Juergen Gall

Optical Flow

MA-INF 2201 - Computer Vision

WS24/25

Visual motion

Lana Lazebnik

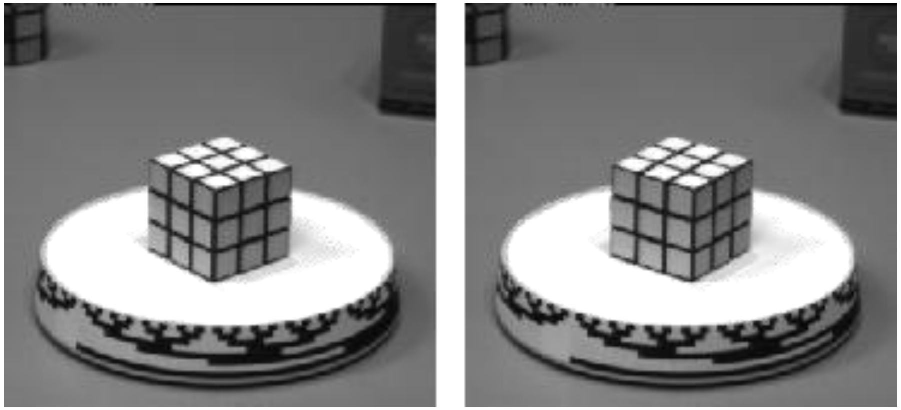
Uses of motion

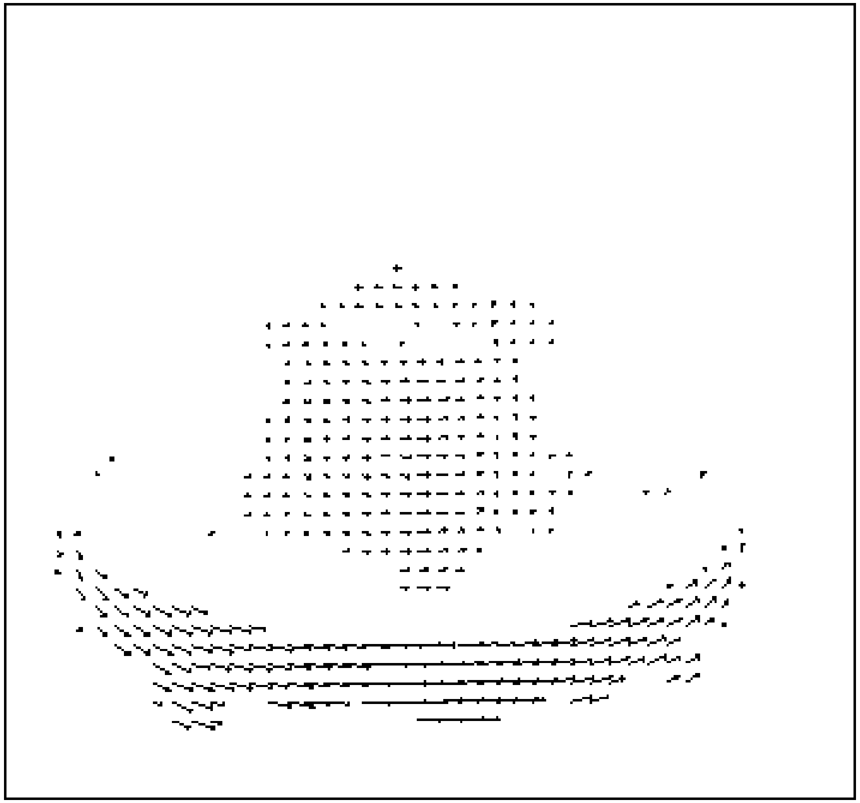
- Estimating 3D structure
- Segmenting objects based on motion cues
- Learning and tracking dynamical models
- Recognizing events and activities

Lana Lazebnik

Motion field

- The motion field is the projection of the 3D scene motion into the image

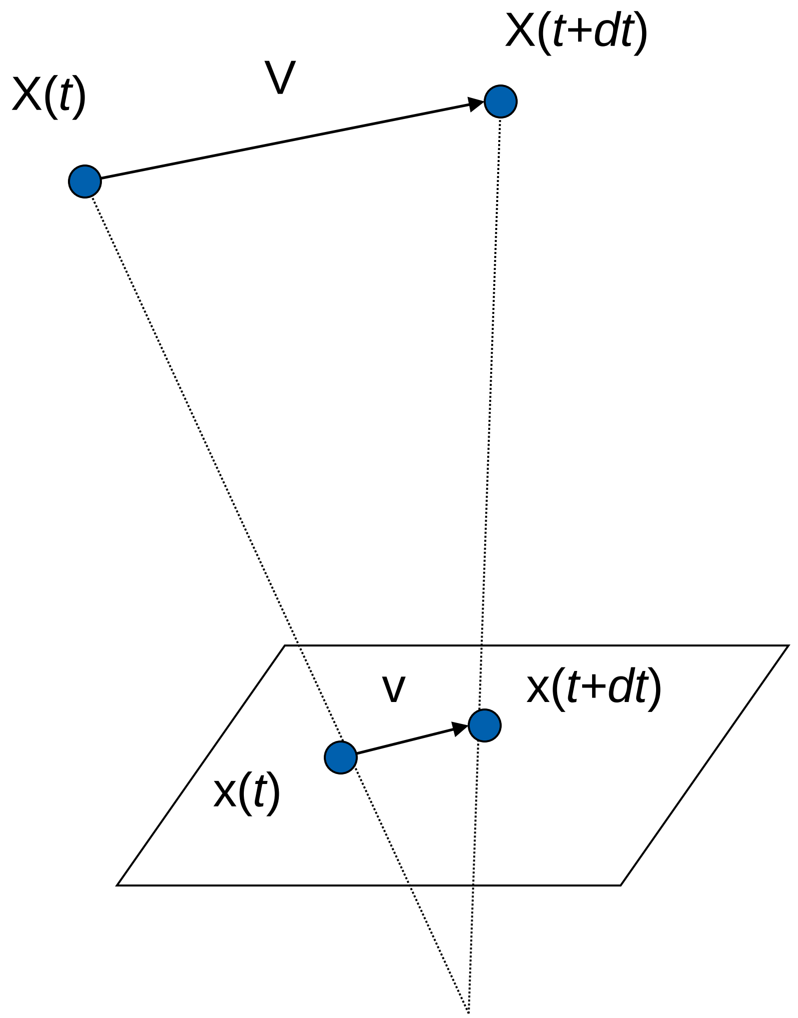




Lana Lazebnik

Motion field and optical flow

- $\mathbf{X}(t)$ is a moving 3D point
- Velocity of scene point: $\mathbf{V} = d\mathbf{X}/dt$
- $\mathbf{x}(t) = (x(t),y(t))$ is the projection of **X** in the image
- Apparent velocity **v** in the image: given by components $v_x = dx/dt$ and $v_y = dy/dt$
- These components are known as the *motion field* of the image



Lana Lazebnik

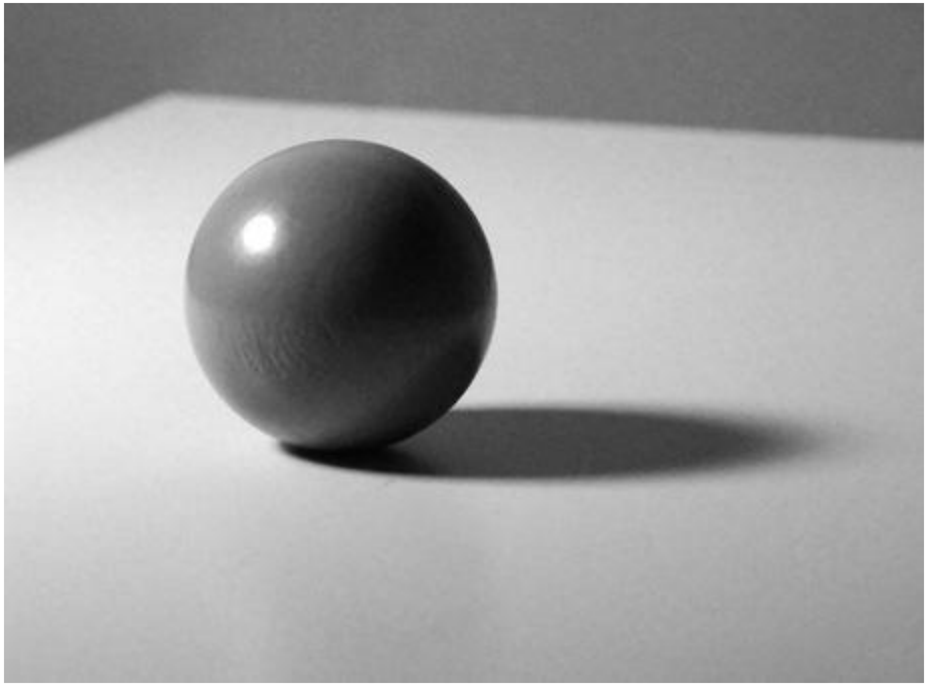
Optical flow

- Definition: optical flow is the *apparent* motion of brightness patterns in the image
- Ideally, optical flow would be the same as the motion field
- Have to be careful: apparent motion can be caused by lighting changes without any actual motion

Lana Lazebnik

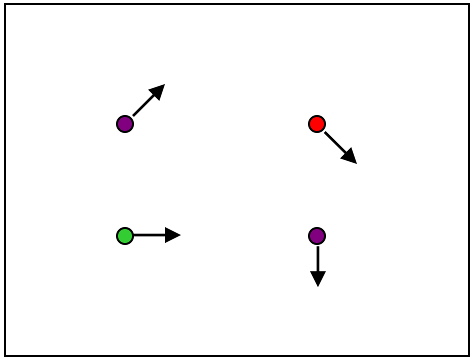
Optical flow

- Have to be careful: apparent motion can be caused by lighting changes without any actual motion

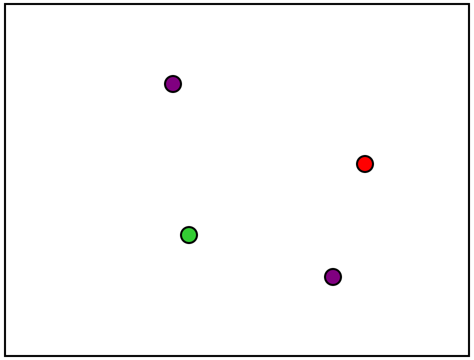


Lana Lazebnik

Estimating optical flow



$I(x,y,t-1)$

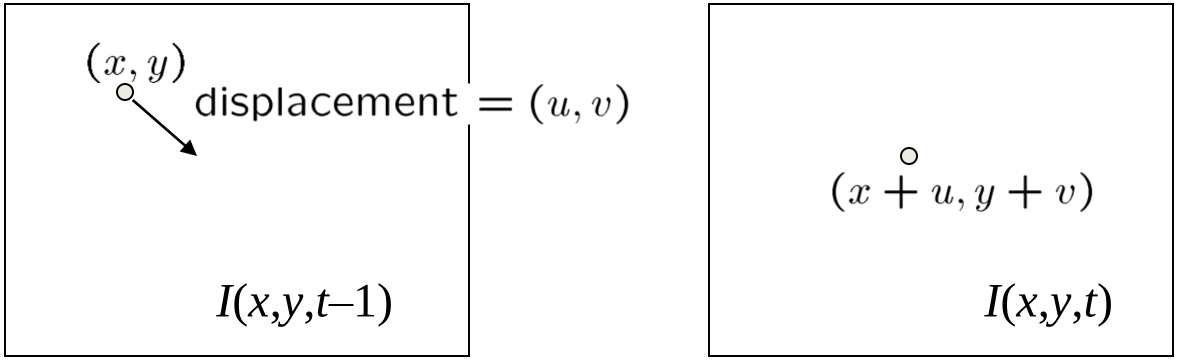


$I(x,y,t)$

- Given two subsequent frames, estimate the apparent motion field $u(x,y)$ and $v(x,y)$ between them

Lana Lazebnik

The brightness constancy constraint



- Brightness Constancy Equation:

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

Linearizing the right side using Taylor expansion:

$$I(x, y, t-1) \approx I(x, y, t) + I_x \cdot u(x, y) + I_y \cdot v(x, y)$$

Hence, $I_x \cdot u + I_y \cdot v + I_t \approx 0$ or $\nabla I \cdot (u, v) + I_t \approx 0$

Lana Lazebnik

The brightness constancy constraint

$$I_x \cdot u + I_y \cdot v + I_t = 0$$

- How many equations and unknowns per pixel?
  – One equation, two unknowns

Lana Lazebnik

The brightness constancy constraint

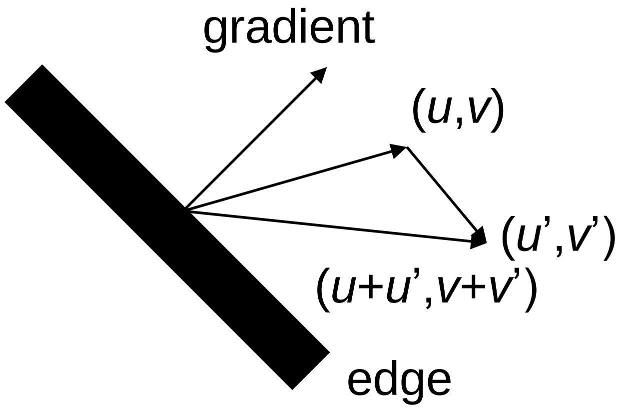
- Intuitively, what does this constraint mean?

$$\nabla I \cdot (u, v) + I_t = 0$$

- The component of the flow perpendicular to the gradient (i.e., parallel to the edge) is unknown

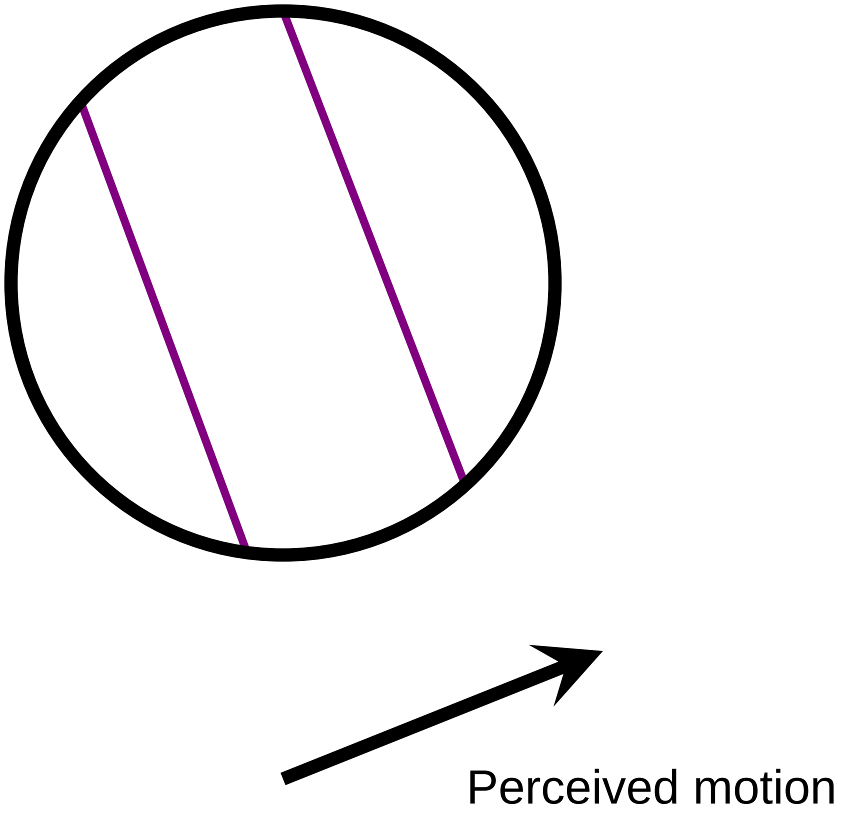
- If (u, v) satisfies the equation, so does (u+u', v+v') if

