Joint Disparity and Optical Flow by Correspondence Growing

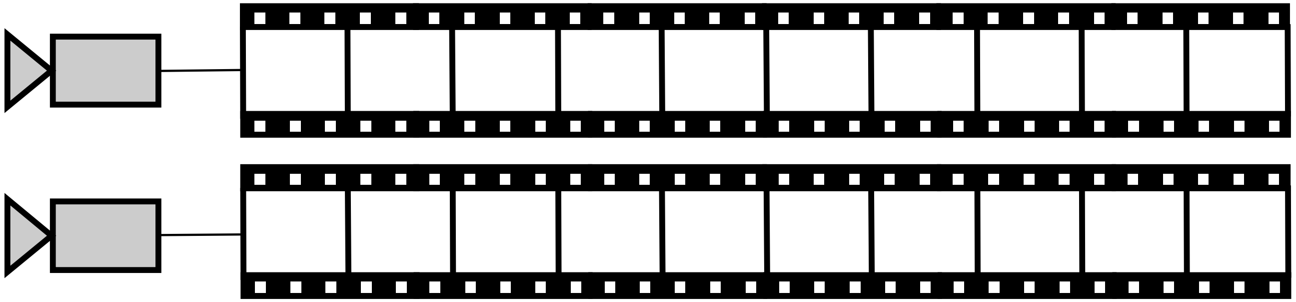
Jan Čech, Radu Horaud

INRIA Rhône-Alpes

Grenoble, France

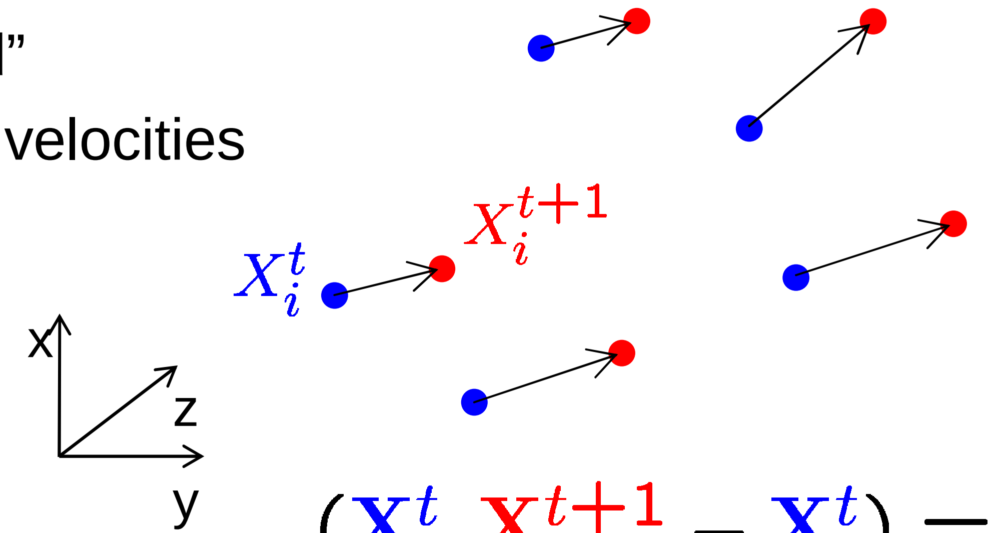
Stereo Disparity and Optical Flow ⇒ Scene Flow

INPUT: two stereo sequences from calibrated and synchronized cameras in a narrow baseline setup



