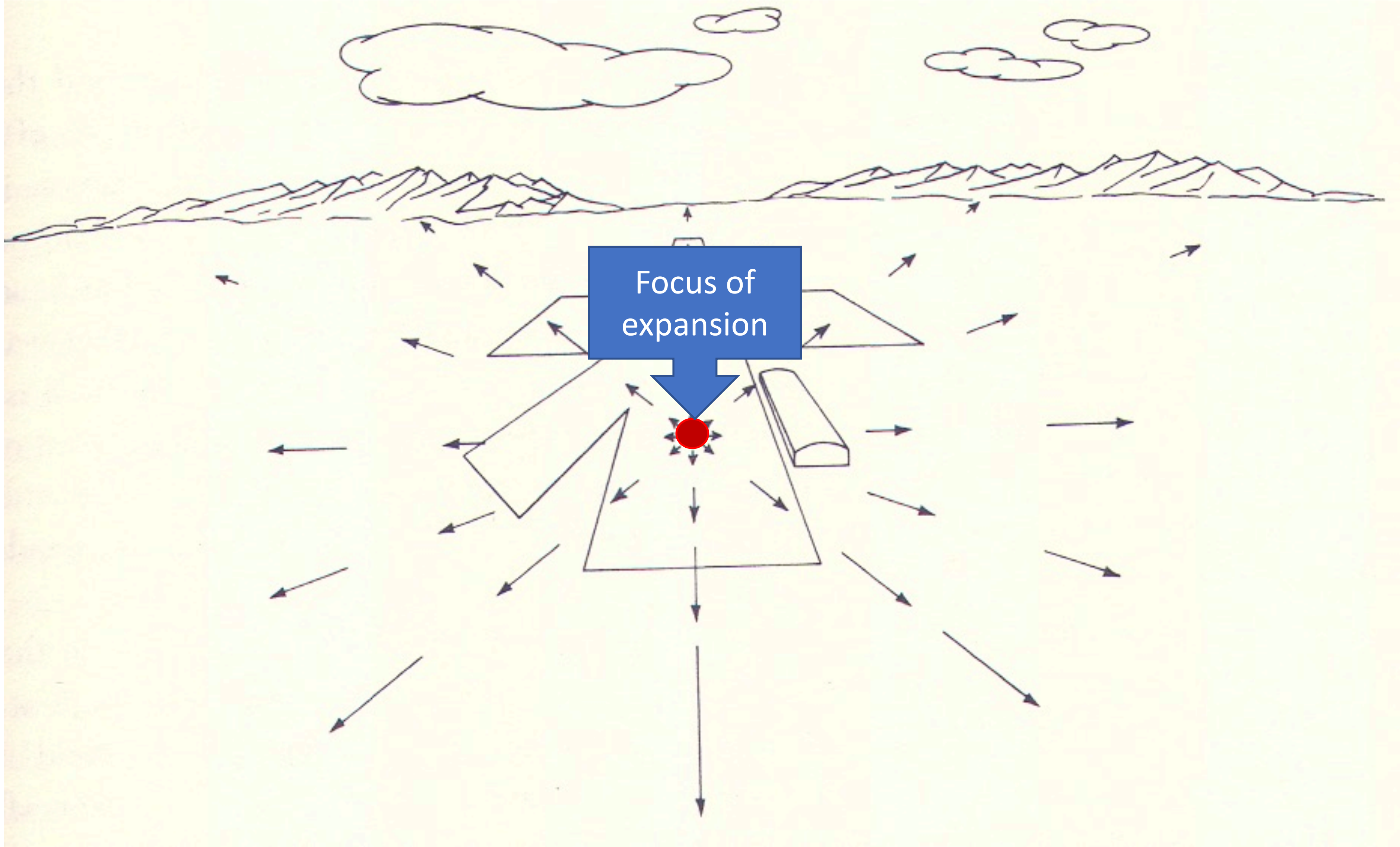
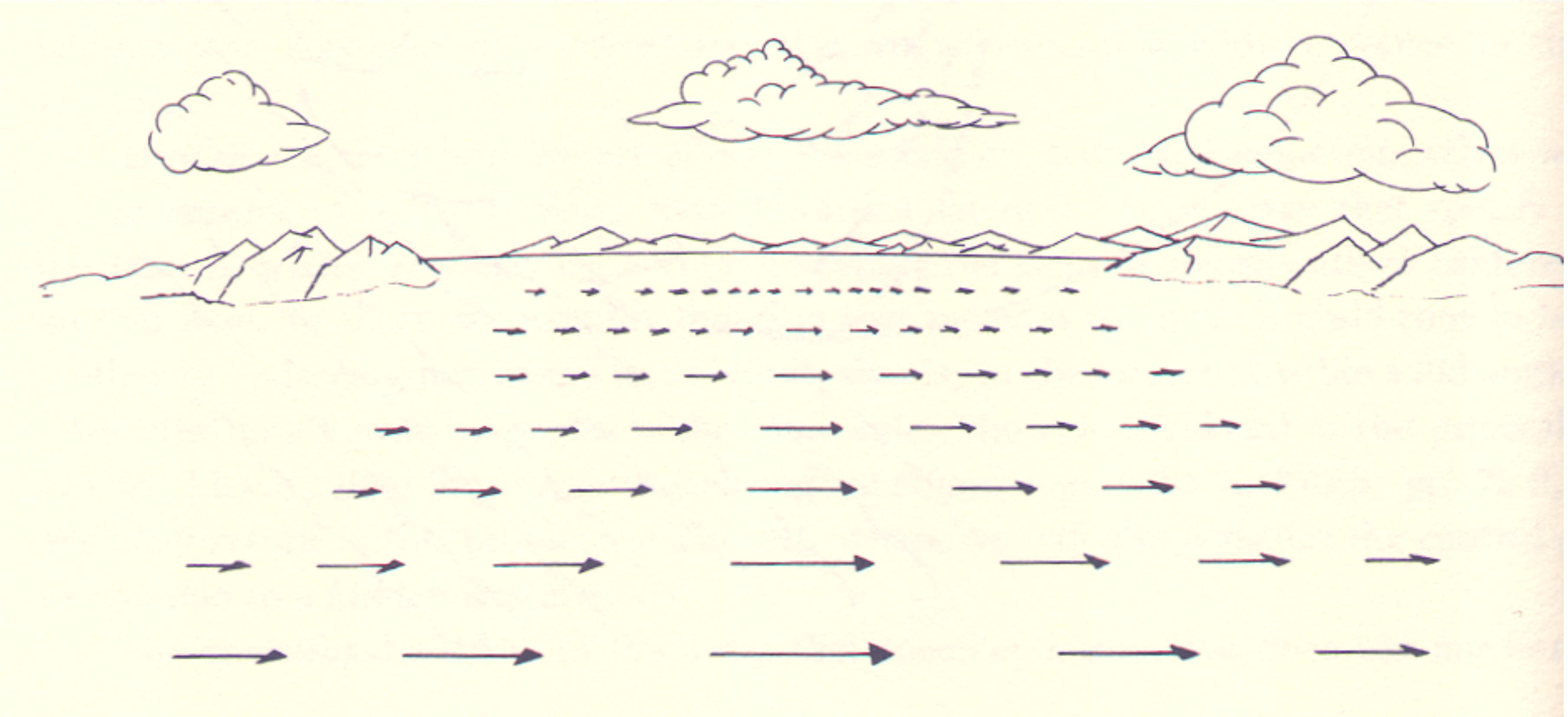
Videos - Optical flow



