Secrets of Optical Flow Estimation and Their Principles

Deqing Sun

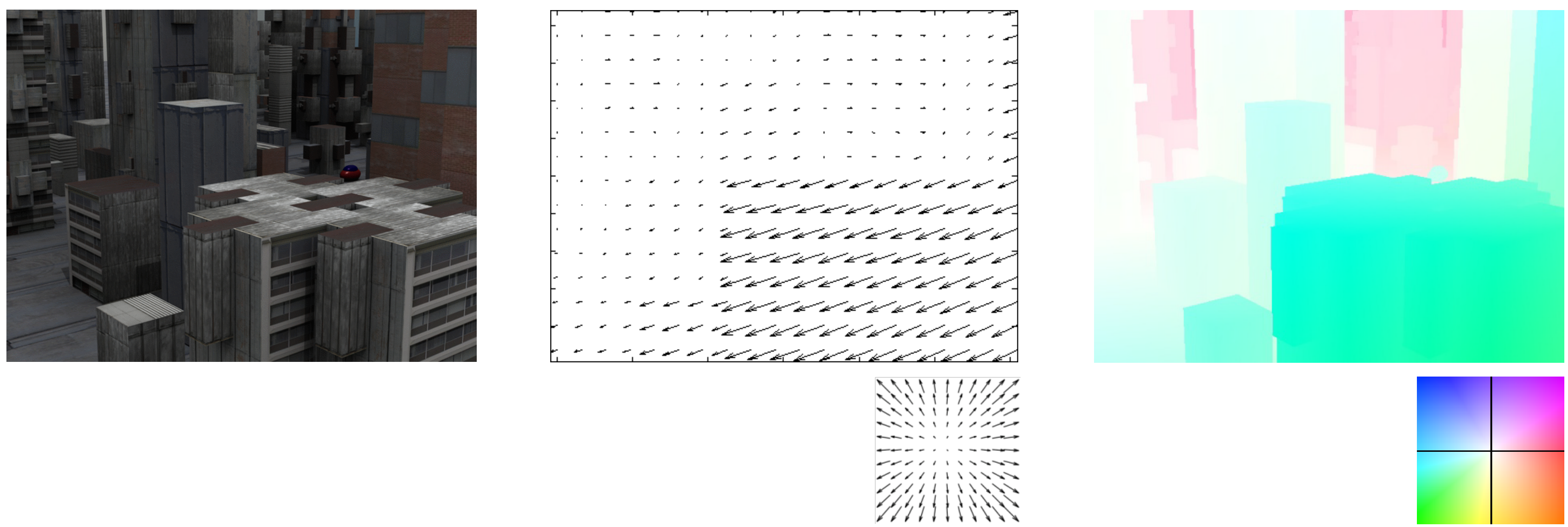
Department of Computer Science

Brown University

Joint work with Stefan Roth and Michael J. Black

Introduction

- Optical flow: motion of image pixels



Introduction

Middlebury benchmark (Baker et al.) 2007 Oct.

Optical flow evaluation results Choose error measures: Average SD R1.0 R3.0 R5.0 A50 A75 A95

Average angle error	avg.	Dimetrodon (Hidden texture) GT im0 im1			Seashell (Hidden texture) GT im0 im1			Rock (Synthetic) GT im0 im1			Grove (Synthetic) GT im0 im1			Yosemite (Synthetic) GT im0 im1			Venus (Stereo) GT im0 im1			Moebius (Stereo) GT im0 im1		
	rank	all	disc	untext	all	disc	untext	all	disc	untext	all	disc	untext	all	disc	untext	all	disc	untext	all	disc	untext
Bruhn et al.	1.6	10.99 3	9.41 1	14.22 3	11.09 2	19.48 2	16.21 2	6.14 1	17.41 1	12.86 2	6.32 1	12.41 1	10.98 1	1.69 1	2.86 1	1.05 1	8.73 2	31.46 2	8.15 2	5.85 1	10.12 2	8.80 2
Black and Anandan	2.1	9.26 1	10.11 3	12.08 1	11.20 3	19.83 3	17.01 3	7.67 3	18.44 3	16.80 4	7.89 2	13.55 2	13.96 4	2.65 2	4.18 2	1.88 2	7.64 1	30.13 1	7.31 1	7.05 2	10.02 1	8.41 1
Pyramid LK	2.8	10.27 2	9.71 2	13.63 2	9.46 1	18.62 1	12.07 1	6.53 2	18.43 2	10.95 1	8.14 3	15.08 3	12.78 2	5.22 3	6.64 3	4.29 3	14.61 4	36.18 4	24.67 5	12.98 5	13.85 4	20.61 5
MediaPlayerTM	4.1	15.82 4	26.42 4	16.96 4	23.18 4	27.71 5	21.78 4	9.44 4	22.25 4	15.03 3	10.99 4	18.15 5	13.64 3	11.09 4	17.16 4	10.66 5	15.48 5	43.56 5	15.09 4	9.98 4	15.04 5	9.47 3
Zitnick et al.	4.2	30.10 5	34.27 5	31.58 5	29.07 5	27.55 4	21.78 4	12.38 5	23.93 5	17.59 5	12.55 5	15.56 4	17.35 5	18.50 5	28.00 5	9.41 4	11.42 3	31.46 2	11.12 3	9.88 3	12.83 3	11.28 4

Introduction

Middlebury benchmark (Baker et al.) 2009 Dec.

Optical flow evaluation results — Statistics: Average SD R0.5 R1.0 R2.0 A50 A75 A95 — Error type: endpoint angle interpolation normalized interpolation

