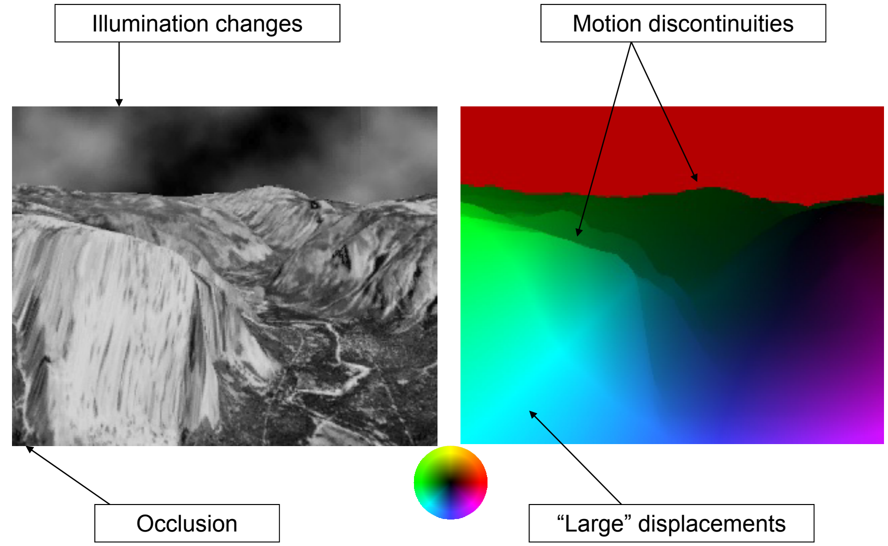
Variational Optical Flow Estimation

Part II: Modeling Aspects

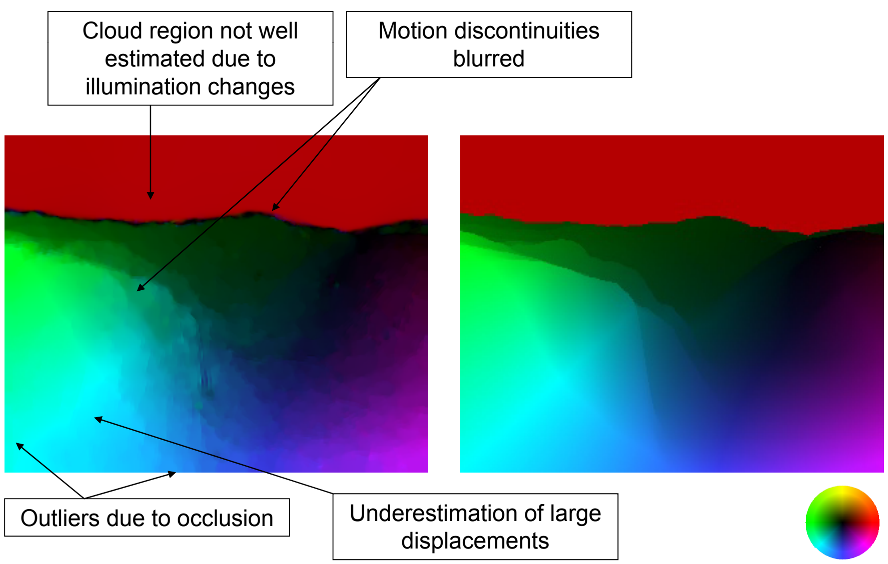
Discontinuity preserving smoothness terms
Robust data terms
Image features for matching
Large displacements

Funded by ONR-MURI and the German Academic Exchange Service (DAAD)

Yosemite test sequence



Limitations of Horn-Schunck method



Agenda of this part: challenges we would like to resolve

Motion discontinuities

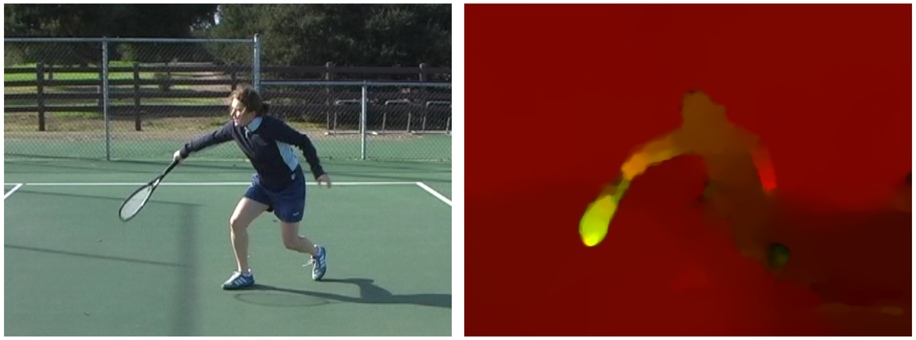
Occlusions

Illumination changes

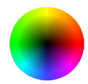
Large displacements

Really large displacements

Motion discontinuities



- Different, independently moving objects cause discontinuities in the correct flow field
- Usually we do not know the object boundaries



Motion discontinuities and the smoothness assumption

- Fact: smoothness assumption not satisfied everywhere
- Quadratic penalizer → Gaussian error distribution

$$E(u,v) = \int_\Omega (I_x u + I_y v + I_z)^2 + \alpha(|\nabla u|^2 + |\nabla v|^2)\, dxdy$$

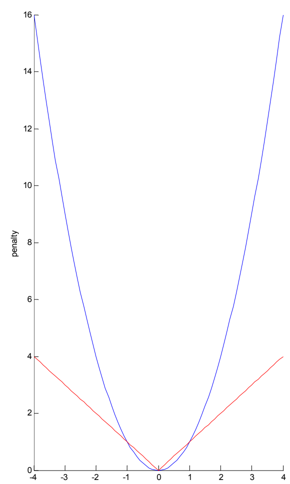
- Too much influence given to outliers → blurring
- Remedy: robust error norm
  (Huber 1981, Cohen 1993, Schnörr 1994, Black-Anandan 1996, Mémin-Pérez 1998)

Non-quadratic smoothness term

- Applying a robust function:

$$E_S(u,v) = \int_\Omega \Psi(|\nabla u|^2 + |\nabla v|^2)\, dxdy$$

- For example: $\Psi(s^2) = \sqrt{s^2 + \epsilon^2}$
  Total variation (TV) norm
- Advantage: still convex
  → global optimum!