J. J. Gibson



Optical flow due to camera motion

- Consider camera translating and rotating

$$\mathbf{P} = (X, Y, Z)^T$$

$$x = \frac{X}{Z} \qquad y = \frac{Y}{Z}$$

$$\dot{\mathbf{P}} = -\mathbf{t} - \omega \times \mathbf{P}$$

Optical flow due to camera motion

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} \dot{x} \\ \dot{y} \end{bmatrix} = \frac{1}{Z}\begin{bmatrix} -1 & 0 & x \\ 0 & -1 & y \end{bmatrix}\begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix} + \begin{bmatrix} xy & -(1+x^2) & y \\ 1+y^2 & -xy & -x \end{bmatrix}\begin{bmatrix} \omega_x \\ \omega_y \\ \omega_z \end{bmatrix}$$

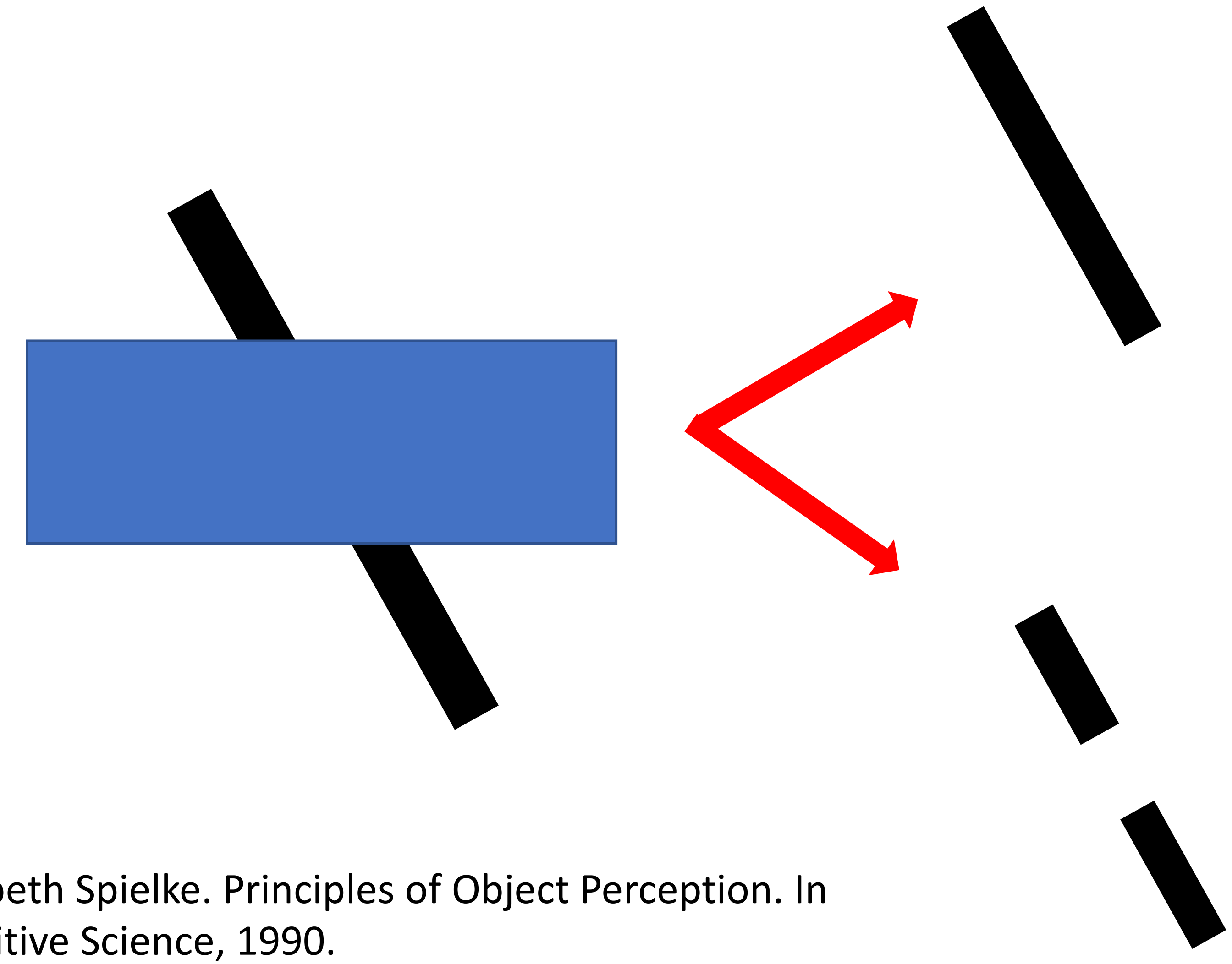
Optical flow for moving scenes

Optical flow for moving scenes

Optical flow for moving scenes

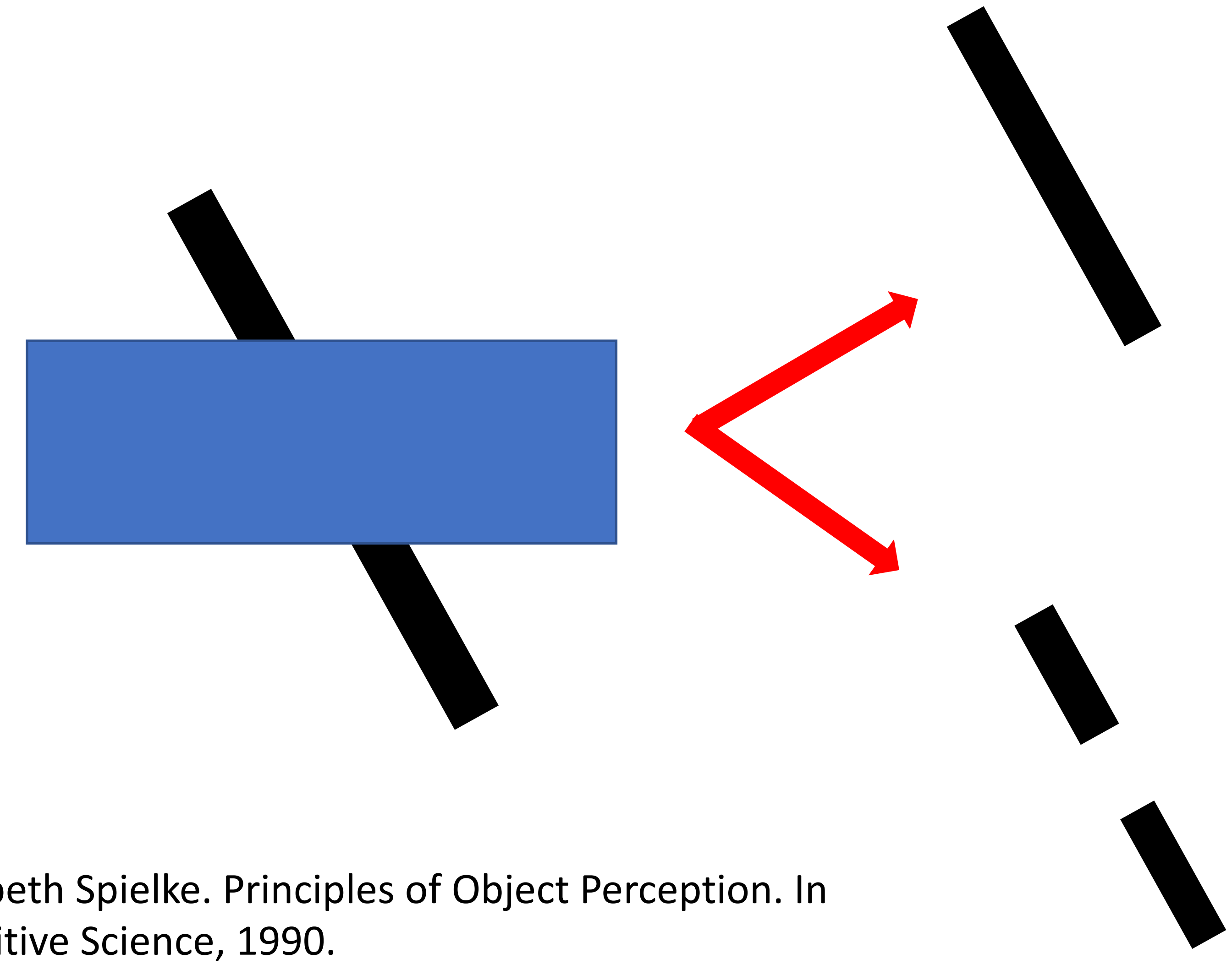
- Optical flow helps *grouping*
- *Gestalt principle of common fate*
  - *Things that move together belong together*