Average endpoint error	avg. rank	Army (Hidden texture) GT im0 im1			Mequon (Hidden texture) GT im0 im1			Schefflera (Hidden texture) GT im0 im1			Wooden (Hidden texture) GT im0 im1			Grove (Synthetic) GT im0 im1			Urban (Synthetic) GT im0 im1			Yosemite (Synthetic) GT im0 im1			Teddy (Stereo) GT im0 im1		
		all	disc	untext	all	disc	untext	all	disc	untext	all	disc	untext	all	disc	untext	all	disc	untext	all	disc	untext	all	disc	untext
Adaptive [20]	4.4	0.09 1	0.26 1	0.06 1	0.23 5	0.78 4	[illegible]	0.54 8	[illegible]	0.21 3	0.18 1	0.91 3	0.10 1	0.88 3	1.25 3	0.73 5	0.50 3	1.28 3	0.31 3	0.14 10	0.16 12	0.22 10	0.65 3	1.37 3	0.79 4
Complementary OF [21]	5.7	0.11 5	0.28 3	0.10 9	0.18 1	0.63	0.12 1	0.31 3	0.75	0.18 1	0.19 2	0.97 5	0.12 3	0.97 10	1.31 6	1.00 11	1.78 20	1.73 7	0.87 14	0.11 4	0.12 2	0.22 10	0.68 4	1.48 4	0.95 8
Aniso. Huber-L1 [22]	5.8	0.10 3	0.28 3	0.08 3	0.31 11	0.88	0.28 12	[illegible]	[illegible]	0.29 12	0.20 4	0.92 4	0.13 5	0.84 2	1.20 2	0.70 2	0.39 1	1.23 1	0.28 1	0.17 15	0.15 9	0.27 16	0.64 2	1.36 2	0.79 4
DPOF [18]	6.1	0.13 12	0.35 12	0.09 4	0.25 6	0.79 5	[illegible]	[illegible]	[illegible]	0.21 3	0.19 2	0.62 1	0.15 11	0.74 1	1.09 1	0.49 1	0.66 7	1.80 10	0.63 8	0.19 17	0.17 14	0.35 20	0.50 1	1.08 1	0.55 1
TV-L1-improved [17]	7.2	0.09 1	0.26 1	0.07 2	0.20 3	0.71 3	0.16 2	0.53 7	1.18 9	0.22 5	0.21 7	1.24 11	0.11 2	0.90 4	1.31 6	0.72 3	1.51 14	1.93 11	0.84 11	0.18 16	0.17 14	0.31 17	0.73 8	1.62 9	0.87 7
CBF [12]	7.8	0.10 3	0.28 3	0.09 4	0.34 12	0.80 6	0.37 13	0.43 5	0.95 5	0.26 8	0.21 7	1.14 8	0.13 5	0.90 4	1.27 4	0.82 7	0.41 2	1.23 1	0.30 2	0.23 22	0.19 20	0.39 21	0.76 9	1.56 6	1.02 9
Brox et al. [5]	8.4	0.11 5	0.32 8	0.11 12	0.27 9	0.93 10	0.22 9	0.39 4	0.94 4	0.24 7	0.24 9	1.25 12	0.13 5	1.10 13	1.39 12	1.43 17	0.89 8	1.77 8	0.55 7	0.10 2	0.13 4	0.11 1	0.91 11	1.83 12	1.13 12
Rannacher [23]	8.5	0.11 5	0.31 6	0.09 4	0.25 6	0.84 7	0.21 8	0.57 12	1.27 15	0.26 8	0.24 9	1.32 14	0.13 5	0.91 7	1.33 8	0.72 3	1.49 13	1.95 13	0.78 9	0.15 12	0.14 7	0.26 13	0.69 6	1.58 8	0.86 6
F-TV-L1 [15]	8.8	0.14 13	0.35 12	0.14 15	0.34 12	0.98 12	0.26 11	0.59 14	1.19 10	0.26 8	0.27 13	1.36 15	0.16 12	0.90 4	1.30 5	0.76 6	0.54 4	1.62 6	0.36 4	0.13 6	0.15 9	0.20 9	0.68 4	1.56 6	0.66 2
Second-order prior [8]	9.0	0.11 5	0.31 6	0.09 4	0.26 8	0.93 10	0.20 7	0.57 12	1.25 14	0.26 8	0.20 4	1.04 6	0.12 3	0.94 8	1.34 9	0.83 9	0.61 6	1.93 11	0.47 6	0.20 18	0.16 12	0.34 19	0.77 10	1.64 10	1.07 10
Fusion [6]	9.4	0.11 5	0.34 10	0.10 9	0.19 2	0.69 2	0.16 2	0.29 2	0.66 2	0.23 6	0.20 4	1.19 10	0.14 9	1.07 11	1.42 13	1.22 13	1.35 10	1.49 5	0.86 13	0.20 18	0.20 21	0.26 13	1.07 14	2.07 16	1.39 16
Dynamic MRF [7]	11.1	0.12 11	0.34 10	0.11 12	0.22 4	0.89 9	0.16 2	0.44 6	1.13 7	0.20 2	0.24 9	1.29 13	0.14 9	1.11 14	1.52 17	1.13 12	1.54 15	2.37 20	0.93 15	0.13 6	0.12 2	0.31 17	1.27 18	2.33 20	1.66 17
SegOF [10]	11.7	0.15 14	0.36 14	0.10 9	0.57 15	1.16 15	0.59 19	0.68 15	1.24 12	0.64 14	0.32 15	0.86 2	0.26 15	1.18 17	1.50 16	1.47 18	1.63 18	2.09 14	0.96 16	0.08 1	0.13 4	0.12 2	0.70 7	1.50 5	0.69 3
Learning Flow [11]	13.3	0.11 5	0.32 8	0.09 4	0.29 10	0.99 13	0.23 10	0.55 9	1.24 12	0.29 12	0.36 16	1.56 17	0.25 14	1.25 19	1.64 21	1.41 16	1.55 17	2.32 19	0.85 12	0.14 10	0.18 18	0.24 12	1.09 15	2.09 18	1.27 13
Filter Flow [19]	14.3	0.17 16	0.39 16	0.13 14	0.43 14	1.09 14	0.38 14	0.75 16	1.34 16	0.78 19	0.70 19	1.54 16	0.68 19	1.13 16	1.38 11	1.51 19	0.57 5	1.32 4	0.44 5	0.22 20	0.23 23	0.26 13	0.96 12	1.66 11	1.12 11
GraphCuts [14]	14.5	0.16 15	0.38 15	0.14 15	0.59 18	1.36 19	0.46 15	0.56 10	1.07 6	0.64 14	0.26 12	1.14 8	0.17 13	0.96 9	1.35 10	0.84 10	2.25 23	1.79 9	1.22 21	0.22 20	0.17 14	0.43 22	1.22 17	2.05 15	1.78 19
Black & Anandan [4]	15.0	0.18 17	0.42 17	0.19 18	0.58 17	1.31 17	0.50 16	0.95 19	1.58 18	0.70 16	0.49 17	1.59 18	0.45 17	1.08 12	1.42 13	1.22 13	1.43 11	2.28 17	0.83 10	0.15 12	0.17 14	0.17 6	1.11 16	1.98 14	1.30 14
SPSA-learn [13]	15.7	0.18 17	0.45 18	0.17 17	0.57 15	1.32 18	0.51 17	0.84 17	1.50 17	0.72 17	0.52 18	1.64 19	0.49 18	1.12 15	1.42 13	1.39 15	1.75 19	2.14 15	1.06 20	0.13 6	0.13 4	0.19 7	1.32 19	2.08 17	1.73 18
GroupFlow [9]	15.9	0.21 19	0.51 19	0.21 19	0.79 21	1.69 21	0.72 21	0.86 18	1.64 19	0.74 18	0.30 14	1.07 7	0.26 15	1.29 22	1.81 22	0.82 7	1.94 21	2.30 18	1.36 22	0.11 4	0.14 7	0.19 7	1.06 13	1.96 13	1.35 15
2D-CLG [1]	17.4	0.28 21	0.62 22	0.21 19	0.67 20	1.21 16	0.70 20	1.12 21	1.80 21	0.99 22	1.07 22	2.06 21	1.12 22	1.23 18	1.52 17	1.62 22	1.54 15	2.15 16	0.96 16	0.10 2	0.11 1	0.16 4	1.38 20	2.26 19	1.83 20
Horn & Schunck [3]	18.6	0.22 20	0.55 20	0.22 21	0.61 19	1.53 20	0.52 18	1.01 20	1.73 20	0.80 20	0.78 20	2.02 20	0.77 20	1.26 20	1.58 19	1.55 20	1.43 11	2.59 22	1.00 18	0.16 14	0.18 18	0.15 3	1.51 21	2.50 21	1.88 21
TI-DOFE [24]	19.6	0.38 23	0.64 23	0.47 23	1.16 22	1.72 22	1.26 22	1.39 23	2.06 24	1.17 23	1.29 23	2.21 23	1.41 23	1.27 21	1.61 20	1.57 21	1.28 9	2.57 21	1.01 19	0.13 6	0.15 9	0.16 4	1.87 22	2.71 22	2.53 22
FOLKI [16]	22.6	0.29 22	0.73 24	0.33 22	1.52 23	1.96 24	1.80 23	1.23 22	2.04 23	0.95 21	0.99 21	2.20 22	1.08 21	1.53 23	1.85 23	2.07 23	2.14 22	3.23 24	1.60 23	0.26 23	0.21 22	0.68 23	2.67 23	3.27 23	4.32 23
Pyramid LK [2]	23.7	0.39 24	0.61 21	0.61 24	1.67 24	1.78 23	2.00 24	1.50 24	1.97 22	1.38 24	1.57 24	2.39 24	1.78 24	2.94 24	3.72 24	2.98 24	3.33 24	2.74 23	2.43 24	0.30 24	0.24 24	0.73 24	3.80 24	5.08 24	4.88 24

Introduction

- Papers differ in
  - Objective function
  - Optimization method
  - Implementation details

- What makes optical flow accurate?

