Layered Image Motion with Explicit Occlusions, Temporal Consistency, and Depth Ordering

Deqing Sun, Erik B. Sudderth, and Michael J. Black

Department of Computer Science, Brown University, Providence, RI, USA

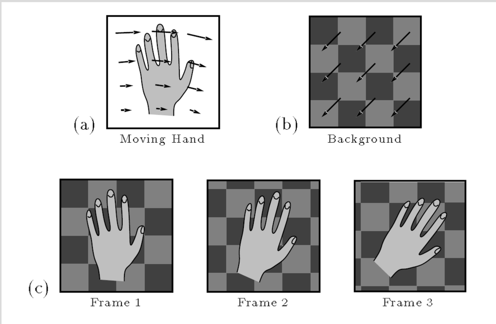
{dqsun, sudderth, black}@cs.brown.edu

Overview

Goal: Motion estimation, segmentation, and occlusion detection

Benefits of Layered Models:

- Within-layer motion can be described more simply
- Motion boundaries estimated separately from smooth flow
- Reasoning about occlusion relationships is easier
- Layers provide a segmentation of scene motion



Wang & Adelson 1994

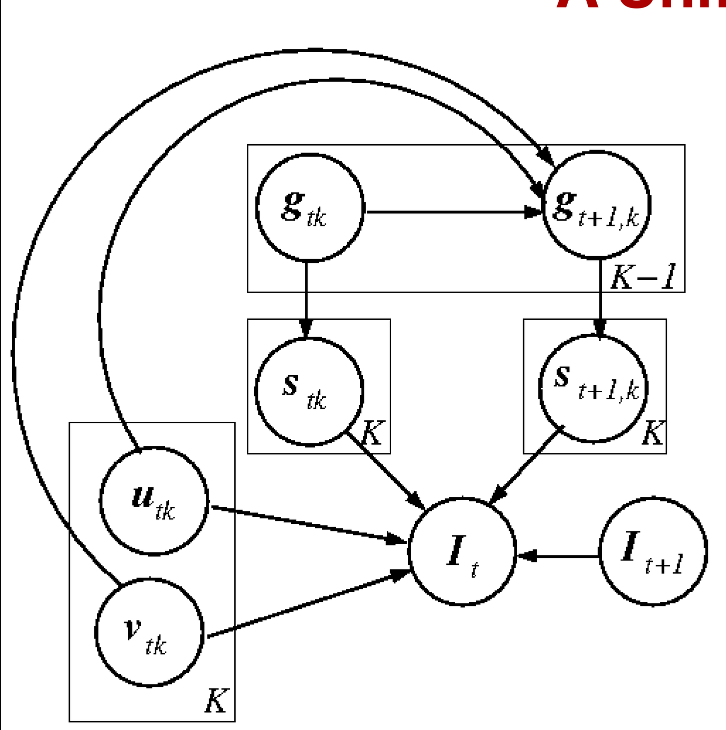
Problem:

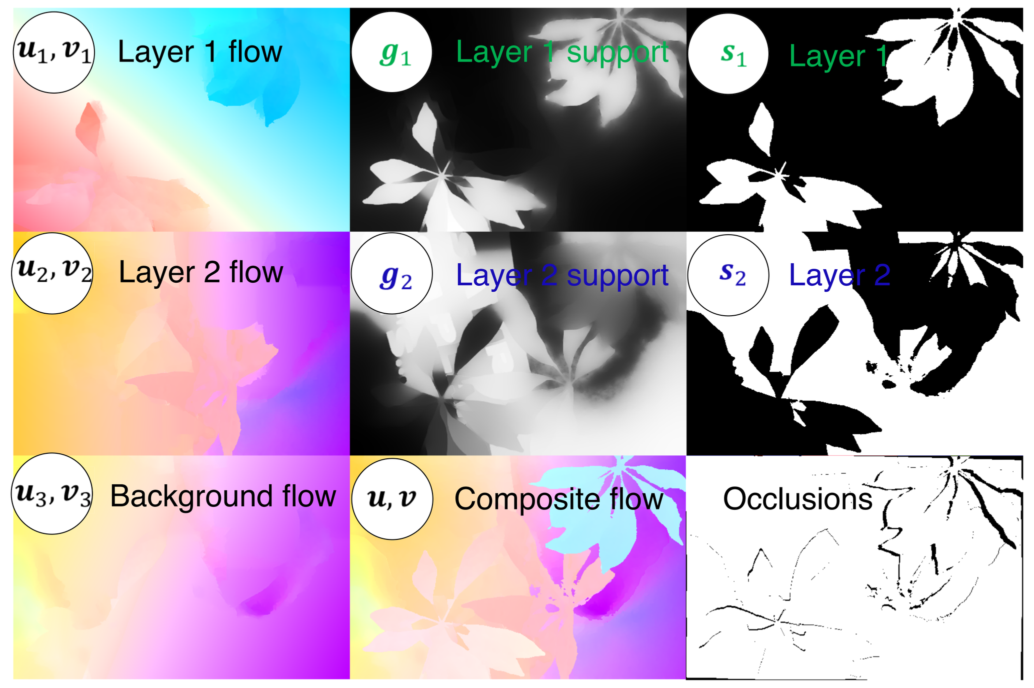
- None of the current top performing optical flow methods use a layered approach
- The most accurate approaches are single-layered, use robust functions to cope with flow discontinuities, but usually make mistakes in occlusion regions