Motion segmentation in humans



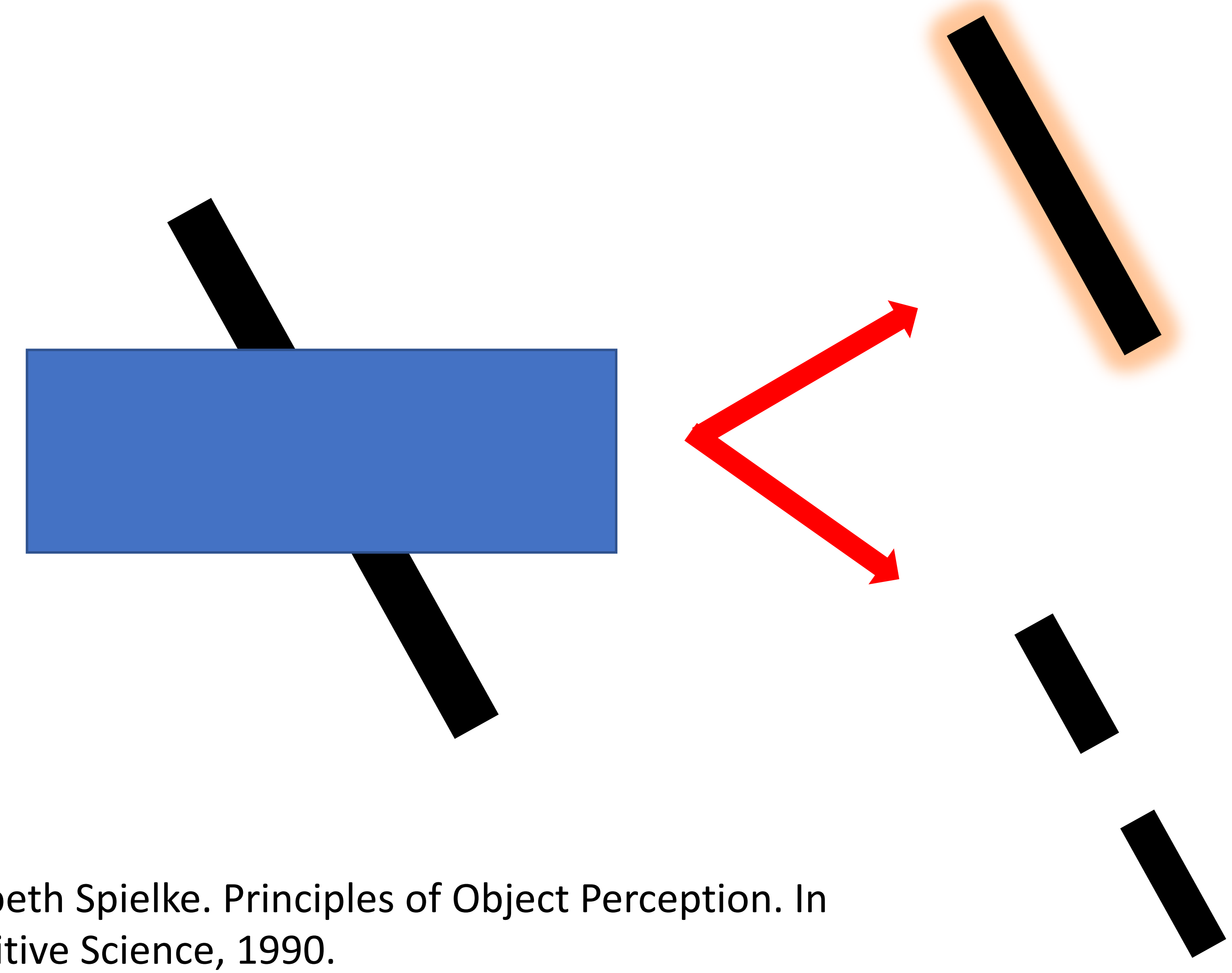
Elizabeth Spielke. Principles of Object Perception. In Cognitive Science, 1990.

