

Image Processing and Computer Graphics

Image Processing

Class 5

Motion estimation

- Given a sequence of images $I(x, y, t)$, what is the motion of each pixel between subsequent frames?



Hamburg taxi sequence

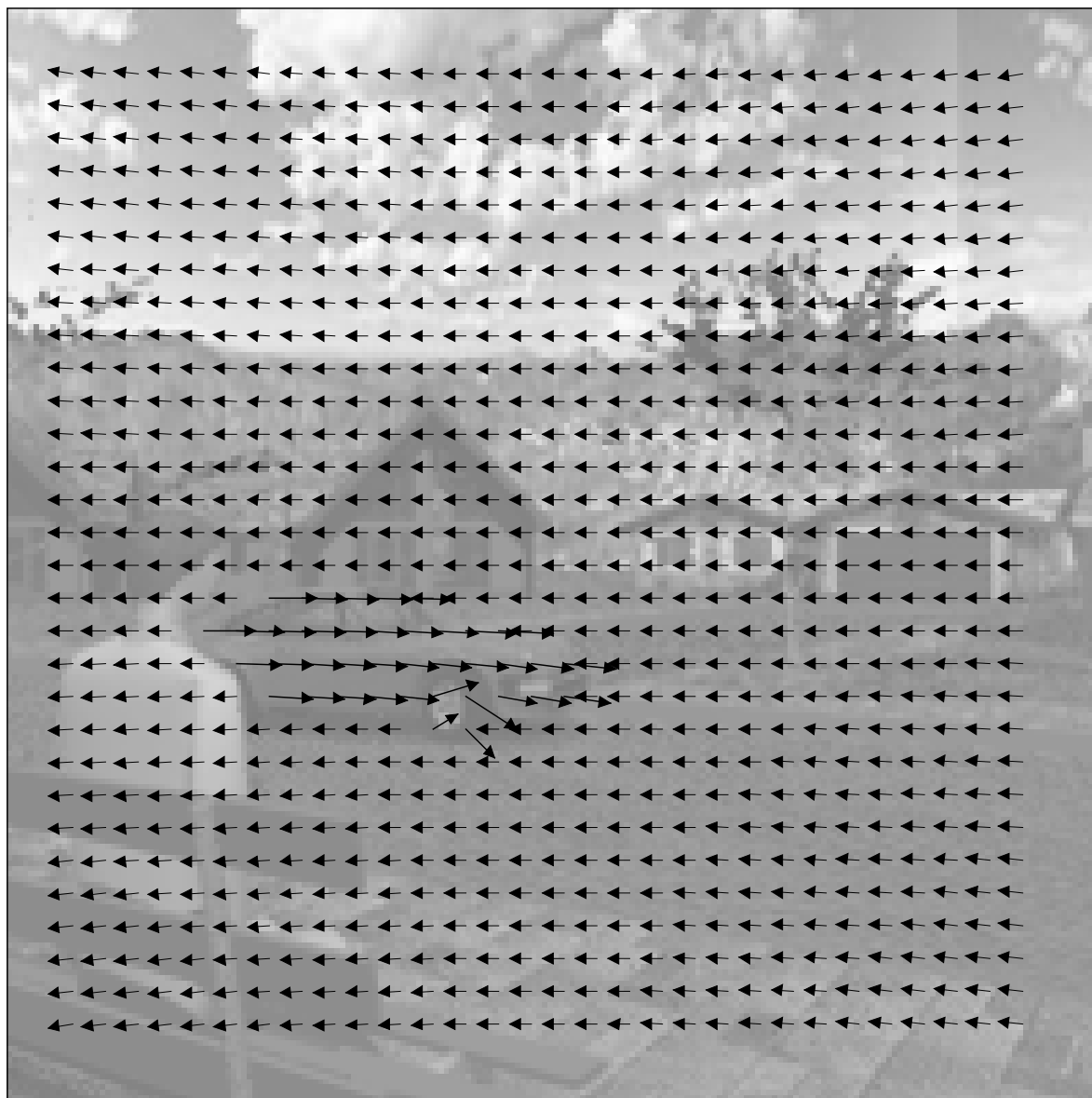


Sequence from VSB100 dataset

- Formally, we seek a vector field $(u, v)(x, y, t)$ that describes the motion of points at a particular time point (a frame).
the vector to think of
Motion at time t
- This vector field is called **optical flow**. It individually moves each point (x, y) at time t such that it fits the point at time $t + 1$.