$$\nabla I \cdot (u', v') = 0$$



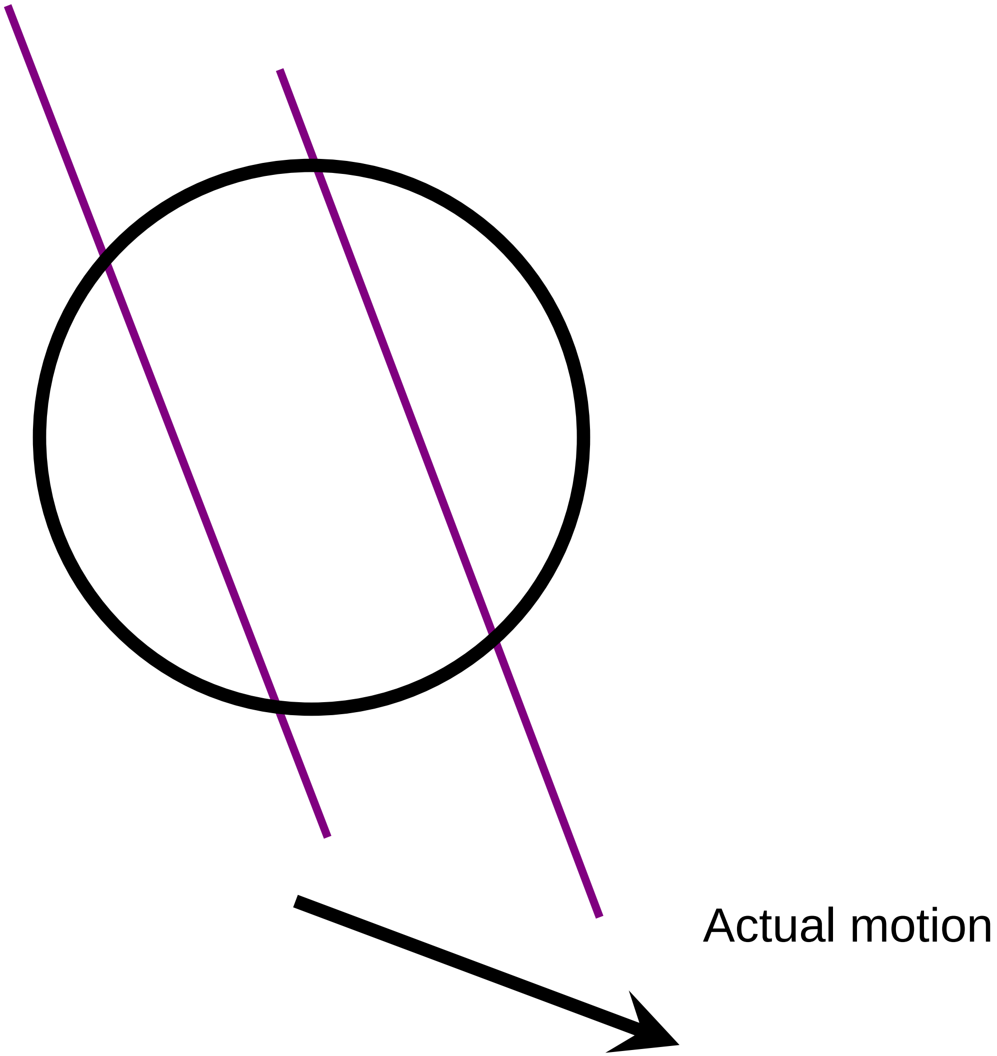
Lana Lazebnik

The aperture problem



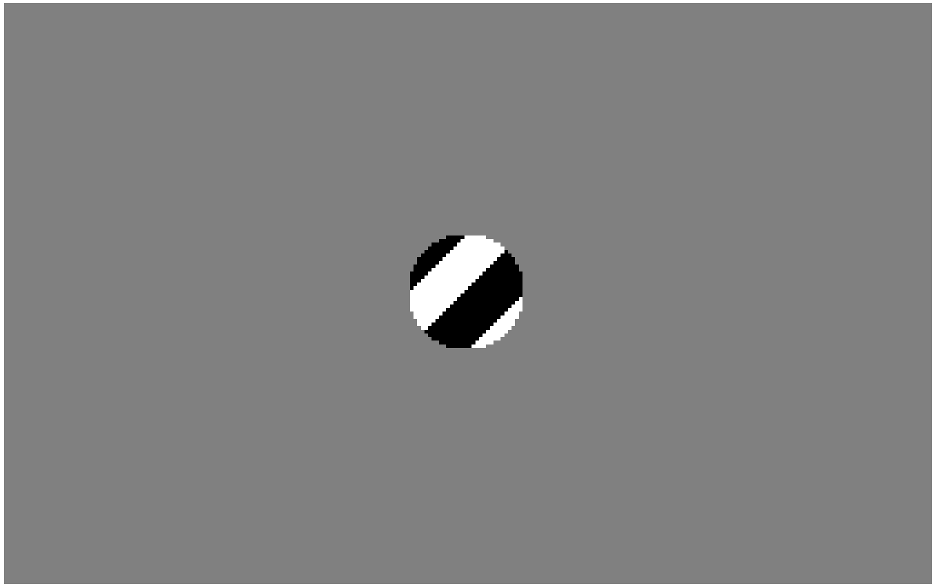
Lana Lazebnik

The aperture problem



Lana Lazebnik

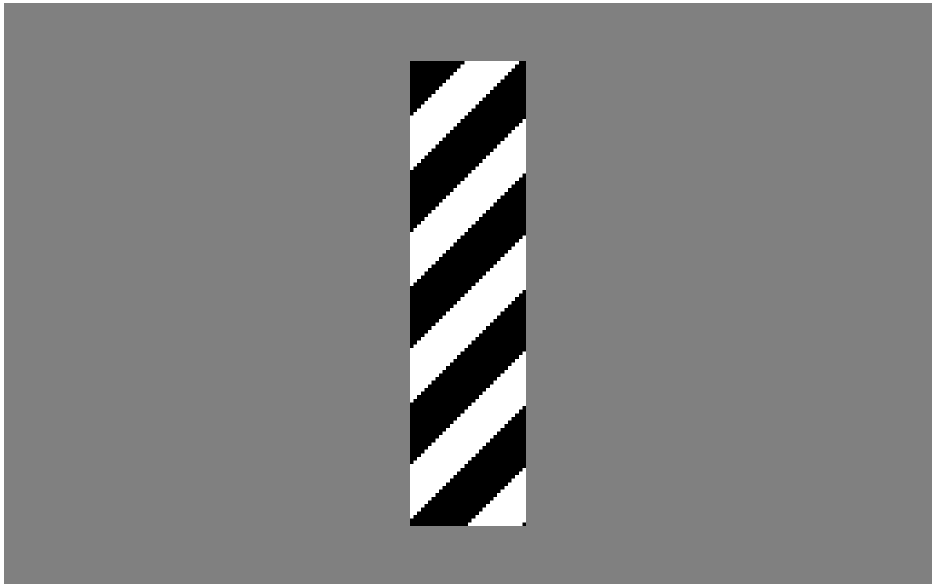
The barber pole illusion



http://en.wikipedia.org/wiki/Barberpole_illusion

Lana Lazebnik

The barber pole illusion



http://en.wikipedia.org/wiki/Barberpole_illusion

Lana Lazebnik

The barber pole illusion

http://en.wikipedia.org/wiki/Barberpole_illusion

Lana Lazebnik

Solving the aperture problem

UNIVERSITÄT BONN

- How to get more equations for a pixel?
- **Spatial coherence constraint:** pretend the pixel's neighbors have the same (u,v)
  - If we use a 5x5 window, that gives us 25 equations per pixel

$$0 = I_t(\mathbf{p_i}) + \nabla I(\mathbf{p_i}) \cdot [u \;\; v]$$

$$\begin{bmatrix} I_x(\mathbf{p}_1) & I_y(\mathbf{p}_1) \\ I_x(\mathbf{p}_2) & I_y(\mathbf{p}_2) \\ \vdots & \vdots \\ I_x(\mathbf{p}_{25}) & I_y(\mathbf{p}_{25}) \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = - \begin{bmatrix} I_t(\mathbf{p}_1) \\ I_t(\mathbf{p}_2) \\ \vdots \\ I_t(\mathbf{p}_{25}) \end{bmatrix}$$

B. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pp. 674–679, 1981.

Lana Lazebnik

Solving the aperture problem

- Least squares problem:

$$\begin{bmatrix} I_x(\mathbf{p}_1) & I_y(\mathbf{p}_1) \\ I_x(\mathbf{p}_2) & I_y(\mathbf{p}_2) \\ \vdots & \vdots \\ I_x(\mathbf{p}_{25}) & I_y(\mathbf{p}_{25}) \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = - \begin{bmatrix} I_t(\mathbf{p}_1) \\ I_t(\mathbf{p}_2) \\ \vdots \\ I_t(\mathbf{p}_{25}) \end{bmatrix} \qquad \underset{25\text{x}2}{A} \; \underset{2\text{x}1}{d} = \underset{25\text{x}1}{b}$$

- Is this system always solvable?
- Only if rank of A is 2
- Not the case for edges

B. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pp. 674–679, 1981.

Lana Lazebnik

Conditions for solvability

Bad case

Good case

Lucas-Kanade flow

UNIVERSITÄT BONN

- Linear least squares problem

$$\begin{bmatrix} I_x(\mathbf{p}_1) & I_y(\mathbf{p}_1) \\ I_x(\mathbf{p}_2) & I_y(\mathbf{p}_2) \\ \vdots & \vdots \\ I_x(\mathbf{p}_{25}) & I_y(\mathbf{p}_{25}) \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = - \begin{bmatrix} I_t(\mathbf{p}_1) \\ I_t(\mathbf{p}_2) \\ \vdots \\ I_t(\mathbf{p}_{25}) \end{bmatrix} \qquad \underset{25\text{x}2}{A} \ \underset{2\text{x}1}{d} = \underset{25\text{x}1}{b}$$

Solution given by $(A^T A)\ d = A^T b$

$$\underbrace{\begin{bmatrix} \sum I_x I_x & \sum I_x I_y \\ \sum I_x I_y & \sum I_y I_y \end{bmatrix}}_{A^T A} \begin{bmatrix} u \\ v \end{bmatrix} = - \underbrace{\begin{bmatrix} \sum I_x I_t \\ \sum I_y I_t \end{bmatrix}}_{A^T b}$$

The summations are over all pixels in the window

B. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pp. 674–679, 1981.

Lana Lazebnik

Lucas-Kanade flow

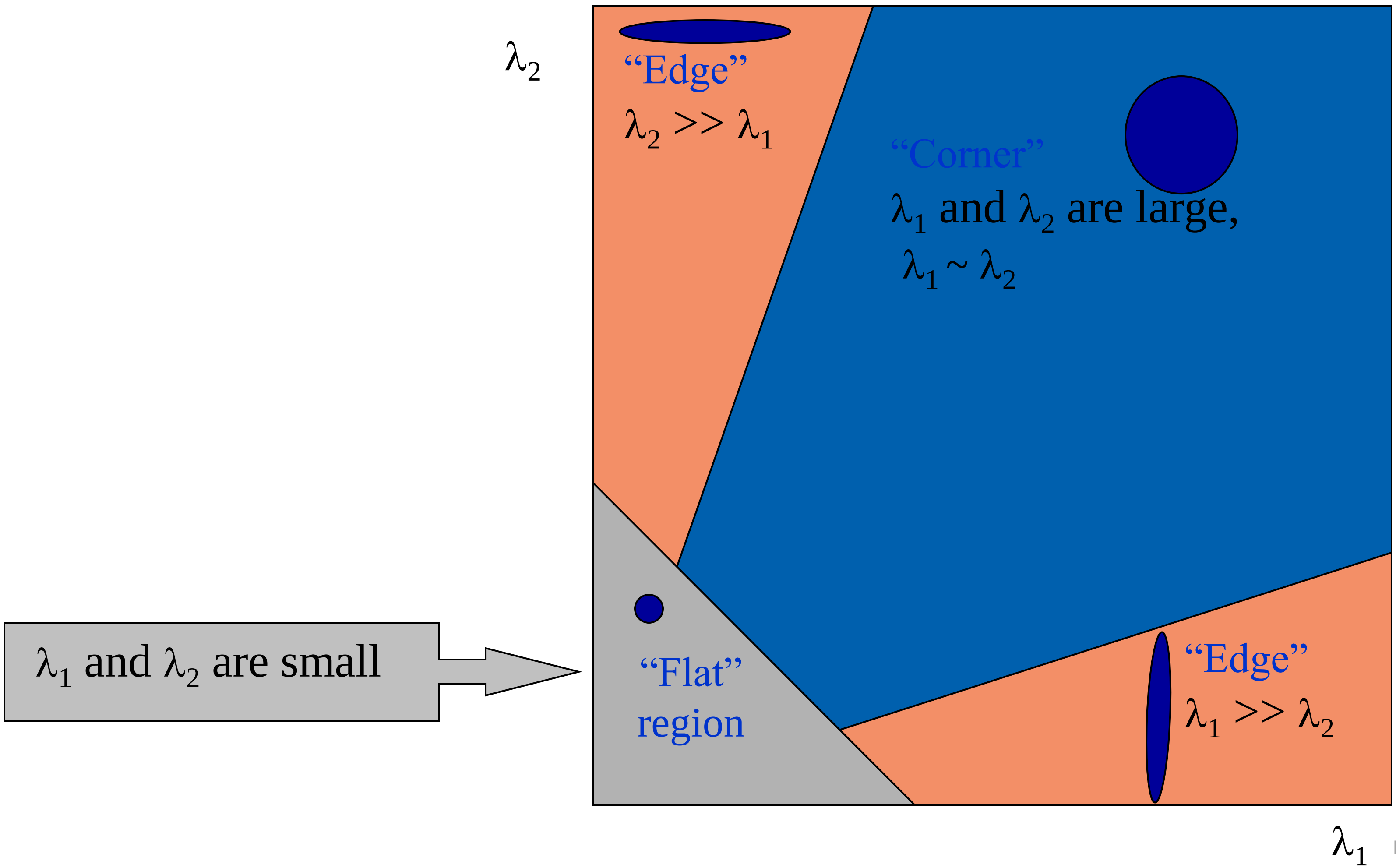
$$\underbrace{\begin{bmatrix} \sum I_x I_x & \sum I_x I_y \\ \sum I_x I_y & \sum I_y I_y \end{bmatrix}}_{A^T A} \begin{bmatrix} u \\ v \end{bmatrix} = - \underbrace{\begin{bmatrix} \sum I_x I_t \\ \sum I_y I_t \end{bmatrix}}_{A^T b}$$

- Recall the Harris corner detector: $M = A^T A$ is the *second moment matrix*

Lana Lazebnik

Interpreting the eigenvalues

Classification of image points using eigenvalues of the second moment matrix:



Lana Lazebnik

Uniform region

– gradients have small magnitude
– small λ_1, small λ_2
– system is ill-conditioned

Lana Lazebnik

Edge

– gradients have one dominant direction

– large λ_1, small λ_2

– system is ill-conditioned

Lana Lazebnik

High-texture or corner region

– gradients have different directions, large magnitudes
– large λ_1, large λ_2
– system is well-conditioned

Lana Lazebnik

Errors in Lucas-Kanade

- The motion is large (larger than a pixel)
  - Iterative refinement
  - Coarse-to-fine estimation
  - Exhaustive neighborhood search (feature matching)
- A point does not move like its neighbors
  - Motion segmentation
- Brightness constancy does not hold
  - Exhaustive neighborhood search with normalized correlation

Lana Lazebnik

Feature tracking

- So far, we have only considered optical flow estimation in a pair of images
- If we have more than two images, we can compute the optical flow from each frame to the next
- Given a point in the first image, we can in principle reconstruct its path by simply “following the arrows”

Lana Lazebnik

Shi-Tomasi feature tracker

- Find good features using eigenvalues of second-moment matrix
  - Key idea: “good” features to track are the ones whose motion can be estimated reliably
- From frame to frame, track with Lucas-Kanade
  - This amounts to assuming a translation model for frame-to-frame feature movement
- Check consistency of tracks by *affine* registration to the first observed instance of the feature
  - Affine model is more accurate for larger displacements
  - Comparing to the first frame helps to minimize drift

J. Shi and C. Tomasi. Good Features to Track. CVPR 1994.

Lana Lazebnik

Tracking example

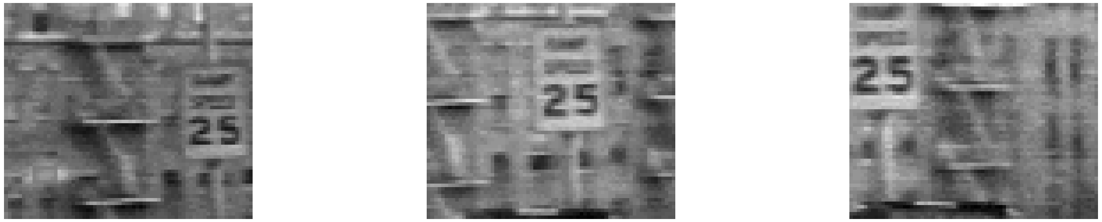


Figure 1: Three frame details from Woody Allen's *Manhattan*. The details are from the 1st, 11th, and 21st frames of a subsequence from the movie.