Motion segmentation in humans

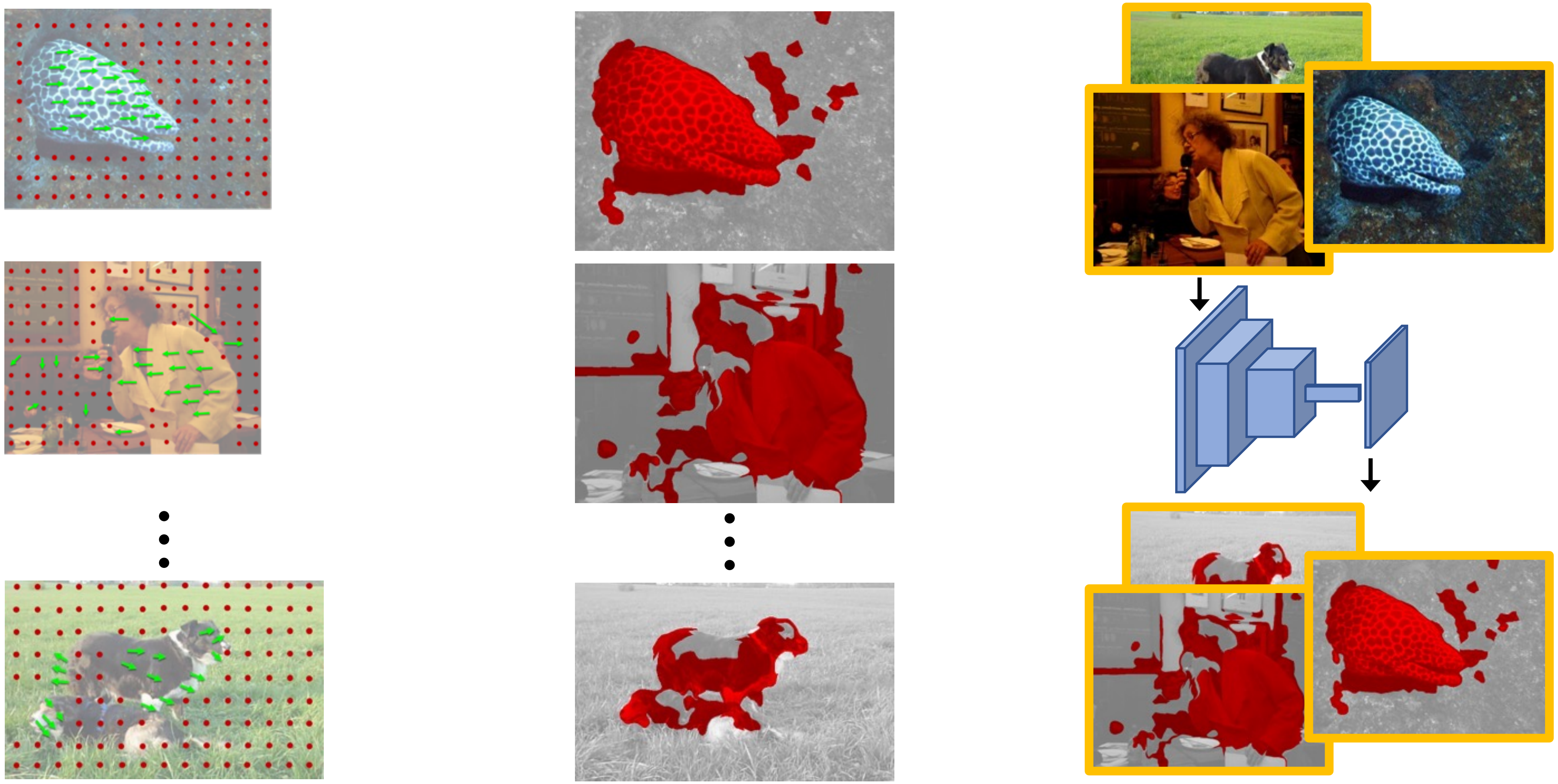


Elizabeth Spielke. Principles of Object Perception. In Cognitive Science, 1990.

Motion segmentation in humans



Elizabeth Spielke. Principles of Object Perception. In Cognitive Science, 1990.



1. Collect videos

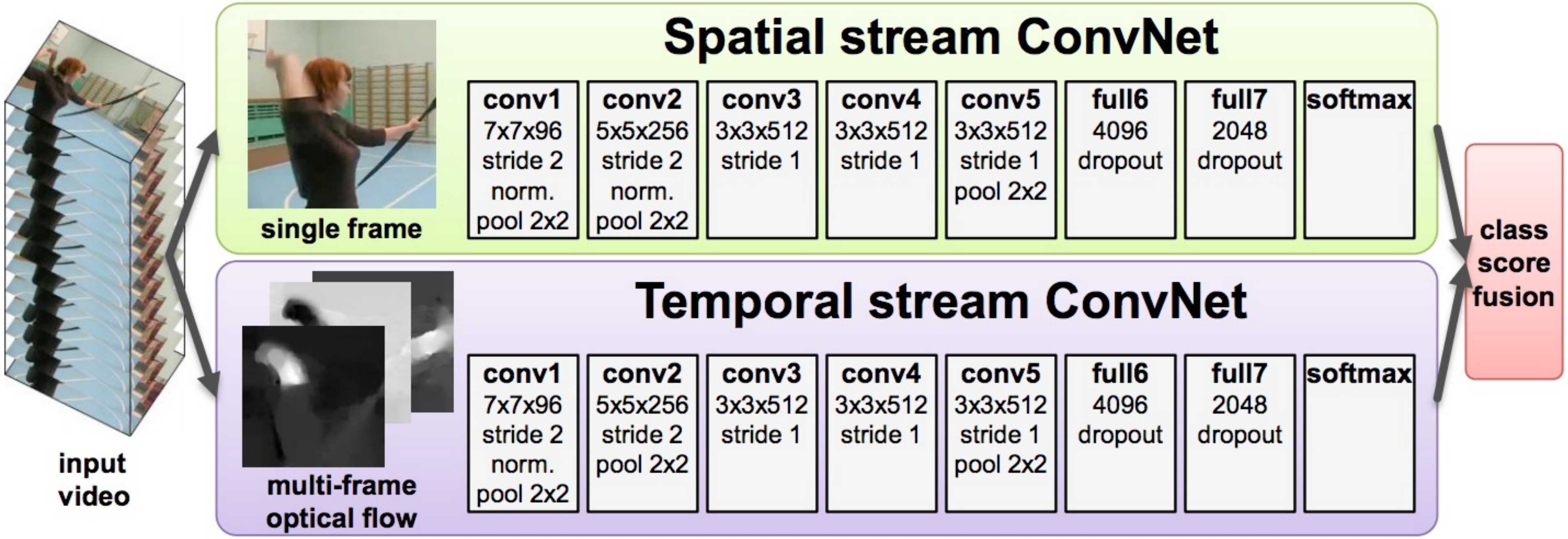
2. Segment using motion

3. Train ConvNet

Learning Features by Watching Objects Move. D. Pathak, R. Girshick, P. Dollar, T. Darrell, B. Hariharan. *CVPR,* 2017.

Optical flow for moving scenes

- Motion is cue for recognition
  - Gestures, actions, …



Two-Stream Convolutional Networks for Action Recognition in Videos. Simonyan and Zisserman. In *NIPS 2014.*

Optical flow for moving scenes

- Motion is cue for recognition
  - Gestures, actions, …

Model	Accuracy
Without optical flow	73.0%
With optical flow	**88.0%**

Two-Stream Convolutional Networks for Action Recognition in Videos. Simonyan and Zisserman. In *NIPS 2014.*

Estimating optical flow

- Yet another correspondence problem!
- But:
  - Bad: scene can move
  - Good: changes are usually very small (often sub-pixel)

Optical flow constraint equation

- Image intensity *continuous* function of x, y, t
- In time dt, pixel (x,y,t) moves to (x + u dt, y + v dt, t + dt)

$$\min_{u,v}(I(x+u\Delta t, y+v\Delta t, t+\Delta t) - I(x,y,t))^2$$

$$\equiv \min_{u,v}(I(x,y,t) + I_x u\Delta t + I_y v\Delta t + I_t\Delta t - I(x,y,t))^2$$

$$\equiv \min_{u,v}(I_x u\Delta t + I_y v\Delta t + I_t\Delta t)^2$$

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

- Optical flow constraint equation: One equation, two variables

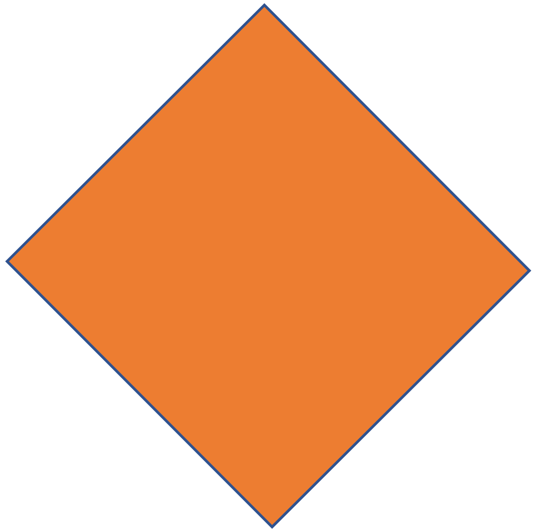
Lucas-Kanade

- Assume all pixels in patch have the same flow
- When will this produce a unique solution?