OUTPUT: disparity and optical flow maps ⇒ 3D scene flow

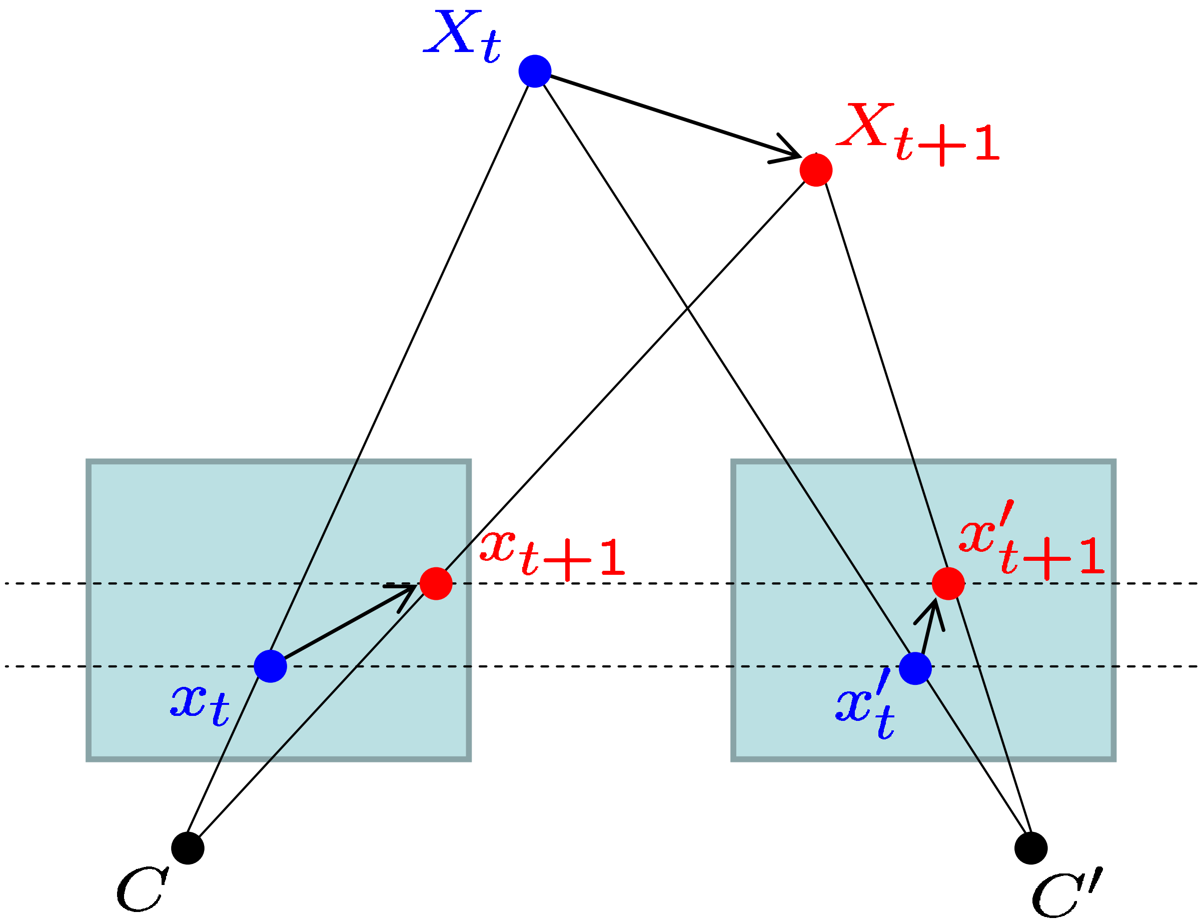
- Scene flow = "motion field"
  - 3D scene points + 3D velocities



