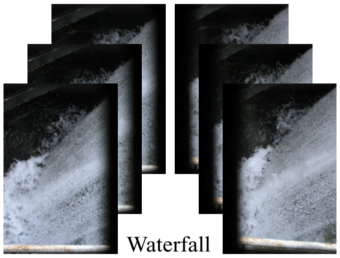
Spacetime Stereo: Shape Recovery for Dynamic Scenes

Appears in *Proc. of CVPR 2003*, pp. 367-374.



Waterfall

Facial deformation under
time-varying structured light.

Li Zhang **Brian Curless** **Steven M. Seitz**
University of Washington, Seattle
{lizhang,curless,seitz}@cs.washington.edu
http://grail.cs.washington.edu/projects/ststereo

Introduction

Goal: stereo reconstruction for scenes with abruptly changing appearance but smoothly changing geometry.

Problem: applying standard stereo algorithms **frame by frame** results in temporally inconsistent and noisy/over-smoothed shapes.

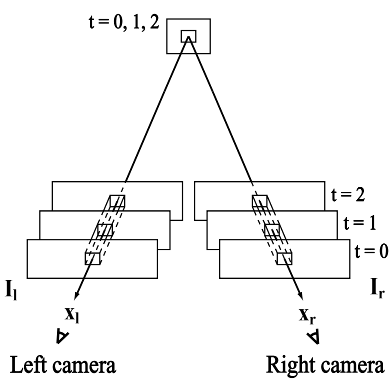
Solution: match **sequence to sequence**, yielding temporally consistent and spatially accurate reconstruction.

Spacetime Stereo Metrics and Algorithms

We extend traditional 2D window matching in image plane to 3D window matching in video volume.

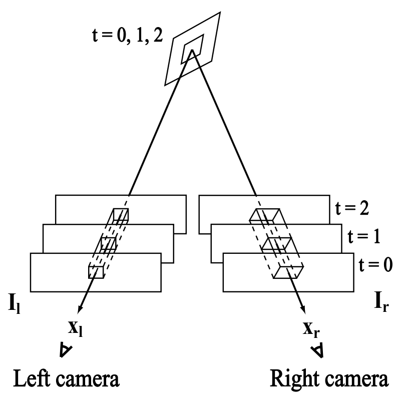
Static cases:

A fronto-parallel surface



Solve for x shift.

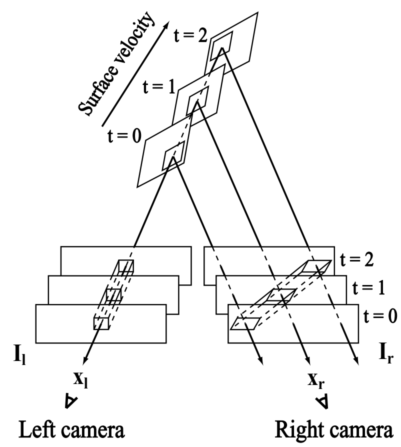
An oblique surface



Solve for x shift, x scale, y shear.

Moving case:

An oblique surface



Solve for x shift, x scale, y shear, t shear.

Standard stereo matching algorithms can be adapted to use spacetime window metric.
We use Dynamic Programming (DP) for initialization and Lucas-Kanade (LK) for subpixel refinement.

Related work

Structured light: [Sato87][Kanade91][Curless95][Pulli98][Bouguet98][Hall-Holt01]

- Idea: illuminate scenes with special time-varying light patterns encoding correspondences
- Limitation: previous methods do not fully exploit temporal information for moving scenes.

Motion Stereo: [Vedula99][Mandelbaum99][Tao01][Zhang01][Carceroni01][Strecha02]

- Idea: compute stereo correspondence and optical flow simultaneously
- Limitation: lighting, shadowing, and texture variations violate brightness constancy

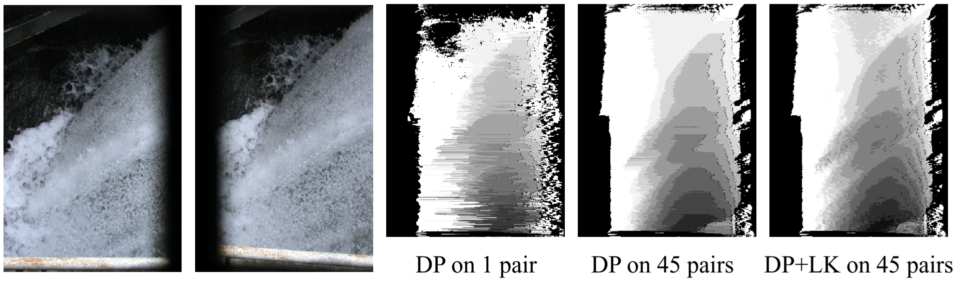
Sequence to Sequence alignment: [Caspi00]

- Idea: register two video sequences with a global transformation
- Limitation: it works only for concentric camera motion or planar objects

We generalize stereo matching to incorporate arbitrary appearance variations to estimate depth more accurately at each pixel in each frame. Our technique applies a form of sequence to sequence alignment for a spacetime window around each pixel in the video. In the same proceedings, Davis et al. [03] also propose using a similar spacetime stereo framework for static scenes. Our primary motivation is to recover moving scenes.