$$\begin{pmatrix} \nabla I(x_1, y_1)^T \\ \vdots \\ \nabla I(x_n, y_n)^T \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} = - \begin{pmatrix} I_t(x_1, y_1) \\ \vdots \\ I_t(x_n, y_n) \end{pmatrix}$$

Aperture problem

Aperture problem



Lucas-Kanade

$$\begin{pmatrix} \nabla I(x_1, y_1)^T \\ \vdots \\ \nabla I(x_n, y_n)^T \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} = - \begin{pmatrix} I_t(x_1, y_1) \\ \vdots \\ I_t(x_n, y_n) \end{pmatrix}$$

- Equation of the form Ax = b
- Solve using Normal equations: $x = (A^T A)^{-1} A^T b$
- Need $A^T A$ to be invertible - corners!

Lucas-Kanade

- What if we consider the whole image as one patch?
  - Constant optical flow for the entire image?
- Better: what if we consider flow as a *parametric function* of pixel location?
  - e.g. affine $\begin{bmatrix} u \\ v \end{bmatrix} = A\mathbf{x} + b$
  - More generally: $\mathbf{x}' = W(\mathbf{x}; \mathbf{p})$
    - W is some 2D → 2D parametric warp function
    - **p** is a parameter vector
  - “Motion models”

Lucas-Kanade

$$\min_{\mathbf{p}} \sum_{\mathbf{x}} (I(W(\mathbf{x};\mathbf{p})) - T(\mathbf{x}))^2$$

- T is the previous frame, also called *template*
- I is the current frame
- Goal is to find **p**

Baker, Simon, and Iain Matthews. "Lucas-kanade 20 years on: A unifying framework." *International journal of computer vision* 56.3 (2004): 221-255.

Lucas-Kanade

- Iterative process
- Assume that we have a current iterate $\boldsymbol{p}$ and we want to find the next iterate $\boldsymbol{p} + \Delta\boldsymbol{p}$
- Find $\Delta\boldsymbol{p}$ by optimizing $\min_{\Delta\mathbf{p}} \sum_{\mathbf{x}} (I(W(\mathbf{x};\mathbf{p}+\Delta\mathbf{p})) - T(\mathbf{x}))^2$
- Hard because I and W are both non-linear
- Assume $\Delta\boldsymbol{p}$ is small and linearize:
  - Linearize W: $W(\mathbf{x};\mathbf{p}+\Delta\mathbf{p}) \approx W(\mathbf{x};\mathbf{p}) + \frac{\partial W}{\partial \mathbf{p}}\Delta\mathbf{p}$
  - Linearize I: $I(W(\mathbf{x};\mathbf{p}+\Delta\mathbf{p})) \approx I(W(\mathbf{x};\mathbf{p})) + \nabla I \frac{\partial W}{\partial \mathbf{p}}\Delta\mathbf{p}$

Baker, Simon, and Iain Matthews. "Lucas-kanade 20 years on: A unifying framework." *International journal of computer vision* 56.3 (2004): 221-255.

Lucas Kanade

- Iterative process
- At each step, find $\Delta \boldsymbol{p}$ that optimizes

$$\min_{\Delta \mathbf{p}} \left(I(W(\mathbf{x};\mathbf{p})) + \nabla I \frac{\partial W}{\partial \mathbf{p}} \Delta \mathbf{p} - T(\mathbf{x}) \right)^2$$

  - Warped image
  - Gradient of warped image
  - Jacobian of warp function
  - Template

- Quadratic in $\Delta \boldsymbol{p}$, solve exactly

Lucas-Kanade

- Solve by iterating on parameters
- Equivalent to Newton iteration + linearization
- Can we remove the parametric assumption?

Baker, Simon, and Iain Matthews. "Lucas-kanade 20 years on: A unifying framework." *International journal of computer vision* 56.3 (2004): 221-255.

Horn-Schunk

$$E(\mathbf{u},\mathbf{v}) = E_{data}(\mathbf{u},\mathbf{v}) + E_{smoothness}(\mathbf{u},\mathbf{v})$$

$$E(\mathbf{u},\mathbf{v}) = \int\int (I(x+u(x,y)\Delta t, y+v(x,y)\Delta t, t+\Delta t) - I(x,y,t))^2$$

Data

$$+\alpha(\|\nabla u\|^2 + \|\nabla v\|^2)dxdy$$

Smoothness

Horn-Schunk

$$E(\mathbf{u},\mathbf{v}) = E_{data}(\mathbf{u},\mathbf{v}) + E_{smoothness}(\mathbf{u},\mathbf{v})$$

$$E(\mathbf{u},\mathbf{v}) = \int\int (I_x u + I_y v + I_t)^2 + \alpha(\|\nabla u\|^2 + \|\nabla v\|^2) dx dy$$

Data

Smoothness

Variational minimization

- u and v are *functions*
- Euler-lagrange equations
  - Similar to “gradient=0”

$$\min_q \int L(t, q(t), \dot{q}(dt))dt$$

$$\frac{\partial L}{\partial q} - \frac{d}{dt}\frac{\partial L}{\partial \dot{q}} = 0$$

Variational minimization

$$\min_q \int L(t, q(t), \dot{q}(dt))dt$$

$$\frac{\partial L}{\partial q} - \frac{d}{dt}\frac{\partial L}{\partial \dot{q}} = 0$$