Minimization: Euler-Lagrange equations

$$-\text{div}\left(\Psi'(|\nabla u|^2 + |\nabla v|^2)\nabla u\right) = 0$$

$$-\text{div}\left(\Psi'(|\nabla u|^2 + |\nabla v|^2)\nabla v\right) = 0$$

- Interpretation:
  adaptive smoothing with a joint **diffusivity** $\Psi'(s^2)$

  TV norm: $\Psi'(s^2) = \frac{1}{\sqrt{s^2 + \epsilon^2}}$

- Diffusivity depends on the unknown flow
  → nonlinear system of equations

Resolve nonlinearity: fixed point iterations

1. Keep Ψ' fixed for an (initial) solution (u, v) (fixed point)

2. Solve resulting <u>linear system</u> to obtain a new fixed point (see Part I+III)

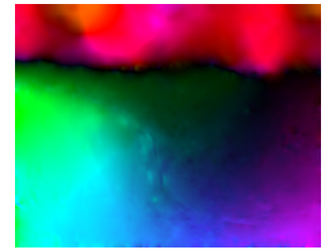
3. Update Ψ', iterate

- Linear system need not be solved exactly in each iteration

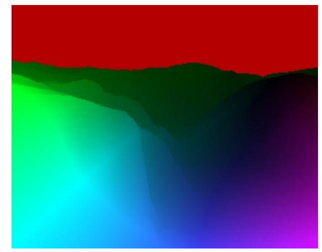
- Scheme also called **lagged diffusivity** or **lagged nonlinearity**

Result: improvements at motion discontinuities

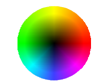
Non-quadratic smoothness term (6.36°)



Horn-Schunck (7.17°)



Ground truth



Occlusions

- Some pixels of frame 1 no longer visible in frame 2

- Pixels matched to most similar pixels in frame 2
  → erroneous motion vectors

Dealing with occlusions and non-Gaussian noise

- Robust function applied to the data term:
  (Black-Anandan 1996, Mémin-Pérez 1998)

$$E(u,v) = \int_\Omega \Psi\left((I_x u + I_y v + I_z)^2\right) + \alpha \Psi\left(|\nabla u|^2 + |\nabla v|^2\right) \, dx dy$$

- Euler-Lagrange equations:

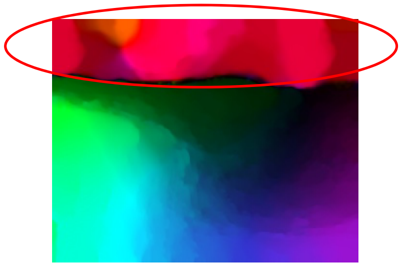
$$\Psi'\left((I_x u + I_y v + I_z)^2\right)(I_x u + I_y v + I_z) I_x - \alpha \, \mathrm{div}\left(\Psi'(|\nabla u|^2 + |\nabla v|^2)\nabla u\right) = 0$$

$$\Psi'\left((I_x u + I_y v + I_z)^2\right)(I_x u + I_y v + I_z) I_y - \alpha \, \mathrm{div}\left(\Psi'(|\nabla u|^2 + |\nabla v|^2)\nabla v\right) = 0$$

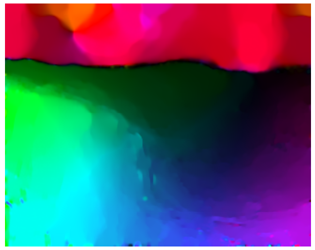
- Interpretation: less importance given to pixels with high matching cost

- Nonlinear system can again be solved with fixed point iterations (lagged nonlinearity)

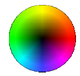
Result: Better treatment of occlusions

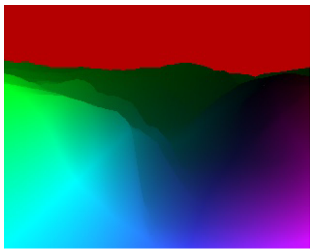


Both terms non-quadratic (5.97°)

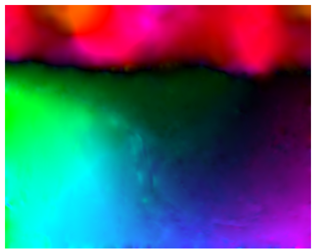


Non-quadratic smoothness term (6.36°)





Ground truth



Horn-Schunck (7.17°)

Illumination changes

“Schefflera” from Middlebury benchmark

Two frames from a Miss Marple movie

Typically caused by:

- Shadows
- light source flickering
- self-adaptive cameras
- different viewing angles and non-Lambertian surfaces

Image features for matching

- Gray value constancy:
  matching the **gray value** of a single pixel

  $$I(x+u, y+v, t+1) - I(x, y, t) = 0$$

- Linearization leads to the well-known data term

  $$E_D(u,v) = \int_\Omega (I_x u + I_y v + I_z)^2 dxdy$$

- We can make use of the pixel’s **color**:

  $$E_D(u,v) = \int_\Omega \sum_{k=1}^{3} (I_x^k u + I_y^k v + I_z^k)^2 dxdy$$

- Does it make sense to use more complex image features?

Matching the gradient

- **Gradient** constancy assumption
  (Uras et al. 1988, Schnörr 1994, Brox et al. 2004)

  $$\nabla I(x+u, y+v, z+1) - \nabla I(x, y, z) = 0$$

- Two important positive effects:
  1. More information: two more equations
  2. Invariant to additive brightness changes

- Negative effects:
  1. Not rotation invariant
  2. More noise sensitive

Gradient constancy assumption

- Linearization:

$$I_{xx}u + I_{xy}v + I_{xz} = 0$$

$$I_{xy}u + I_{yy}v + I_{yz} = 0$$

- New data term:

$$E_{\mathrm{GC}} = \int_\Omega \Psi\left((I_{xx}u + I_{xy}v + I_{xz})^2 + (I_{xy}u + I_{yy}v + I_{yz})^2\right)\,dxdy$$

- Euler-Lagrange equations and numerical scheme are straightforward extensions of the versions with classic OFC
- Can be written in compact form using the **motion tensor notation**

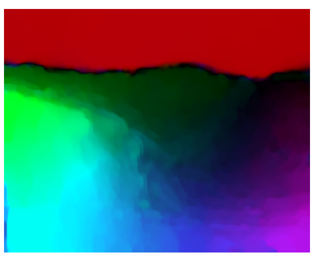
Combination of features

- Combine with classic optic flow constraint:

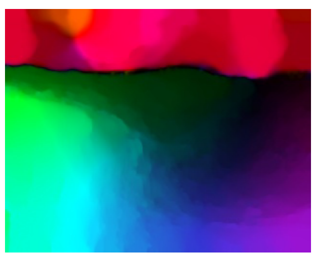
$$E(u,v) = \int_\Omega \Psi\left((I_x u + I_y v + I_z)^2\right)\,dxdy$$
$$+\gamma \int_\Omega \Psi\left((I_{xx}u + I_{xy}v + I_{xz})^2 + (I_{xy}u + I_{yy}v + I_{yz})^2\right)\,dxdy$$
$$+\alpha E_{\mathrm{Smooth}}$$