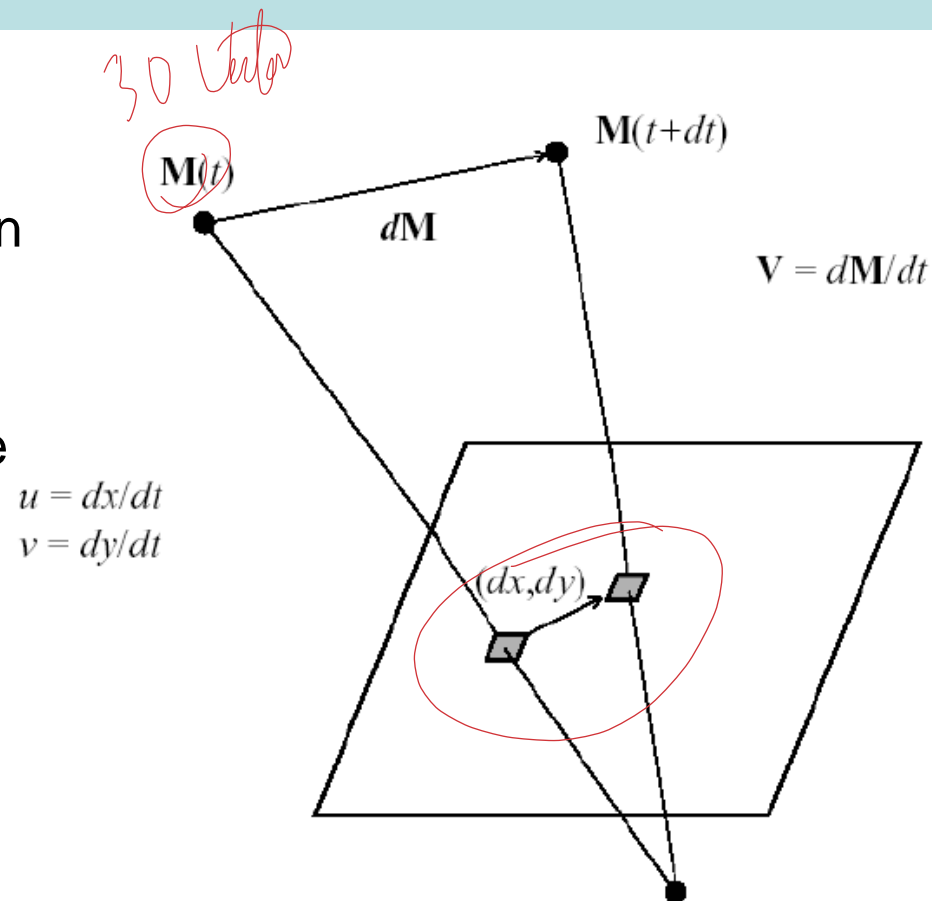
Street sequence



*Segmentation:
to left - background
to right - object.*

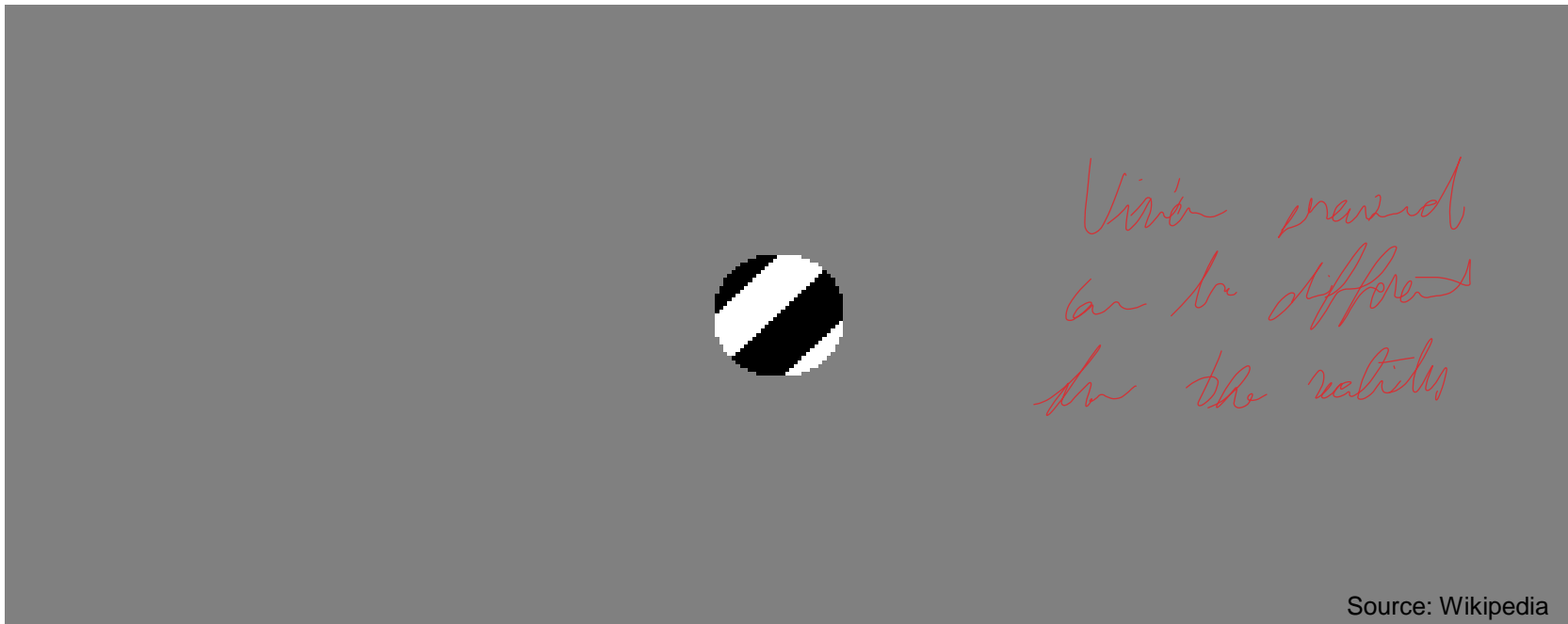
- Optical flow provides important information for many other tasks in vision.
- It can be used for detection or segmentation of objects by comparing the object motion and the background motion → **motion segmentation**
Separate objects from the background
- Optical flow is also useful for **label propagation in videos**
- It can be used for estimating the 3D motion of objects, the so-called **scene flow**. It is important, e.g., for catching a fly or for avoiding collisions.
- Optical flow can also help estimate the 3D structure of objects. This is called **structure from motion**.
 - Moving your head, the image of near objects moves faster than the image of distant objects (motion parallax).
 - For many animals, motion is the only way to see 3D.