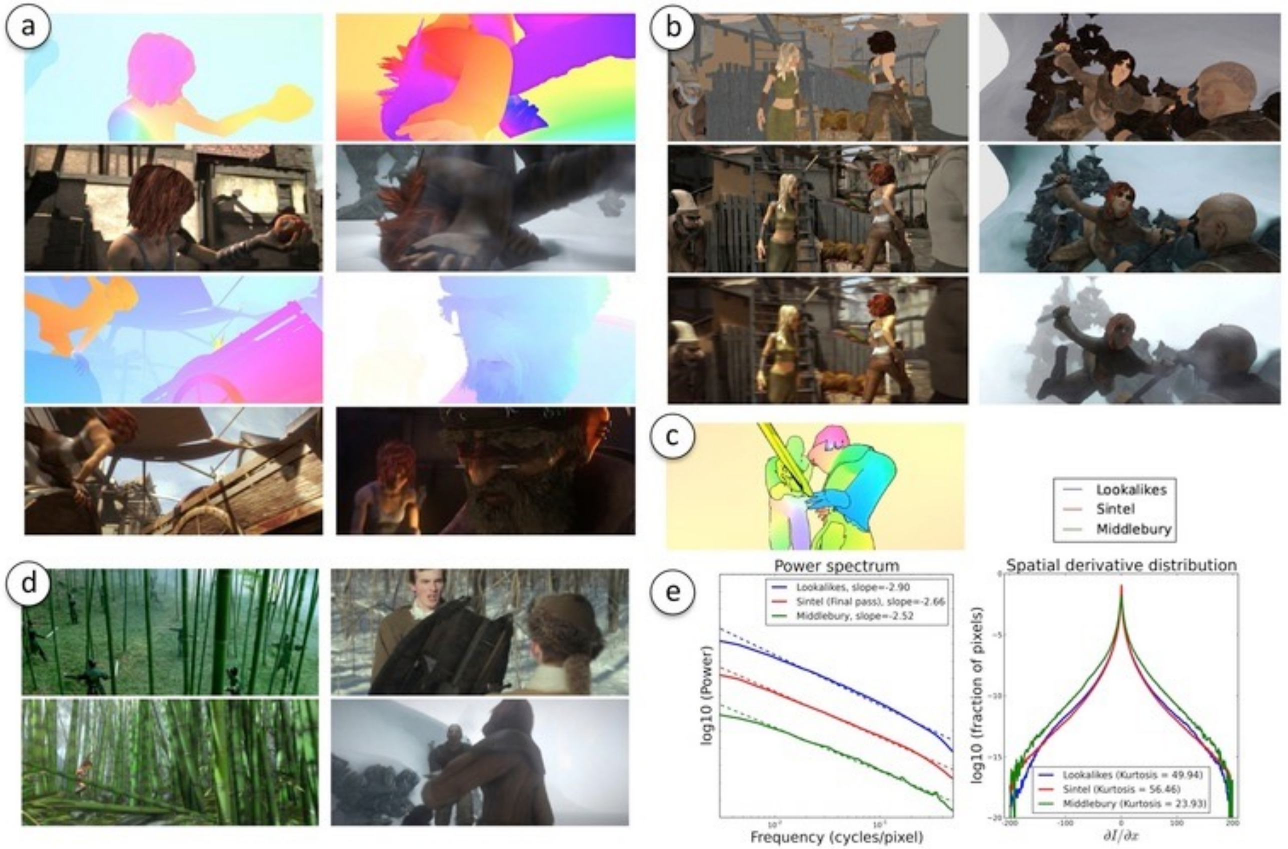
$$\min_{u,v} \int\int f(x, y, u, v, u_x, u_y, v_x, v_y)dxdy$$

$$\frac{\partial f}{\partial u} - \frac{d}{dx}\frac{\partial f}{\partial u_x} - \frac{d}{dy}\frac{\partial f}{\partial u_y} = 0$$

$$\frac{\partial f}{\partial v} - \frac{d}{dx}\frac{\partial f}{\partial v_x} - \frac{d}{dy}\frac{\partial f}{\partial v_y} = 0$$

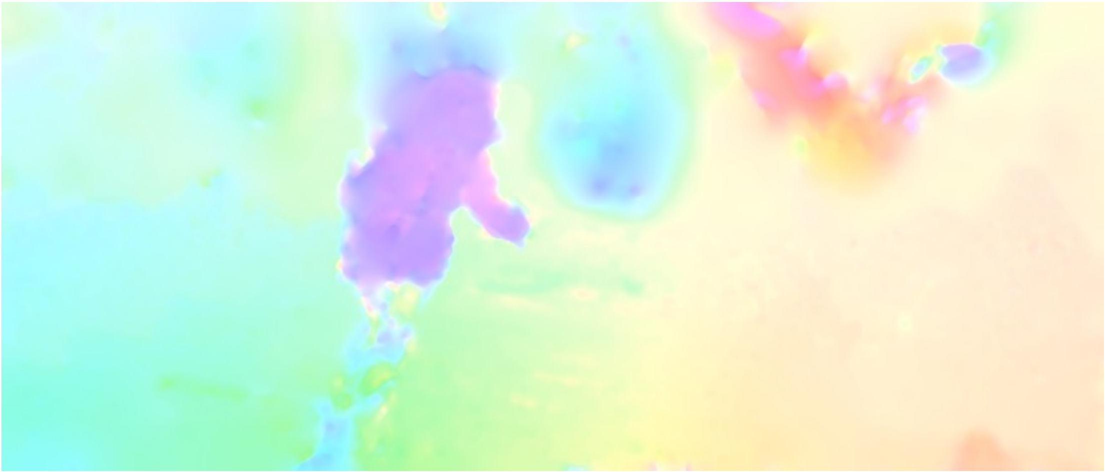
MPI-Sintel

- Open-source animated movie "Sintel"
- "Naturalistic" video
- Ground truth optical flow
- Large motions
- Complex scenes



Butler, D. J. and Wulff, J. and Stanley, G. B. and Black, M. J. A naturalistic open source movie for optical flow evaluation. *ECCV,* 2012

MPI-Sintel results

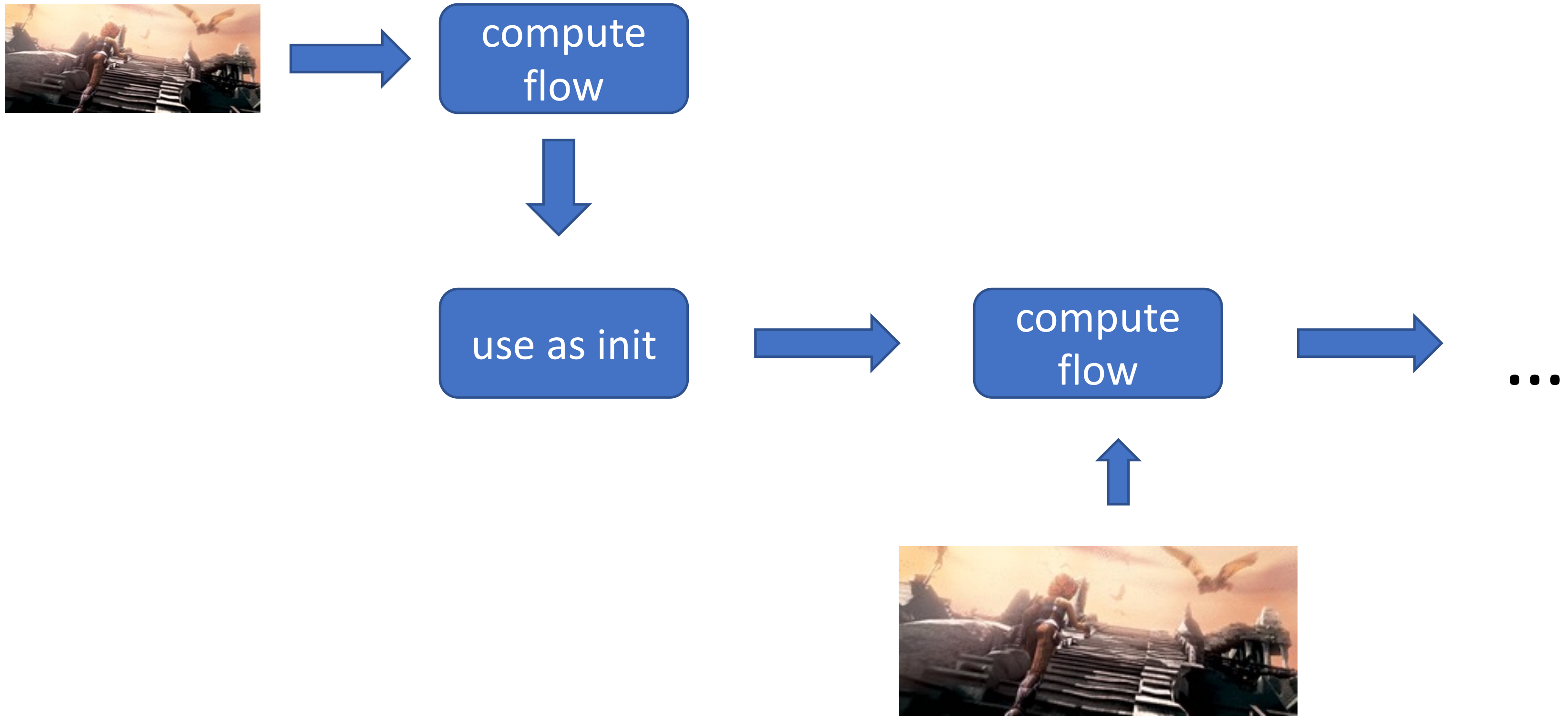


Optical flow with large displacements

- Optical flow constraint equation assumes differential optical flow
- “Large displacement”?
- Key idea: reducing resolution reduces displacement
- Reduce resolution, then upsample?
  - will lose fine details