Contributions of Our Approach:

- New probabilistic generative model of motion in layers
- True "layered" model of depth ordered occlusions
- Non-rigidly enforce temporal consistency of layer support
- Robust flow priors allow "roughness in layers"
- Spatial layer support modeled via an image-dependent Gaussian field prior
- Lowest average endpoint and angular error in Middlebury evaluation

A Unified Generative Model





"Roughness in Layers"

What is an appropriate model for the flow within each layer?

MRF models [2]: locally smooth but globally complex

$$p(\mathbf{u}_{tk}) \propto \exp\{-E_{\mathrm{mrf}}(\mathbf{u}_{tk})\} = \exp\left\{-\frac{1}{2}\sum_{(i,j)}\sum_{(i',j')\in\Gamma_{(i,j)}} \rho_s(u_{tk}^{ij} - u_{tk}^{i'j'})\right\}$$

Parametric (affine) models: globally coherent but too restrictive

$$u_{\boldsymbol{\theta}_{tk}}^{ij} = \theta_{tk1} + \theta_{tk2}\cdot i + \theta_{tk3}\cdot j, \qquad v_{\boldsymbol{\theta}_{tk}}^{ij} = \theta_{tk4} + \theta_{tk5}\cdot i + \theta_{tk6}\cdot j$$

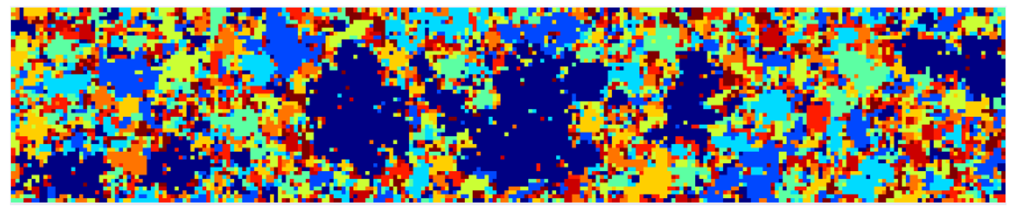
Affine + deformation: our approach is globally coherent but locally complex

$$E_{\mathrm{aff}}(\mathbf{u}_{tk}) = \frac{1}{2}\sum_{(i,j)}\sum_{(i',j')\in\Gamma_{(i,j)}} \rho_s\left((u_{tk}^{ij} - u_{\boldsymbol{\theta}_{tk}}^{ij}) - (u_{tk}^{i'j'} - u_{\boldsymbol{\theta}_{tk}}^{i'j'})\right)$$

Acknowledgements: DS and MJB were supported in part by the NSF Collaborative Research in Computational Neuroscience Program (IIS-0904875) and a gift from Intel Corporation.