- Separate robust penalizer for each feature (Bruhn-Weickert 2005)
  → automatically picks the <u>most reliable</u> feature
- Other higher order constancy assumptions feasible, but usually not advantageous (Papenberg et al. 2006)

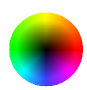
Result: Illumination changes no longer a problem

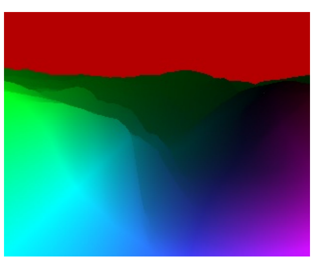


Gradient constancy (3.5°)

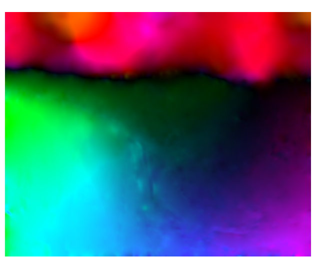


Both terms non-quadratic (5.97°)





Ground truth



Horn-Schunck (7.17°)

Brightness changes: alternatives

1. Structure-texture-decomposition
   (Aujol et al. 2005, Wedel et al. 2008)

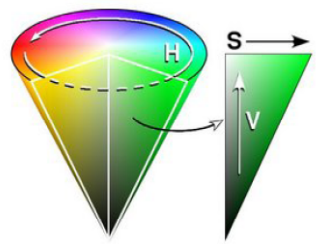
Input image

Texture residual for matching

- Smooth the image by edge preserving diffusion (TV flow)
- Consider residual image for matching

Brightness changes: alternatives

2. Spherical color model (HSV)… (Mileva et al. 2007)
   - Hue (H) denotes the color
     Invariant to multiplicative illumination changes, shadows, shading, and specularities
   - Saturation (S) is invariant to multiplicative illumination changes, shadows, and shading
   - Value (V) is the actual brightness (no invariance)



…with separate robust functions

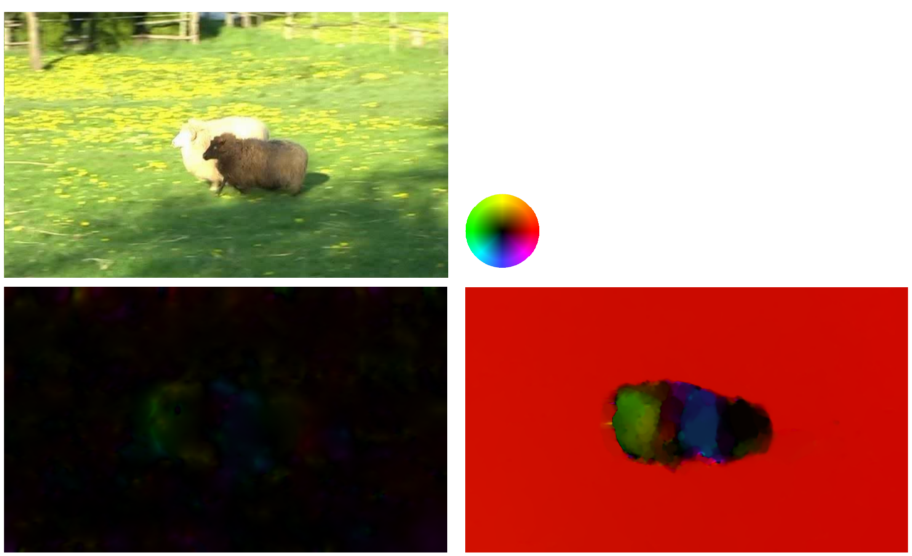
$$E_D = \int_\Omega \Psi\left((H_x u + H_y v + H_t)^2\right) dxdy + \int_\Omega \Psi\left((S_x u + S_y v + S_t)^2\right) dxdy + \int_\Omega \Psi\left((V_x u + V_y v + V_t)^2\right) dxdy$$

→ picks invariant components when needed (Zimmer et al. 2009)

Why not using more descriptive features (Patches,SIFT,HOG)?

- Possible, but not advantageous:
  - Lower accuracy at motion discontinuities
  - Descriptive power of no use in the gradient descent like optimization
- Exploiting descriptive features for large displacement matching needs special treatment in the optimization
  - Liu et al. ECCV 2008 (belief propagation)
  - Brox et al. CVPR 2009
  - Steinbruecker et al. ICCV 2009

Large displacements



Horn&Schunck

Brox et al. 2004

Linearization and large displacements

- Constancy assumptions have been linearized

  $I(x+u, y+v, z+1) - I(x, y, z) = 0 \quad \rightarrow \quad I_x u + I_y v + I_z = 0$
- Linearization only valid for very small displacements (subpixel displacements)
- Fails in practice if displacements are larger than ~ 5 pixels (depends on smoothness of image data)
- What happens if we do not linearize?
  (Nagel-Enkelmann 1986, Alvarez et al. 2000)

Non-linearized constancy assumption

- Energy functional:

$$E(u,v) = \int_\Omega \Psi\left((I(\mathbf{x}+\mathbf{w}) - I(\mathbf{x}))^2\right) + \alpha\Psi\left(|\nabla u|^2 + |\nabla v|^2\right)\,\mathbf{dx}$$

$$\mathbf{x} := (x,y,z) \qquad \mathbf{w} := (u,v,1)$$

- Correct description of the matching criterion (no approximation)
- Two problems:
  1. Not convex in $\mathbf{w}$ → multiple local minima
  2. Further source of nonlinearity in Euler-Lagrange equations

Non-linearized constancy assumption

- Euler-Lagrange equations:

$$I_x := \partial_x I(\mathbf{x}+\mathbf{w}) \qquad I_y := \partial_y I(\mathbf{x}+\mathbf{w}) \qquad I_z := I(\mathbf{x}+\mathbf{w}) - I(\mathbf{x})$$