$$(\mathbf{X}^t, \mathbf{X}^{t+1} - \mathbf{X}^t) = (\mathbf{X}, \mathbf{v})$$

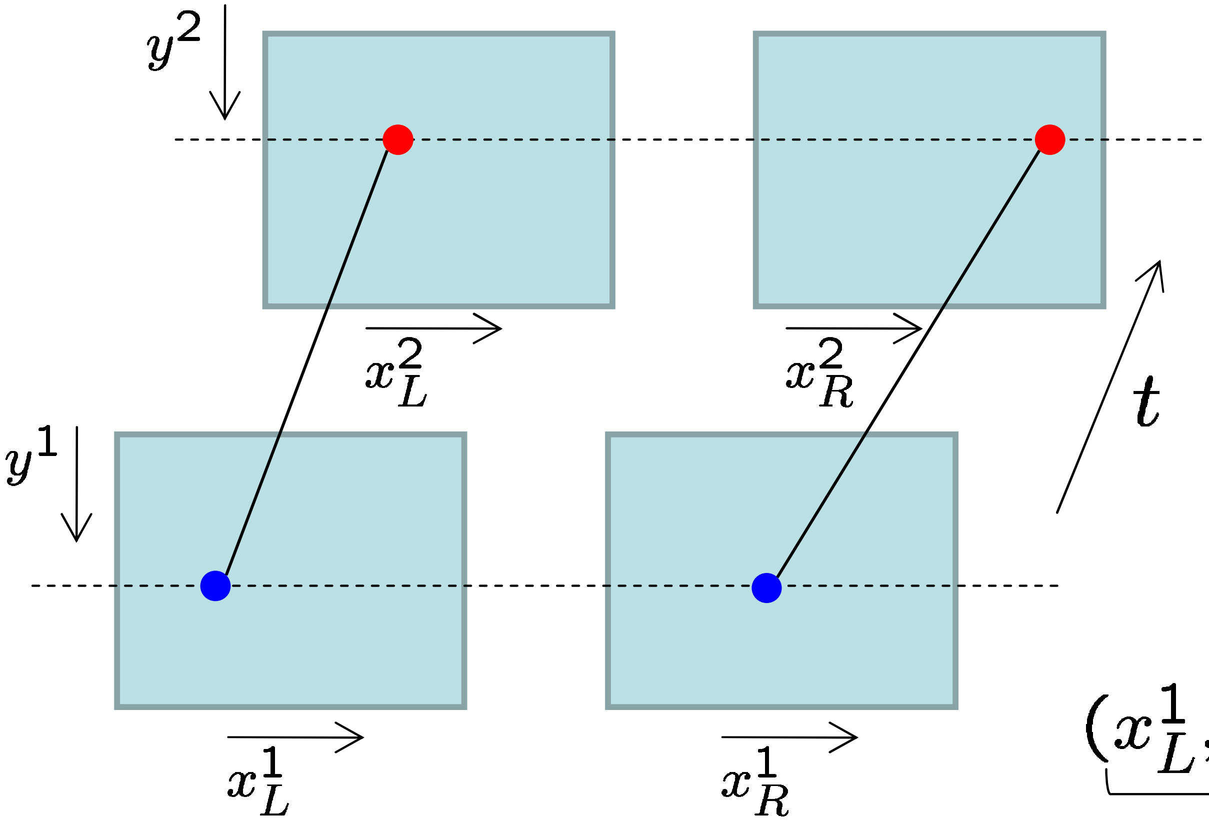
Scene Flow Estimation

- Stereo disparity and optical flow – image correspondence problem
  - Assuming calibration of cameras $\Rightarrow$ 3D scene flow
- No assumptions on rigidity of the scene
- Camera static or moving
  (but stereo-rig geometry fixed, calibration known)



Scene Flow Estimation

- The task: given X^t , find X^{t+1}
  - Reconstruct as many points as possible



$$d^1 = x_L^1 - x_R^1$$

$$d^2 = x_L^2 - x_R^2$$

$$f_x = x_L^2 - x_L^1$$

$$f_y = y_L^2 - y_L^1$$

$$(\underbrace{x_L^1, x_R^1, y^1}_{\rightarrow X^t},\ \underbrace{x_L^2, x_R^2, y^2}_{\rightarrow X^{t+1}})$$

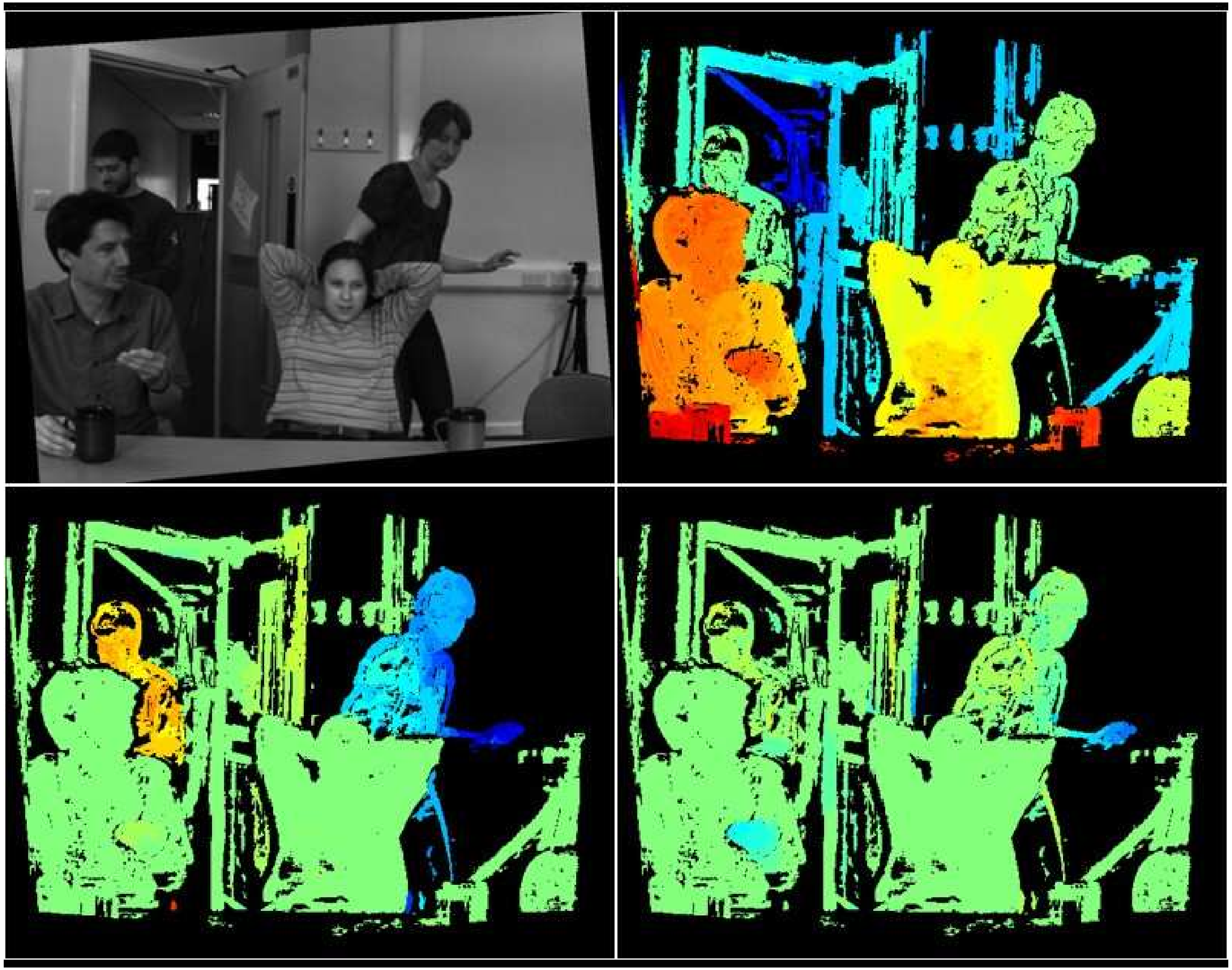
$\Rightarrow$ Stereo and optical flow problems constraint each other

