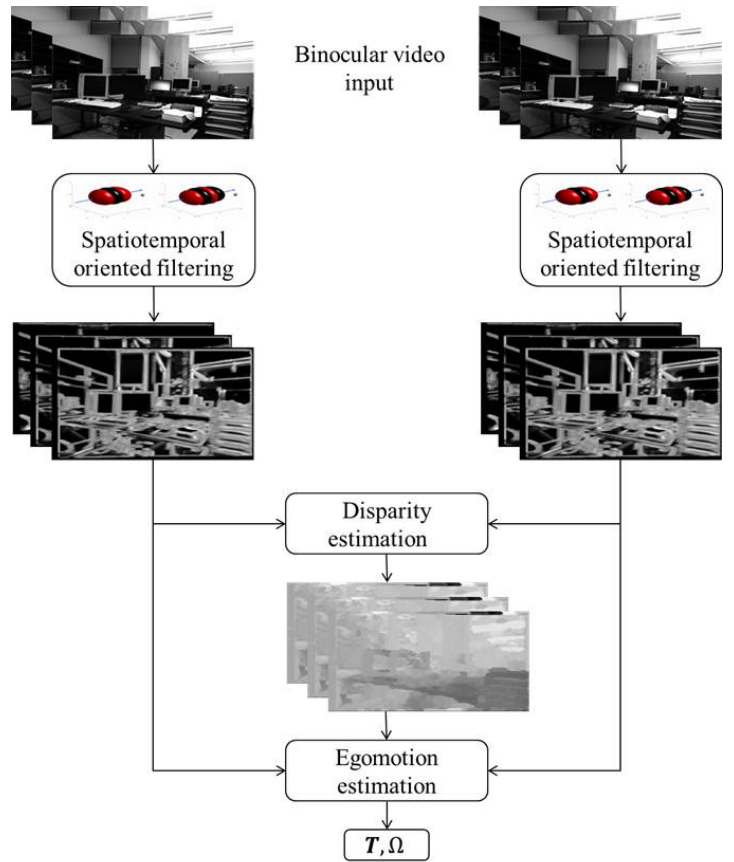


Figure 1: Overview of the proposed approach to egomotion estimation. Binocular video serves as input. The video streams are filtered with a bank of spatiotemporal filters that are tuned for 3D orientation in visual spacetime,  $(x, y, t)$ . The filtered images are used as input to a stereo disparity estimator [3]. Recovered disparity is combined with the oriented filtering results to yield egomotion estimates as instantaneous translation,  $\mathbf{T}$ , and rotation,  $\Omega$ .

was acquired in a calibrated laboratory setting and includes groundtruth egomotion. This dataset is available to the community. The second and third datasets were captured in more naturalistic settings, both indoor (an office) and outdoor (building exterior with foreground ground cover). For the sake of comparison, two alternative egomotion estimation algorithms [1, 2] also were executed on the same datasets. Both qualitative and quantitative results show that the developed approach is competitive with and even exceeds the accuracy of the alternative algorithms.

- [1] K. Derpanis and P. Chang. Closed-form linear solution to motion estimation in disparity space. In *Proceedings of the IEEE Intelligent Vehicles Symposium*, pages 268–275, 2006.
- [2] B. Kitt, A. Geiger, and H. Lategahn. Visual odometry based on stereo image sequences with RANSAC-based outlier rejection scheme. In *Proceedings of the IEEE Intelligent Vehicles Symposium*, pages 486–492, 2010.
- [3] M. Sizintsev and R. Wildes. Spatiotemporal oriented energies for spacetime stereo. In *Proceedings of the IEEE International Conference on Computer Vision*, 2011.