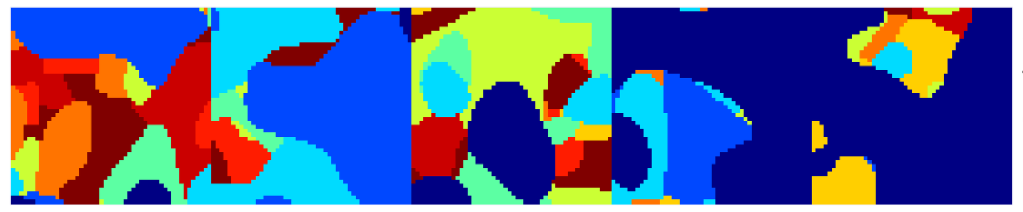
Layer Support and Depth Ordering

Ising & Potts MRFs favor unnatural segmentations, and do not model relationships between regions:


Samples from Potts model

Thresholded continuous layer support functions better match the statistics of natural scenes [3]:

$$k_{t*}^{ij} = \min(\{k \mid 1 \le k \le K-1, g_{tk}(i,j) \ge 0\} \cup \{K\})$$


Samples from thresholding Gaussian processes

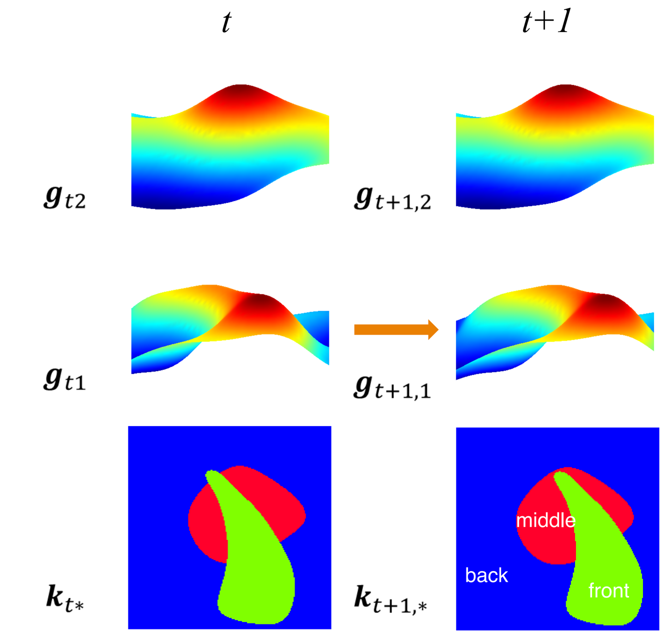
Spatial and Temporal Layer Consistency

Spatial prior uses layer-specific Gaussian conditional random fields:

$$E_{\mathrm{space}}(\mathbf{g}_{tk}) = \frac{1}{2}\sum_{(i,j)}\sum_{(i',j')\in\Gamma_{(i,j)}} w_{i'j'}^{ij}(g_{tk}(i,j) - g_{tk}(i',j'))^2$$

Temporal prior uses flow field to ensure consistency of support:

$$E_{\mathrm{time}}(\mathbf{g}_{tk}, \mathbf{g}_{t+1,k}, \mathbf{u}_{tk}, \mathbf{v}_{tk}) = \sum_{(i,j)}\left(g_{tk}(i,j) - g_{t+1,k}(i+u_{tk}^{ij}, j+v_{tk}^{ij})\right)^2$$



Temporal consistency

Occlusion Reasoning

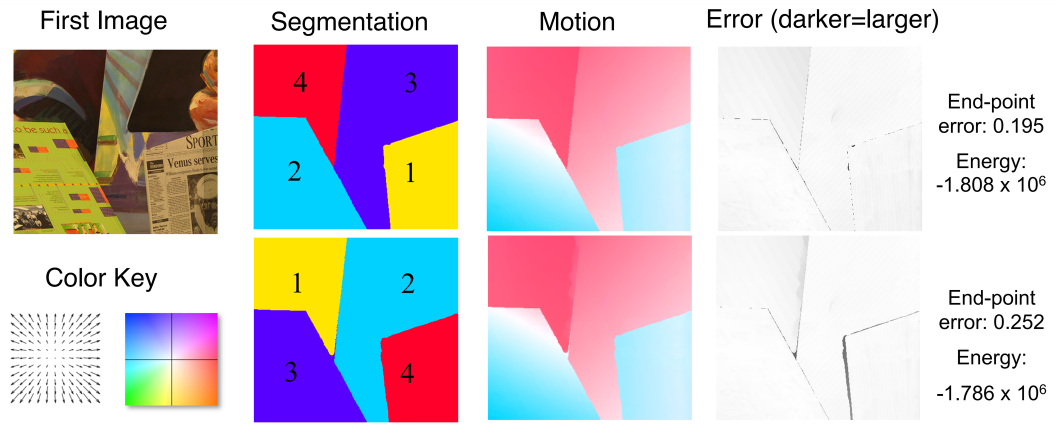
For pixel (i,j), if its corresponding pixel at the next frame is visible