Example of the Results

left image | disparity map

$\mathbf{I}_L$

$\mathbf{D}$



$\mathbf{F}_x$

$\mathbf{F}_y$

horizontal optical flow map | Vertical optical flow map

Correspondence Problem

- Ambiguity of the correspondence problem – What helps?
  - Epipolar constraint
  - Spatial smoothness (but occlusions, object and motion boundaries)
    - Neigbouring pixels usually have similar disparities, optical flow
  - Temporal smoothness (but occlusions)
    - None of the disparity and optical flow maps change abruptly

- Explicit regularization – Global energy minimization
  - Often used in the literature $\Rightarrow$ MRF, variational formulation
  - Computationally intensive
  - Artifacts caused by the prior model winning over the weak data

- Discriminability based
  - Finds unambiguous parts of the solution $\Rightarrow$ semi-dense matching

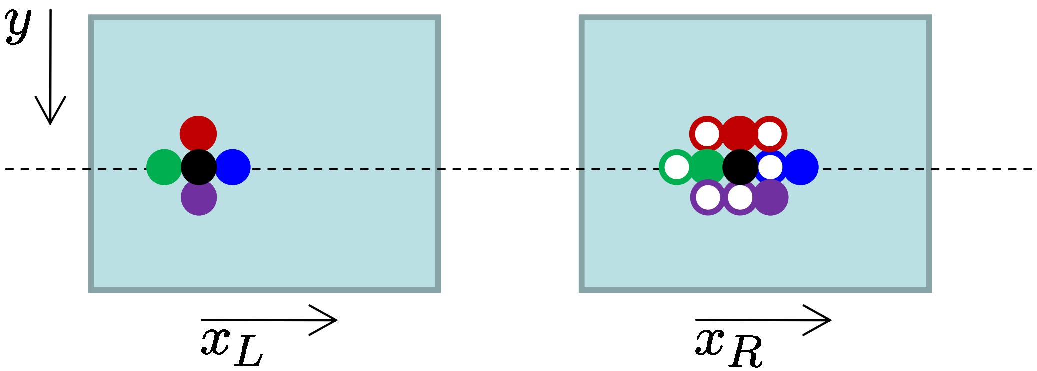
Seed Growing Approach

- Basic Ideas:
  - Set of initial correspondences (seeds)
  - Other correspondences found in a neigbourhood of these seeds (points which are close have similar disparity and optical flow)
  - New correspondences found become new seeds

- Seed growing **stereo** algorithm [Cech-PAMI-2010]
  - Tradeoff between global energy methods and greedy local methods, the growing process constraints possible solution
  - Robust against wrong seeds
  - Very fast (search space reduction), $\mathcal{O}(n^3) \rightarrow \mathcal{O}(n^2)$
  - Semi-dense (not fully dense), but identifies ambiguous (textureless) regions