- There is a difference between 3D motion (in the scene) and 2D motion (in the image).
- Moreover, there can be a difference between the true motion in the image and the apparent motion.
- How comes that there is a difference between the true and the apparent motion?



Author: Martial Hebert

- The optical flow is usually not uniquely determined from the image data alone. A black pixel potentially could have moved to any black pixel in the other frame.
- Prior assumptions (e.g. that neighboring pixels move the same way) are needed for a unique solution. The neighborhood (aperture) determines the solution. Therefore, the ambiguity is also called the **aperture problem**.



- The most common assumption for optical flow estimation is that the gray value of a moving pixel stays constant over time.

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

- This constraint is nonlinear in the unknown flow, which makes estimation difficult.
- To ease optical flow estimation, we can linearize the first expression with a Taylor expansion:

Linearization x-derivative of the image

$$I(x+u, y+v, t+1) = I(x, y, t) + \underbrace{I_x u + I_y v + I_t}_{\text{x-derivative of the image}} + O(u^2, v^2)$$

- This leads to the (linearized) optic flow constraint:

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

- This linearization is only sufficiently precise, if the images are smooth (locally linear) and the flow is small.

- Optic flow constraint: $I_x u + I_y v + I_t = 0$
- We see the aperture problem: regarding a single pixel, we cannot uniquely determine its flow vector. There is only one equation, but two unknowns.
- We need a second assumption to obtain a unique solution.
- The **Lucas-Kanade method** assumes that the motion is constant within a local neighborhood.
- We then obtain multiple equations for the two unknowns, i.e., an over-determined system of equations:

$$I_x(x', y', t)u + I_y(x', y', t)v + I_t(x', y', t) = 0 \quad \forall x', y' \in \mathcal{R}(x, y)$$

- In general, such an over-determined system has not a solution anymore.
- Instead we seek a solution that minimizes the total squared error:

$$\operatorname{argmin}_{u,v} \sum_{x,y \in \mathcal{R}} (I_x(x,y)u + I_y(x,y)v + I_t(x,y))^2$$

Optical flow constraint

- The necessary condition for a minimum is:

$$\frac{\partial E}{\partial u} = 2 \sum_{x,y \in \mathcal{R}} (I_x(x,y)u + I_y(x,y)v + I_t(x,y))I_x(x,y) = 0$$

$$\frac{\partial E}{\partial v} = 2 \sum_{x,y \in \mathcal{R}} (I_x(x,y)u + I_y(x,y)v + I_t(x,y))I_y(x,y) = 0$$

- This leads to a linear system at each pixel

$$\begin{pmatrix} \sum I_x^2 & \sum I_x I_y \\ \sum I_x I_y & \sum I_y^2 \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} -\sum I_x I_t \\ -\sum I_y I_t \end{pmatrix}$$

- Instead of a uniformly weighted, box-shaped window, we can also use a Gaussian window.

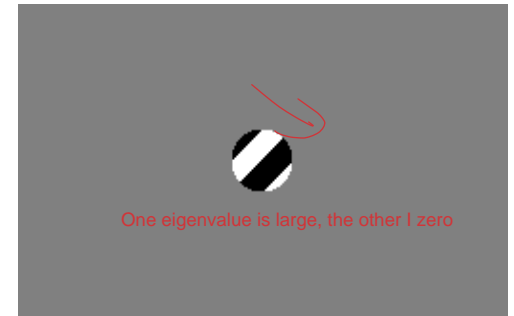
Window ↑ Reliability ↓

- The sums weighted by the window function come down to a convolution:

$$\begin{pmatrix} K_{\rho} * I_x^2 & K_{\rho} * I_x I_y \\ K_{\rho} * I_x I_y & K_{\rho} * I_y^2 \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} -K_{\rho} * I_x I_t \\ -K_{\rho} * I_y I_t \end{pmatrix}$$

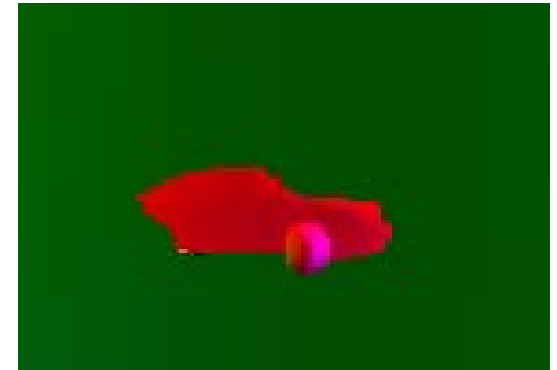
*2 Non-zero
Eigenvalues*

- A unique solution will be obtained only if the system is not singular.
- What does this mean?
 - The local neighborhood must contain non-parallel gradients.
 - Otherwise the aperture problem is still present.