- Approach: start from classical formulation

Evaluation

- Classical formulation (Horn & Schunck 1981, Black & Anadan 1996)

$$E(\mathbf{u},\mathbf{v}) = E_D(\mathbf{u},\mathbf{v}) + \lambda E_S(\mathbf{u},\mathbf{v})$$

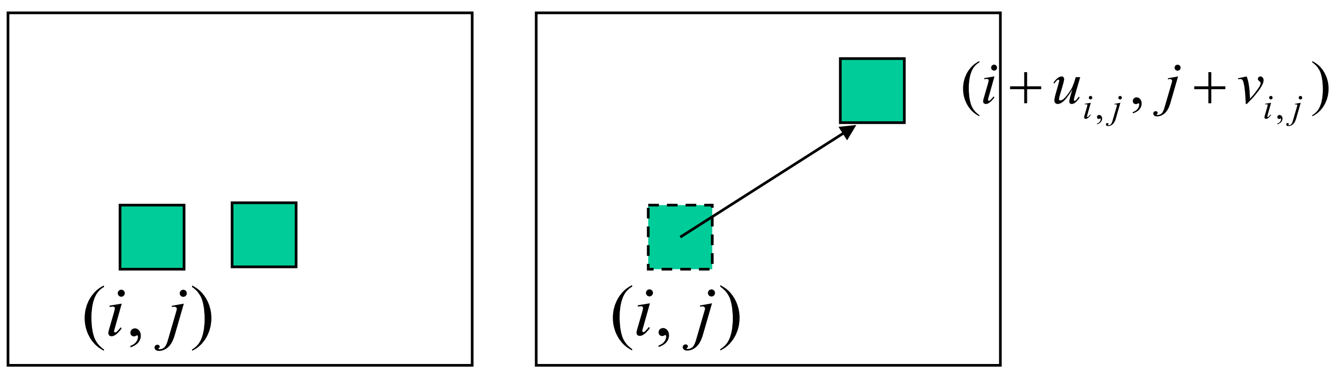
Evaluation

- Classical formulation (Horn & Schunck 1981, Black & Anadan 1996)

$$E(\mathbf{u},\mathbf{v}) = E_D(\mathbf{u},\mathbf{v}) + \lambda E_S(\mathbf{u},\mathbf{v})$$

- Data term: constancy of image property

$$E_D(\mathbf{u},\mathbf{v}) = \sum_{i,j} \rho_D(I_1(i,j) - I_2(i+u_{i,j}, j+v_{i,j}))$$



First image   Second image

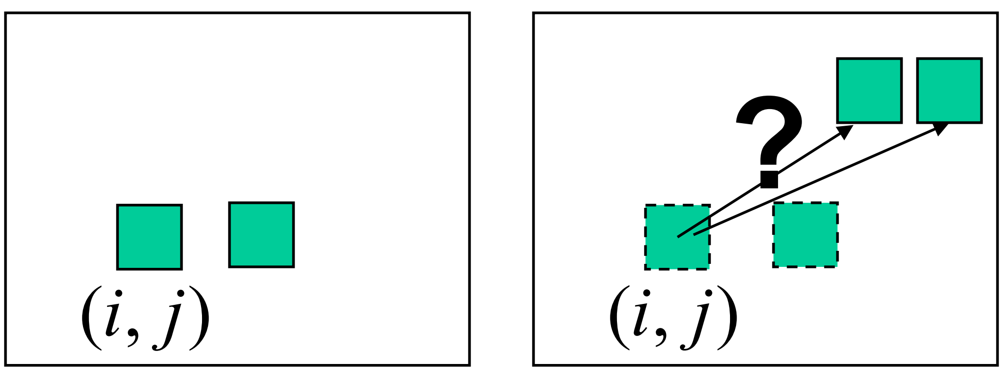
Evaluation

- Classical formulation (Horn & Schunck 1981, Black & Anadan 1996)

$$E(\mathbf{u},\mathbf{v}) = E_D(\mathbf{u},\mathbf{v}) + \lambda E_S(\mathbf{u},\mathbf{v})$$

- Spatial term: smoothness (pairwise MRF)

$$E_{Su}(\mathbf{u}) = \sum_{i,j} \rho_S(u_{i,j} - u_{i+1,j}) + \rho_S(u_{i,j} - u_{i,j+1})$$