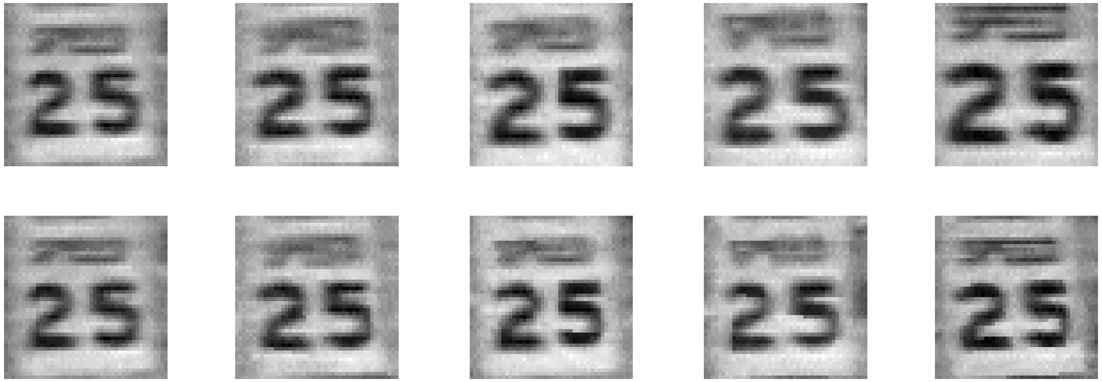


Figure 2: The traffic sign windows from frames 1,6,11,16,21 as tracked (top), and warped by the computed deformation matrices (bottom).

J. Shi and C. Tomasi. Good Features to Track. CVPR 1994.

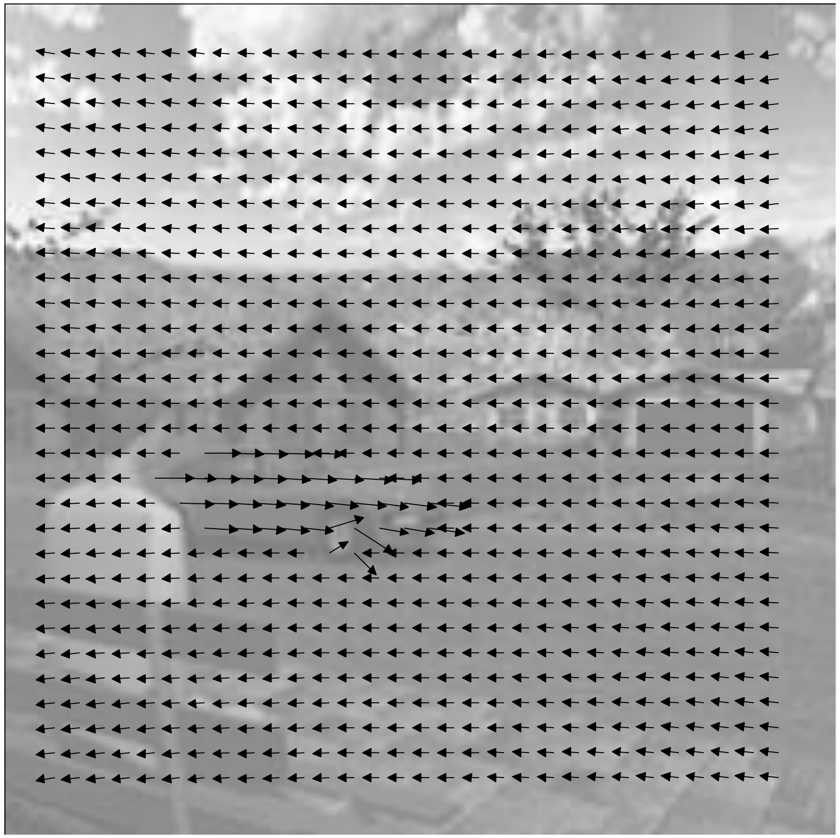
Lana Lazebnik

Dense motion estimation

UNIVERSITÄT BONN

- For each pixel in an image, optical flow determines the corresponding pixel in the next image
- Can be cast as a variational problem

Image sequence $I(x, y, t)$



Optical flow field $(u, v)(x, y, t)$.

Thomas Brox

Horn-Schunck model

- Most basic model → good to start with

- Assumption I:
  Gray value constancy (optic flow constraint)

  $I_x u + I_y v + I_t = 0$ reminder: subscripts denote partial derivatives

  Linearized version of

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

- Assumption II:
  Smoothness of optic flow field
  $|\nabla u|^2 + |\nabla v|^2 \rightarrow \min$

- Combined in an energy functional:

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

Thomas Brox

Analysis of Horn-Schunck model

UNIVERSITÄT BONN

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

- First term penalizes deviations from optic flow constraints
- Second term penalizes deviations from smoothness
- Parameter α steers relative importance of assumptions

Thomas Brox

Theoretical analysis

- Both terms of the energy are convex in (u, v)

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

- Linear combination of two convex functionals is again convex

- There exists one unique minimum (Schnörr 1994)

- Further properties (Schnörr 1994):
  – Solution depends continuously on the input data
  – Smoothness: $u, v \in H^2(\Omega)$ (twice differentiable)

Thomas Brox

Euler-Lagrange equations

UNIVERSITÄT BONN

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

- Necessary conditions for a minimum: Euler-Lagrange equations

$$\frac{d}{d\epsilon} E(u(x) + \epsilon h(x), v(x))\Big|_{\epsilon=0} = 0 \qquad \forall h(x)$$

$$\frac{d}{d\epsilon} E(u(x), v(x) + \epsilon h(x))\Big|_{\epsilon=0} = 0 \qquad \forall h(x)$$

- This leads to:

$$(I_x u + I_y v + I_t) I_x - \alpha \Delta u = 0$$

$$(I_x u + I_y v + I_t) I_y - \alpha \Delta v = 0$$

- Discretization needed to solve for (u, v)

Thomas Brox

Spatial discretization

UNIVERSITÄT BONN

- Euler-Lagrange equation (for u):

$$(I_x u + I_y v + I_t) I_x - \alpha \Delta u = 0$$

- Approximate Laplacian by

0	1	0
1	-4	1
0	1	0

$$\Delta u = u_{xx} + u_{yy}$$

$$\approx \frac{u_{i+1,j} + u_{i-1,j} + u_{i,j+1} + u_{i,j-1} - 4u_{\boxed{i,j}}}{h^2}$$

Pixel indices

- Grid size usually set to $h = 1$

- Relevant in multi-grid schemes

Thomas Brox

Gradient descent

UNIVERSITÄT BONN

- Gradient descent $$-\frac{dE}{du} = \Delta u - \frac{1}{\alpha}(I_x u + I_y v + I_t) I_x$$
- Most simple iterative scheme (explicit scheme)

$$u_{i,j}^{0} = 0$$

Iteration index

$$u_{i,j}^{k+1} = u_{i,j}^{\boxed{k}} + \tau \left(\Delta u_{i,j}^{k} - \frac{1}{\alpha}(I_x u_{i,j}^{k} + I_y v_{i,j}^{k} + I_t) I_x \right)$$

- Right hand side contains only known values
  → easy to evaluate
- Stable for step sizes $\tau \leq \frac{1}{4}$
- Drawbacks:
  – Terribly slow
  – Updates eventually become too small (number precision)

Thomas Brox

Linear system of equations

UNIVERSITÄT BONN

- Euler-Lagrange equations

$$(I_x u + I_y v + I_t) I_x - \alpha \Delta u = 0$$

$$(I_x u + I_y v + I_t) I_y - \alpha \Delta v = 0$$

- Discretized version

$$\frac{u_{i+1,j} + u_{i-1,j} + u_{i,j+1} + u_{i,j-1} - 4u_{i,j}}{h^2} - \frac{1}{\alpha}(I_x u_{i,j} + I_y v_{i,j} + I_t) I_x = 0$$

$$\frac{v_{i+1,j} + v_{i-1,j} + v_{i,j+1} + v_{i,j-1} - 4v_{i,j}}{h^2} - \frac{1}{\alpha}(I_x u_{i,j} + I_y v_{i,j} + I_t) I_y = 0$$

- Linear in (u, v)

Concatenation of u and v

- Can be written in matrix-vector notation $A^h w = b$

Thomas Brox

Linear system of equations

- Sparse linear system

$$A^h = \left(\begin{array}{ccccc|ccccc} d &&&&& d &&&& \\ & d &&&&& d &&& \\ && d &&&&& d && \\ &&& d &&&&& d & \\ &&&& d &&&&& d \\ \hline d &&&&& d &&&& \\ & d &&&&& d &&& \\ && d &&&&& d && \\ &&& d &&&&& d & \\ &&&& d &&&&& d \end{array}\right) + \left(\begin{array}{ccccc|ccccc} s & s & & s & & & & & & \\ s & s & s & & s & & & & & \\ & s & s & s & & & & & & \\ s & & s & s & s & & & & & \\ & s & & s & s & & & & & \\ \hline & & & & & s & s & & s & \\ & & & & & s & s & s & & s \\ & & & & & & s & s & s & \\ & & & & & s & & s & s & s \\ & & & & & & s & & s & s \end{array}\right)$$

- Blocks due to coupling of u and v
- Data term contributes only to block main diagonals
- Smoothness term has four additional block off-diagonals (coupling of neighboring pixels)

Thomas Brox

Solving the linear system

UNIVERSITÄT BONN

- Gauß-elimination
  Not recommendable for sparse matrices (very slow)
- Iterative methods: Jacobi method (quite slow)
- Faster: Gauß-Seidel method
- Even faster: Successive over-relaxation (SOR)
- Other approaches:
  - Multi-grid methods
  - Conjugate gradient (CG) with a preconditioner (good for GPU implementation)
  - Primal-dual methods (require another numerical approach)

Thomas Brox

SOR for Horn-Schunck method

UNIVERSITÄT BONN

- Split the matrix into the diagonal part D, the lower triangle L and the upper triangle U

$$Aw = b \quad \Leftrightarrow \quad Dw = b - Lw - Uw$$

- Update scheme:

$$u_i^{k+1} = (1-\omega)\, u_i^k + \omega \frac{\sum_{j \in \mathcal{N}^-(i)} u_j^{k+1} + \sum_{j \in \mathcal{N}^+(i)} u_j^k - \frac{1}{\alpha}((I_x I_y)_i v_i^k + (I_x I_t)_i)}{\sum_{j \in \mathcal{N}^-(i) \cup \mathcal{N}^+(i)} 1 + \frac{1}{\alpha}(I_x^2)_i}$$

- Relaxation parameter $\omega \in (0, 2)$

- Gauß-Seidel method for $\omega = 1$

- Optimal ω (fastest convergence) depends on the problem
  Horn-Schunck method: $\omega = 1.98$

[Thomas Brox. From Pixels to Regions: Partial Differential Equations in Image Analysis. 2005]

Thomas Brox

Computation speed with various solvers

UNIVERSITÄT BONN

- Results from Bruhn et al. 2005, CPU

	Iterations	Frames/sec	Speedup
Gauß-Seidel	21931	0.026	1
SOR	286	1.381	54
Cascadic GS	237	1.667	65
Cascadic SOR	25	8.125	316
Full multi-grid	1	18.229	708

- Speedup of multi-grid techniques much smaller for extended models

Thomas Brox

Average angular error

- Established measure of how accurate an optic flow method is

- Computes the angle between the correct $(u_c, v_c)^\top$ and estimated $(u, v)^\top$ flow vector

$$\text{AAE} := \frac{1}{n}\sum_{i=1}^{n} \arccos\left(\frac{(u_c)_i u_i + (v_c)_i v_i + 1}{\sqrt{((u_c)_i^2 + (v_c)_i^2 + 1)(u_i^2 + v_i^2 + 1)}}\right)$$

- Alternative: average endpoint error:

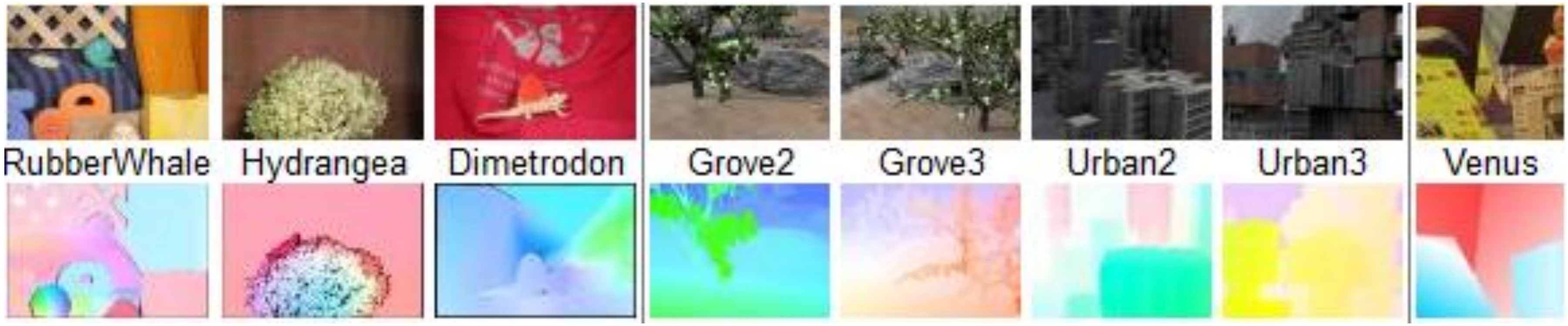
$$\text{AEE} := \frac{1}{n}\sum_{i=1}^{n} \left(((u_c)_i - u_i)^2 + ((v_c)_i - v_i)^2\right)$$

Thomas Brox

Middlebury flow benchmark

Popular optical flow benchmark
http://vision.middlebury.edu/flow/
Poor correlation with real-world problems

