Problem statement

Real-time people segmentation in stereo video

Input

Two stable video streams (left and right)
in a building corridor (constant conditions)
and a depth interval

Output

Dense (in both time and space) segmentation
on one of the streams (left)
of all the people (two-layer)
in the given depth interval

Planned approach

1. Dense depth estimation (DP, max-flows, etc.)
2. Ad-hod segmentation by energy formulation:
   - Contrast
   - Depth
   - Motion
   - Machine learning for people detection

Current situation

Somehow working

Middlebury Stereo Matcher C-code for depth estimation for a pair of frames

Nikita's *BackgroundCut* framework modified for depth

Testing material taken

15min video in MediaLab's corridor

Workflow

1. Middlebury Stereo Matcher (best results using graphcuts)

left + right → disparity (depth map)

Workflow

2. Learning (local & global background models)

left + true segmentation mask + background → BG model

Workflow

3. BackgroundCut

left + disparity + BG model → segment

Workflow

4. Precision estimation (TODO)

segment + true-mask → precision