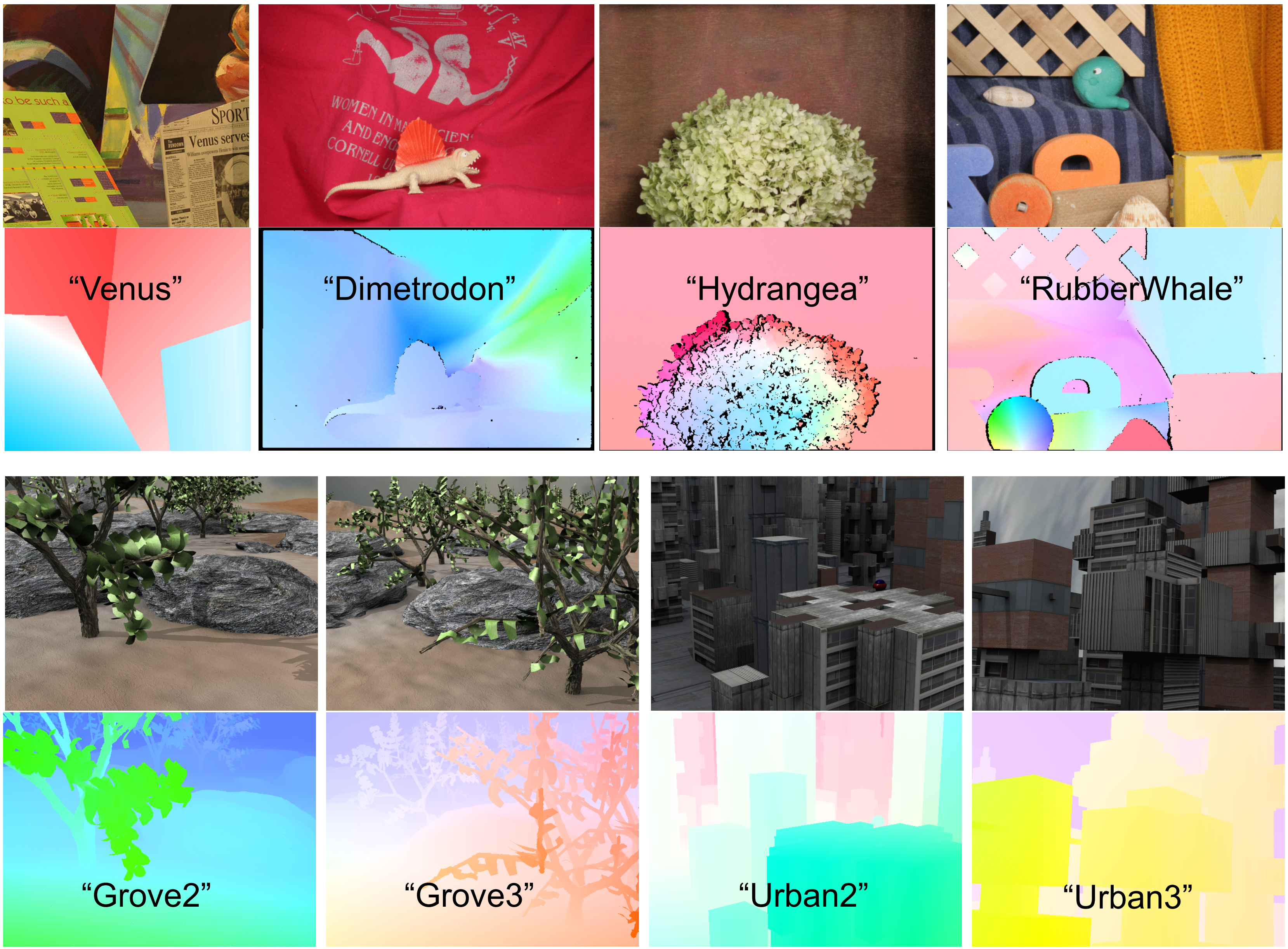


First image Second image

Evaluation

- Modern implementation
  - Preprocessing: Rudin-Osher-Fatemi (ROF) structure texture decomposition
  - Standard incremental multi-resolution technique, 10 warping steps per level
  - Graduate non-convexity (GNC) for non-quadratic penalty
  - Bicubic interpolation to warp image and its derivatives
  - 5-point image derivative filter
  - Temporal averaging of image derivatives
  - 5X5 median filtering of intermediate flow field per warping step

Training set (Baker et al. 2007)

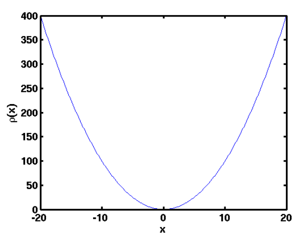


Evaluation

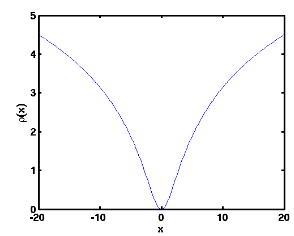
$$E_{Su}(\mathbf{u},\mathbf{v}) = \sum_{i,j} \rho_S(u_{i,j} - u_{i+1,j}) + \rho_S(u_{i,j} - u_{i,j+1})$$

$$E_D(\mathbf{u},\mathbf{v}) = \sum_{i,j} \rho_D(I_1(i,j) - I_2(i+u_{i,j}, j+v_{i,j}))$$

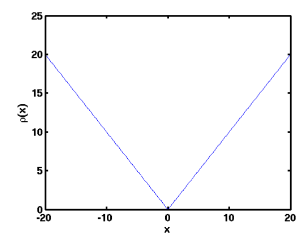
Training	HS	Classic-L	**Classic-C**
	(Quadratic)	(Lorentzian)	(Charbonnier)



$$\rho(x) = x^2$$



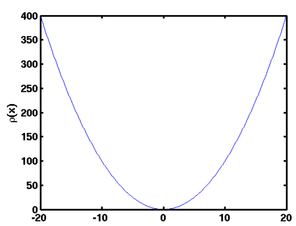
$$\rho(x) = \log(1 + \frac{x^2}{2\sigma^2})$$

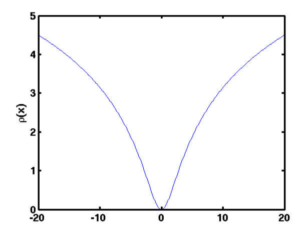


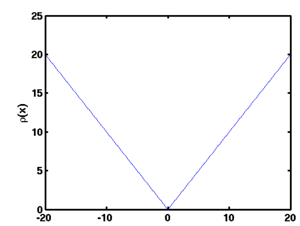
$$\rho(x) = \sqrt{x^2 + \varepsilon^2}$$

Evaluation

Training	HS (Quadratic)	Classic-L (Lorentzian)	**Classic-C** (Charbonnier)
Avg. EPE	0.384	0.319	0.298



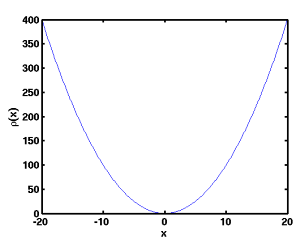


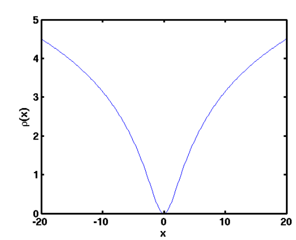


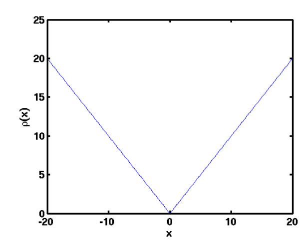
End-point error $$\mathrm{EPE} = \sqrt{\| \mathbf{u} - \mathbf{u}_{GT} \|^2 + \| \mathbf{v} - \mathbf{v}_{GT} \|^2}$$

Evaluation

Training	HS (Quadratic)	Classic-L (Lorentzian)	**Classic-C** (Charbonnier)
Avg. EPE	0.384	0.319	0.298
Significant*	1	1	-
p-value*	0.0078	0.0078	-







*Wilcoxon signed rank test: *p*-value < 0.05 leads to rejection (significant=1)

Test set (Baker et al. 2007)

Evaluation

Test	**Classic-C**	Adaptive*	Complementary OF*
Avg. EPE	0.408	0.401	0.485
Rank	14.9	11.5	10.1

* Two top published methods (Dec. 2009.)

- What is important?
- Approach: change one property of baseline (Classic-C) at a time, compare avg. EPE, and test statistical significance

Evaluation

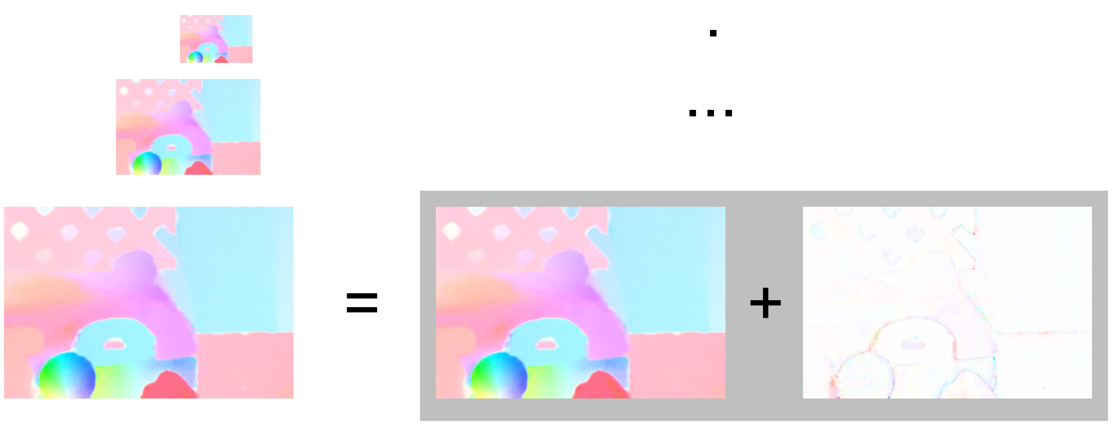
	Middlebury ***training*** set	**Avg. EPE**	signif.	*p*-value
Baseline	Classic-C	0.298	-	-
Preprocessing	Brightness constancy	0.288	0	0.2969
	Gradient constancy	0.305	0	0.4609
Interpolation and image derivatives	Bilinear interpolation	0.302	0	0.1016
	Central difference filter	0.300	0	0.7266
	7-point derivative filter	0.302	0	0.3125
	Spline-based bicubic interpolation	**0.290**	**1**	**0.0391** ☺
	No temporal average of derivatives	0.306	0	0.1562
Coarse-to-fine estimation and GNC	Downsampling factor 0.5	0.298	0	1.0000
	3 warping steps per level	0.304	0	0.9688
	No graduated non-convexity (GNC)	0.354	0	0.1094
Penalty function	**Generalized Charbonnier-0.45**	**0.292**	**1**	**0.0156** ☺
	Generalized Charbonnier-0.25	0.298	0	1.0000
Median filtering	Median filter size 3 X 3	0.305	0	0.1016
	Median filter size 7 X 7	0.305	0	0.5625
	Median filtering twice	0.300	0	1.0000
	No median filtering	**0.352**	**1**	**0.0078** ☹
Best practices	**Classic++**	**0.285**	**1**	**0.0078** ☺