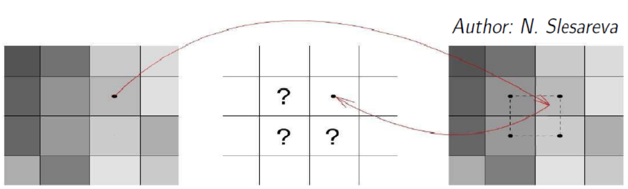
$$\Psi'\left(I_z^2\right) I_z I_x - \alpha\,\mathrm{div}\left(\Psi'(|\nabla u|^2 + |\nabla v|^2)\nabla u\right) = 0$$

$$\Psi'\left(I_z^2\right) I_z I_y - \alpha\,\mathrm{div}\left(\Psi'(|\nabla u|^2 + |\nabla v|^2)\nabla u\right) = 0$$

- Highly nonlinear system of equations
- Another fixed point iteration loop necessary
- Local minima → combination with a multi-scale strategy

Image warping (dates back to Lucas-Kanade 1981)

- Compute expressions of type $I(\mathbf{x}+\mathbf{w})$ for given $\mathbf{w}$
- Second image (at $t+1$), warped by means of (u,v)
- Compute $\tilde{I}(\mathbf{x}) := I(\mathbf{x}+\mathbf{w}) = I(x+u, y+v, t+1)$
- Points may fall between grid points → bilinear interpolation



Gauss-Newton method

- Goal: solve nonlinear system:

$$I_x := \partial_x I(\mathbf{x}+\mathbf{w}) \qquad I_y := \partial_y I(\mathbf{x}+\mathbf{w}) \qquad I_z := I(\mathbf{x}+\mathbf{w}) - I(\mathbf{x})$$

$$\Psi'\left(I_z^2\right) I_z I_x - \alpha\,\mathrm{div}\left(\Psi'(|\nabla u|^2 + |\nabla v|^2)\nabla u\right) = 0$$

$$\Psi'\left(I_z^2\right) I_z I_y - \alpha\,\mathrm{div}\left(\Psi'(|\nabla u|^2 + |\nabla v|^2)\nabla v\right) = 0$$

- Start with an initial fixed point $\mathbf{w}^0$
- Linearize with a first order Taylor expansion

$$I_z^{k+1} := I(\mathbf{x}+\mathbf{w}^{k+1}) - I(\mathbf{x}) \approx I(\mathbf{x}+\mathbf{w}^k) + I_x^k du^k + I_y^k dv^k - I(\mathbf{x})$$
$$= I_x^k du^k + I_y^k dv^k + I_z^k$$

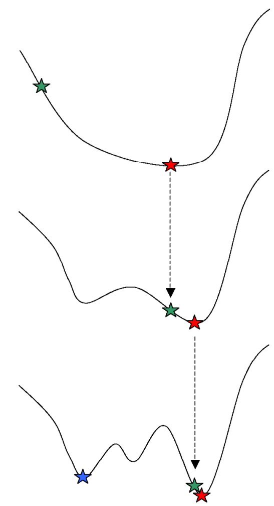
$$u^{k+1} := u^k + du^k \qquad v^{k+1} := v^k + dv^k$$

Gauss-Newton method

- Yields nonlinear system in increment (du^k, dv^k)
- Solve for (du^k, dv^k) by an inner fixed point iteration loop (as in case of the linearized OFC)
- Yields a new fixed point $\mathbf{w}^{k+1} = \mathbf{w}^k + (du^k, dv^k, 0)$
- Difference to linearization in the energy functional: linearization must hold only for the **increment,** not the flow vector itself

Continuation method

- Energy functional has multiple local minima
- How to find the global minimum?
- Heuristic: continuation method
  - Start with a smoothed version of the energy (by smoothing the input images)
  - Find minimum of this energy
  - Use solution as initialization for refinement



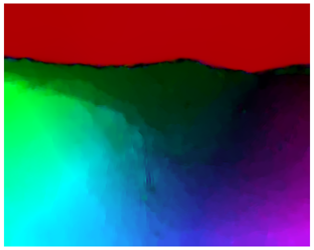
Optical flow with an efficient continuation method

- Create a pyramid of downsampled images
- Using downsampling factors around 0.95 instead of 0.5 increases accuracy (Brox et al. 2004)
- Start outer fixed point iteration loop at coarsest level
- In each such iteration, refine the level
- Two birds killed with one stone:
  - Continuation method avoids local minima
  - Multi-scale implementation provides speedup

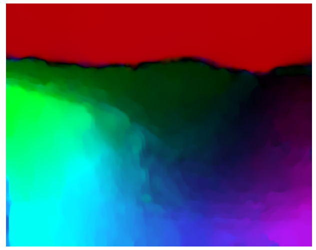
Summary of numerical scheme (Brox et al. 2004)

- Two nested fixed point loops + linear solver
- Outer loop:
  - Removes nonlinearity due to non-linearized constancy assumptions
  - Leads to solving for an increment (du^k, dv^k) in each iteration
  - Requires warping of the second input image
  - Coupled with coarse-to-fine strategy:
    - First iteration starts at the coarsest level
    - Every new iteration is executed at the next finer level
- Inner loop:
  - Removes the remaining nonlinearity in the increment (du^k, dv^k) due to nonquadratic penalizers
  - Lagged nonlinearity: factors kept fixed in each iteration
- Iterative solver for the remaining linear system

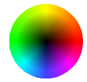
Result: higher accuracy, large displacements possible

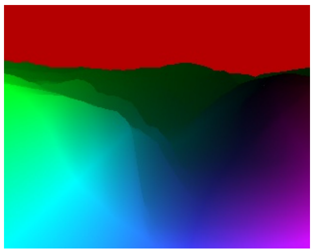


Non-linearized constancy (2.44°)

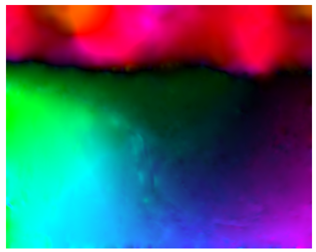


Gradient constancy (3.5°)



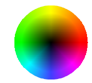


Ground truth

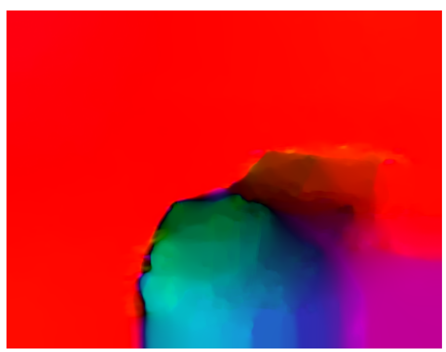


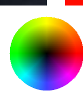
Horn-Schunck (7.17°)