8 test sequences with ground truth

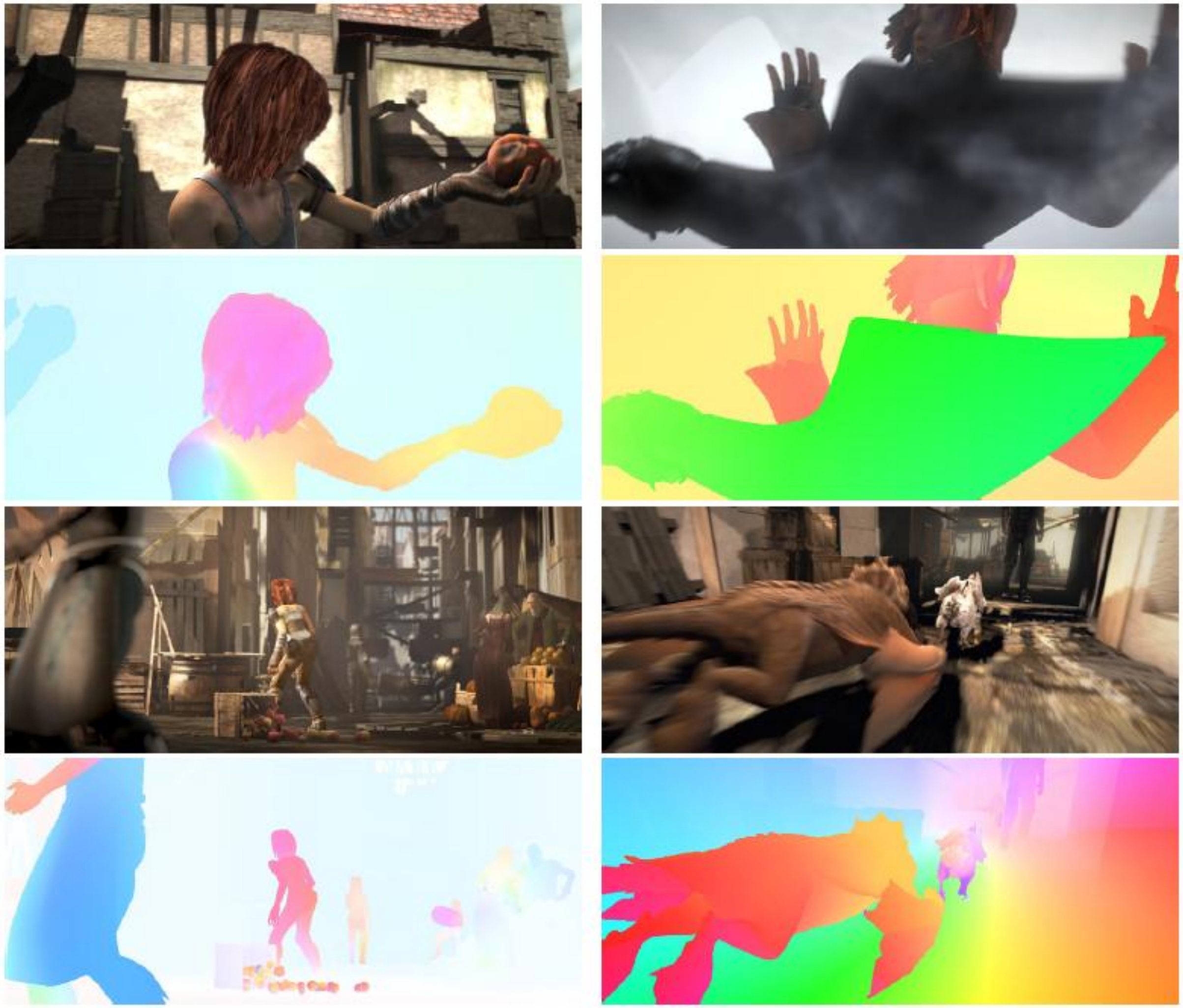


8 training sequences with ground truth

Thomas Brox

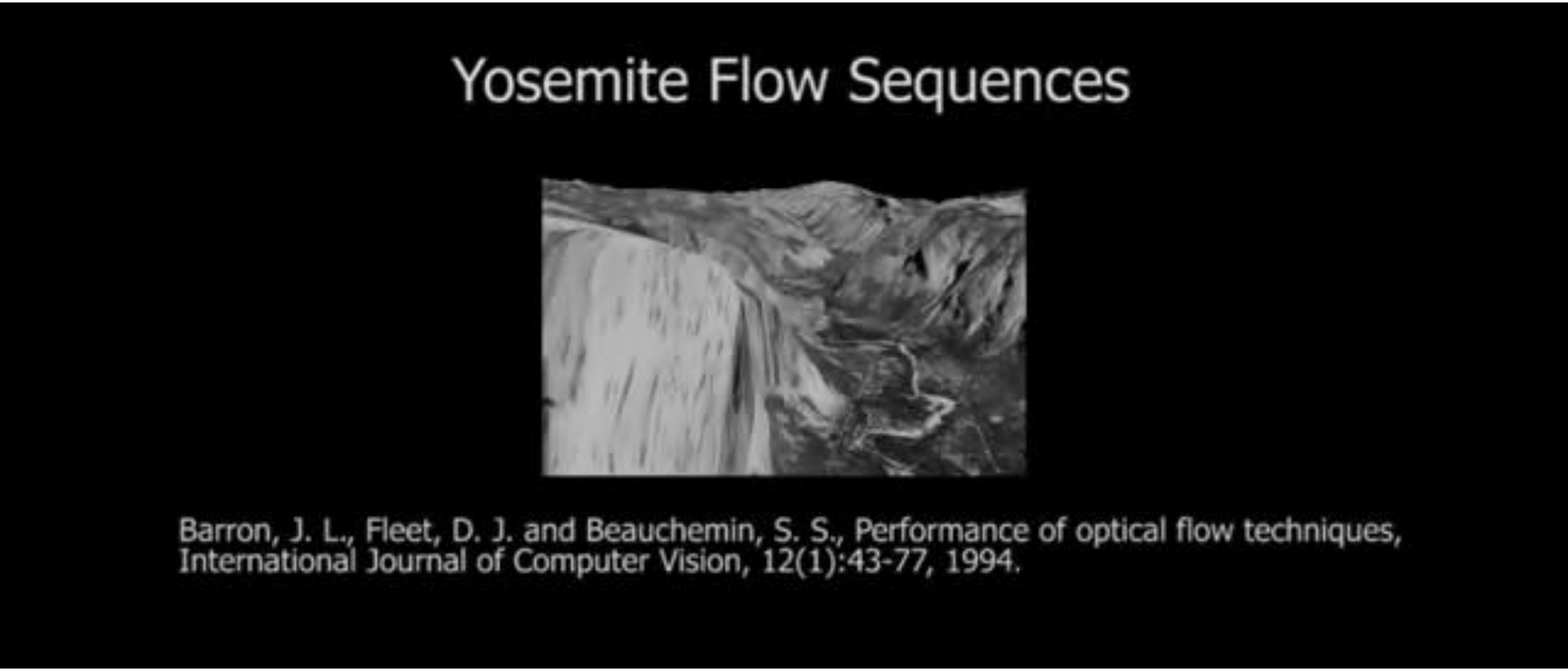
MPI-Sintel Optical Flow Dataset

Ground-truth from open-source movie Sintel (Butler et al. ECCV 2012)

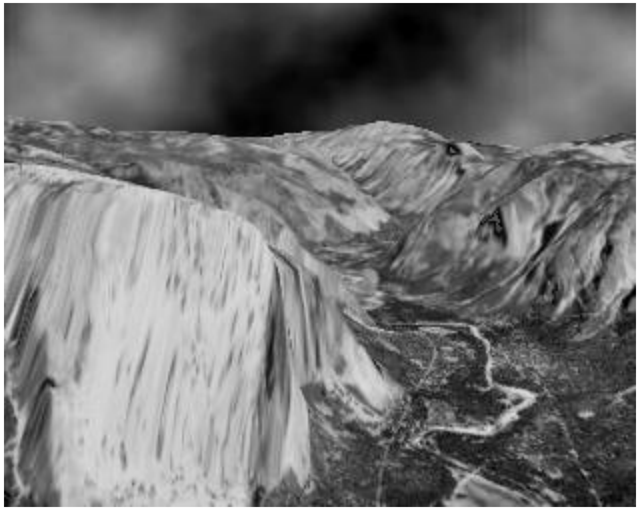


Thomas Brox

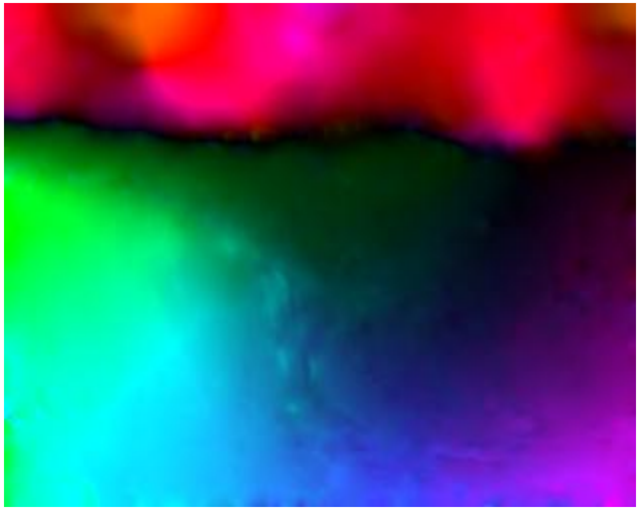
Optical Flow Dataset



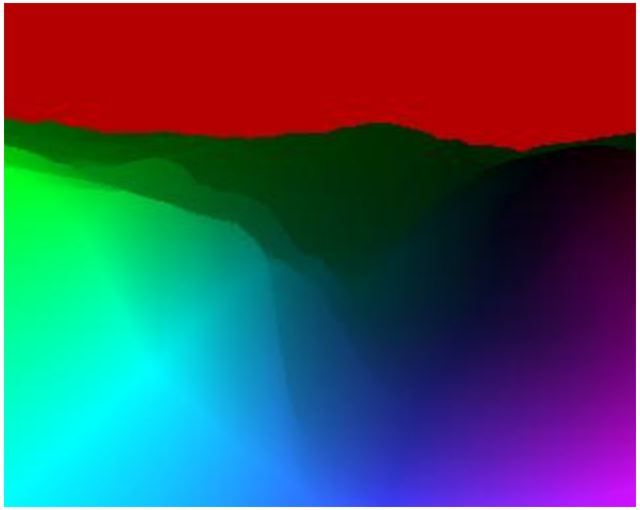
Results with Horn-Schunck method



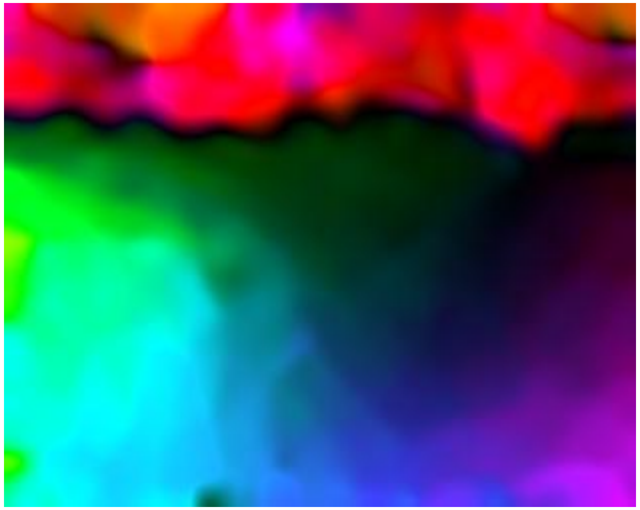
Input sequence



Horn-Schunck (7.17°)



Ground truth



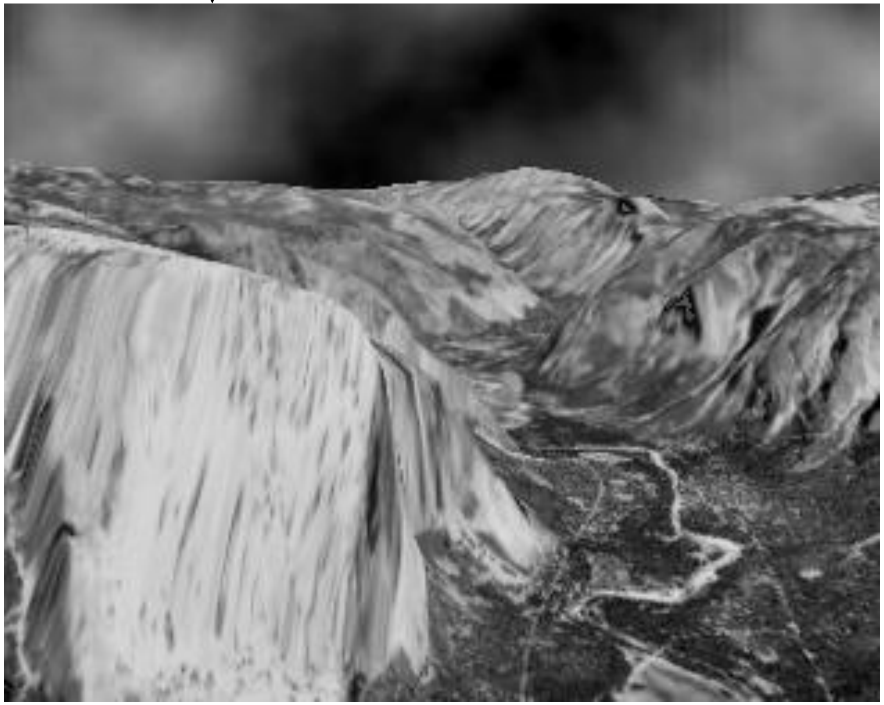
Lucas-Kanade (8.78°)

