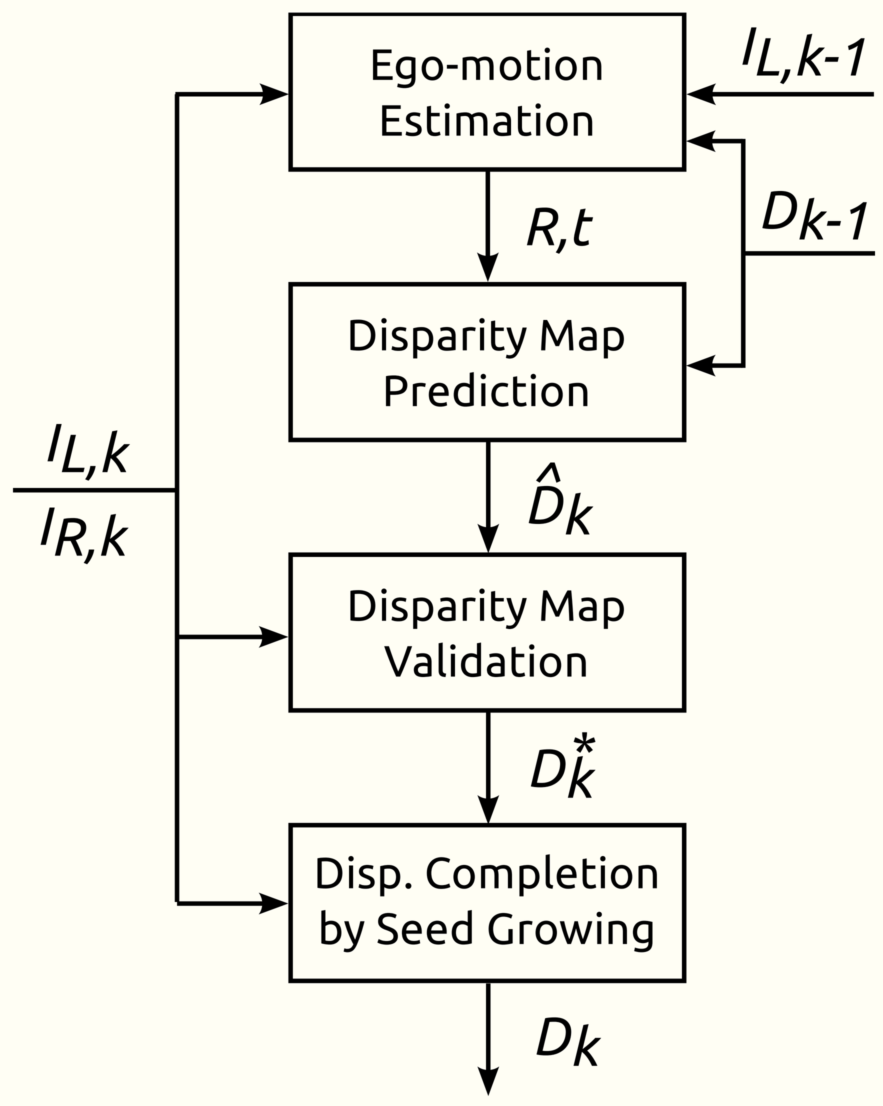
Real-time Global Prediction for Temporally Stable Stereo

Martin Dobias and Radim Sara
Center for Machine Perception
Czech Technical University in Prague
http://cmp.felk.cvut.cz

We present a method for calculation of disparity maps from stereo sequences. Disparity map from previous frame is first transferred to the new frame using estimated motion of the calibrated stereo rig. The predicted disparities are validated for the new frame and areas where prediction failed are matched with a traditional stereo matching algorithm. This method produces very fast and temporally stable stereo matching suitable for real-time applications even on non-parallel hardware.

Main Processing Cycle



$I_{L,k}, I_{R,k}$ k-th frame left and right image
$I_{L,k-1}$ previous left image
D_k disparity map for frame k
D_{k-1} previous disparity map

Assumptions

- Stereo rig calibration is known ($\mathtt{K}$, focal length f, baseline $\mathbf{b}$)
- Scene is static, camera may be moving

Algorithms

- Disparity prediction (per pixel)
  $\mathbf{x}'_\mathbf{d} = \mathtt{M}\mathbf{x}_\mathbf{d}$ where $\mathbf{x}_\mathbf{d} = (x, y, w, d)$ consists of homogenous image coordinates (x, y, w) and disparity d.

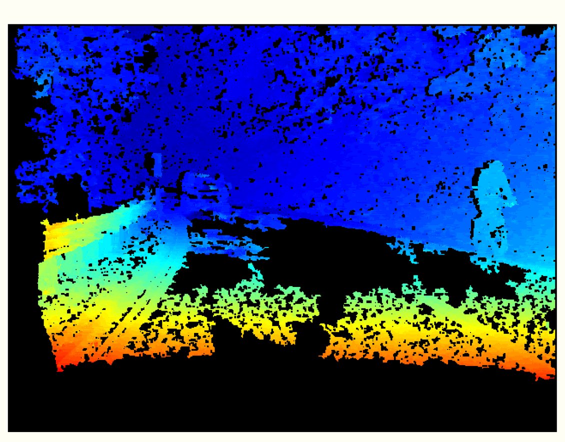
$$\mathtt{M} = \begin{bmatrix} \mathtt{KRK}^{-1} & \frac{\mathtt{K}\mathbf{t}}{f\|\mathbf{b}\|} \\ \mathbf{0}^\top & 1 \end{bmatrix}$$