Other results: zoom into a scene

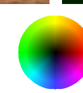


Car in a movie

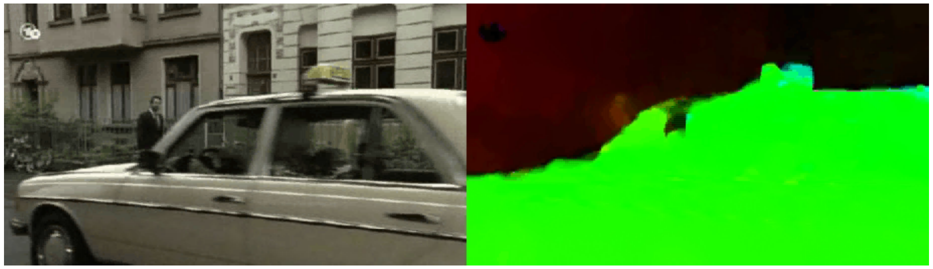


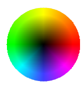


Persons in another movie



Taxi sequence anno 2009





Middlebury benchmark: quantitative comparison

Average angle 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
Complementary OF [26]	5.2	4.44 11	11.2 8	4.04 12	2.51 2	9.77 3	**1.74** 1	3.93 4	10.6 4	**2.04** 1	3.87 7	18.8 5	2.19 6	**3.17** 1	**4.00** 1	2.92 5	4.64 8	13.8 3	3.64 7	2.17 5	3.36 3	2.51 13	3.08 2	7.04 2	3.65 10
Adaptive [25]	5.5	**3.29** 1	**9.43** 1	**2.28** 1	3.10 7	11.4 8	2.46 7	6.58 11	15.7 11	2.52 6	**3.14** 1	15.6 2	**1.56** 1	3.67 10	4.46 6	3.48 9	**3.32** 1	13.0 2	**2.38** 1	2.76 12	4.39 13	1.93 8	3.58 5	8.18 4	2.88 3
Aniso. Huber-L1 [27]	6.9	3.71 3	10.1 3	3.08 3	4.36 14	13.0 10	3.77 14	6.92 13	15.3 9	3.60 15	3.54 3	15.9 3	2.04 4	3.38 3	4.45 5	2.47 2	3.88 3	**12.9** 1	2.74 2	3.37 17	4.36 12	2.85 16	3.16 3	7.52 3	2.90 4
DPOF [20]	7.6	5.12 15	12.9 15	3.49 7	3.07 6	10.3 4	2.44 6	**3.09** 1	7.47 2	2.43 5	3.42 2	**12.9** 1	2.41 11	3.55 7	4.56 10	3.35 7	4.69 9	14.2 4	5.14 10	3.59 19	4.67 17	3.83 22	**2.00** 1	**4.93** 1	**1.65** 1
Spatially variant [21]	7.8	3.73 4	10.2 5	3.33 4	3.02 5	11.0 6	2.67 8	5.36 7	13.8 8	2.35 2	3.67 4	19.3 7	1.84 3	3.81 14	4.81 18	3.69 12	4.48 7	16.0 10	3.90 8	2.11 3	3.26 2	2.12 10	4.66 14	9.41 12	4.35 14
TV-L1-improved [19]	8.7	3.36 2	9.63 2	2.62 2	2.82 4	10.7 5	2.23 2	6.50 10	15.8 12	2.73 7	3.80 6	21.3 14	1.76 2	3.34 2	4.38 4	**2.39** 1	5.97 12	18.1 16	5.67 16	3.57 18	4.92 21	3.43 20	4.01 10	9.84 13	3.44 8
Occlusion bounds [28]	9.5	4.42 9	12.4 11	3.90 10	3.86 11	13.2 11	3.32 12	5.00 6	13.0 5	3.30 10	4.45 12	20.7 11	2.37 9	3.84 15	4.67 14	4.39 21	3.75 2	15.9 9	3.30 5	2.19 6	4.00 11	1.17 2	4.33 12	9.20 8	3.19 5
Rannacher [29]	9.9	4.13 7	11.0 7	3.61 9	3.39 9	12.3 9	2.80 9	7.26 15	17.4 17	3.59 14	4.40 11	23.1 17	2.24 7	3.43 5	4.54 8	2.56 3	5.41 10	18.5 17	4.23 9	2.92 14	3.91 8	2.82 15	3.45 4	9.14 7	3.27 6
Brox et al. [7]	10.2	4.44 11	12.4 11	4.22 16	3.72 10	13.5 12	3.06 10	4.97 5	13.3 7	3.11 8	4.58 14	22.0 15	2.37 9	3.79 12	4.60 11	4.33 20	3.91 4	17.0 13	3.45 6	2.22 7	3.79 6	1.19 3	4.62 13	10.0 14	3.38 7
Multicue MRF [23]	10.2	4.50 13	10.1 3	4.18 15	2.52 3	**7.07** 1	2.36 5	**3.09** 1	**7.41** 1	2.36 3	4.46 13	20.8 12	2.73 14	3.51 6	4.11 2	4.06 16	6.08 14	15.6 8	5.40 14	5.25 28	5.36 23	9.02 27	3.63 6	8.39 5	4.15 13
F-TV-L1 [17]	10.9	5.44 16	12.5 14	5.69 19	5.46 16	15.0 16	4.03 15	7.48 16	16.3 13	3.42 12	5.08 16	23.3 18	2.81 15	3.42 4	4.34 3	3.03 6	4.05 5	15.1 7	3.18 4	2.43 9	3.92 9	1.87 7	3.90 8	9.35 11	2.61 2
CBF [14]	11.2	3.88 5	10.2 5	3.50 8	4.60 15	11.3 7	5.06 16	5.43 8	13.1 6	3.39 11	4.09 8	21.2 13	2.16 5	3.80 13	4.72 17	3.52 10	4.33 6	14.4 5	3.01 3	4.97 26	5.51 26	4.93 26	3.99 9	9.27 10	3.91 12
Fusion [8]	13.6	4.43 10	13.7 17	4.08 13	**2.47** 1	8.91 2	2.24 3	3.70 3	9.68 3	3.12 9	3.68 5	19.8 8	2.54 13	4.26 21	5.16 20	4.31 19	6.32 16	16.8 12	6.15 18	4.55 24	5.78 26	3.10 19	7.12 21	13.6 21	7.86 22
Dynamic MRF [9]	13.6	4.58 14	12.4 11	4.14 14	3.25 8	13.9 13	2.27 4	6.02 9	16.8 14	2.36 3	4.39 10	22.6 16	2.51 12	3.61 8	4.55 9	3.46 8	6.81 20	22.2 24	6.78 21	2.41 8	3.48 4	3.69 21	9.26 25	17.8 25	10.2 25
Second-order prior [10]	13.6	4.03 6	11.6 9	3.35 5	3.88 12	14.0 14	3.08 11	7.21 14	17.6 18	3.57 13	4.14 9	19.9 9	2.31 8	3.66 9	4.86 19	2.73 4	7.32 21	21.2 21	6.76 20	4.02 22	4.58 16	4.01 24	4.27 11	10.4 16	5.12 15