Optical flow with large displacements

- Key idea 2: Use upsampled flow as *initialization*
- *Changes to initialization will be infinitesimal*



Brox, Thomas, et al. "High accuracy optical flow estimation based on a theory for warping." *Computer Vision-ECCV 2004* (2004)

Optical flow for large displacements

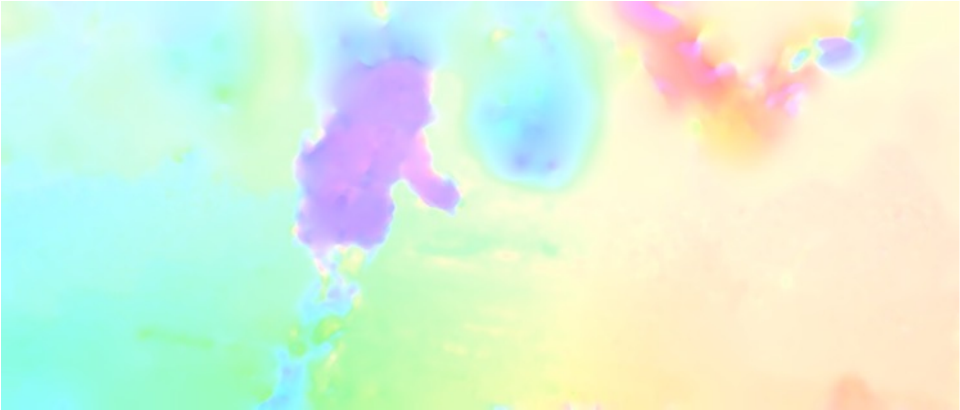
- Possible issue: large appearance change => incorrect matching based on color alone
- Use descriptor matching (e.g. SIFT) on sparse points
- Use smoothness to propagate to all pixels:
  - Flow is weighted average of nearest neighbors: $F(p) = \frac{\sum_q k(p,q)F(q)}{\sum_q k(p,q)}$
  - *Kernel k* dependent on relative position + edges
- Use this as initialization for optimization

Brox, Thomas, Christoph Bregler, and Jitendra Malik. "Large displacement optical flow." *CVPR,* 2009.

Revaud, Jerome, et al. "Epicflow: Edge-preserving interpolation of correspondences for optical flow." CVPR. 2015.

LDOF and EpicFlow

Video



Basic Horn-Schunk (Error = 2.069)

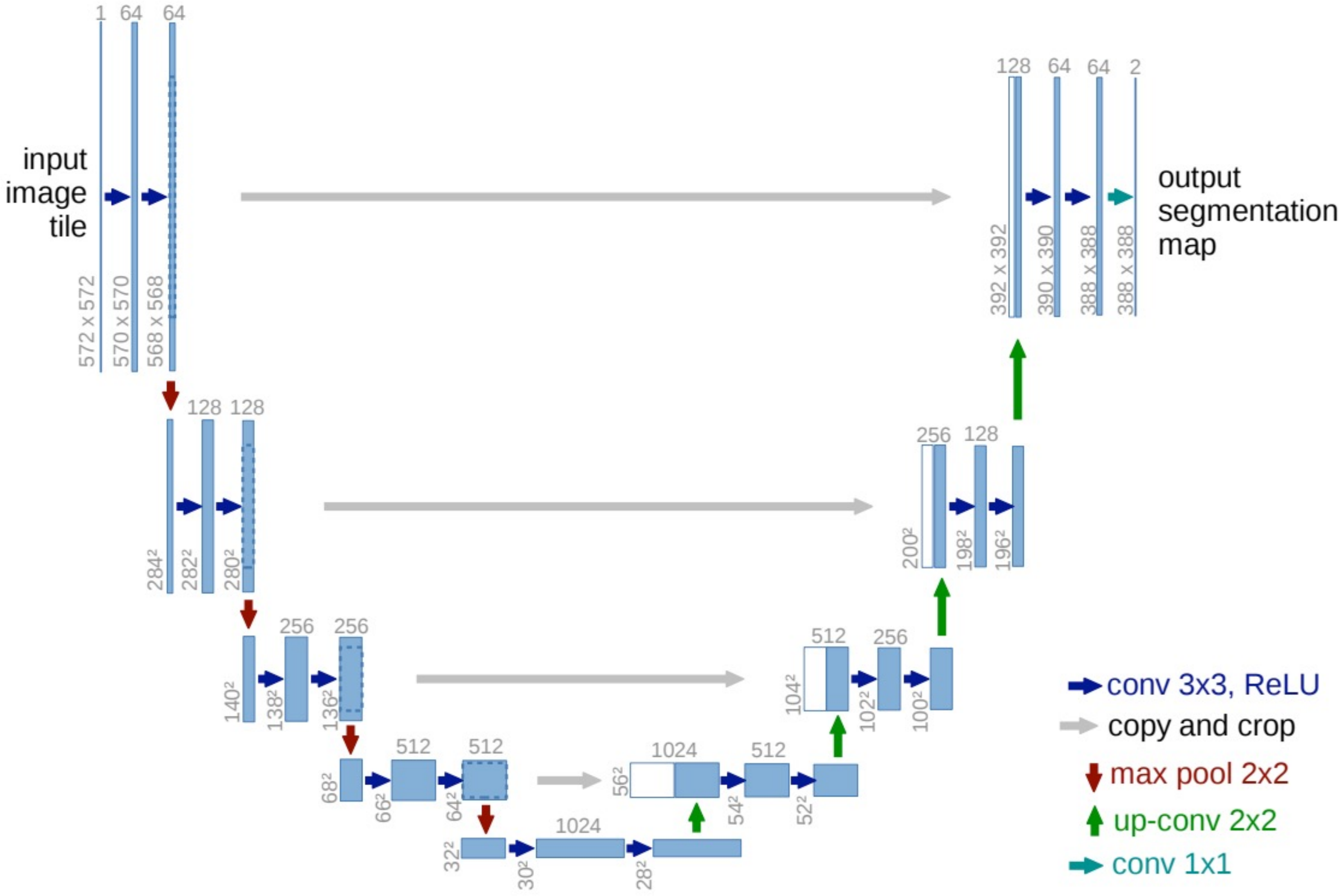
LDOF (Brox et al, 2009) (Error = 1.606)