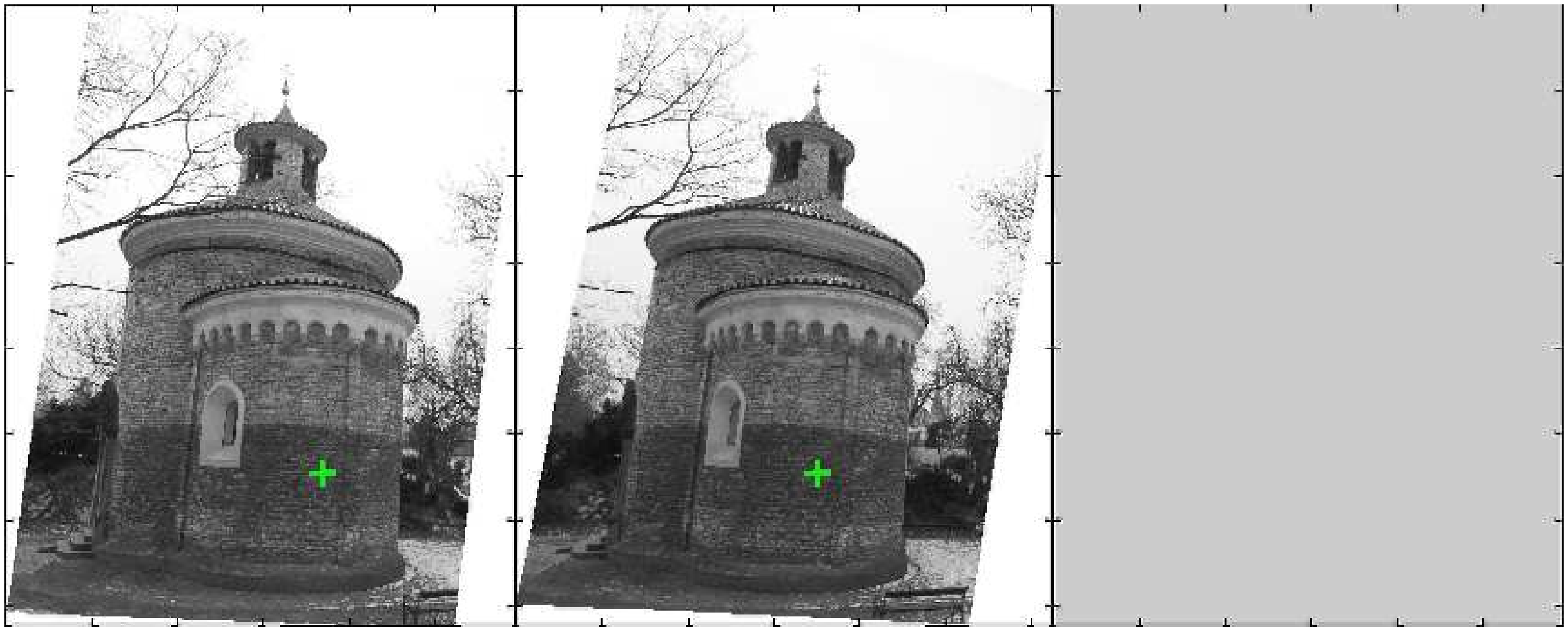
Correspondence growing in Stereo

- Stereo (2 still images only)
- a correspondence seed: $s = (x_L, x_R, y)$



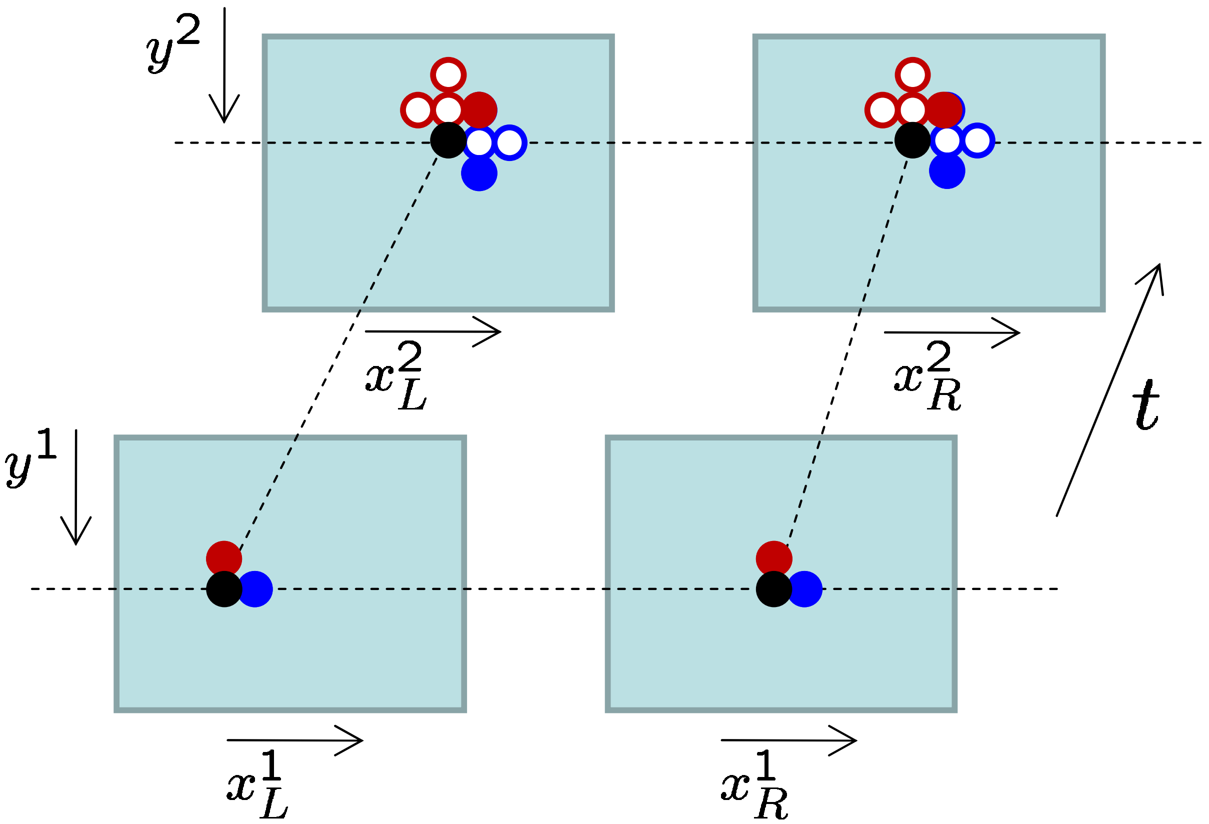
$$p^* = \arg \max_{p \in \mathcal{N}(s)} \mathrm{corr}(p)$$

If $\mathrm{corr}(p^*) > \tau$,
accept a new match.



