A NATIONAL SCIENCE FOUNDATION
ENGINEERING RESEARCH CENTER

Perceptual Grouping from Motion Cues Using Tensor Voting

Changki Min, Mircea Nicolescu
Gérard Medioni

Research Goal

Establish a voting-based computational framework for visual motion analysis.

Recover the dense velocity field, motion boundaries and regions from a sequence of multiple images.

Role in IMSC

Ability to extract velocity information, motion regions and boundaries from image sequences.

Useful for motion capture, avatars, face and body animation, and video compression.

Uniqueness & Related Work

Strong support within motion layers, weak influence across layers or from isolated tokens.

Temporal smoothness in conjunction with spatial smoothness for the enhanced dense velocity field and motion boundaries.

Related Works

[1] J. Barron, D. Fleet, S. Beauchemin, "Performance of Optical Flow Techniques", *IJCV*, 1994.
[2] F. Heitz, P. Bouthemy, "Multimodal Estimation of Discontinuous Optical Flow Using Markov Random Fields", *PAMI*, 1993.
[3] S. Hsu, P. Anandan, S. Peleg, "Accurate Computation of Optical Flow by Using Layered Motion Representations", *ICPR*, 1994.

Accomplishments

Disparity error rate has been reduced (e.g. error rate of Teddy bear sequence has been reduced to 8.1% by using multiple frames – 15.4% for two frames).

Publications

- Mircea Nicolescu, Changki Min, Gérard Medioni, "Analysis and Interpretation of Multiple Motions through Surface Saliency", *SCVMA* (in conjunction with the *ECCV*), 2004.
- Mircea Nicolescu, Gérard Medioni, "Voting-Based Grouping and Interpretation of Visual Motion", *Proceedings of the IEEE Southwest Symposium on Image Analysis and Interpretation*, 2004.
- Mircea Nicolescu, Gérard Medioni, "Motion Segmentation with Accurate Boundaries – A Tensor Voting Approach", *CVPR*, 2003.

5-Year Plan

2004-2005
Multiple frame processing

2005-2006
Computationally efficient algorithm

2006 - 2008
3D reconstruction for video stream
Multiple scale analysis

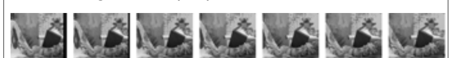
Research Approach

Voting-based computational framework

- *Layered 4-D representation* of data
- Token affinity is propagated through voting, in the 4-D space of image positions (*x*,*y*) and velocities (v_x,v_y)
- Distinct moving regions emerge as *smooth surface layers* in the 4-D space
- The only constraint is smoothness of image motion

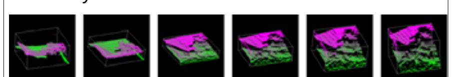
Input Data

- One frame is set to a reference frame for the given *n* frames
- The reference frame is paired with each of remaining frames (*n-1*)



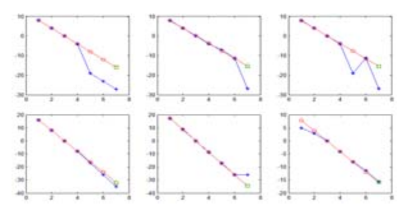
Layer Extraction for Each Pair

- Encode potential matches as 4-D *ball tensors*
- Propagate affinities through sparse voting
- Correct matches identified by maximal surface saliency



Temporal Smoothness Constraint

Every (*x*,*y*) of the reference frame has (*n-1*) matches, and they are plotted in a (*x*,*y*,*t*) space for the temporal smoothing process.



Ground truth disparity | Two frame tensor voting | Multiple frame tensor voting

Approaches	Error Rate
Tensor Voting with multiple frames	**8.1 %**
Tensor Voting with two frames	15.4 %
Sum of Squared Differences	26.5 %
Dynamic Programming	30.1 %
Graph Cuts	29.3 %