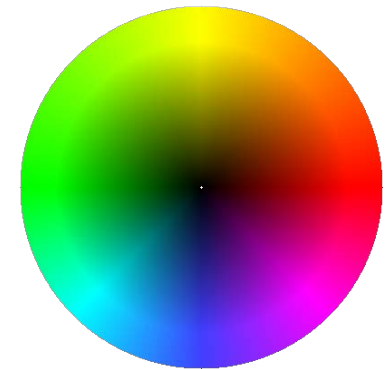


Input sequence

Optical flow estimated
with Lucas-Kanade
method

Correct flow field

- Drawbacks of the Lucas-Kanade method:
 - The assumption of locally constant motion often is not realistic (e.g. at motion boundaries).
 - There are no direct constraints on the smoothness of the resulting flow field.
- Advantages: fast and simple



- The Lucas-Kanade method is a “parametric” or “local” method
 - It assumes a parametric motion model.
 - The parameters are estimated from data in a local neighborhood of each point.
 - ➔ Point estimates are computed independently.
- In contrast, global methods have a global dependency between points due to a global smoothness assumption. Not pixel by pixel anymore
- A global solution is usually derived with variational methods.
- Most basic variational model: Horn-Schunck method
- This model can be further extended to yield highly accurate estimates in many practical situations.

- The first assumption is the same as in the Lucas-Kanade method: gray value constancy leading to the optic flow constraint.

$$I_x u + I_y v + I_z = 0$$

- The second assumption is different. The Horn-Schunck model assumes global smoothness of the flow field:

$$|\nabla u|^2 + |\nabla v|^2 \rightarrow \min$$

- Both assumptions can be combined in an energy functional:

with image

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

*Quadratic in u, v so
linear.*

- The sought optical flow is the minimizer of this energy.
- Note the similarities to the denoising model from the previous class.

- We can find the minimizer of this functional with the calculus of variation.
- The energy is convex → we get a unique global optimum.
- With the Gâteaux derivatives

$$\left. \frac{d}{d\epsilon} E(u(x) + \epsilon h(x), v(x)) \right|_{\epsilon=0} = 0 \quad \forall h(x)$$

$$\left. \frac{d}{d\epsilon} E(u(x), v(x) + \epsilon h(x)) \right|_{\epsilon=0} = 0 \quad \forall h(x)$$

we obtain the Euler-Lagrange equations

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

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

Laplacian: $\Delta v = v_{xx} + v_{yy}$

*Second order
Derivative.*

- The discretized versions of these equations read

neighbors

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

neighbors

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

- The equations are linear in (u, v) . Thus we obtain a large linear system to solve (as in the denoising case).
- This system can be written in matrix-vector notation: $A^h w = b$, where $w := (u, v)$.
- As in the denoising case, the system matrix is sparse with only few diagonals different from zero.

$$A^h = \begin{pmatrix} d & & & & \\ & d & & & \\ & & d & & \\ & & & d & \\ & & & & d \\ d & & & & \\ & d & & & \\ & & d & & \\ & & & d & \\ & & & & d \end{pmatrix} + \begin{pmatrix} s & s & & s & \\ s & s & s & & s \\ & s & s & s & \\ s & & s & s & s \\ & s & & s & s \\ s & & & s & s \\ & s & & & s \\ s & s & & s & \\ s & s & s & & s \\ & s & s & s & \\ s & & s & s & s \\ & s & & s & s \end{pmatrix}$$

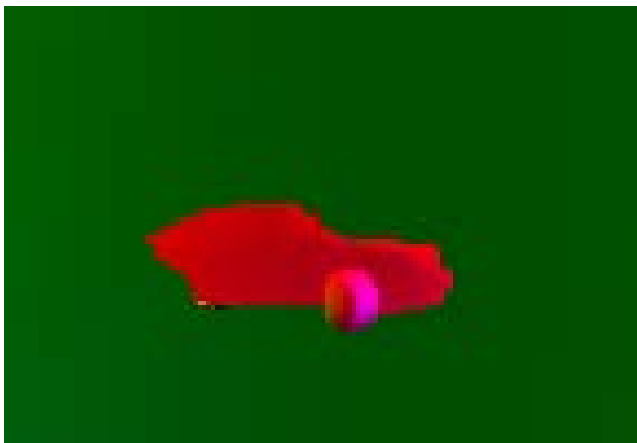
- The blocks structure is due to the coupling of u and v .
- The **data term** contributes only to block main diagonals.
- The **smoothness term** has four additional block off-diagonals (coupling of neighboring pixels).
- We can solve this linear system with the linear solvers from class 4 (Gauss-Seidel, SOR, Multigrid).



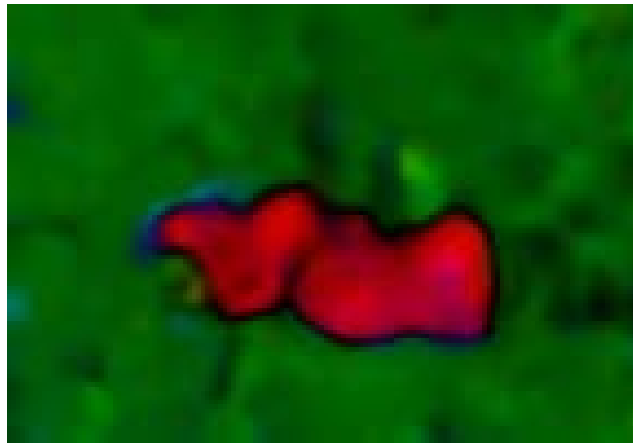
Input sequence



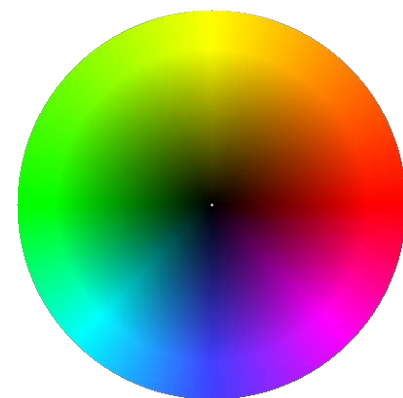
Horn-Schunck (10.4°)