SegOF [12]	14.0	5.85 17	13.5 16	3.98 11	7.40 18	14.9 15	8.13 22	8.55 18	17.3 16	9.01 17	6.50 19	18.1 4	5.14 19	3.90 18	4.53 7	4.81 24	6.57 19	21.7 23	6.81 22	**1.65** 1	3.49 5	**1.08** 1	3.71 7	9.23 9	3.63 9
Learning Flow [13]	16.2	4.23 8	11.7 10	3.41 6	4.16 13	15.3 17	3.42 13	6.78 12	16.9 15	3.83 16	6.41 18	25.3 20	4.25 17	4.66 24	6.01 26	4.00 15	6.33 18	20.7 20	5.30 11	3.09 15	4.84 18	2.91 17	7.08 20	15.0 23	5.27 17
GraphCuts [16]	16.9	6.25 18	14.3 18	5.53 18	8.60 20	20.1 22	6.61 18	7.91 17	15.4 10	10.9 18	4.88 15	19.0 6	3.05 16	3.78 11	4.71 15	3.94 14	8.74 24	16.4 11	5.39 13	4.04 23	4.87 19	4.85 25	6.35 18	12.2 17	6.05 19
Filter Flow [22]	17.2	6.48 19	14.6 19	4.96 17	5.73 17	15.7 18	5.07 17	10.1 19	18.6 19	14.3 22	9.04 22	23.3 18	7.80 22	3.98 19	4.71 15	4.21 18	5.86 11	15.0 6	5.41 15	4.98 27	6.87 28	2.78 14	4.82 15	8.66 6	3.65 10
SPSA-learn [15]	18.1	6.84 21	16.7 21	6.74 21	8.47 19	19.4 20	7.49 20	12.5 20	23.1 21	13.1 21	8.40 21	25.8 21	7.08 21	3.87 17	4.66 13	4.10 17	6.32 16	18.8 18	6.89 23	2.56 10	3.85 7	1.79 5	7.29 22	12.5 19	7.47 21
Black & Anandan 3 [6]	18.5	6.81 20	15.4 20	7.43 22	8.77 21	19.5 21	7.35 19	13.0 22	22.9 20	12.5 19	8.29 20	26.1 22	6.77 20	4.18 20	5.28 21	3.69 12	6.19 15	20.0 19	5.34 12	3.63 20	5.05 22	1.79 5	6.45 19	12.2 17	5.17 16
2D-CLG [3]	19.2	10.1 26	22.6 27	7.59 23	9.84 24	16.9 19	11.1 25	16.9 25	28.2 26	18.8 27	14.1 26	31.1 25	13.1 26	3.86 16	4.62 12	4.53 22	5.98 13	21.2 21	5.97 17	1.76 2	**3.14** 1	1.46 4	6.29 17	12.9 20	5.81 18
GroupFlow [11]	19.4	8.00 23	18.6 22	8.09 24	11.1 25	23.7 27	10.3 24	12.6 21	25.6 23	12.8 20	5.84 17	20.3 10	4.39 18	4.69 25	5.81 24	3.67 11	9.29 26	22.4 25	10.1 26	2.11 3	3.99 10	2.29 12	5.75 16	10.0 14	7.39 20
Black & Anandan 2 [2]	20.9	7.83 22	18.7 23	6.41 20	9.70 23	21.9 23	8.60 23	13.7 23	23.7 22	18.1 25	10.9 23	30.0 23	9.44 23	4.60 22	5.55 22	5.06 25	7.85 22	17.6 14	6.38 19	2.61 11	4.44 14	2.15 11	8.58 23	14.3 22	8.54 23
Horn & Schunck [5]	23.1	8.01 24	19.9 25	8.38 25	9.13 22	23.2 26	7.71 21	14.2 24	25.9 24	14.6 23	12.4 24	30.6 24	11.3 24	4.64 23	5.64 23	4.60 23	8.21 23	24.4 26	8.45 24	4.01 21	5.41 24	1.95 9	9.16 24	17.5 24	8.86 24
Black & Anandan [1]	24.6	9.32 25	19.4 24	10.0 26	13.5 26	22.5 24	14.3 26	17.2 26	27.4 25	18.9 28	14.0 25	32.0 26	12.9 25	5.89 27	6.74 28	8.03 27	8.99 25	17.9 15	8.77 25	3.10 16	4.88 20	3.96 23	13.2 26	18.9 26	15.2 26
STOB [24]	25.8	11.6 28	26.0 29	14.6 28	15.3 27	25.0 28	17.5 27	17.8 28	30.1 28	18.1 25	25.4 29	33.6 28	28.0 29	5.25 26	5.90 25	7.03 26	10.3 27	27.4 28	10.6 27	2.89 13	4.47 15	2.94 18	14.9 27	20.7 27	18.8 27
FOLKI [18]	27.5	10.5 27	25.6 28	11.9 27	20.9 28	26.2 29	26.1 28	17.6 27	31.1 29	16.5 24	15.4 27	32.6 27	16.0 27	6.16 28	6.53 27	9.07 28	12.2 28	29.7 29	13.0 28	4.67 25	5.83 27	9.41 28	18.2 28	22.8 28	25.1 28
Pyramid LK [4]	28.5	13.9 29	20.9 26	21.4 29	24.1 29	23.1 25	30.2 29	20.9 29	29.5 27	21.9 29	22.2 28	34.6 29	25.0 28	18.7 29	23.1 29	20.2 29	21.2 29	24.5 27	21.0 29	6.41 29	7.02 29	10.8 29	25.6 29	31.5 29	34.5 29

This tutorial | Variational approaches | ICCV 2009 papers

How to reach first rank on Middlebury (Zimmer et al. EMMCVPR 2009)

- Build upon previous Brox et al. 2004 model
- Use HSV color model with separate robust functions:

$$E_D = \int_\Omega \Psi\left((H_x u + H_y v + H_t)^2\right) dxdy$$
$$+ \int_\Omega \Psi\left((S_x u + S_y v + S_t)^2\right) dxdy$$
$$+ \int_\Omega \Psi\left((V_x u + V_y v + V_t)^2\right) dxdy$$