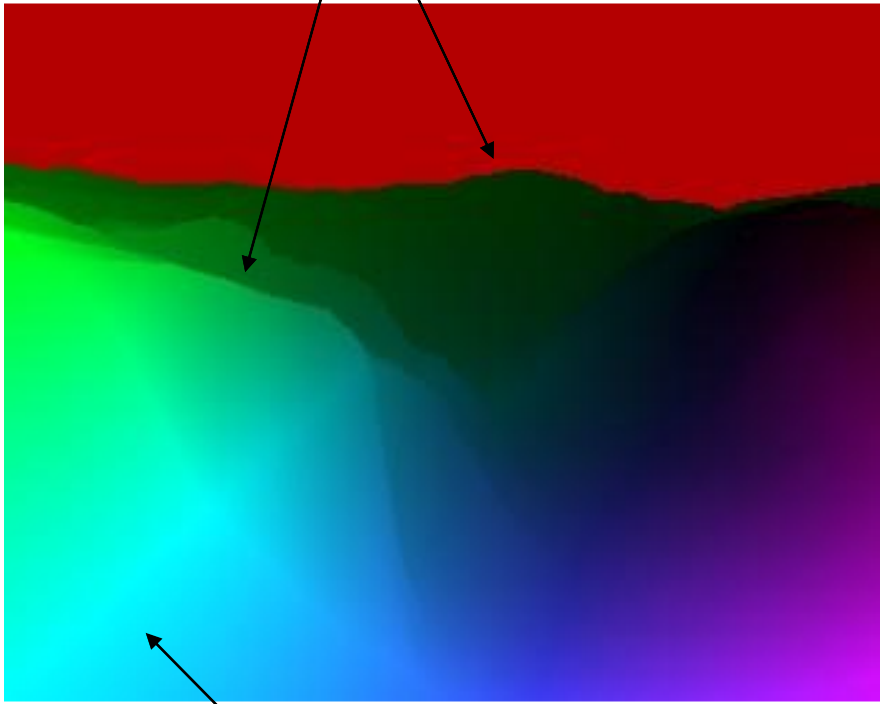
Thomas Brox

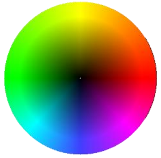
Yosemite test sequence

Illumination changes

Motion discontinuities





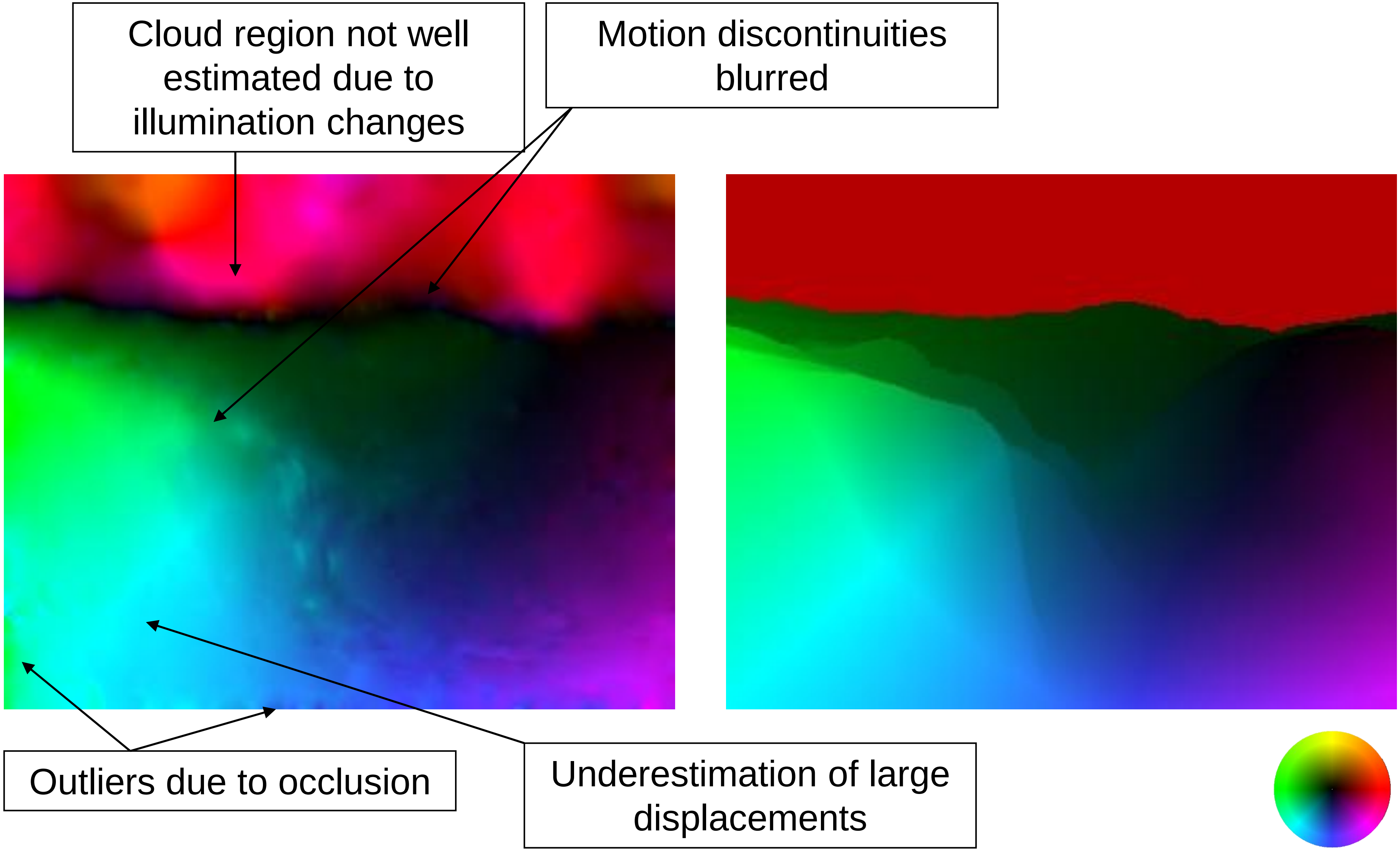


Occlusion

“Large” displacements

Thomas Brox

Limitations of Horn-Schunck method



Thomas Brox

Motion discontinuities

Occlusions

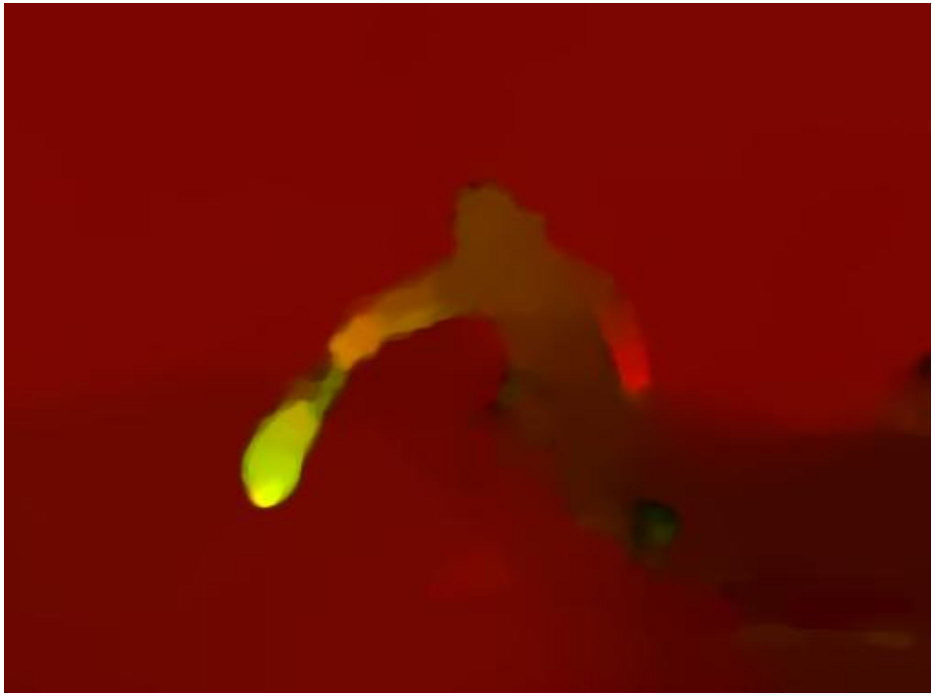
Illumination changes

Large displacements

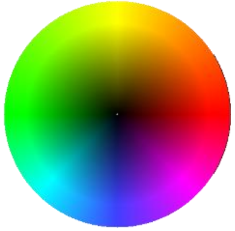
Really large displacements

Thomas Brox

Motion discontinuities



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



Thomas Brox

Motion discontinuities and the smoothness assumption

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

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

- Too much influence given to outliers → blurring

Thomas Brox

Non-quadratic smoothness term

Applying a robust function:

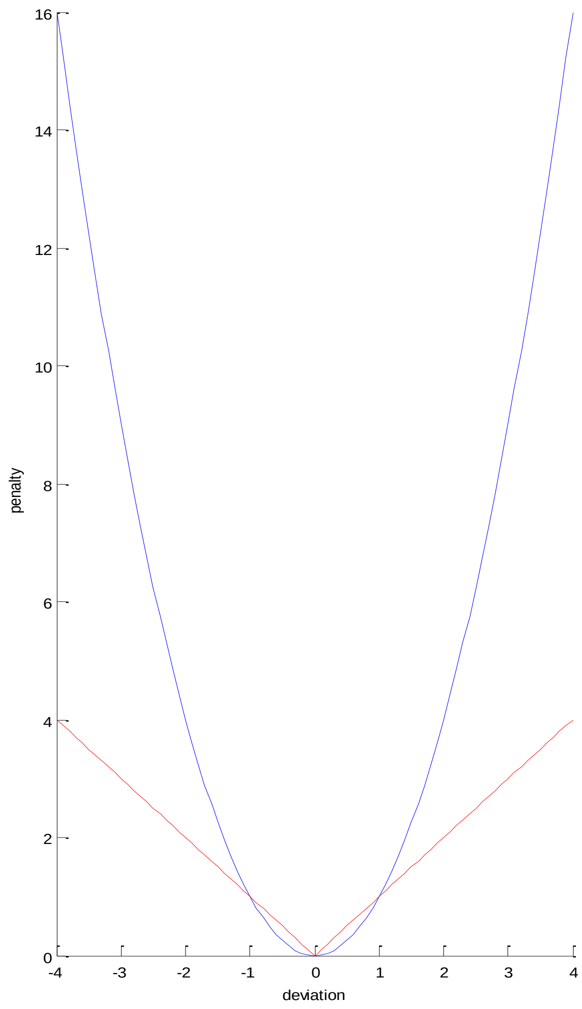
$$E_S(u,v) = \int_\Omega \Psi(|\nabla u|^2 + |\nabla v|^2)\, dxdy$$

For example:

Total variation (TV)

$$\Psi(s^2) = \sqrt{s^2 + \epsilon^2}, \quad \epsilon = 0.001$$

Advantage: still convex
→ global optimum!



Thomas Brox

Fixed-point iteration

Lagrange equations for regularizer

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

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

TV penalizer depends on the unknown flow

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

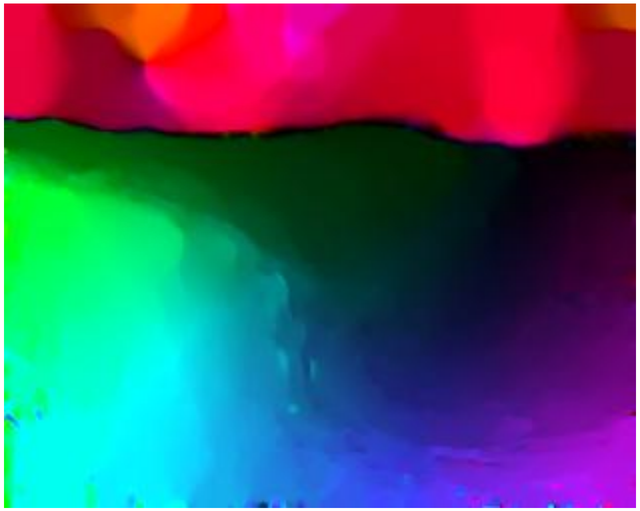
Solution:

- Keep Ψ' fixed for an (initial) solution (u, v) (fixed point)
- Solve resulting linear system to obtain a new fixed point
- Update Ψ', iterate

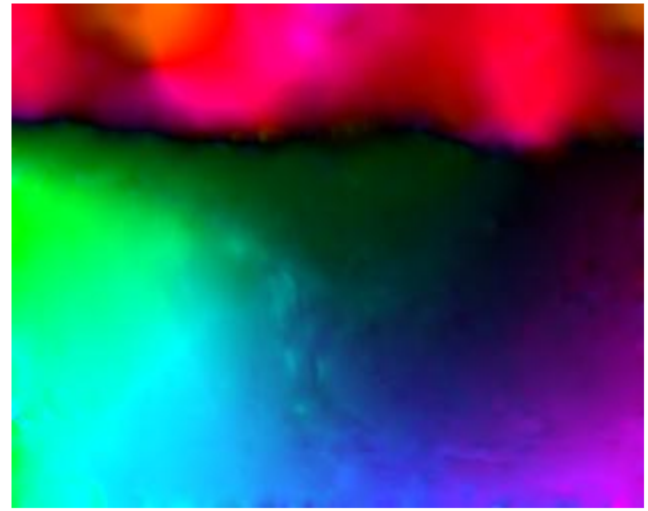
[Brox et al. **High Accuracy Optical Flow Estimation Based on a Theory for Warping**. ECCV 2004]

Thomas Brox

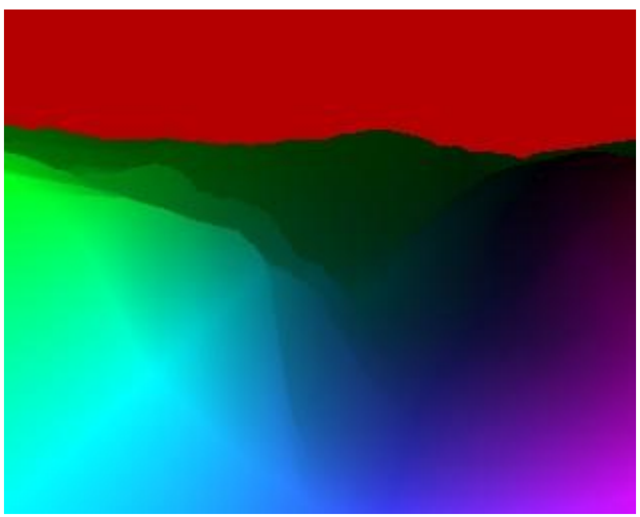
Result: improvements at motion discontinuities



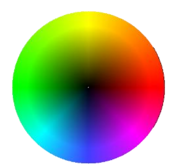
Non-quadratic smoothness term (6.36°)



Horn-Schunck (7.17°)



Ground truth



Thomas Brox

Occlusions

Some pixels of frame 1 no longer visible in frame 2

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

Thomas Brox

Dealing with occlusions and non-Gaussian noise

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

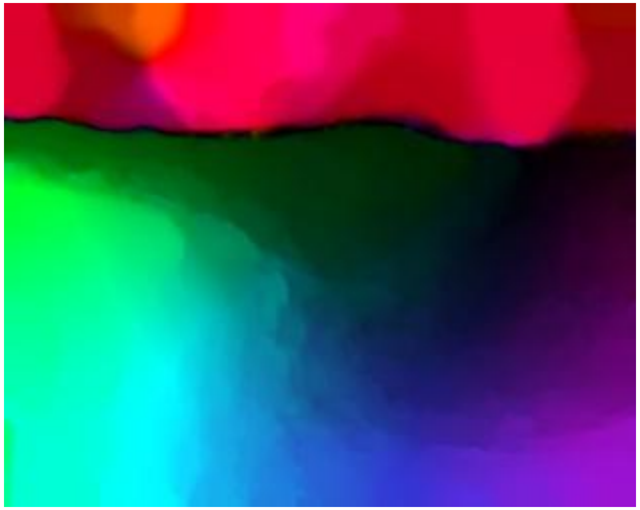
$$E(u,v) = \int_{\Omega} \Psi\left((I_x u + I_y v + I_t)^2\right) + \alpha \Psi\left(|\nabla u|^2 + |\nabla v|^2\right) \, dx dy$$

Interpretation: less importance given to pixels with high matching cost → mainly occlusions

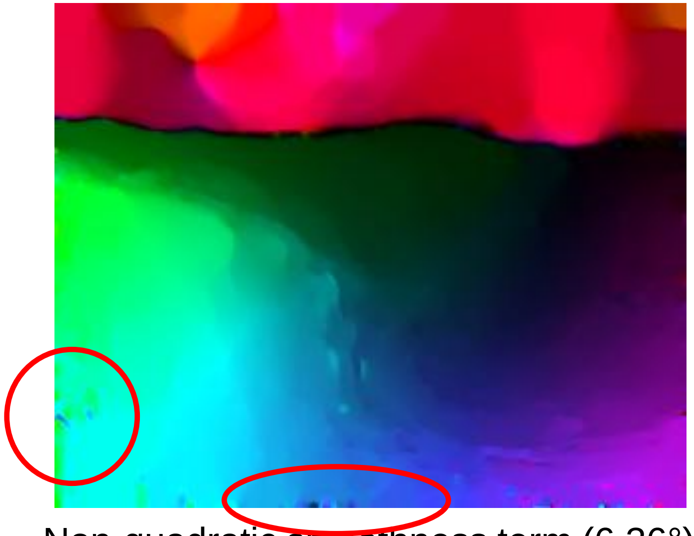
Nonlinear system can again be solved with fixed point iterations

Thomas Brox

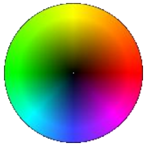
Result: Better treatment of occlusions

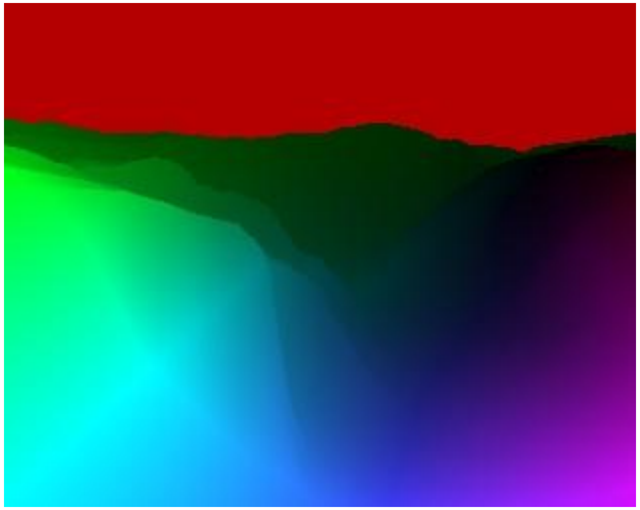


Both terms non-quadratic (5.97°)

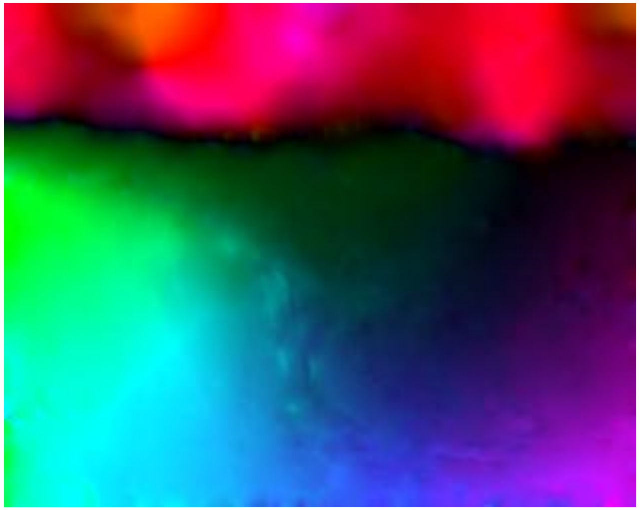


Non-quadratic smoothness term (6.36°)





Ground truth



Horn-Schunck (7.17°)

Thomas Brox

Illumination changes

Typically caused by:

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

“Schefflera” from Middlebury benchmark

Two frames from a Miss Marple movie

Thomas Brox

Matching the gradient

UNIVERSITÄT BONN

- Brightness constancy assumption:

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

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

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

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

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

Thomas Brox

Combination of features

- Linearization:

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

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

- Combine with gray value constancy assumption:

$$E(u,v) = \int_\Omega \Psi\left((I_x u + I_y v + I_t)^2\right) dxdy$$

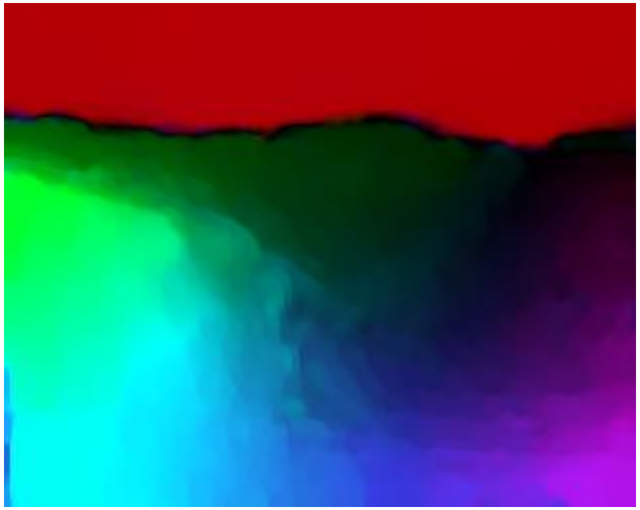
$$+\gamma \int_\Omega \Psi\left((I_{xx}u + I_{xy}v + I_{xt})^2 + (I_{xy}u + I_{yy}v + I_{yt})^2\right) dxdy$$

$$+ \alpha E_{\text{Smooth}}$$

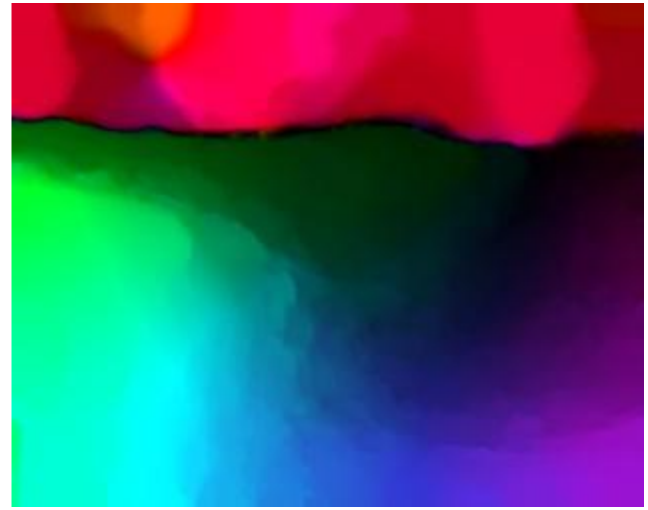
- Separate robust penalizer Ψ for each feature (Bruhn-Weickert 2005)

Thomas Brox

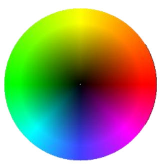
Result: Illumination changes no longer a problem

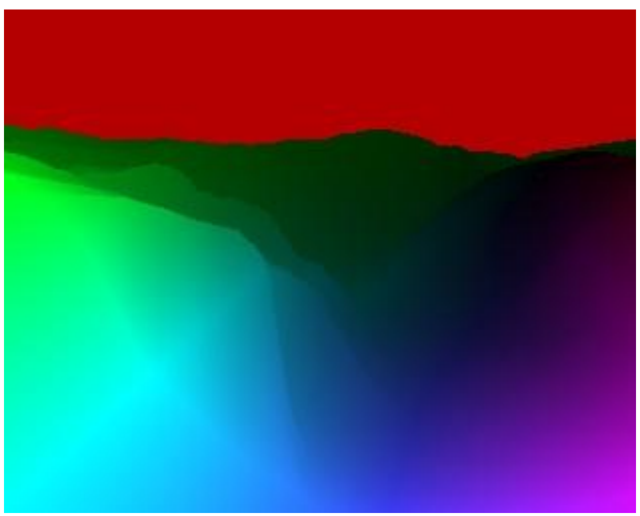


Gradient constancy (3.5°)

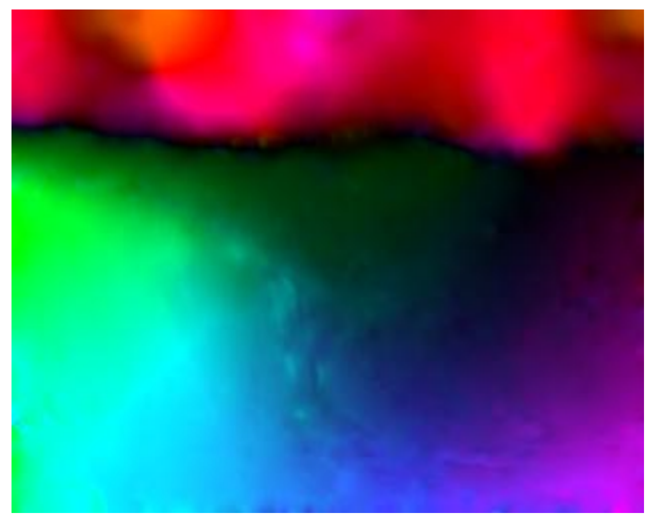


Both terms non-quadratic (5.97°)





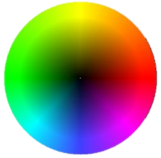
Ground truth

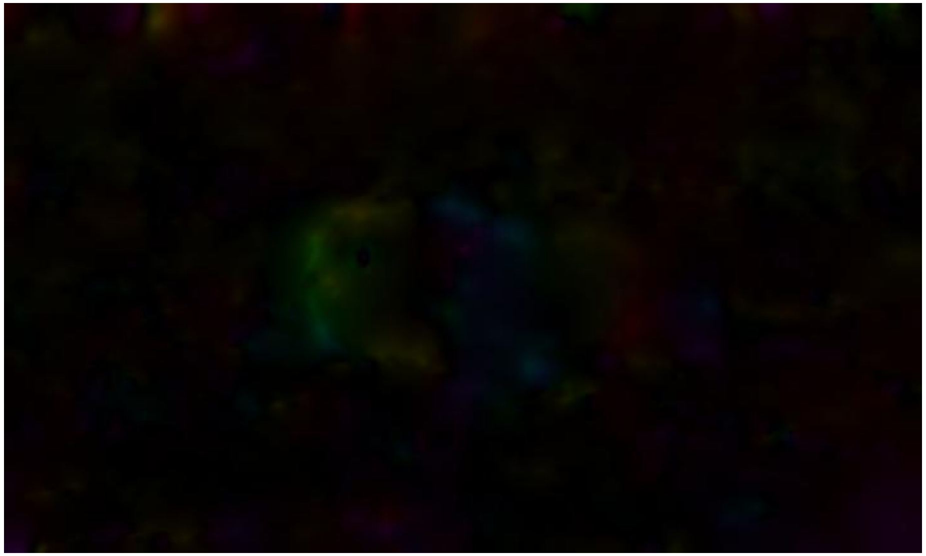


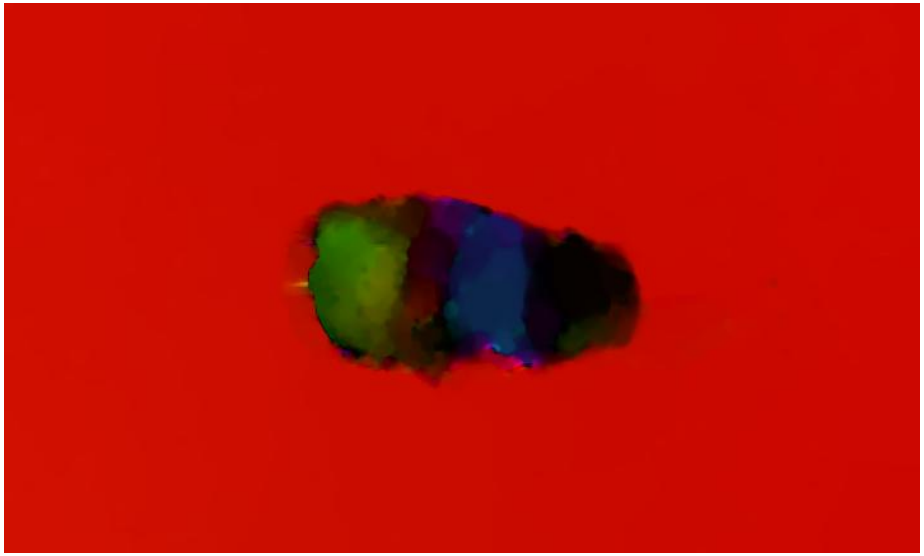
Horn-Schunck (7.17°)

Thomas Brox

Large displacements







Horn&Schunck

Brox et al. 2004

Thomas Brox

Linearization and large displacements

- Constancy assumptions have been linearized

$$I(x+u, y+v, t+1) - I(x, y, t) = 0 \quad \rightarrow \quad I_x u + I_y v + I_t = 0$$

- Assumes that the image is a linear gradient

- Locally (between two pixels), this is true
  → Linearization only valid for very small displacements (subpixel displacements)

- If images are quite smooth, approximation is good enough for larger displacements

- But typical images are not smooth at all

Thomas Brox

Non-linearized constancy assumption

UNIVERSITÄT BONN

- Energy functional:

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

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

- Correct description of the matching criterion (no approximation)

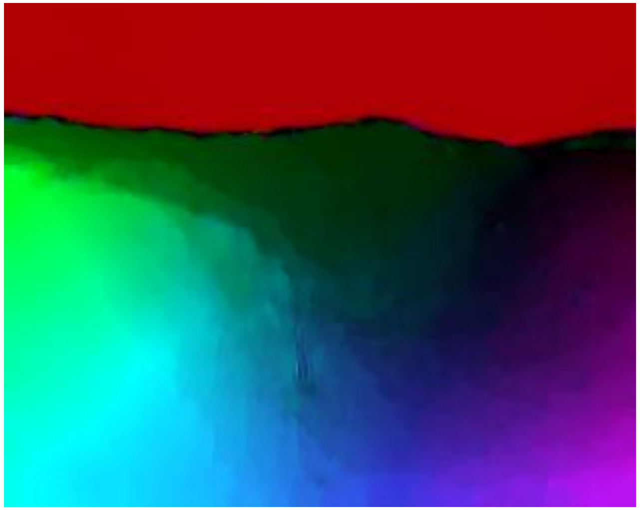
- Problem:
  - Not convex in $\mathbf{w}$ → multiple local minima

- One can work with the non-linearized assumptions and iteratively estimate an increment using the Gauss-Newton method

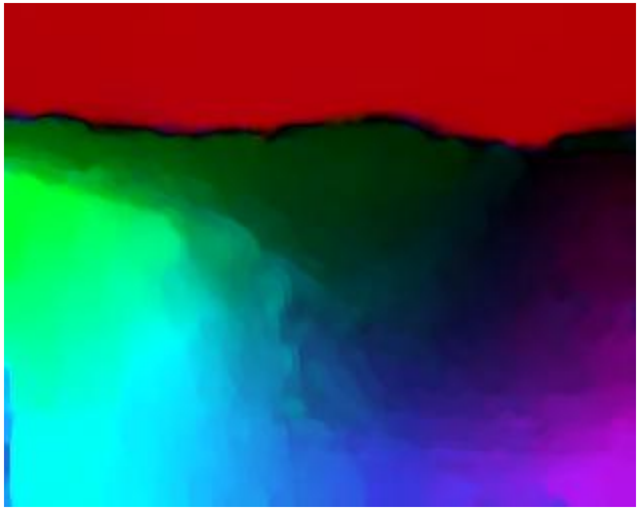
- Computing the increment requires computation of warped images

Thomas Brox

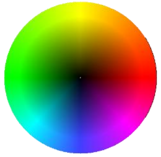
Result: higher accuracy, large displacements possible

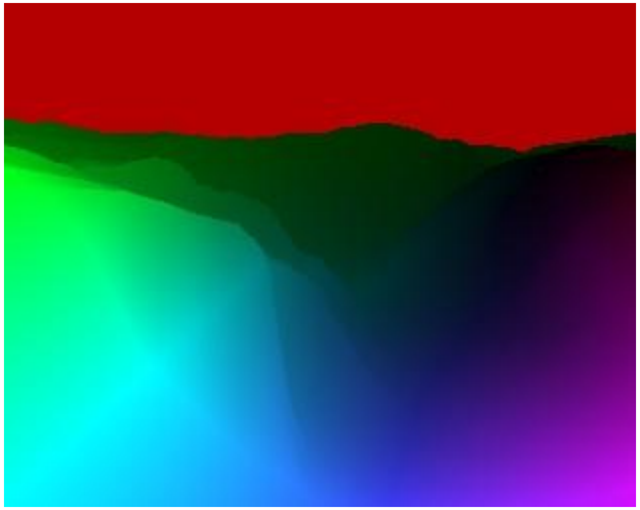


Non-linearized constancy (2.44°)

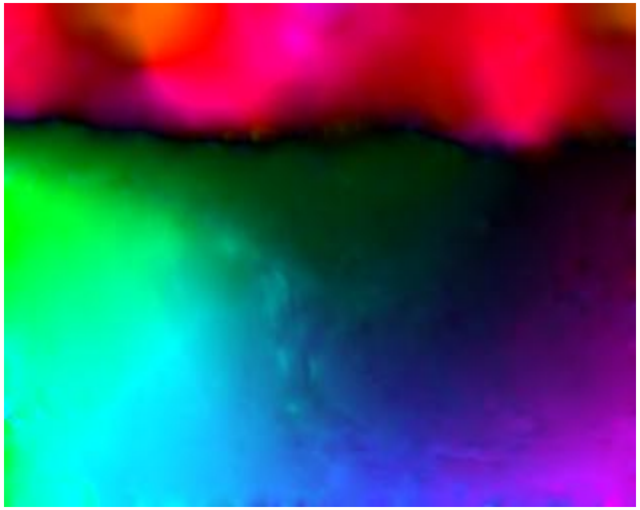


Gradient constancy (3.5°)





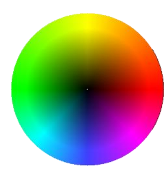
Ground truth



Horn-Schunck (7.17°)

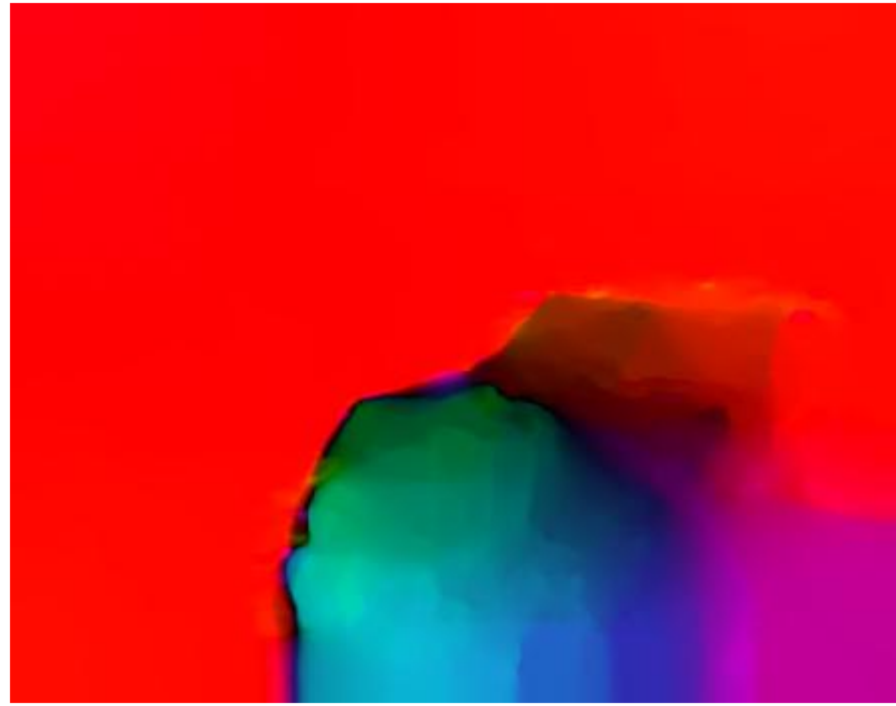
Thomas Brox

Other results: zoom into a scene

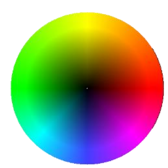


Thomas Brox

Car in a movie



from “Miss Marple: A Pocket Full of Rye”



Thomas Brox

SIFT flow (Liu et al. 2008)