Ground truth



Lucas-Kanade (10.54°)



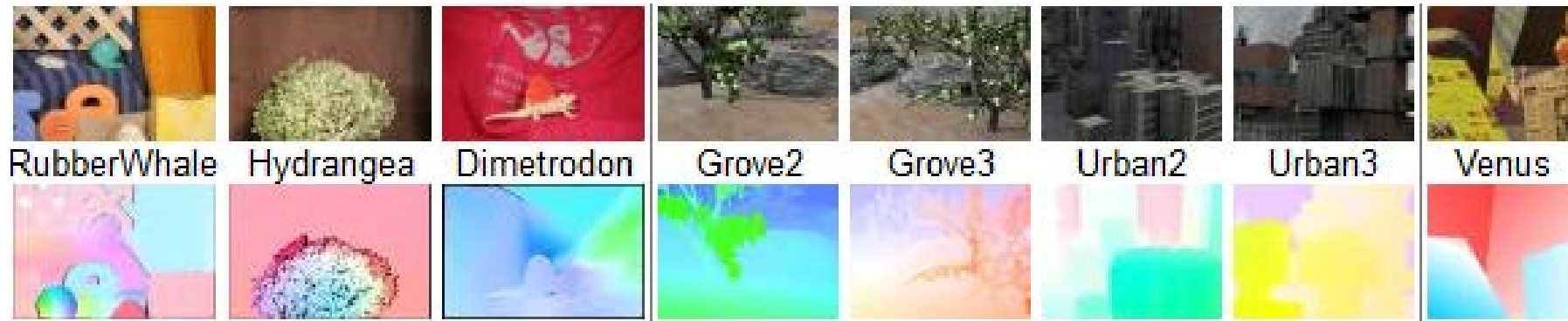
- An established measure of how accurately the optical flow has been estimated is the **average angular error**.
- This requires, of course, knowledge about the true motion field, the so-called **ground truth**. For some synthetic sequences such ground truth data is available.
- The average angular error measures the 3D angle between the correct and the estimated flow vectors:

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

- Note that this also captures errors in the length of the 2D flow vector.
- An alternative measure is the **end point error**:

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

- Available at <http://vision.middlebury.edu/flow/>

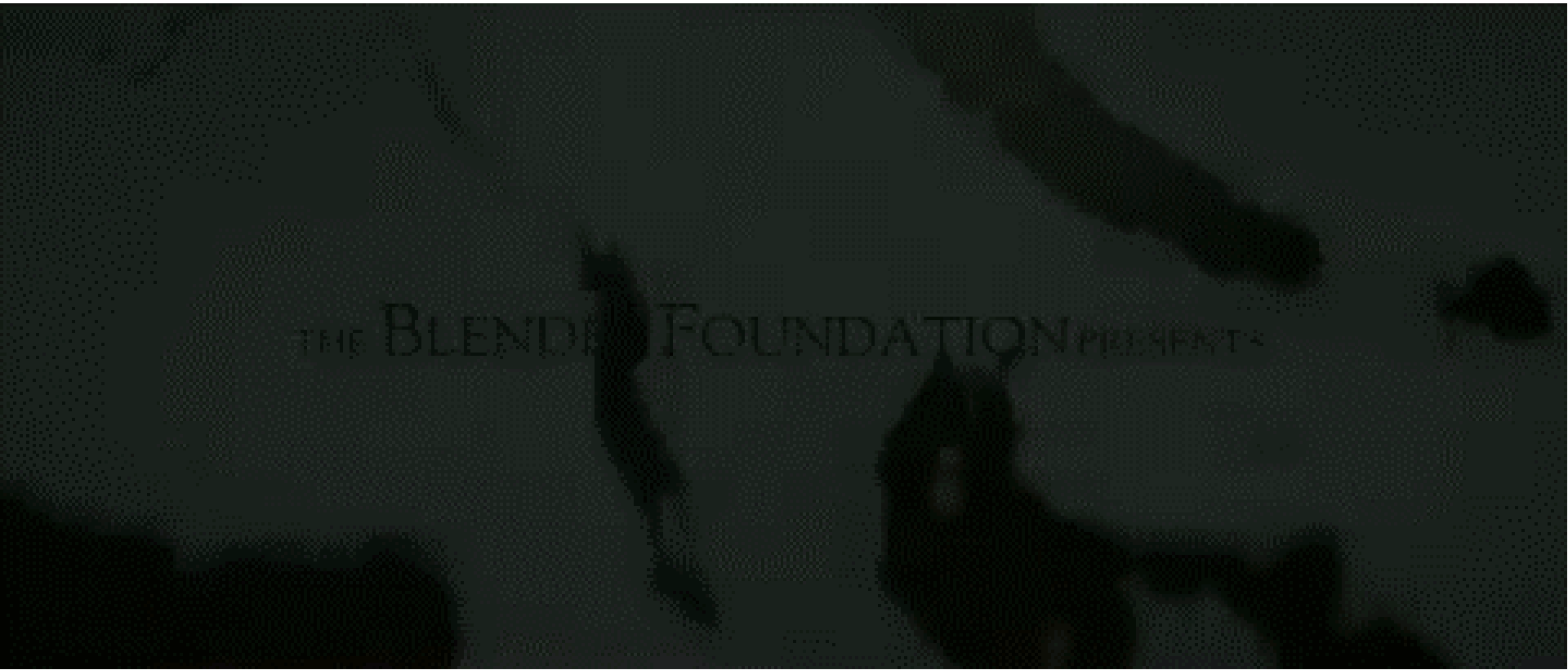


8 training sequences with ground truth (for parameter tuning)