Evaluation

$$E_D(\mathbf{u},\mathbf{v}) = \sum_{i,j} \rho_D(I_1(i,j) - I_2(i+u_{i,j}, j+v_{i,j}))$$

$$I_2(i+u_{i,j}^k + du_{i,j}, j+v_{i,j}^k + dv_{i,j})$$

- Coarse-to-fine, warping-based estimation for nonlinear data term (Bergen et al. 1992, Brox et al. 2004)
  - Median filter flow field per warping (Wedel et al. 2008)



$(\boldsymbol{u}^{k+1}, \boldsymbol{v}^{k+1})$ $(\boldsymbol{u}^k, \boldsymbol{v}^k)$ $(d\boldsymbol{u}, d\boldsymbol{v})$

Evaluation

Median filtering improves results significantly

Training	Classic-C	w/o MF
Avg. EPE	0.298	0.352
Significance	-	1
p-value	-	0.0078

EPE 0.093

EPE 0.113

Evaluation

- Summary
  - Classical formulation competitive with modern implementation (8^{th} out of ~40, June 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	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
Classic+NL [38]	5.8	**0.08** 1	**0.23** 1	0.07 2	0.22 8	0.74 9	0.18 10	0.29 7	0.65 7	0.19 7	**0.15** 1	0.73 3	**0.09** 1	**0.64** 1	**0.93** 1	**0.47** 1	0.52 9	1.12 3	0.33 6	0.16 21	0.13 6	0.29 27	**0.49** 1	**0.98** 1	0.74 5
MDP-Flow [30]	6.4	0.09 2	0.25 3	0.08 5	0.19 2	0.54 2	0.18 10	**0.24** 1	0.55 3	0.20 9	0.16 4	0.91 8	**0.09** 1	0.74 2	1.06 2	0.61 5	0.46 5	1.02 2	0.35 8	0.12 7	0.14 11	0.17 7	0.78 17	1.68 20	0.97 17
NL-TV-NCC [28]	7.4	0.10 7	0.26 6	0.08 5	0.22 8	0.72 8	0.15 3	0.35 9	0.85 9	**0.16** 1	**0.15** 1	0.70 2	**0.09** 1	0.79 5	1.16 7	0.51 3	0.78 17	1.38 8	0.48 14	0.16 21	0.15 16	0.26 19	0.55 3	1.16 3	**0.55** 1
Layer+dense [37]	8.1	0.09 2	0.28 12	0.08 5	0.22 8	0.74 9	0.19 13	0.25 4	0.58 4	0.21 11	0.17 6	0.92 10	**0.09** 1	0.87 9	1.17 8	0.94 21	**0.35** 1	**0.95** 1	0.31 4	0.16 21	0.13 6	0.28 25	0.55 3	1.17 4	0.79 6
Complementary OF [24]	9.1	0.10 7	0.26 6	0.09 10	0.20 6	0.70 6	0.14 2	0.35 9	0.85 9	**0.16** 1	0.19 9	1.05 14	0.10 7	0.87 9	1.25 10	0.71 8	1.46 28	1.61 13	0.73 22	0.11 5	0.12 4	0.21 13	0.60 5	1.37 8	0.80 9
Adaptive [23]	10.2	0.09 2	0.26 6	**0.06** 1	0.23 13	0.78 12	0.18 10	0.54 20	1.19 22	0.21 11	0.18 7	0.91 8	0.10 7	0.88 12	1.25 10	0.73 13	0.50 8	1.28 6	0.31 4	0.14 14	0.16 21	0.22 15	0.65 8	1.37 8	0.79 6
Adapt-Window [36]	11.5	0.10 7	0.24 2	0.09 10	0.19 2	0.59 3	0.15 3	0.27 5	0.64 5	0.17 4	0.18 7	0.82 5	0.11 11	0.74 2	1.07 3	0.56 4	1.78 35	1.73 18	0.95 30	0.22 33	0.16 21	0.45 36	0.70 11	1.28 5	0.88 13
Classic++ [39]	2.2	0.09 2	0.25 3	0.07 2	0.23 13	0.78 12	0.19 13	0.43 12	1.00 14	0.22 14	0.20 12	1.11 17	0.10 7	0.87 9	1.30 13	0.66 6	0.47 6	1.62 15	0.33 6	0.17 25	0.14 11	0.32 31	0.79 19	1.64 17	0.92 15

  - Median filtering intermediate flow field is key

  - Any principle behind?

Principles

Effect of median filtering

- Free from outliers and increased accuracy

Classic-C (EPE 0.093)

Without median filtering (EPE 0.113)

- But Higher energy!

502,384 449,290

Principles

- What is being minimized?

Principles

- What is being minimized?

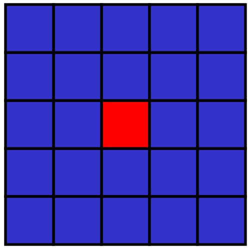
$$E_A(\mathbf{u},\mathbf{v},\hat{\mathbf{u}},\hat{\mathbf{v}}) = \sum_{i,j} \{\rho_D(I_1(i,j) - I_2(i+u_{i,j}, j+v_{i,j}))$$

$$+\lambda[\rho_S(u_{i,j} - u_{i+1,j}) + \rho_S(u_{i,j} - u_{i,j+1}) +$$

$$\rho_S(v_{i,j} - v_{i+1,j}) + \rho_S(v_{i,j} - v_{i,j+1})]\}$$

$$+\lambda_2(\|\mathbf{u} - \hat{\mathbf{u}}\|^2 + \|\mathbf{v} - \hat{\mathbf{v}}\|^2)$$

$$+\sum_{i,j} \sum_{(i',j') \in N_{i,j}} \lambda_3(|\hat{u}_{i,j} - \hat{u}_{i',j'}| + |\hat{v}_{i,j} - \hat{v}_{i',j'}|)$$



Non-local term to integrate information over large neighborhood (Buades et al. 2005, Gilboa et al. 2008, Li&Osher 2009)