- Label transfer from training to test data

Training image

Training annotation

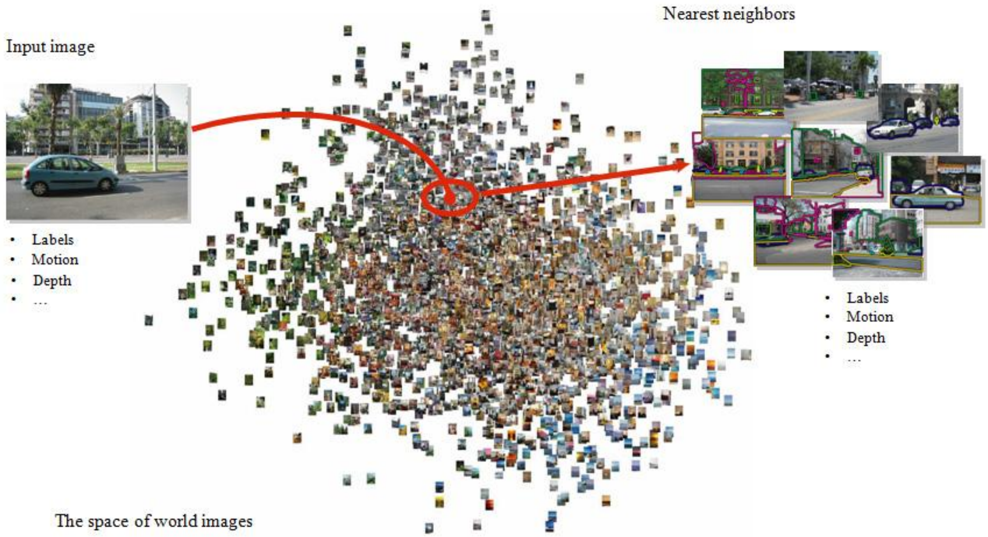
Test image

Warped annotation

- Requires flow from test image to training image

- Only feasible if the two images are sufficiently similar

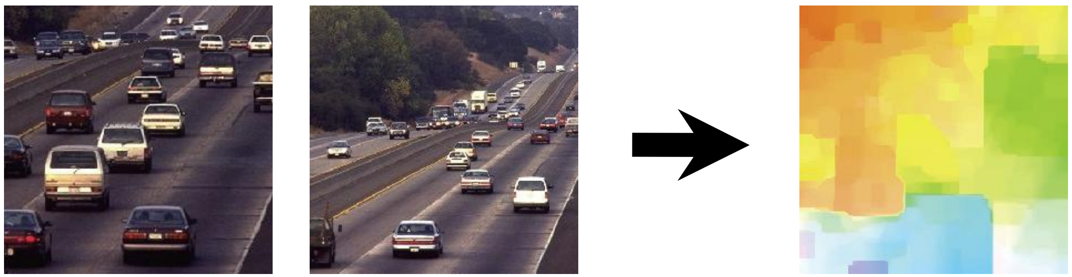
- Search most similar training image in a database using global image descriptors



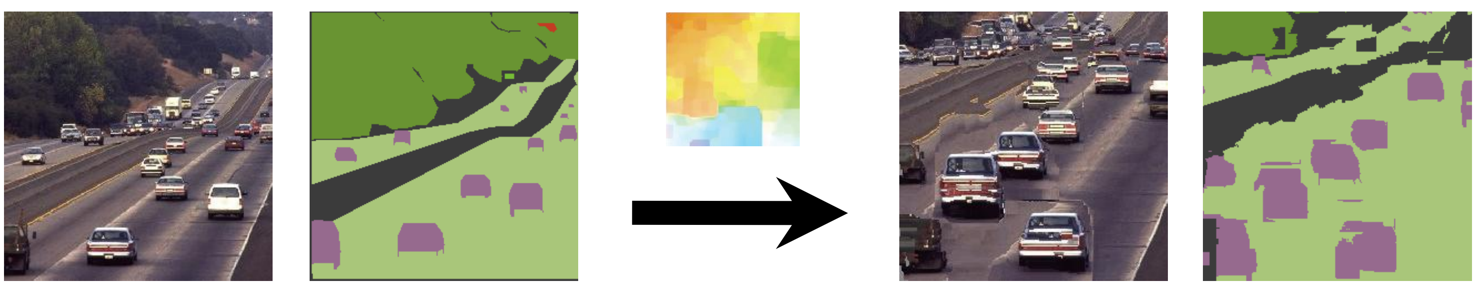
Thomas Brox

SIFT flow

- Compute SIFT descriptors densely in both images
- Apply combinatorial method (Shekhovtsov et al. Efficient MRF Deformation Model for Non-Rigid Image Matching. CVIU 2008) to estimate the flow between both images



- Warping the annotation of the training image yields the transferred annotation of the test image



Thomas Brox

More results

UNIVERSITÄT BONN

Image 1

Image 2

SIFT flow

Warping

Image 1

Image 2

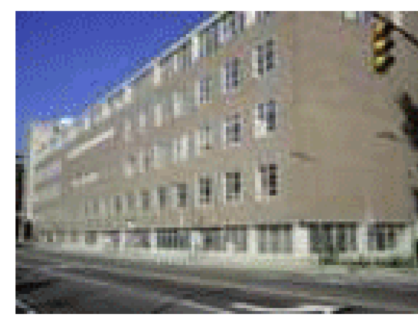
SIFT flow

Warping

Image 1

Image 2

SIFT flow



Warping

Thomas Brox

Incorporating correspondences from descriptor matching

UNIVERSITÄT BONN

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

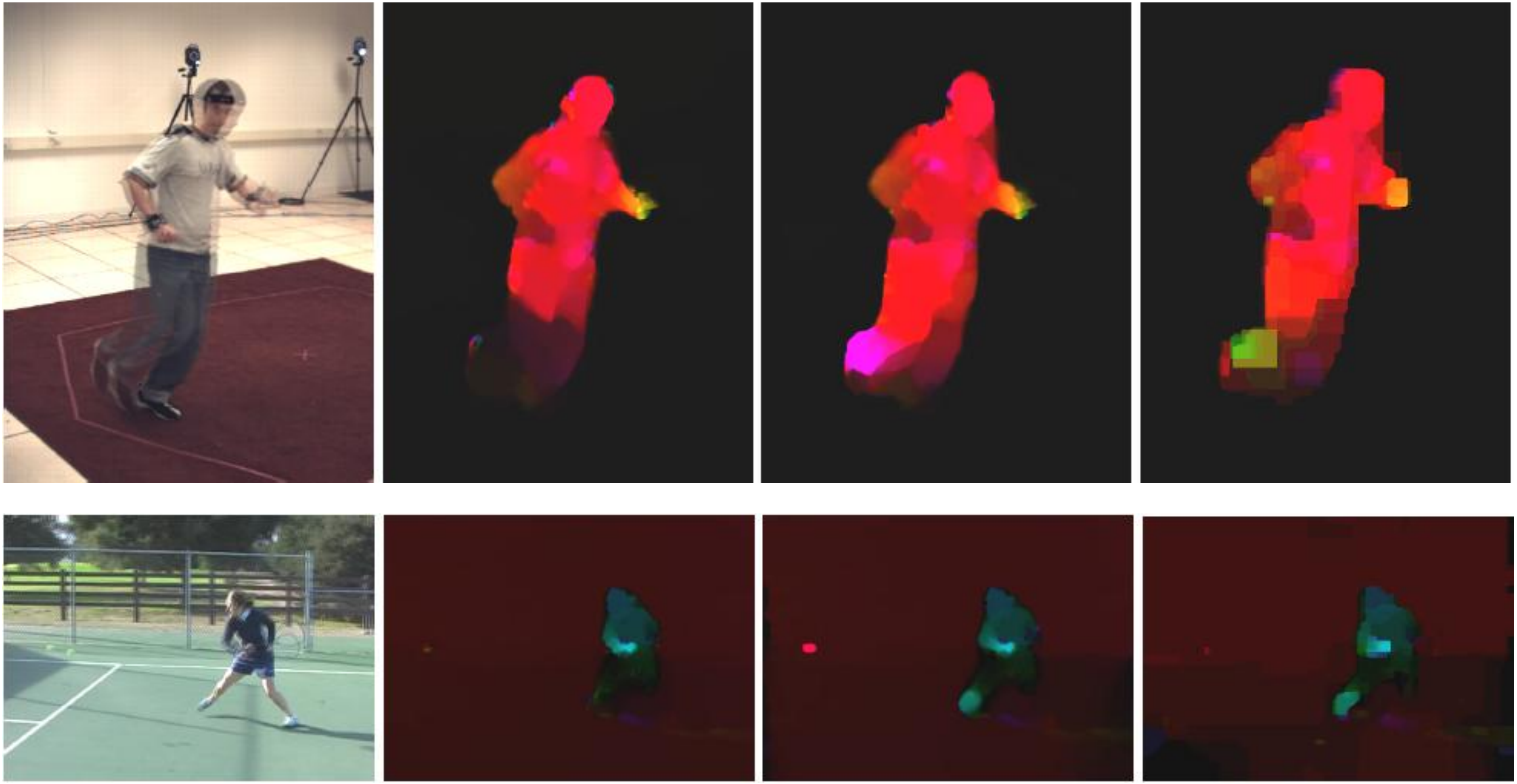
point correspondences from HOG descriptors

$$
+\beta \int \underbrace{\rho(\mathbf{x})}_{\text{Matching score}} \underbrace{\Psi}_{\text{Robust function}}\Big(\underbrace{(u(\mathbf{x}) - u'(\mathbf{x}))^2 + (v(\mathbf{x}) - v'(\mathbf{x}))^2}_{\text{Flow = correspondence vector}}\Big) d\mathbf{x}
$$

[T. Brox and J. Malik. **Large displacement optical flow: descriptor matching in variational motion estimation**. PAMI 2011]

Thomas Brox

Comparison to SIFT flow



Input images (overlaid) | Classical coarse-to-fine method | LDOF | SIFT flow

Thomas Brox

How do we find object regions?

Video analysis/motion segmentation

- Motion cues can be very clear when objects move
- Several indications in biological vision (studies with infants and healed blind people)

Thomas Brox

Motion segmentation

- Goal: separate regions by their motion

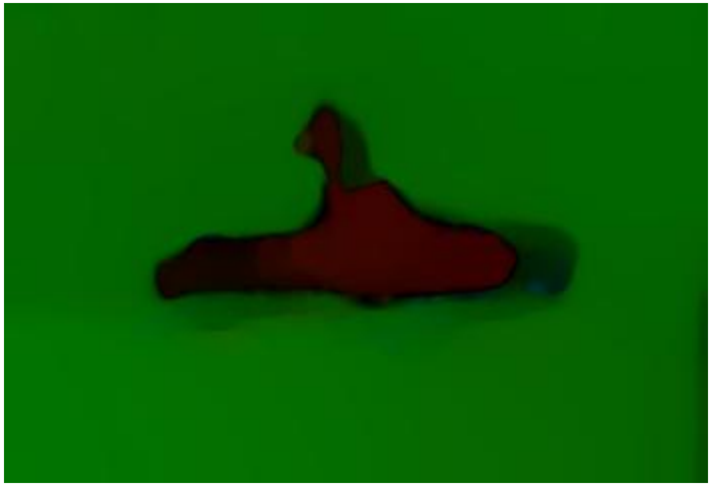
- Typical chicken-egg problem:
  - A motion field is necessary for estimating the partitioning
  - A partitioning helps estimate the motion field

- Standard approach (layered motion: Wang-Adelson 1994):
  1. First compute optical flow
  2. Then compute segmentation based on the optical flow

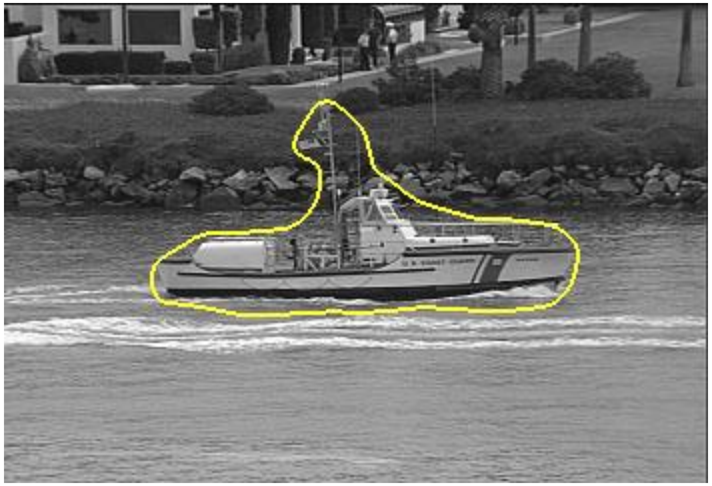
Thomas Brox

Example

Input sequence



Estimated flow field



Flow segmentation

Thomas Brox

Simultaneous flow estimation and segmentation

- So far: two-step approach
  - Decide on the optic flow
  - Use result for segmentation
- Problem: wrong decisions in the first step cannot be corrected → optical flow estimation cannot benefit from segmentation
- Solution: state joint problem in one energy functional
- Optimize in both variables simultaneously

Thomas Brox

Motion competition

- Joint energy functional (Cremers-Soatto 2005):

$$E(u_i, v_i, C) = \sum_{i=1}^{N} \int_{\Omega_i} (I_x u_i + I_y v_i + I_z)^2 \,\mathbf{dx} + \nu |C|$$

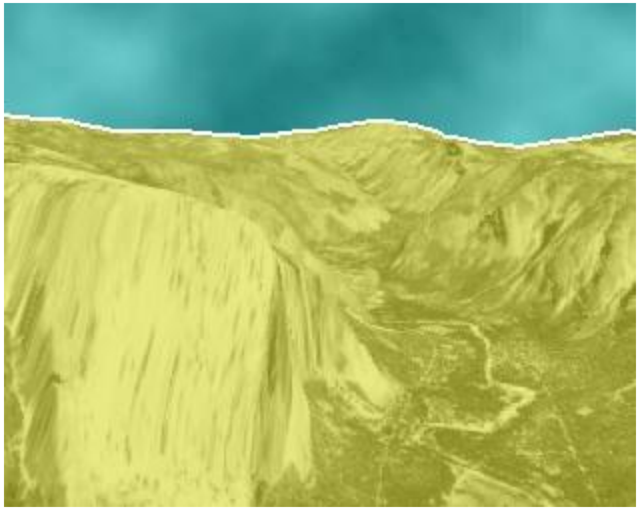
- Remarks:
  – Partitioning of the image domain into N disjoint regions
  – Simple parametric (constant) motion field in each region Ω_i
  – Length constraint prefers smooth contours

- Alternating optimization:
  – Estimate motion field in the initial regions (Lucas-Kanade method restricted to the region's domain)
  – Evolve the contour according to the motion field (e.g. with level sets)
  – Iterate

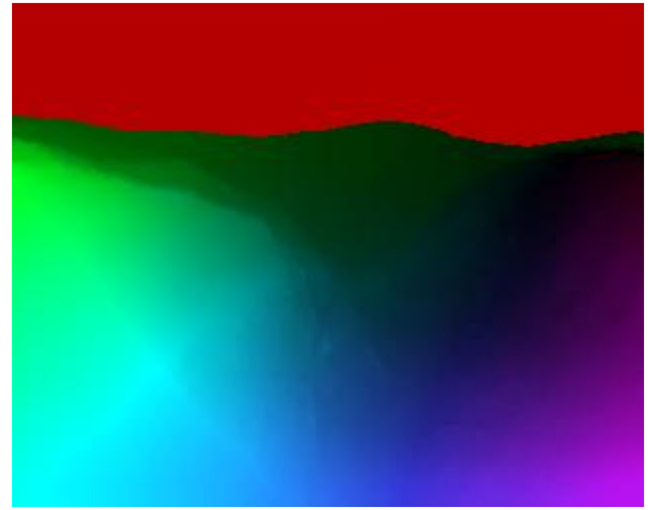
- Segmentation and motion estimation are coupled and benefit from each other