EpicFlow (Revaud et al, 2015) (Error = 1.295)

Coarse-to-fine processing

- A specific instance of a general idea
  - Also coarse-to-fine versions of Lucas-Kanade
- Coarse scales:
  - Global / large structures
  - Long-range relationships
  - But: imprecise localization
- Fine scales:
  - Precise localization
  - But: aperture problem
- Idea: start from coarse scales, add fine scale detail

Coarse-to-fine processing

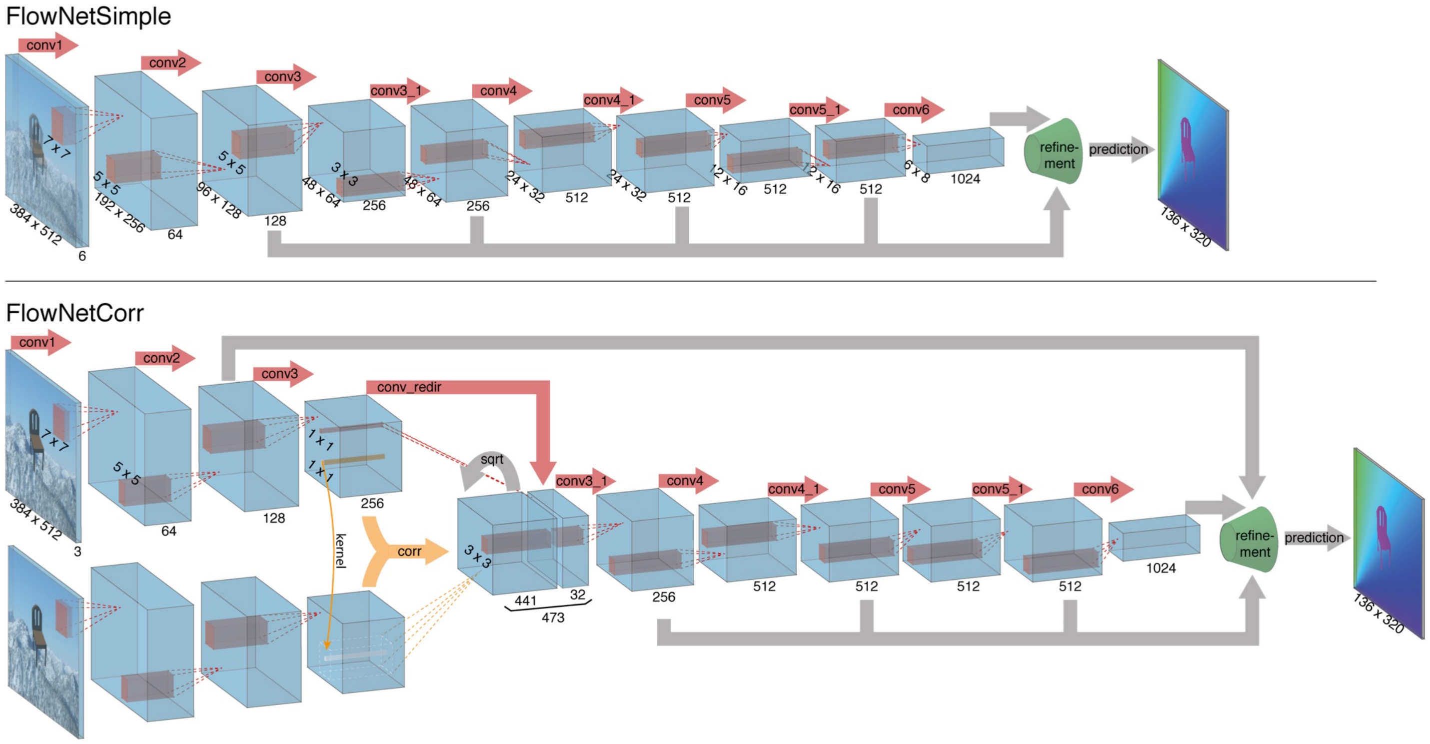


U-Net: Convolutional Networks for Biomedical Image Segmentation. Olaf Ronneberger, Philipp Fischer, and Thomas Brox. In *MICCAI,* 2015.

Dosovitskiy, Alexey, et al. "Flownet: Learning optical flow with convolutional networks." *Proceedings of the IEEE international conference on computer vision*. 2015.

Learning optical flow

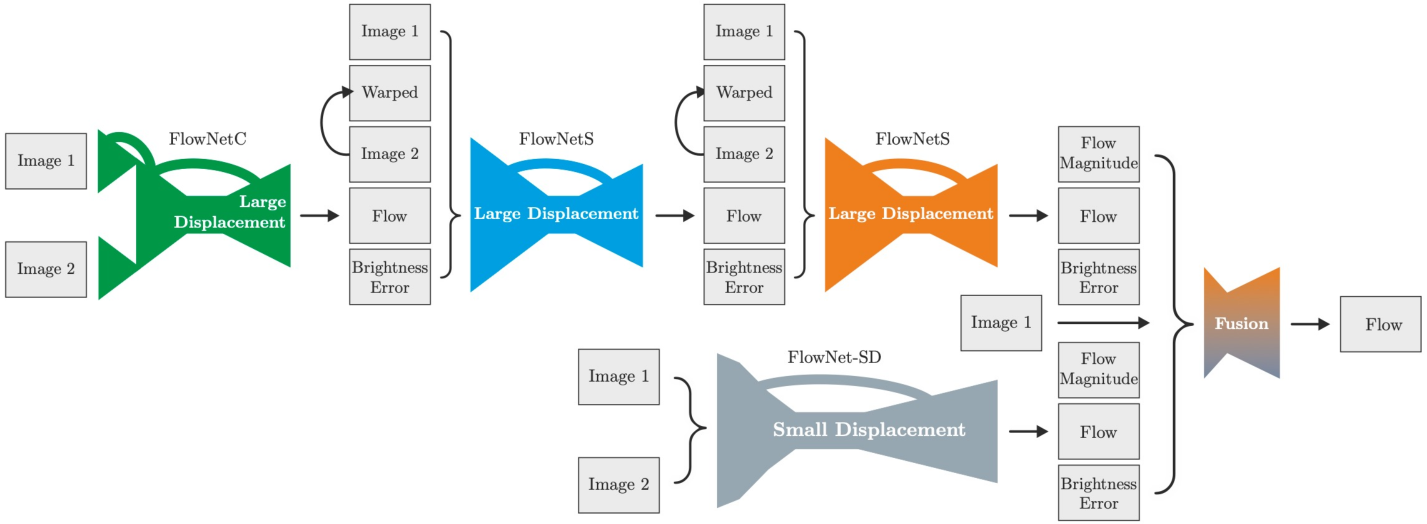
- Training data?
  - Synthetic
- Architecture?
  - Based on segmentation
  - Based on correlating pairs of pixels



Learning optical flow

Ilg, Eddy, et al. "Flownet 2.0: Evolution of optical flow estimation with deep networks." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017.

- Better flow through coarse-to-fine refinement



Learning optical flow

Teed, Zachary, and Jia Deng. "Raft: Recurrent all-pairs field transforms for optical flow." *European conference on computer vision*. Springer, Cham, 2020.

- Have learning architecture mimic actual computation?
- Should iteratively optimize flow
- Should use data term (pairwise similarity of pixels) and prior term