$$p\left(\mathbf{I}_t^s(i,j) \mid \mathbf{I}_{t+1}^s(i+u_{tk}^{ij}, j+v_{tk}^{ij})\right) \propto \exp\left\{-\rho_d\left(\mathbf{I}_t^s(i,j) - \mathbf{I}_{t+1}^s(i+u_{tk}^{ij}, j+v_{tk}^{ij})\right)\right\}$$

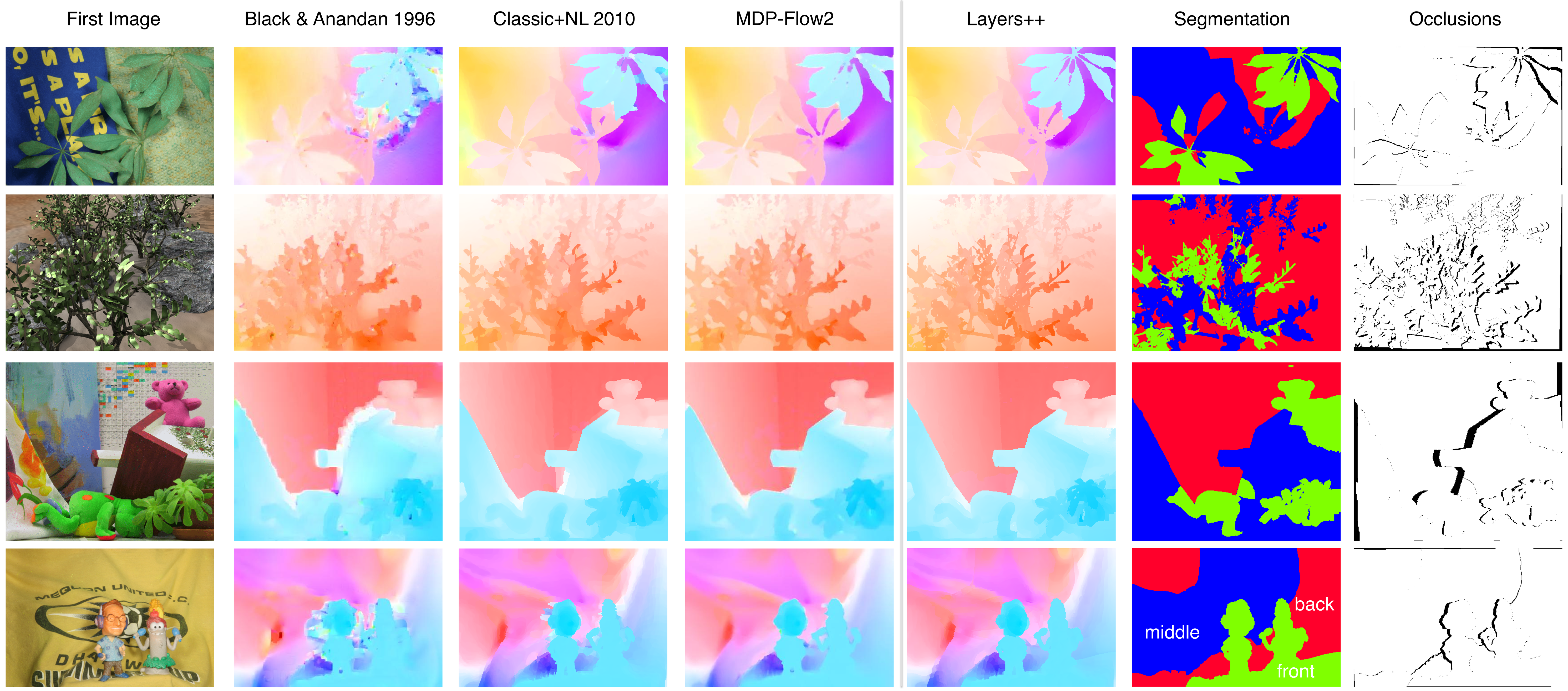
Otherwise

$$p\left(\mathbf{I}_t^s(i,j) \mid \mathbf{I}_{t+1}^s(i+u_{tk}^{ij}, j+v_{tk}^{ij})\right) = \mathrm{Uniform}(0, Z)$$