Growing the Scene Flow

- Growing disparity + optical flow at once
- Correspondence seed: $s = (x_L^1, x_R^1, y^1, x_L^2, x_R^2, y^2)$
- disparity map for $t = 1$ is given
- seeds obtained by matching Harris points + LK tracker
- similarity: $\mathrm{corr}(p) = \frac{1}{3}(c_{LR}^{22} + c_L^{12} + c_R^{12})$



$$p^* = \arg \max_{p \in \mathcal{N}(s)} \mathrm{corr}(p)$$

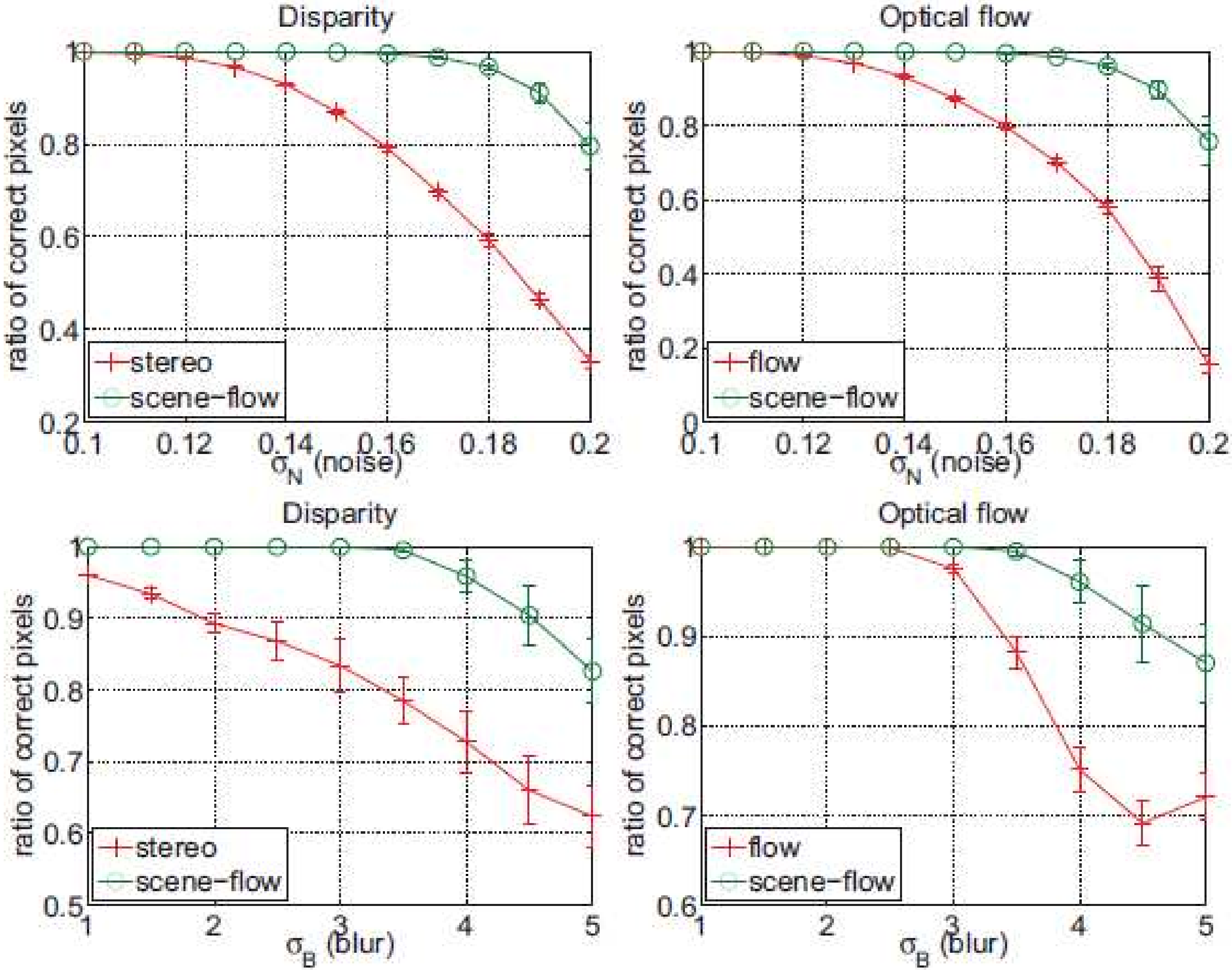
If $\mathrm{corr}(p^*) > \tau$,
accept a new match.