Principles

Alternating optimization

$$E_A(\mathbf{u},\mathbf{v},\hat{\mathbf{u}},\hat{\mathbf{v}}) = \sum_{i,j}\{\rho_D(I_1(i,j) - I_2(i+u_{i,j}, j+v_{i,j}))$$

$$+\lambda[\rho_S(u_{i,j} - u_{i+1,j}) + \rho_S(u_{i,j} - u_{i,j+1}) +$$

$$\rho_S(v_{i,j} - v_{i+1,j}) + \rho_S(v_{i,j} - v_{i,j+1})]\}$$

$$+\lambda_2(\|\mathbf{u}-\hat{\mathbf{u}}\|^2 + \|\mathbf{v}-\hat{\mathbf{v}}\|^2)$$

$$+\sum_{i,j}\sum_{(i',j')\in N_{i,j}}\lambda_3(|\hat{u}_{i,j} - \hat{u}_{i',j'}| + |\hat{v}_{i,j} - \hat{v}_{i',j'}|)$$

Principles

- Results of Alternating optimization

	Classic-C	Classic-C-A
Avg. EPE	0.298	0.296
Significant	-	0
p-value	-	0.7188

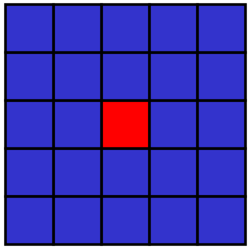
Principles

- Summary
  - Formalize median filtering as approximately optimizing a different objective with a non-local term
  - Can we use such insight to improve flow estimation techniques further?

Improved model

- Non local term

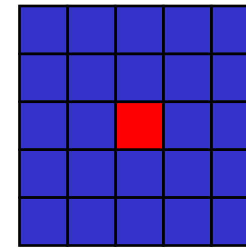
$$\sum_{i,j}\sum_{(i',j')\in N_{i,j}}\left(|\hat{u}_{i,j}-\hat{u}_{i',j'}|+|\hat{v}_{i,j}-\hat{v}_{i',j'}|\right)$$



Improved model

- Non local term

$$\sum_{i,j} \sum_{(i',j') \in N_{i,j}} (|\hat{u}_{i,j} - \hat{u}_{i',j'}| + |\hat{v}_{i,j} - \hat{v}_{i',j'}|)$$



- Destroy fine structures: different from majority

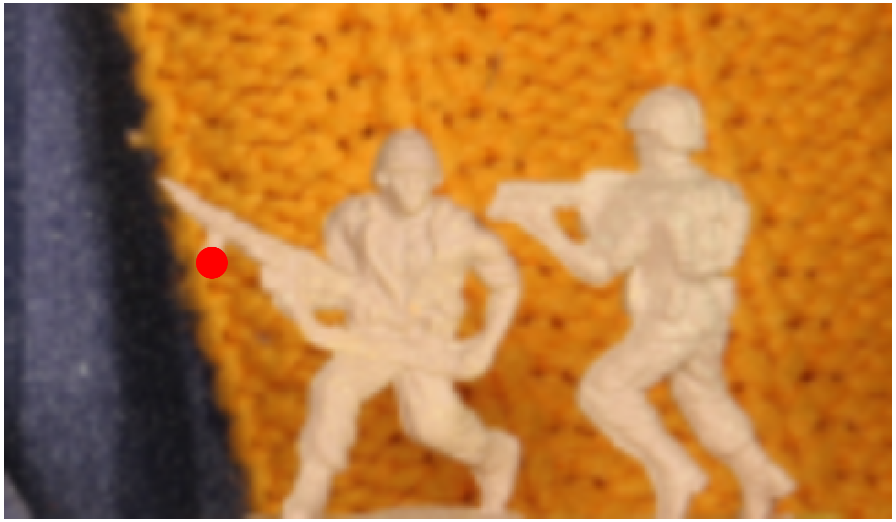
Classic++

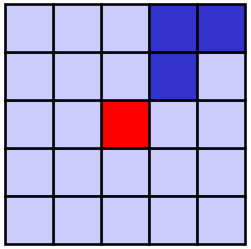
First image

Improved model

- Approach: weight neighbors adaptively

$$\sum_{i,j} \sum_{(i',j') \in N_{i,j}} w_{i,j,i',j'} (|\hat{u}_{i,j} - \hat{u}_{i',j'}| + |\hat{v}_{i,j} - \hat{v}_{i',j'}|)$$





Improved model

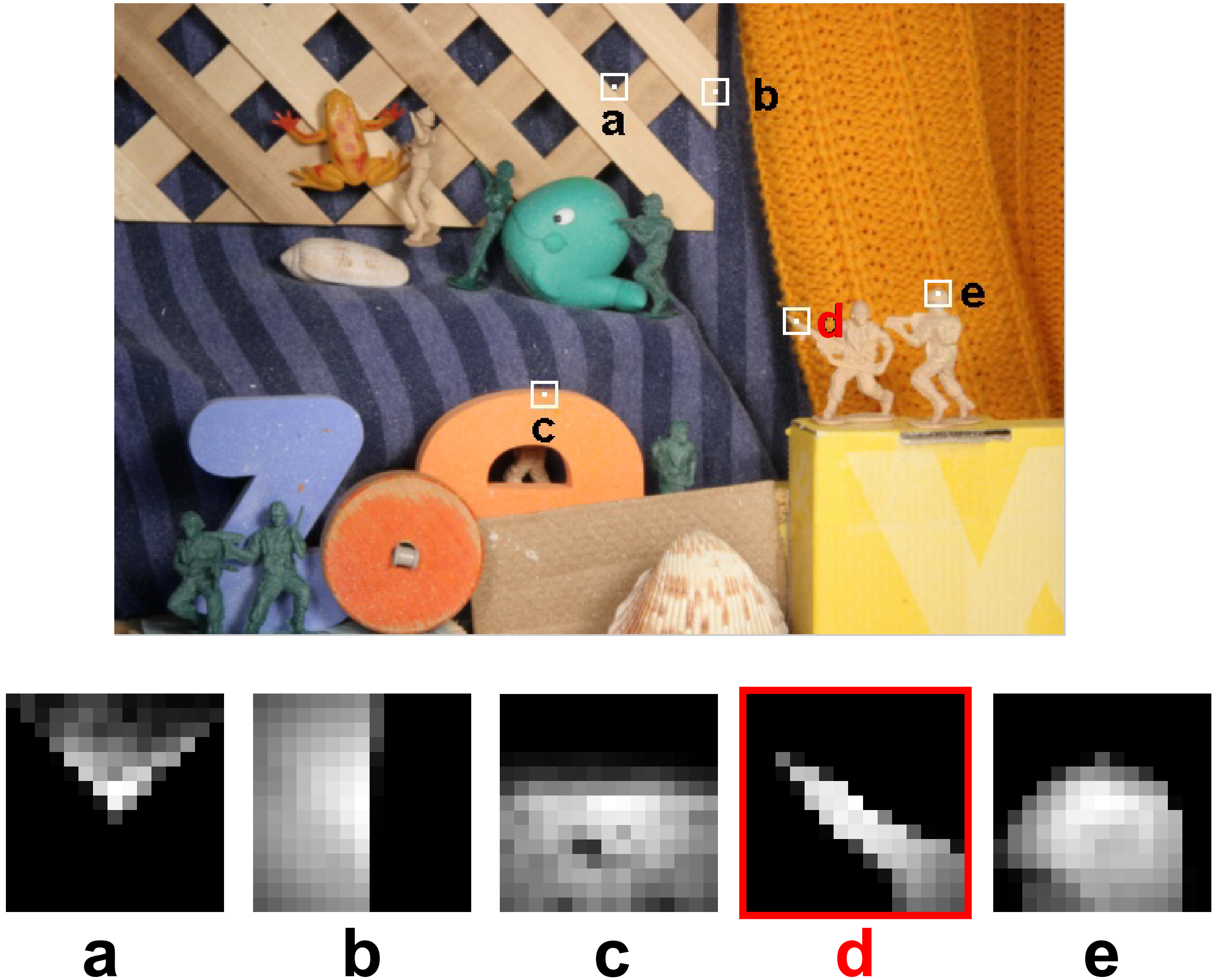
- Approach: weight neighbors adaptively

$$\sum_{i,j}\sum_{(i',j')\in N_{i,j}} w_{i,j,i',j'}(|\hat{u}_{i,j}-\hat{u}_{i',j'}|+|\hat{v}_{i,j}-\hat{v}_{i',j'}|)$$