- Ego-motion estimation
  Harris points, prediction by previous disparity map, Lucas-Kanade tracker and P3P in RANSAC with reprojection error
- Disparity completion and bootstrapping
  Seed growing stereo based on [1], seed queue re-insertion from predicted and validated disparity map

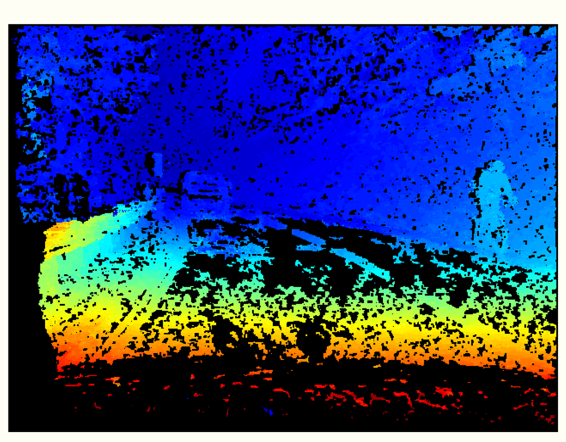
Typical Disparity Maps

Left

Right

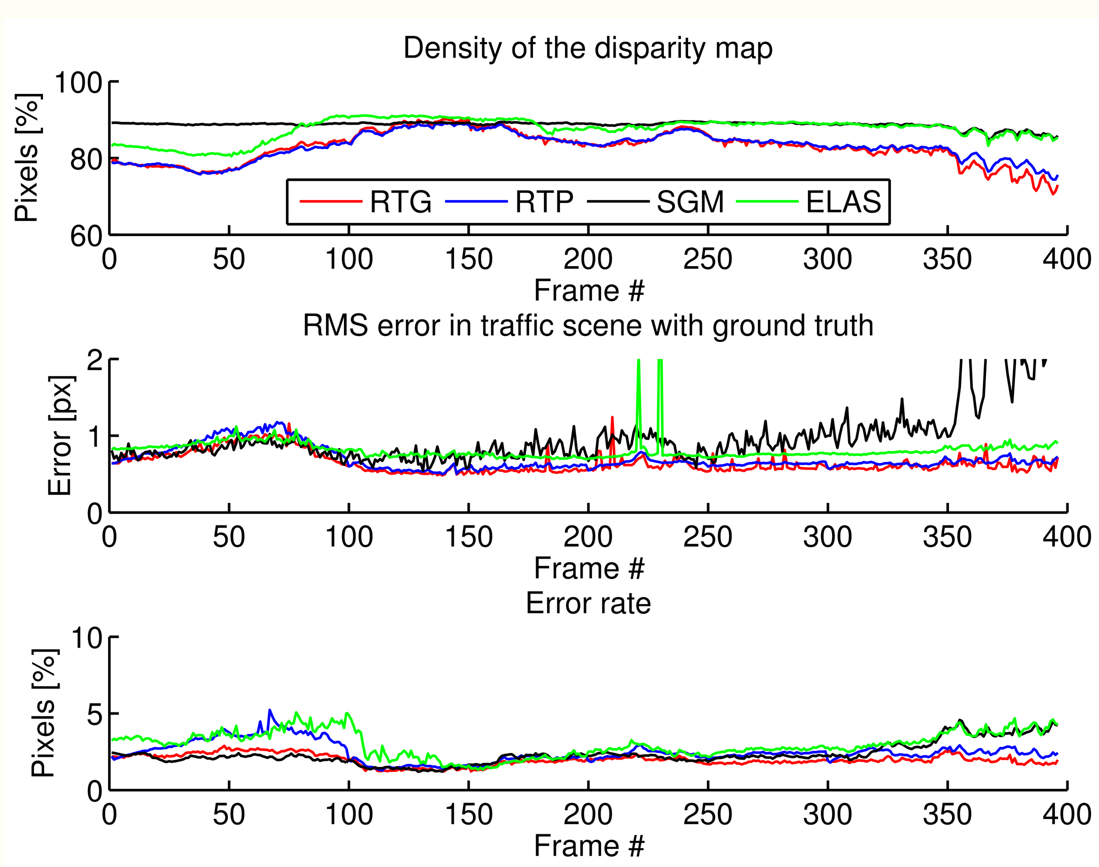


Per Frame Seed Growing (RTG)



Prediction + Growing (RTP)

Comparison with Others



- Compared with ELAS [2] and SGM [3]
- Evaluation on a synthesized sequence
- Error (per pixel) when $|d - d_{GT}| > 1$
- RMS error low for both RTP and RTG
- Challenging data: lower density, but stable error rate

Real-time Implementation

The computation is split into two threads working in a producer-consumer pattern:

- Initialization – precompute MNCC, Harris
- Stereo – ego-motion, prediction, growing

Running 20fps with resolution 640×480 on Dell Latitude E6420 Quad Core (using two cores). Video capture with Bumblebee2 stereo camera (SDK used only for rectification)

Conclusion

- Resulting disparity maps contain less flicker
- Very low computational demands
- Not sensible to disparity search range
- Ego-motion estimate as a bonus
- Prediction failures are not fatal – handled by traditional stereo

Acknowledgement

This work was supported by the Czech Ministry of Education under Project MEB111006.

This work was presented at ICCV 2011 Workshop on Live Dense Reconstruction from Moving Cameras.

Reference

[1] J. Cech and R. Sara. Efficient sampling of disparity space for fast and accurate matching. In *CVPR BenCOS*, 2007.
[2] A. Geiger, M. Roser, and R. Urtasun. Efficient large-scale stereo matching. In *ACCV*, 2010.
[3] H. Hirschmuller. Stereo processing by semiglobal matching and mutual information. *PAMI*, 2008.