- Normalize the linearized constraints (Simoncelli et al. 1991, Lai-Vemuri 1998, Zimmer et al. 2009)
- Use an anisotropic regularizer (Nagel-Enkelmann 1986, Sun et al. ECCV 2008, Zimmer et al. 2009, Werlberger et al. BMVC 2009)

Normalization

- Standard data term comprises implicit weighting by image gradient
- Large gradients might correspond to noise
- Multiply with normalization factors:

$$E_D = \sum_{i=1}^{3} \Psi\left(\theta_0^i (I_x^i du + I_y^i dv + I_z^i)^2\right)$$
$$+ \sum_{i=1}^{3} \Psi\left(\theta_x^i (I_{xx}^i du + I_{xy}^i dv + I_{xz}^i)^2 + \theta_y^i (I_{xy}^i du + I_{yy}^i dv + I_{yz}^i)^2\right)$$

$$\theta_0 = \frac{1}{|\nabla I|^2 + 0.1^2} \qquad \theta_{\{x/y\}} = \frac{1}{|\nabla I_{\{x/y\}}|^2 + 0.1^2}$$

Anisotropic smoothness terms

- Basic idea: reduced/no smoothing across edges, strong smoothing along edges (Nagel-Enkelmann 1986)

- Image-driven: edge information from image cues

$$S_\rho = \underbrace{K_\rho * \left[\nabla I \nabla I^\top\right]}_{\text{structure tensor}} = \underbrace{\lambda_1 \mathbf{s}_1 \mathbf{s}_1^\top + \lambda_2 \mathbf{s}_2 \mathbf{s}_2^\top}_{\text{eigen decomposition}}$$

- Penalize motion discontinuities only along image edges:

$$E_S(u,v) = \int_\Omega u_{\mathbf{s}_2}^2 + v_{\mathbf{s}_2}^2 \, dxdy$$

- Drawback: no smoothing across internal edges

Jointly image- and flow driven regularizer (Sun et al. 2008)

- Smoothing **direction** defined by structure tensor (as before)

$$S_\rho = K_\rho * \left[\nabla I \nabla I^\top\right] = \lambda_1 \mathbf{s}_1 \mathbf{s}_1^\top + \lambda_2 \mathbf{s}_2 \mathbf{s}_2^\top$$

- **Amount** of smoothing defined by flow

$$E_S(u,v) = \int_\Omega \Psi(u_{\mathbf{s}_1}^2) + \Psi(u_{\mathbf{s}_2}^2) + \Psi(v_{\mathbf{s}_1}^2) + \Psi(v_{\mathbf{s}_2}^2) \, dxdy$$

- Rotationally invariant version (Zimmer et al. 2009)

$$E_S(u,v) = \int_\Omega \Psi(u_{\mathbf{s}_1}^2 + v_{\mathbf{s}_1}^2) + \Psi(u_{\mathbf{s}_2}^2 + v_{\mathbf{s}_2}^2) \, dxdy$$

Complementary regularization (Zimmer et al. 2009)

- Directional information from motion tensor (used in data term)

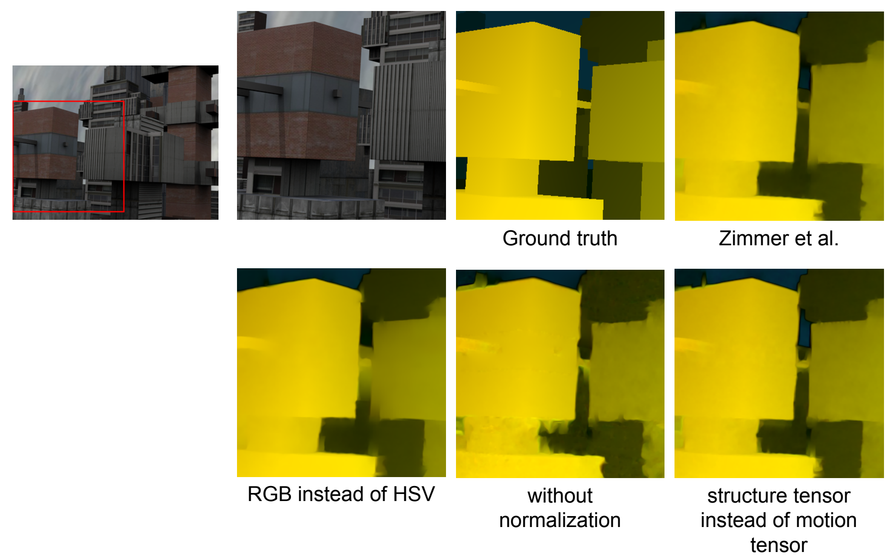
$$R_\rho = K_\rho * \underbrace{\sum_{i=1}^{3} \left[\theta_0^i (\nabla I^i)(\nabla I^i)^\top + \gamma \left(\theta_x^i (\nabla I_x^i)(\nabla I_x^i)^\top + \theta_y^i (\nabla I_y^i)(\nabla I_y^i)^\top\right)\right]}_{\text{Motion tensor}}$$

$$= \sum_{j=1}^{2} \lambda_j \mathbf{r}_j \mathbf{r}_j^\top$$

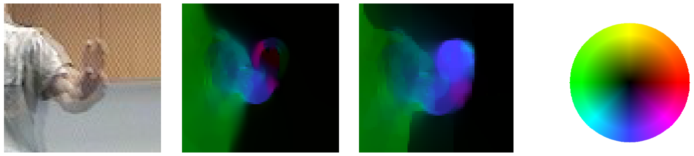
- Amount of smoothing across edge from flow, full smoothing along edge:

$$E_S(u,v) = \int_\Omega \Psi(u_{\mathbf{r}_1}^2 + v_{\mathbf{r}_1}^2) + u_{\mathbf{r}_2}^2 + v_{\mathbf{r}_2}^2 \, dxdy$$

Urban3 sequence from Middlebury dataset

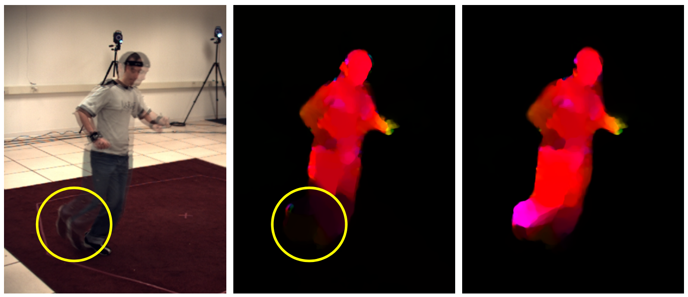


Still problematic: Fast motion of small structures



Warping method

Brox et al. CVPR 2009



Warping method

Brox et al. CVPR 2009

Incorporating correspondences from descriptor matching

$$E(\mathbf{w}(\mathbf{x})) = \int \Psi\left(|I_2(\mathbf{x}+\mathbf{w}(\mathbf{x})) - I_1(\mathbf{x})|^2\right) d\mathbf{x}$$
$$+\gamma \int \Psi\left(|\nabla I_2(\mathbf{x}+\mathbf{w}(\mathbf{x})) - \nabla I_1(\mathbf{x})|^2\right) d\mathbf{x}$$
$$+\alpha \int \Psi\left(|\nabla u(\mathbf{x})|^2 + |\nabla v(\mathbf{x})|^2\right) d\mathbf{x}$$