- According to spatial distance, color difference, and occlusion state (Ren 2008, Sand et al. 2008, Xiao et al. 2006, Yoon et al. 2006)

$$w_{i,j,i',j'} \propto \exp\left\{-\frac{(i-i')^2+(j-j')^2}{2\sigma_1^2}-\frac{\|\mathbf{I}(i,j)-\mathbf{I}(i',j')\|^2}{2\sigma_2^2}\right\}\frac{o(i',j')}{o(i,j)}$$

Improved model



Improved model

Avg EPE	Classic++	Classic+NL	Adaptive*
Training	0.285	0.221	0.264
Test	0.406	0.319	0.401

* Wedel et al. 2009, the previous state of the art (2009 Dec.)

Classic++

Classic+NL

First image

Improved model

- Middlebury public table (June 2010)

Average endpoint error	avg. rank	Army (Hidden texture)			Mequon (Hidden texture)			Schefflera (Hidden texture)			Wooden (Hidden texture)			Grove (Synthetic)			Urban (Synthetic)			Yosemite (Synthetic)			Teddy (Stereo)		
		GT all	im0 disc	im1 untext	GT all	im0 disc	im1 untext	GT all	im0 disc	im1 untext	GT all	im0 disc	im1 untext	GT all	im0 disc	im1 untext	GT all	im0 disc	im1 untext	GT all	im0 disc	im1 untext	GT all	im0 disc	im1 untext
Classic+NL [31]	5.3	0.08 1	0.23 1	0.07 2	0.22 8	0.74 8	0.18 10	0.29 5	0.65 5	0.19 7	0.15 1	0.73 3	0.09 1	0.64 1	0.93 1	0.47 1	0.52 8	1.12 2	0.33 5	0.16 22	0.13 4	0.29 25	0.49 1	0.98 1	0.74 5
MDP-Flow [26]	6.0	0.09 2	0.25 3	0.08 5	0.19 2	0.54 1	0.18 10	0.24 1	0.55 2	0.20 8	0.16 4	0.91 8	0.09 1	0.74 2	1.06 2	0.61 6	0.46 5	1.02 1	0.35 7	0.12 6	0.14 10	0.17 6	0.78 17	1.68 19	0.97 16
NL-TV-NCC [25]	7.0	0.10 6	0.26 6	0.08 5	0.22 8	0.72 6	0.15 3	0.35 9	0.85 8	0.16 1	0.15 1	0.70 2	0.09 1	0.79 5	1.16 6	0.51 3	0.78 14	1.38 7	0.48 12	0.16 22	0.15 16	0.26 20	0.55 3	1.16 3	0.55 1
Complementary OF [21]	8.4	0.10 6	0.26 6	0.09 8	0.21 6	0.73 7	0.15 3	0.34 8	0.85 8	0.16 1	0.19 8	1.06 13	0.10 6	0.87 9	1.25 8	0.72 10	1.46 24	1.62 11	0.73 20	0.11 4	0.12 2	0.21 14	0.60 4	1.37 7	0.80 8
Adaptive [20]	9.3	0.09 2	0.26 6	0.06 1	0.23 11	0.78 9	0.18 10	0.54 17	1.19 19	0.21 10	0.18 6	0.91 8	0.10 6	0.88 11	1.25 8	0.73 13	0.50 7	1.28 5	0.31 3	0.14 14	0.16 21	0.22 16	0.65 8	1.37 7	0.79 6
Adapt-Window [34]	10.4	0.10 6	0.24 2	0.09 8	0.19 2	0.59 2	0.15 3	0.27 3	0.64 3	0.17 4	0.18 6	0.82 5	0.11 9	0.74 2	1.07 3	0.56 4	1.78 32	1.73 15	0.95 27	0.22 31	0.16 21	0.45 34	0.70 11	1.28 4	0.88 13
ComplOF-FED-GPU [36]	10.9	0.11 11	0.29 12	0.10 17	0.21 6	0.78 9	0.14 2	0.32 7	0.79 7	0.17 4	0.19 8	0.99 11	0.11 9	0.89 12	1.29 12	0.73 13	1.25 19	1.74 16	0.64 17	0.14 14	0.13 4	0.30 26	0.64 6	1.50 10	0.83 9
DPOF [18]	11.3	0.13 22	0.35 20	0.09 8	0.25 13	0.79 12	0.19 13	0.24 1	0.49 1	0.21 10	0.19 8	0.62 1	0.15 19	0.74 2	1.09 4	0.49 2	0.66 12	1.80 19	0.63 16	0.19 28	0.17 25	0.35 31	0.50 2	1.08 2	0.55 1
Classic++ [32]	11.3	0.09 2	0.25 3	0.07 2	0.23 11	0.78 9	0.19 13	0.43 11	1.00 13	0.22 12	0.20 11	1.11 15	0.10 6	0.87 9	1.30 13	0.66 7	0.47 6	1.62 11	0.33 5	0.17 25	0.14 10	0.32 29	0.79 18	1.64 16	0.92 15
ACK-Prior [27]	11.4	0.11 11	0.25 3	0.09 8	0.18 1	0.59 2	0.13 1	0.27 3	0.64 3	0.16 1	0.15 1	0.78 4	0.09 1	0.82 6	1.14 5	0.71 9	1.90 33	1.90 20	0.99 30	0.23 34	0.17 25	0.49 35	0.77 15	1.44 9	0.91 14
Aniso. Huber-L1 [22]	11.6	0.10 6	0.28 10	0.08 5	0.31 20	0.88 16	0.28 22	0.56 19	1.13 16	0.29 21	0.20 11	0.92 10	0.13 13	0.84 7	1.20 7	0.70 8	0.39 1	1.23 3	0.28 1	0.17 25	0.15 16	0.27 24	0.64 6	1.36 6	0.79 6
TriangleFlow [30]	13.1	0.11 11	0.29 12	0.09 8	0.26 16	0.95 20	0.17 9	0.47 15	1.07 14	0.18 6	0.16 4	0.87 7	0.09 1	1.07 20	1.47 26	1.10 21	0.87 15	1.39 8	0.57 14	0.15 18	0.19 32	0.23 17	0.63 5	1.33 5	0.84 10
TV-L1-improved [17]	14.3	0.09 2	0.26 6	0.07 2	0.20 5	0.71 5	0.16 6	0.53 16	1.18 18	0.22 12	0.21 15	1.24 20	0.11 9	0.90 13	1.31 15	0.72 10	1.51 26	1.93 21	0.84 23	0.18 27	0.17 25	0.31 27	0.73 13	1.62 15	0.87 12
CBF [12]	14.7	0.10 6	0.28 10	0.09 8	0.34 22	0.80 13	0.37 23	0.43 11	0.95 11	0.26 16	0.21 15	1.14 16	0.13 13	0.90 13	1.27 11	0.82 16	0.41 2	1.23 3	0.30 2	0.23 34	0.19 32	0.39 32	0.76 14	1.56 12	1.02 17
Brox et al. [5]	15.8	0.11 11	0.32 16	0.11 21	0.27 18	0.93 18	0.22 18	0.39 10	0.94 10	0.24 15	0.24 17	1.25 21	0.13 13	1.10 24	1.39 21	1.43 29	0.89 16	1.77 17	0.55 13	0.10 2	0.13 4	0.11 1	0.91 20	1.83 21	1.13 22