The correct layer ordering produces flow fields that are more accurate and more probable (lower energy)



Experimental Results on Middlebury Optical Flow Benchmark



Middlebury Optical Flow Benchmark: Screen shot of public table (Dec. 2010)

Average endpoint error	avg. rank	Army (Hidden texture) GT all	Army im0 disc	Army im1 untext	Mequon (Hidden texture) GT all	Mequon im0 disc	Mequon im1 untext	Schefflera (Hidden texture) GT all	Schefflera im0 disc	Schefflera im1 untext	Wooden (Hidden texture) GT all	Wooden im0 disc	Wooden im1 untext
MDP-Flow2 [40]	3.9	0.09 4	0.23 2	0.07 2	**0.16** 1	**0.52** 1	0.13 2	0.22 2	0.46 3	0.17 3	0.17 8	0.93 14	0.09 3
Layers++ [38]	4.9	**0.08** 1	**0.21** 1	0.07 2	0.19 4	0.56 3	0.17 12	**0.20** 1	**0.40** 1	0.18 7	**0.13** 1	**0.58** 1	**0.07** 1
LSM [41]	7.4	**0.08** 1	0.23 2	0.07 2	0.22 12	0.73 13	0.18 14	0.28 8	0.64 6	0.19 11	0.14 2	0.70 2	0.09 3
Classic+NL [31]	8.0	**0.08** 1	0.23 2	0.07 2	0.22 12	0.74 14	0.18 14	0.29 9	0.65 9	0.19 11	0.15 3	0.73 4	0.09 3
MDP-Flow [26]	9.1	0.09 4	0.25 6	0.08 8	0.19 4	0.54 2	0.18 14	0.24 4	0.55 5	0.20 13	0.16 6	0.91 11	0.09 3

Average endpoint error	Grove (Synthetic) GT all	Grove im0 disc	Grove im1 untext	Urban (Synthetic) GT all	Urban im0 disc	Urban im1 untext	Yosemite (Synthetic) GT all	Yosemite im0 disc	Yosemite im1 untext	Teddy (Stereo) GT all	Teddy im0 disc	Teddy im1 untext	Average EPE
MDP-Flow2 [40]	0.65 4	0.98 5	0.43 4	**0.29** 1	**0.91** 1	**0.26** 1	0.11 5	0.13 5	0.17 6	0.51 5	1.11 6	0.72 6	0.275
Layers++ [38]	**0.48** 1	**0.70** 1	0.33 2	0.47 8	1.01 2	0.33 7	0.15 20	0.14 13	0.24 20	**0.46** 1	**0.88** 1	0.72 6	0.270
LSM [41]	0.66 5	0.97 4	0.48 6	0.50 10	1.06 4	0.33 7	0.15 20	0.12 2	0.29 28	0.50 4	0.99 3	0.73 8	0.316
Classic+NL [31]	0.64 3	0.93 3	0.47 5	0.52 12	1.12 5	0.33 7	0.16 26	0.13 5	0.29 28	0.49 3	0.98 2	0.74 9	0.319
MDP-Flow [26]	0.74 7	1.06 7	0.61 10	0.46 7	1.02 3	0.35 11	0.12 8	0.14 13	0.17 6	0.78 22	1.68 24	0.97 21	0.348

References

[1] Wang & Adelson, Representing Moving Images with Layers. IEEE Trans. on Image Proc., 1994.
[2] Weiss, Smoothness in Layers: Motion Segmentation Using Nonparametric Mixture Estimation. CVPR, 1997.
[3] Sudderth & Jordan, Shared Segmentation of Natural Scenes Using Dependent Pitman-Yor Processes. NIPS, 2009.
[4] Sun, Roth, & Black, Secrets of Optical Flow Estimation and Their Principles. CVPR, 2010.
[5] Xu, Jia, & Matsushita. Motion Detail Preserving Optical Flow Estimation. Submitted to PAMI 2010.