Thomas Brox

Two easy examples



Segmentation



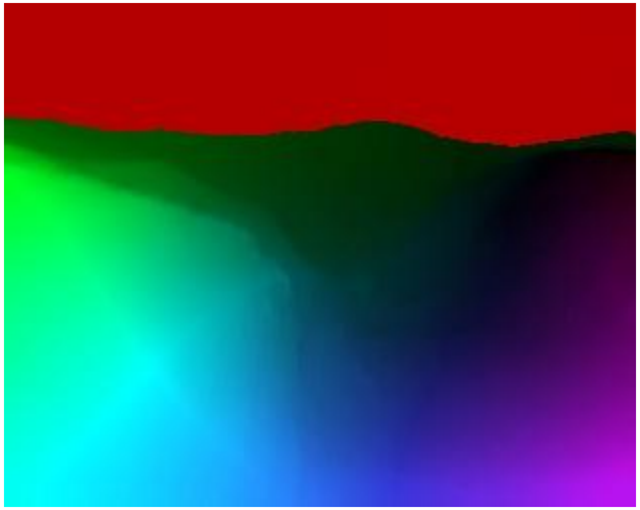
Optical flow

Segmentation

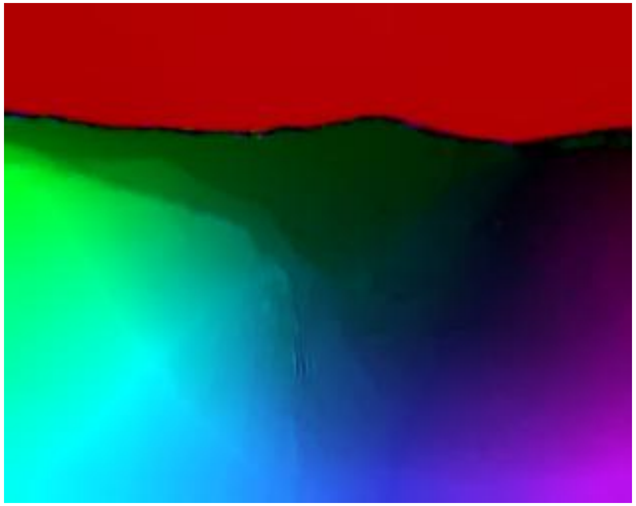
Optical flow

Thomas Brox

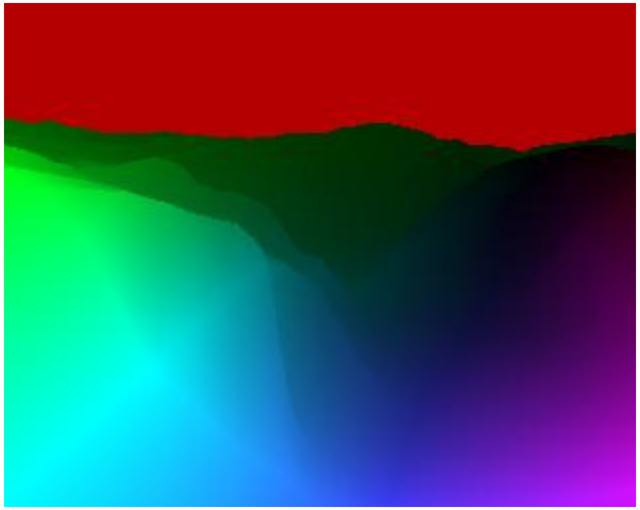
Accurate motion fields



With segmentation (1.22°)



Without segmentation (1.78°)

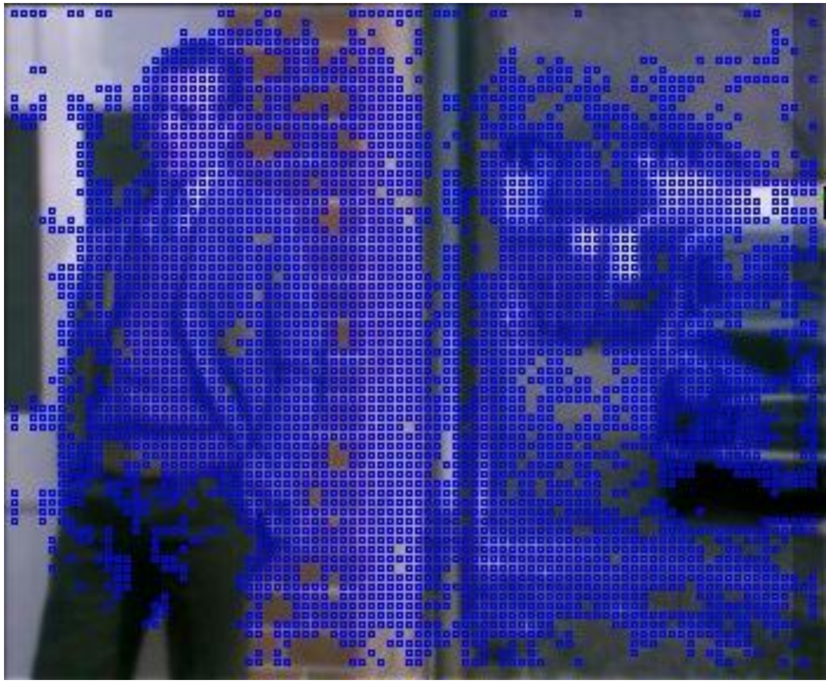


Correct flow field

Thomas Brox

Motion segmentation based on point trajectories

- Two-frame optical flow can only separate objects with different motion in these two particular frames
- Idea: make use of the temporal context
  → point trajectories over multiple frames
- Trajectories obtained via point tracking
- Clustering based on this information can separate objects even at times where they do not move

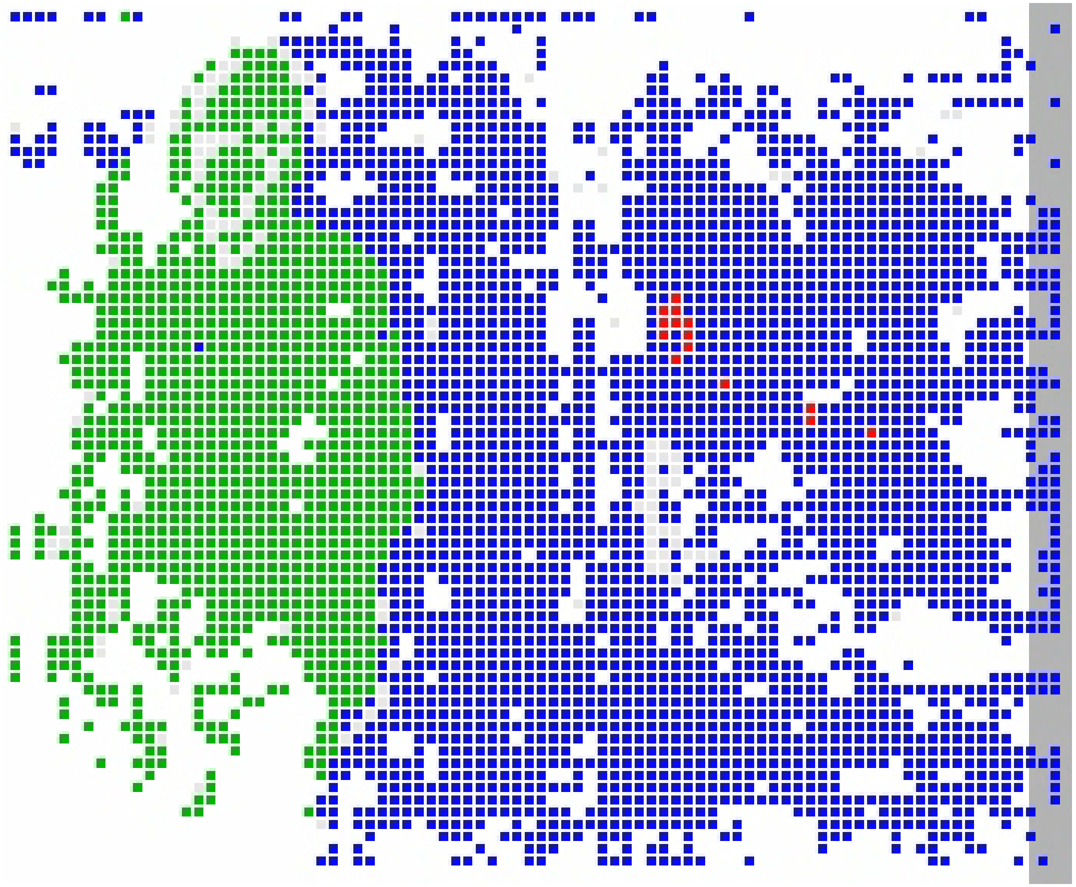


Trajectories obtained with tracker from Sundaram et al. 2010

Thomas Brox

Clustering of point trajectories

Point trajectories → Single graph, pairwise distances → Spectral clustering → Clustered trajectories



[P. Ochs. **Segmentation of moving objects by long term video analysis**. PAMI 2014]

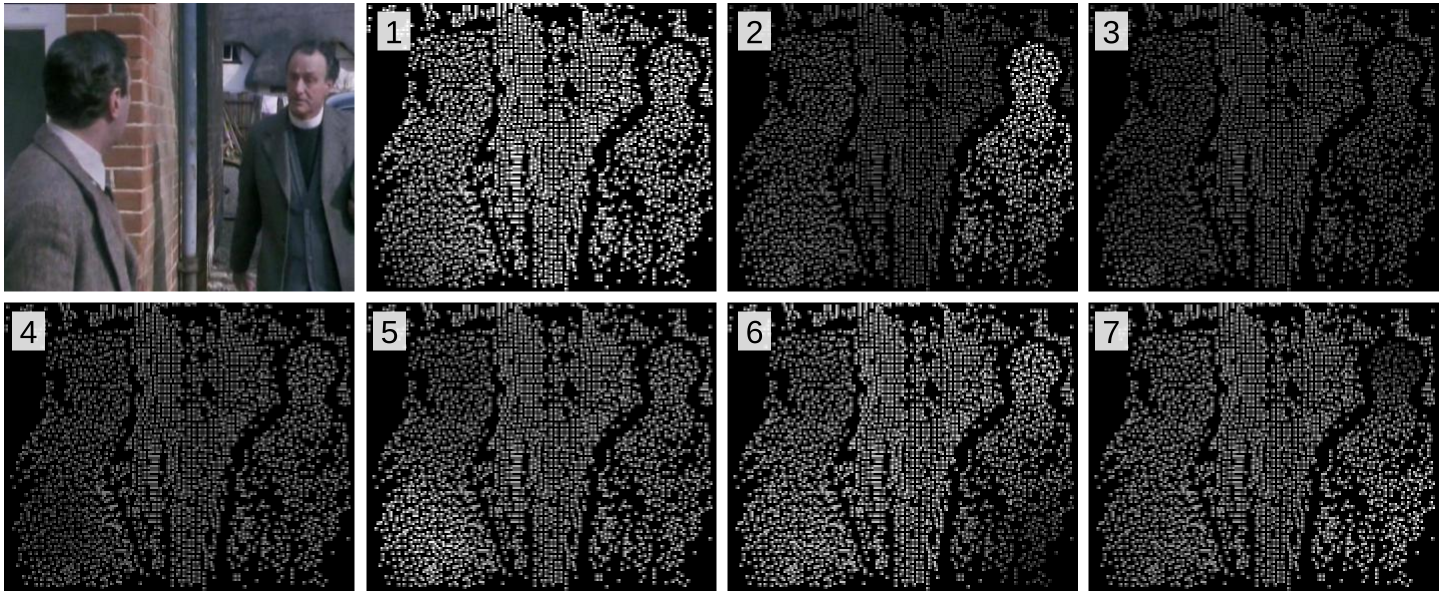
Thomas Brox

Recall: Spectral clustering

1. Compute Laplacian eigenmap

$$V^\top \Lambda V = D^{-1/2}(D - W)D^{-1/2}$$

2. Consider all eigenvectors with eigenvalue $\lambda < 0.04$

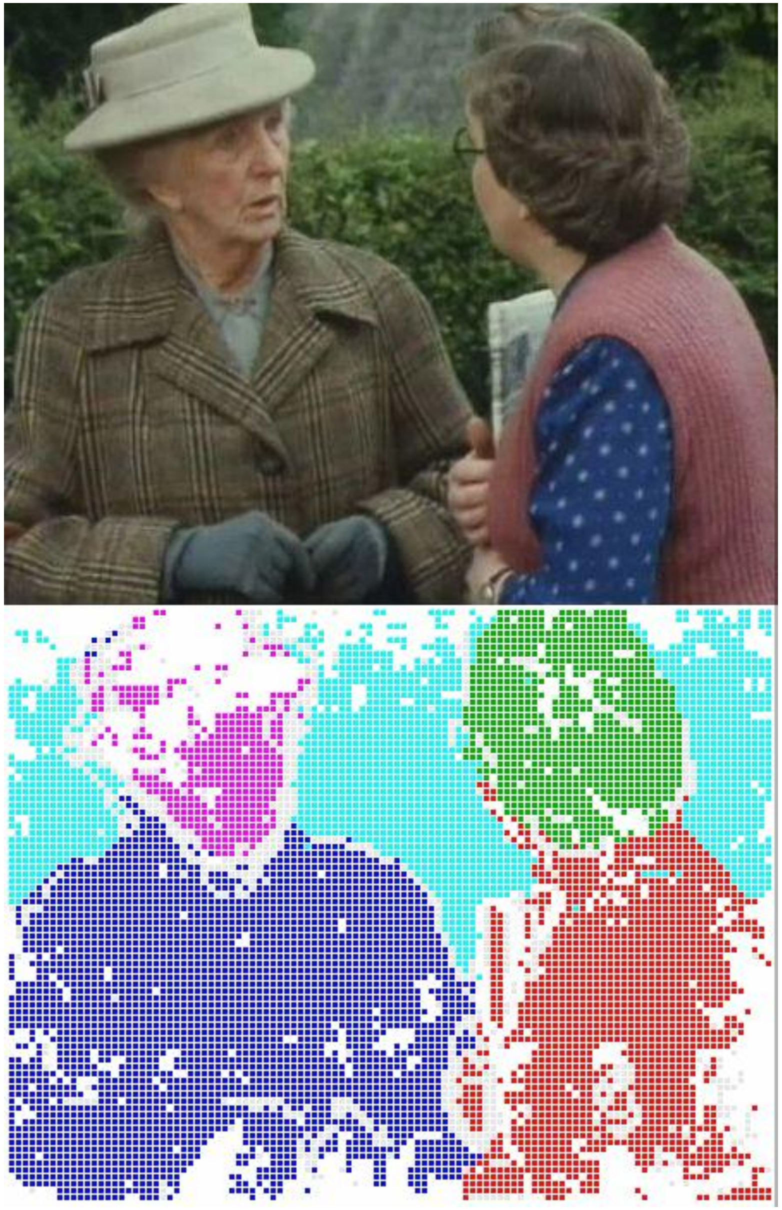


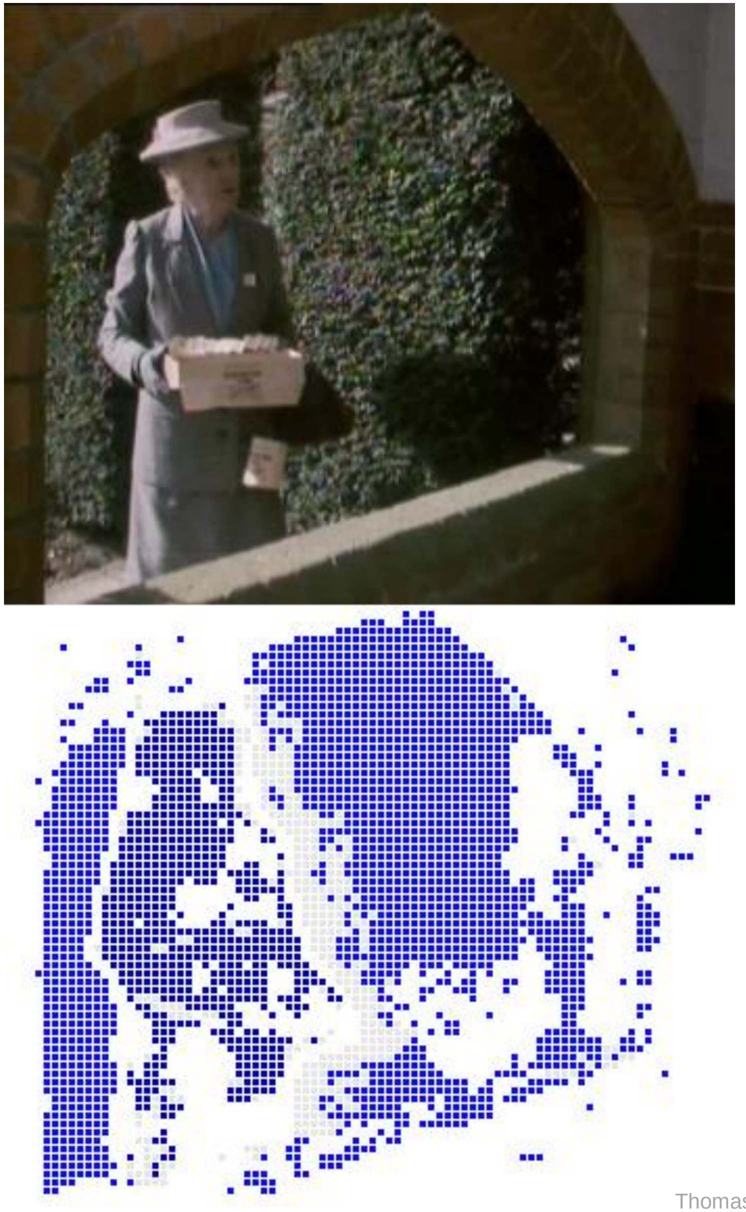
…

3. Run multi-region segmentation on eigenvectors

Thomas Brox

Results on movie shots





Thomas Brox

UNIVERSITÄT BONN