Rannacher [23]	16.1	0.11 11	0.31 14	0.09 8	0.25 13	0.84 15	0.21 16	0.57 21	1.27 25	0.26 16	0.24 17	1.32 24	0.13 13	0.91 16	1.33 16	0.72 10	1.49 25	1.95 23	0.78 21	0.15 18	0.14 10	0.26 20	0.69 10	1.58 14	0.86 11
F-TV-L1 [15]	16.2	0.14 23	0.35 20	0.14 25	0.34 22	0.98 21	0.26 21	0.59 24	1.19 19	0.26 16	0.27 22	1.36 25	0.16 20	0.90 13	1.30 13	0.76 15	0.54 9	1.62 11	0.36 8	0.13 10	0.15 16	0.20 11	0.68 9	1.56 12	0.66 3
Second-order prior [8]	16.8	0.11 11	0.31 14	0.09 8	0.26 16	0.93 18	0.20 15	0.57 21	1.25 24	0.26 16	0.20 11	1.04 12	0.12 12	0.94 17	1.34 17	0.83 18	0.61 11	1.93 21	0.47 11	0.20 29	0.16 21	0.34 30	0.77 15	1.64 16	1.07 19
Fusion [6]	17.1	0.11 11	0.34 18	0.10 17	0.19 2	0.69 4	0.16 6	0.29 5	0.66 6	0.23 14	0.20 11	1.19 18	0.14 17	1.07 20	1.42 23	1.22 23	1.35 21	1.49 10	0.86 25	0.20 29	0.20 34	0.26 20	1.07 24	2.07 27	1.39 26
p-harmonic [29]	18.0	0.12 19	0.36 23	0.11 21	0.25 13	0.82 14	0.21 16	0.57 21	1.24 21	0.28 20	0.26 20	1.20 19	0.19 22	1.07 20	1.39 21	1.31 25	0.44 3	1.65 14	0.37 9	0.15 18	0.16 21	0.21 14	0.87 19	1.76 20	1.06 18
SegOF [10]	19.8	0.15 24	0.36 23	0.10 17	0.57 26	1.16 26	0.59 30	0.68 25	1.24 21	0.64 24	0.32 24	0.86 6	0.26 24	1.18 30	1.50 29	1.47 31	1.63 30	2.09 25	0.96 28	0.08 1	0.13 4	0.12 2	0.70 11	1.50 10	0.69 4
Dynamic MRF [7]	19.8	0.12 19	0.34 18	0.11 21	0.22 8	0.89 17	0.16 6	0.44 14	1.13 16	0.20 8	0.24 17	1.29 23	0.14 17	1.11 25	1.52 30	1.13 22	1.54 27	2.37 33	0.93 26	0.13 10	0.12 2	0.31 27	1.27 29	2.33 33	1.66 28
LDOF [28]	19.8	0.12 19	0.35 20	0.10 17	0.32 21	1.06 23	0.24 20	0.43 11	0.98 12	0.30 23	0.45 26	2.48 37	0.26 24	1.01 19	1.37 19	1.05 20	1.10 18	2.08 24	0.67 18	0.12 6	0.15 16	0.24 18	0.94 21	2.05 24	1.10 20
Ad-TV-NDC [37]	21.3	0.23 33	0.40 27	0.31 34	0.92 34	1.42 32	0.93 33	1.05 32	1.60 30	0.74 28	0.48 27	1.27 22	0.49 29	0.85 8	1.25 8	0.60 5	0.44 3	1.47 9	0.32 4	0.12 6	0.13 4	0.19 8	1.59 34	2.06 26	2.87 35
Learning Flow [11]	22.8	0.11 11	0.32 16	0.09 8	0.29 19	0.99 22	0.23 19	0.55 18	1.24 21	0.29 21	0.36 25	1.56 27	0.25 23	1.25 32	1.64 34	1.41 27	1.55 29	2.32 32	0.85 24	0.14 14	0.18 30	0.24 18	1.09 25	2.09 29	1.27 23
Filter Flow [19]	23.8	0.17 26	0.39 26	0.13 24	0.43 24	1.09 25	0.38 24	0.75 26	1.34 26	0.78 30	0.70 31	1.54 26	0.68 31	1.13 27	1.38 20	1.51 32	0.57 10	1.32 6	0.44 10	0.22 31	0.23 36	0.26 20	0.96 22	1.66 18	1.12 21
GraphCuts [14]	24.5	0.16 25	0.38 25	0.14 25	0.59 29	1.36 31	0.46 25	0.56 19	1.07 14	0.64 24	0.26 20	1.14 16	0.17 21	0.96 18	1.35 18	0.84 19	2.25 36	1.79 18	1.22 34	0.22 31	0.17 25	0.43 33	1.22 28	2.05 24	1.78 30
Modified CLG [35]	24.8	0.19 29	0.46 30	0.17 27	0.49 25	1.08 24	0.51 27	0.93 29	1.59 29	0.82 32	0.49 28	1.65 30	0.42 27	1.14 28	1.48 27	1.42 28	1.06 17	2.16 29	0.68 19	0.12 6	0.14 10	0.20 11	1.12 27	2.17 30	1.52 27
Black & Anandan [4]	25.0	0.18 27	0.42 28	0.19 29	0.58 28	1.31 29	0.50 26	0.95 30	1.58 28	0.70 26	0.49 28	1.59 28	0.45 28	1.08 23	1.42 23	1.22 23	1.43 22	2.28 30	0.83 22	0.15 18	0.17 25	0.17 6	1.11 26	1.98 23	1.30 24
SPSA-learn [13]	25.4	0.18 27	0.45 29	0.17 27	0.57 26	1.32 30	0.51 27	0.84 27	1.50 27	0.72 27	0.52 30	1.64 29	0.49 29	1.12 26	1.42 23	1.39 26	1.75 31	2.14 27	1.06 33	0.13 10	0.13 4	0.19 8	1.32 30	2.08 28	1.73 29

Conclusions

- Classical formulations competitive with modern practices
- Median filtering is key to accuracy, but increases energy
- Formalize median filtering as non-local term that integrates information over large spatial neighborhood
- Weighting neighbors adaptively preserves motion details
- MATLAB code: http://www.cs.brown.edu/~dqsun/

Acknowledgement

- Gift from *Intel Corp*. and NSF CRCNS award IIS-0904875
- Thanks to
  - *Payman Yadollahpour* for his early work on implementing HS
  - *Thomas Brox, Andreas Wedel, and Manuel Werlberger* for clarifying details about their paper
  - *CVPR reviewers* for constructive comments, esp. connections to non local regularization and Yoon's work

http://www.cs.brown.edu/~dqsun/research/software.html

References

- D. Sun, S. Roth, and M. Black: *Secrets of optical flow estimation and their principles*. CVPR 2010.
- Recent related work:
  - M. Werlberger, T. Pock, and H. Bischof: *Motion estimation with non-local total variation regularization*. CVPR 2010.