Results: Synthetic ground-truth experiment

- Synthetic stereo sequence (textured plane moving)

(1) noise

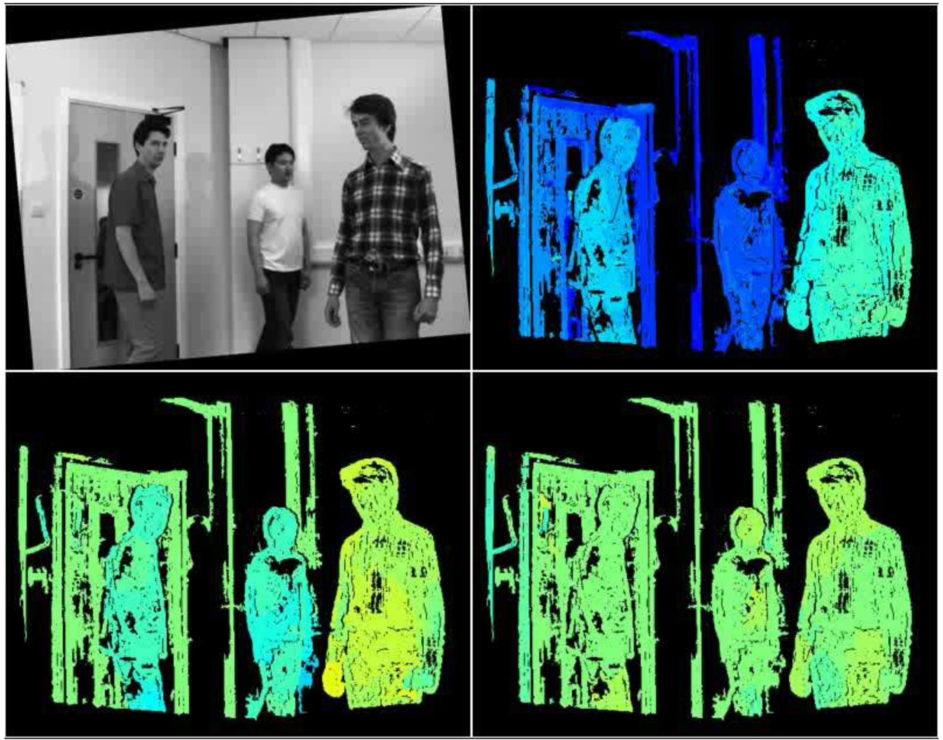
(2) blur

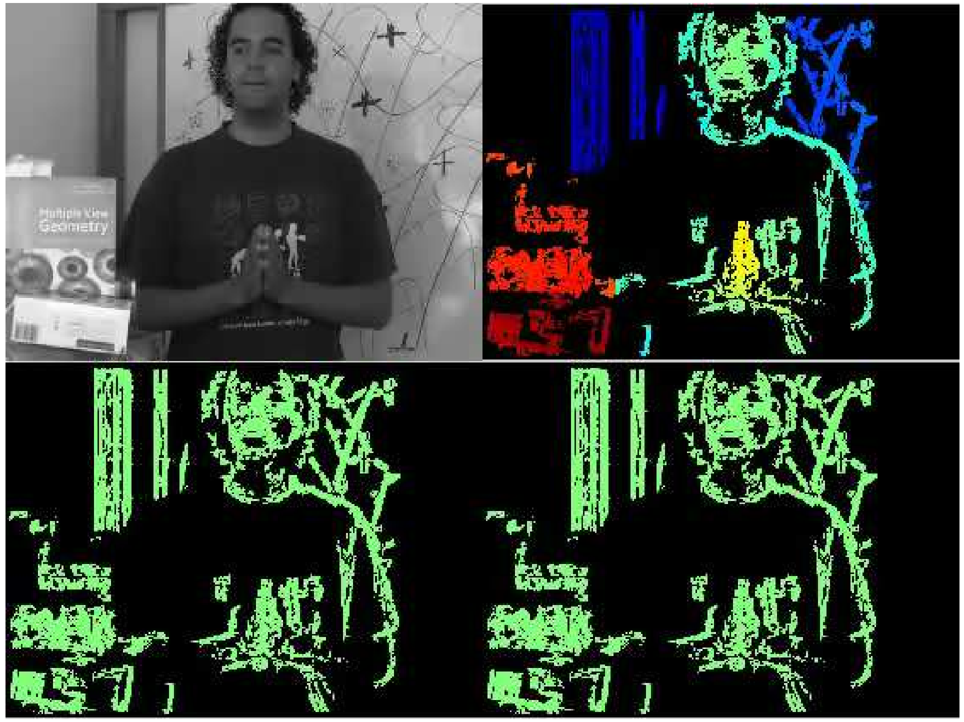


- Joint disparity and optical flow better performance than independently

Results: Real data (1)