8 test sequences with ground truth (for evaluation)

Too few test data → overfitting

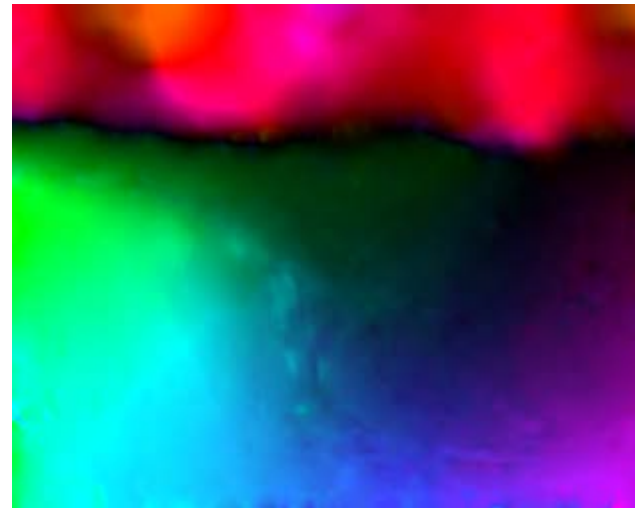


Sample frames and optical flow ground truth (Butler et al. ECCV 2012)

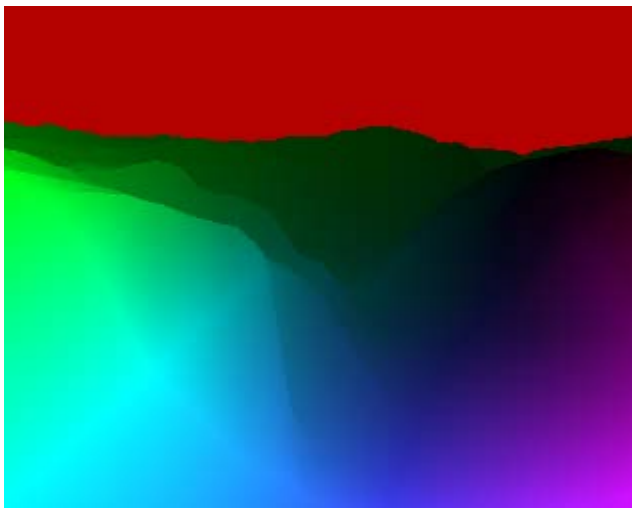




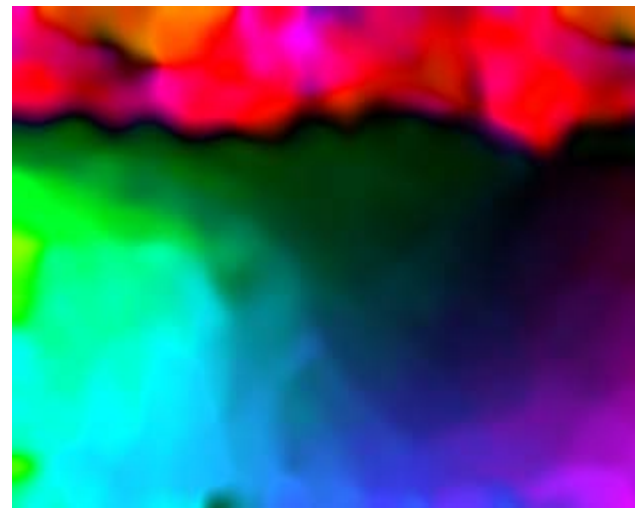
Input sequence



Horn-Schunck (7.17°)