Bruhn-Weickert 2005

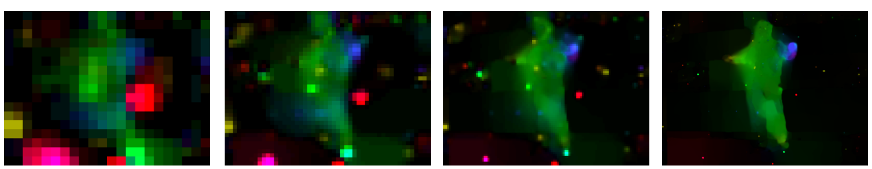
point correspondences by descriptor matching

$$+\beta \sum_{j=1}^{K} \int \rho_j(\mathbf{x}) \Psi\left((u(\mathbf{x})-u_j(\mathbf{x}))^2 + (v(\mathbf{x})-v_j(\mathbf{x}))^2\right) d\mathbf{x}$$

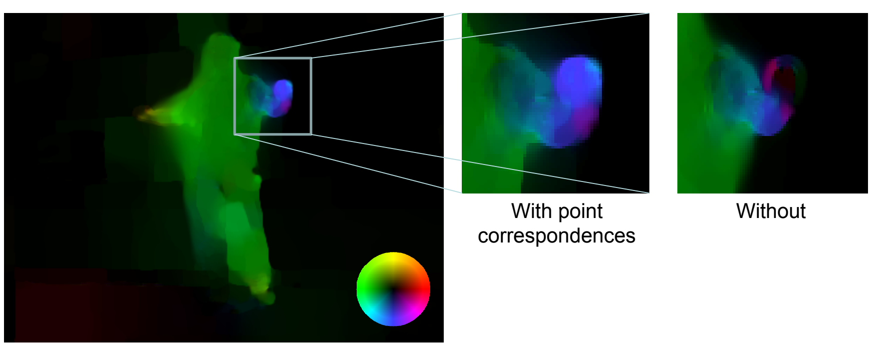
Multiple hypotheses | Matching score | Robust function | Flow = correspondence vector

Brox et al. CVPR 2009

Coarse-to-fine optimization



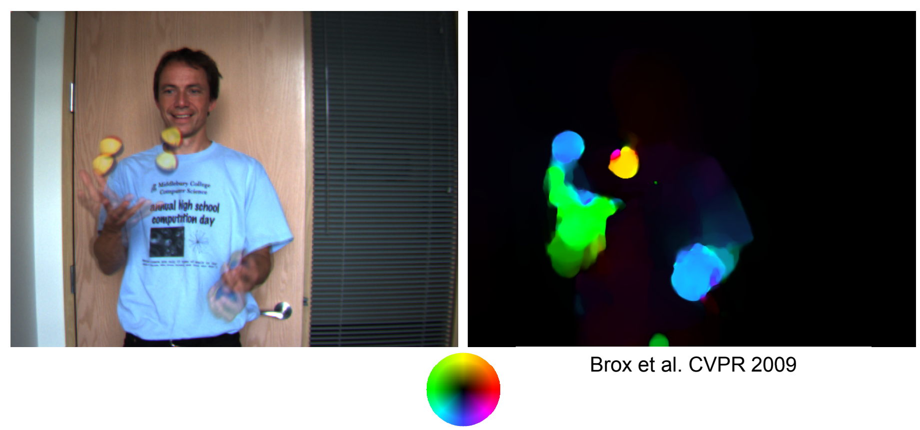
Evolution over scales



With point correspondences

Without

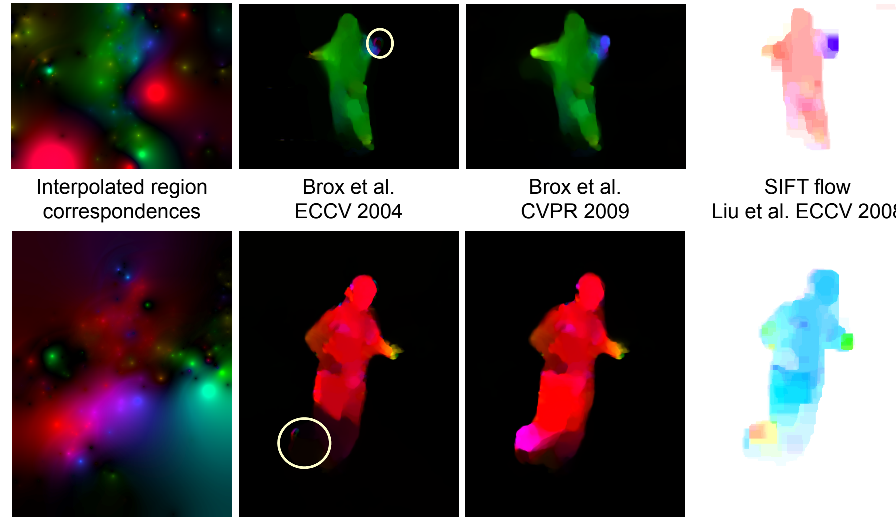
Ball example from Middlebury dataset



Brox et al. CVPR 2009

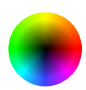
Analyzing challenging sport sequences

Comparison of ways to compute large displacement flow



Shot from a movie

Miss Marple “A pocket full of rye” (slow motion)



Other modeling aspects not considered here

- Spatio-temporal smoothness
  (Nagel ECCV 1990, Weickert-Schnörr JMIV 2001)
  - Process a whole image sequence at once
  - Enforce smoothness in temporal direction
- Over-parametrization (Nir et al. IJCV 2008)
  - Represent flow vector by an affine model
  - Enforce smoothness of affine parameters
  - → Affine transformations not penalized
- Enforcing rigid body motion
  (Valgaerts et al. DAGM 2008, Wedel et al. ICCV 2009)
  - Estimate fundamental matrix from optical flow
  - Soft constraint: correspondences close to epipolar lines