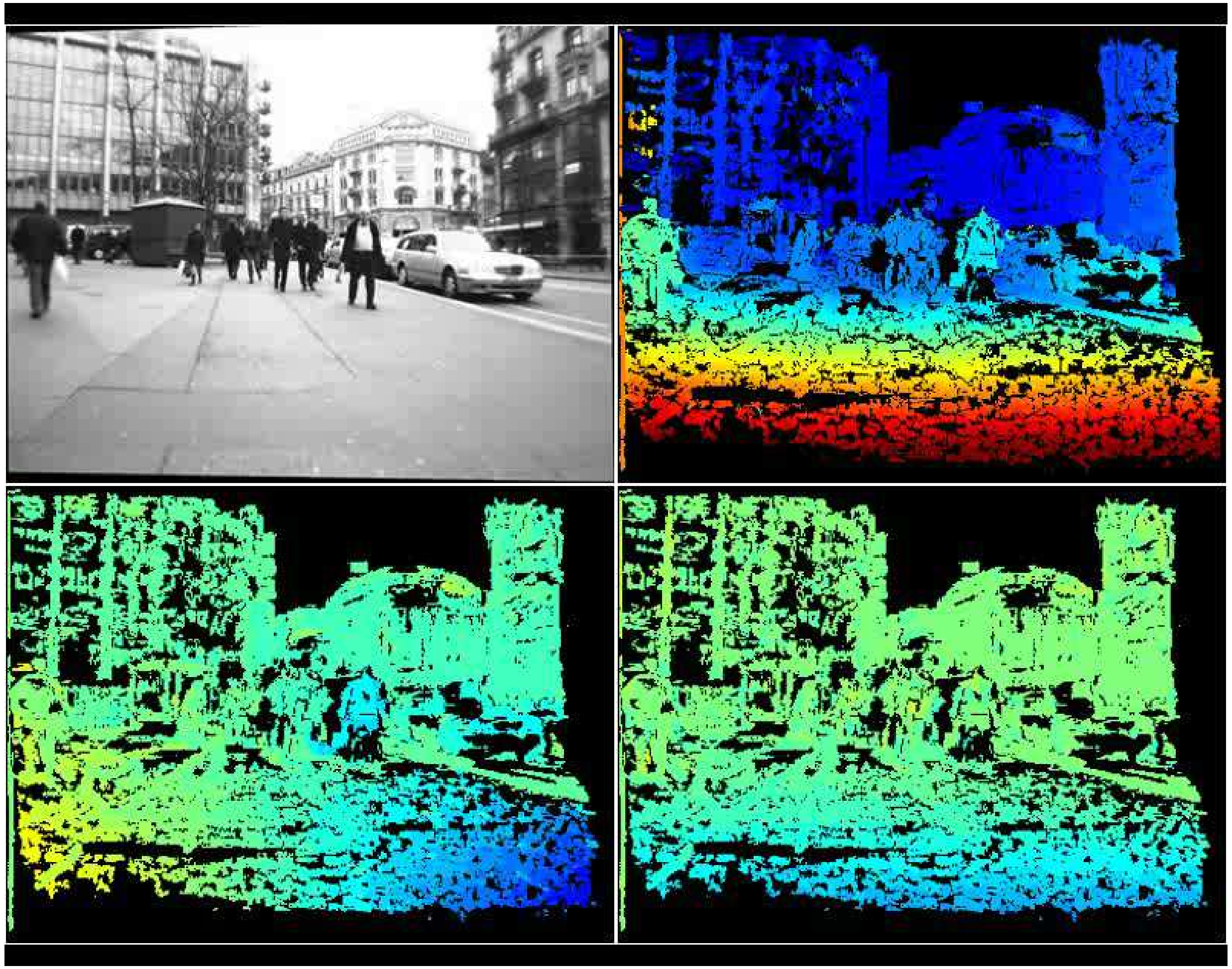
- Static camera
  - INRIA dataset





Results: Real data (2)

- Moving camera
  - pedestrian dataset by ETH Zurich



Conclusions

- Allows large displacement between frames (>30 pixels)
- Occlusion boundaries well preserved, no smoothing artifacts
- Disparity maps are temporally coherent (better results than frame by frame stereo)
- Optical flow maps are more precise than in a single camera setup
- Fast (VGA image in ~ 1.5 s), not a significant extra cost w.r.t. the stereo
  – Low algorithmic complexity due to search space reduction:
    $$\mathcal{O}(n^5) \rightarrow \mathcal{O}(n^2)$$
- Results are not fully dense, but semi-dense only

- Extended version of this algorithm in CVPR 2011
  – Algorithm incorporated in the pipeline for processing a stereo sequence with motion prediction
  – Experimental comparison with State-of-the-art methods of scene-flow, spatiotemporal stereo, and optical flow