Ground truth

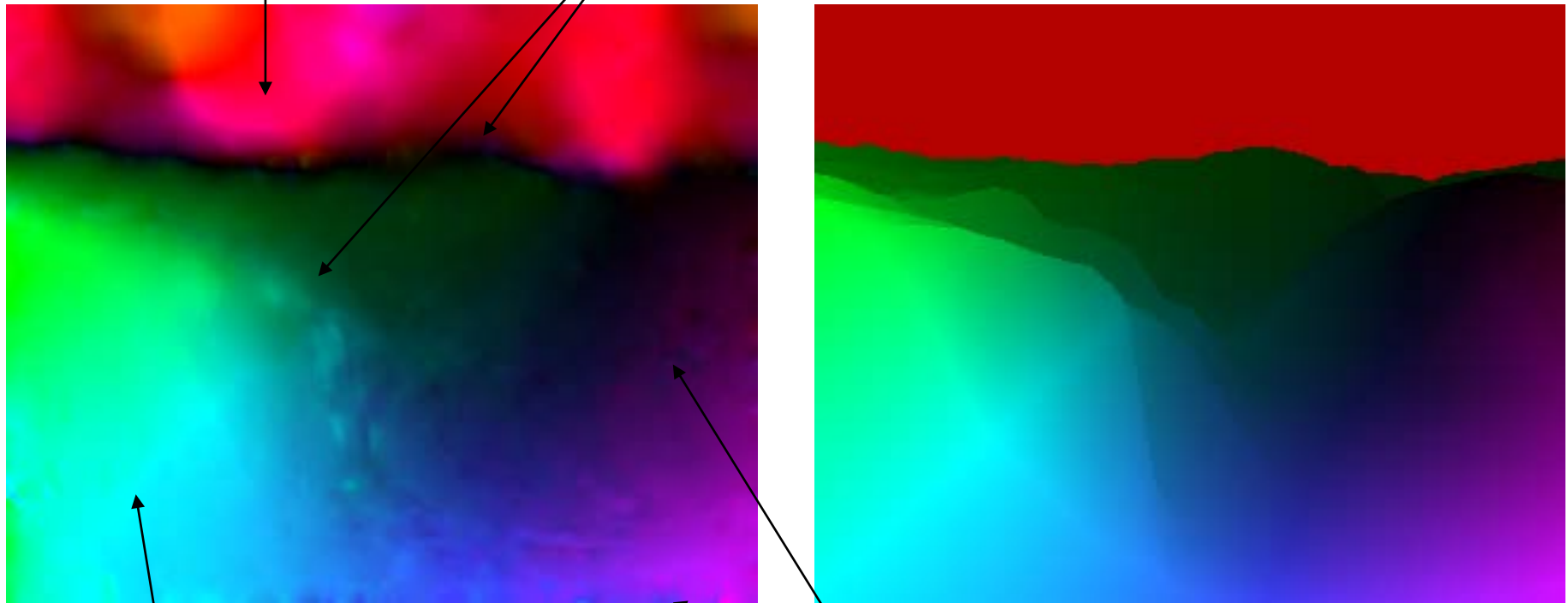


Lucas-Kanade (8.78°)

Limitations of the Horn-Schunck method

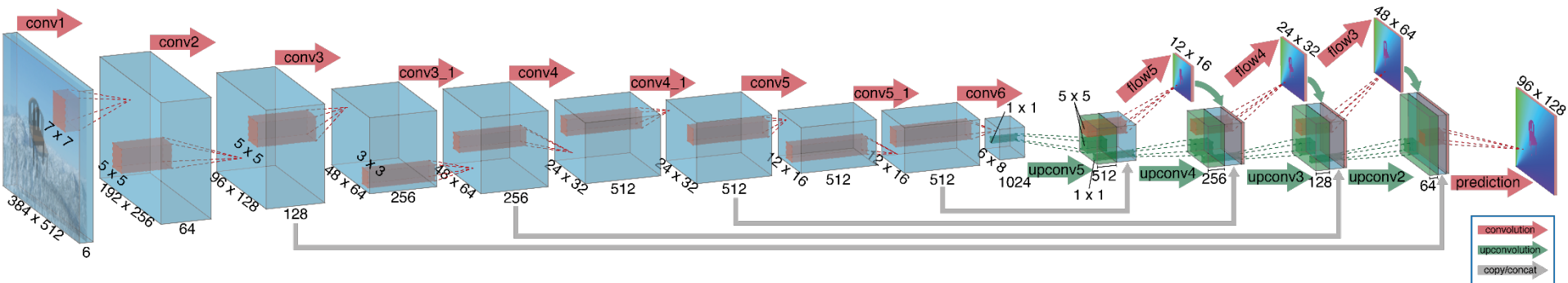
Cloud region not well
estimated due to
illumination changes

Motion discontinuities
blurred



Underestimation of large
displacements

Outliers due to non-Gaussian
noise or occlusion



- Has become the dominant approach for optical flow estimation
- Requires training data (usually rendered data)
- Runs faster than most variational methods (on GPUs)
- Can learn more sophisticated priors from data than the handcrafted smoothness priors