Results

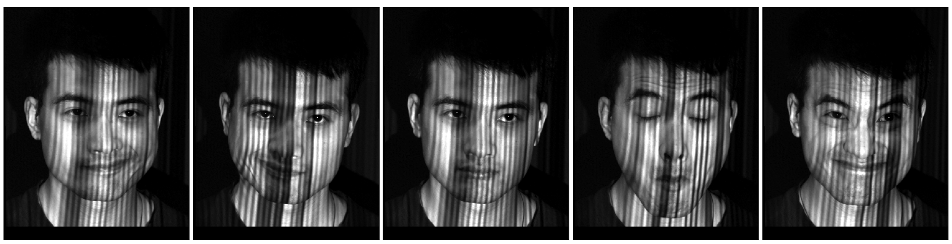
Outdoor scene: waterfall



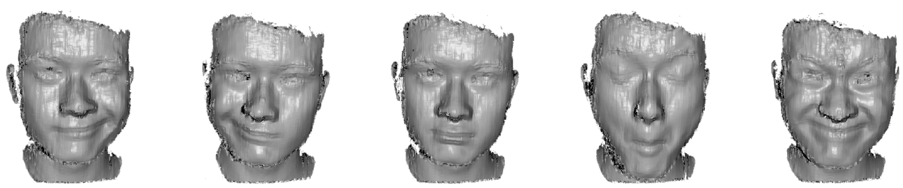
DP on 1 pair DP on 45 pairs DP+LK on 45 pairs

Input: 45 stereo pairs (1 pair shown here) Disparity maps comparison

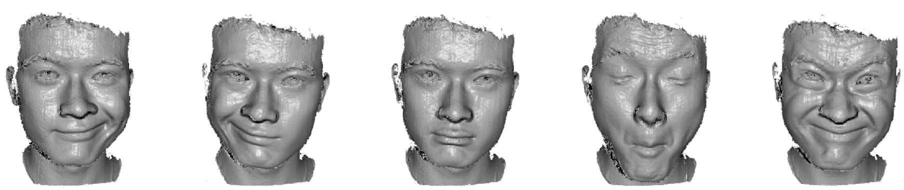
Structured light: facial deformation



Input: 400 stereo pairs (5 left camera imagest shown here)

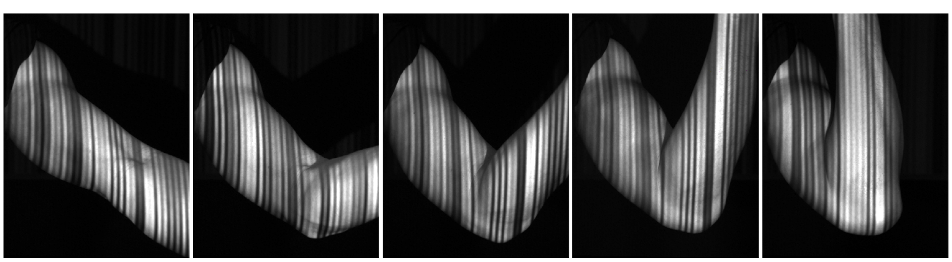


Frame-by-frame stereo reconstruction with 15x15 window

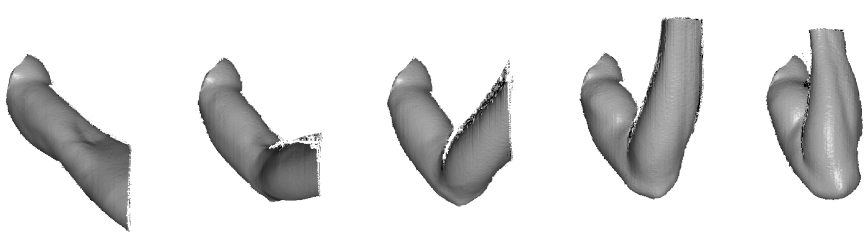


Spacetime stereo reconstruction with 9x5x5 window

Structured light: bending arm



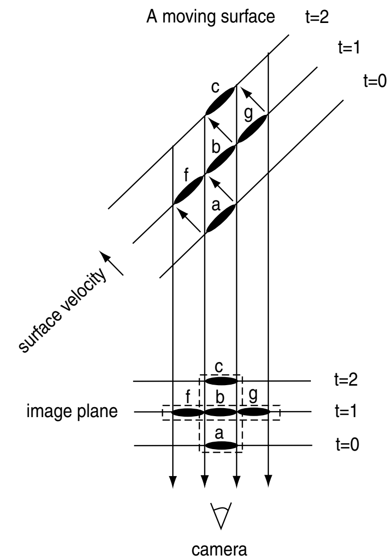
Input: 400 stereo pairs (5 left camera images shown here)



Spacetime stereo reconstruction with 9x5x5 window

Discussion

Tradeoff between spatial and temporal window size:

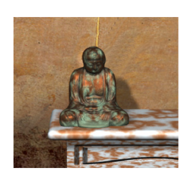


For a diffuse surface moving under static ambient light, the spacetime window a-b-c over t=0,1,2 is equivelent to a larger spatial one f-b-g at t=1.

Future work

- Adaptive windows for spacetime stereo
- Graph cut and belief propagation implementations
- Linear dynamic models for temporal disparity variation
- Reconstruction for full human body in motion
- Other applications:

shape from insects?



shape from weathering?

landscape from waving plants?