Mayer et al. CVPR 2016

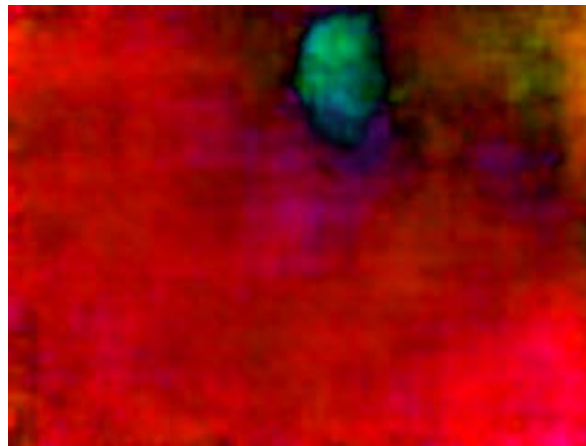


Driving, Monkaa, FlyingThings3D datasets publicly available

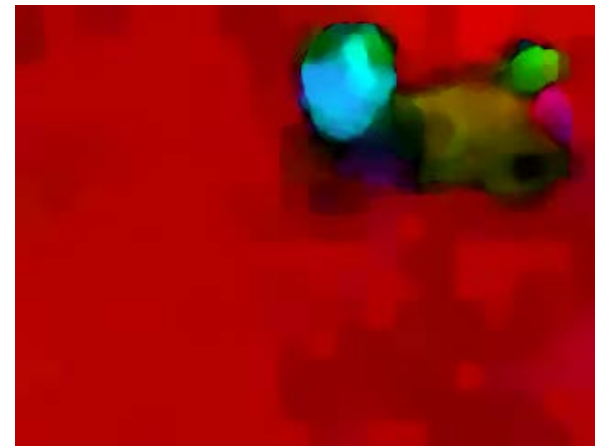




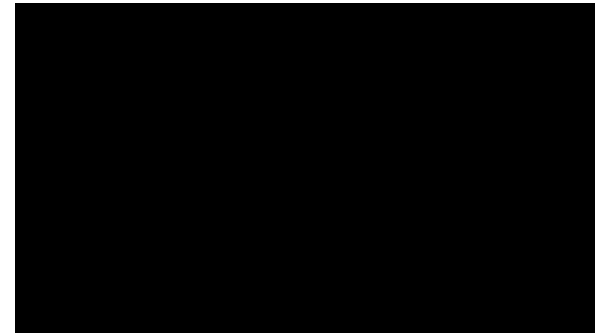
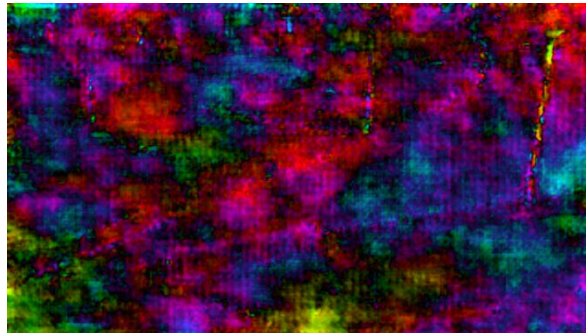
Input



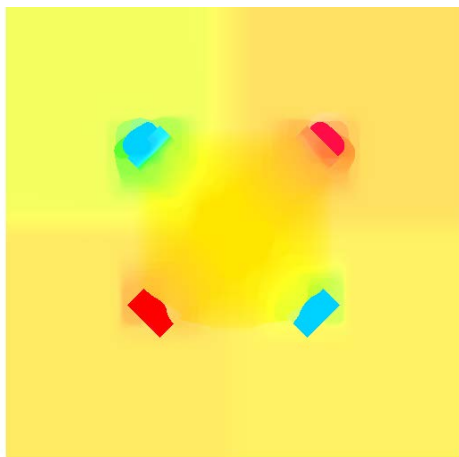
FlowNet



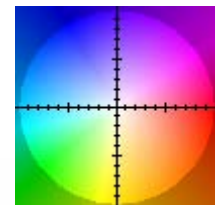
Variational method



A hard toy example



Variational method

Unsupervised
learning method

FlowNet 2.0

- The optical flow is the apparent motion we can measure in the image plane.
- Optical flow estimation is usually based on the gray value constancy assumption.
- Local methods like the Lucas-Kanade method estimate motion parameters independently for each pixel by considering a local neighborhood.
- Global methods rather assume smoothness of the flow field and estimate the whole displacement field in one piece. The solution is found by variational techniques.
- Deep networks are an alternative to variational methods

- B. D. Lucas, T. Kanade: An iterative image registration technique with an application to stereo vision, *Proc. International Joint Conference on Artificial Intelligence*, pp. 674-679, 1981.
- B. Horn, B. Schunck: Determining optical flow, *Artificial Intelligence*, 17:185-203, 1981.
- E. Mémin, P. Pérez: Dense estimation and object-based segmentation of the optical flow with robust techniques, *IEEE Transactions on Image Processing*, 7(5):703-719, 1998.
- T. Brox, A. Bruhn, N. Papenberg, J. Weickert: Highly accurate optical flow estimation based on a theory for warping, *European Conference on Computer Vision*, Springer LNCS 3024, pages 25-36, 2004.
- D. Butler, J Wulff, G. Stanley, M. Black: A naturalistic open source movie for optical flow evaluation. *European Conference on Computer Vision*, 2012.
- A. Dosovitskiy, P. Fischer, E. Ilg, P. Häusser, C. Hazirbas, V. Golkov, P. van der Smagt, D. Cremers, T. Brox: FlowNet: learning optical flow with convolutional networks, *International Conference on Computer Vision*, 2015.
- E. Ilg, N. Mayer, T. Saikia, M. Keuper, A. Dosovitskiy, T. Brox. FlowNet 2.0: Evolution of Optical Flow Estimation with Deep Networks, *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.

- Implement the Lucas-Kanade method for optical flow estimation. For your convenience, make use of the convolution function `NFilter::filter` and the predefined filter masks `CSmooth` and `CDerivative` in `CFilter.h`.
Try your implementation with the Street sequence. For visualization you can make use of the function `flowToImage`.
Adjust the size of the Gaussian neighborhood and play with this parameter.
Presmooth the input images before computing the gradients. Play with the amount of smoothing. Can you explain why the results get better with some presmoothing?
- As an exercise for the exam, derive the Euler-Lagrange equations of the Horn and Schunck functional using Gateaux derivatives.
- Implement the basic Horn-Schunck method. You can implement it via any linear solver discussed in class 3. The SOR solver is recommended. Find a regularization parameter that gives good results. Compare the Horn-Schunck results to the Lucas-Kanade results.
- Extra assignment (optional): Extend your Horn-Schunck method by an edge preserving penalizer.