Video Analysis

LV 188.329, WS 2016/17

Optical Flow Estimation

Ao. Univ.-Prof. Margrit Gelautz

margrit.gelautz@tuwien.ac.at

FAKULTÄT
FÜR !NFORMATIK

Faculty of Informatics

Acknowledgement

These slides were developed by Nicole Brosch in previous years. We wish to thank Nicole for making the slides available for further use.

Overview

- **Basics**
  - Motion Field, Applications, Key Problems, Motion Estimation Techniques
- **Matching**
  - Block Matching, Brightness Constancy, 2D Motion vs. Optical Flow
- **Gradient-based approaches**
  - Optical Flow Equation, Aperture Problem, Lucas-Kanade, Horn-Schunck
- **Challenges**
  - Large Motion, Illumination Changes, ...
- **Evaluation**
  - Ground Truth, Middlebury Data Set

Basics

- **Goal of Optical Flow Estimation**

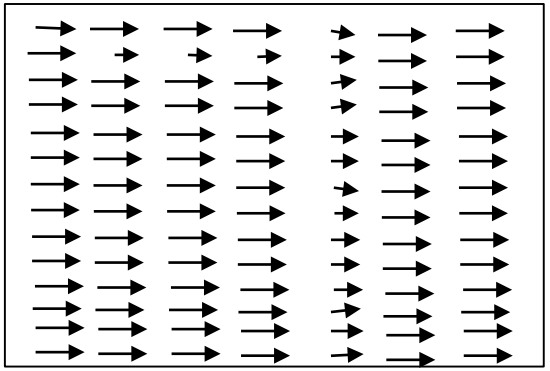
  The goal of OF estimation is to compute an approximation of the *motion field* (optical flow field) from a video sequence.

- **Motion Field**

  A vector field, which vectors indicate *direction* (orientation) and *velocity* (length) of movement.

Frame 1

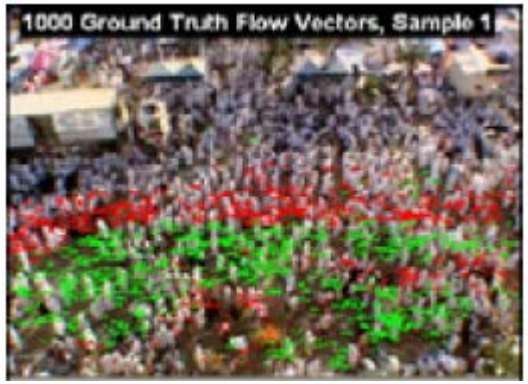
Frame 2



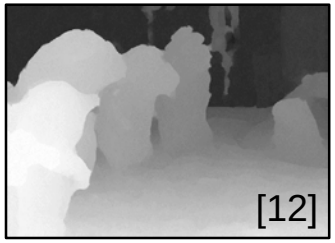
Motion Field

Applications

- **Why estimate visual motion?**
  - Motion detection (e.g., surveillance)
  - Video understanding (e.g., video segmentation)
  - Compositing (e.g., retiming)
  - Interactive 2D-to-3D conversion
  - Video compression (e.g., MPEG)
  - Obstacle avoidance
  - Image registration
  - …



[8]


[12]

[7]

Key Problems

- **Motion representation**
  - E.g. Pixel, sub-pixel
  - E.g. Regions, blocks, segments, …
- **Choose motion estimation criteria & formulate it as function**
  - E.g. Intensities / colors are conserved
  - E.g. Intensity / color changes are conserved
  - E.g. Smooth vector fields
- **Minimize function with an optimization method**
  - Exhaustive search / continuous / discrete / variational
  - Depends on function

Motion estimation techniques

- **Block- / Feature- Matching**

  simply *searches* for the block/patch/feature which fits best.

- **Gradient-based methods**

  formulates equation(s) and *solves* for the estimates.

- **Variational Methods**
- **Spatiotemporal Methods**
- **Segmentation-based Methods**
- **Probabilistic Methods**
- **Learning-based Methods**
- **...**

Block- / Feature-Matching

Block Matching

- **Representation:** Blocks centred at a pixel
- **Constraint**: Intensities are shifted -> values are conserved
- **Optimization:** Exhaustive search of block with minimal difference

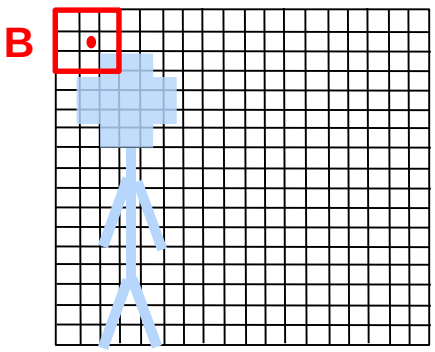
Brightness Constancy Assumption

costs, which should be minimized

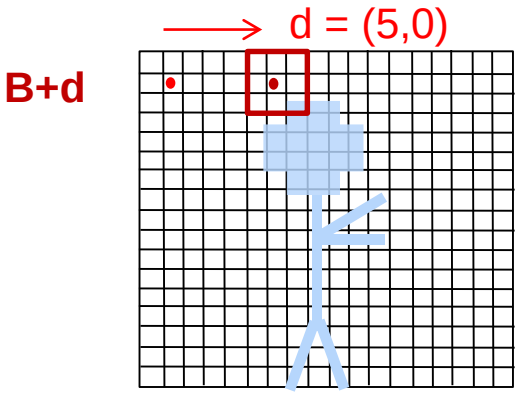
$$E(d) = \sum_{p \in B} c(p, p+d)$$

c(.) .... Similarity measure e.g., squared differences

d ... OF vector
B .. block
p .. pixel



Frame 0



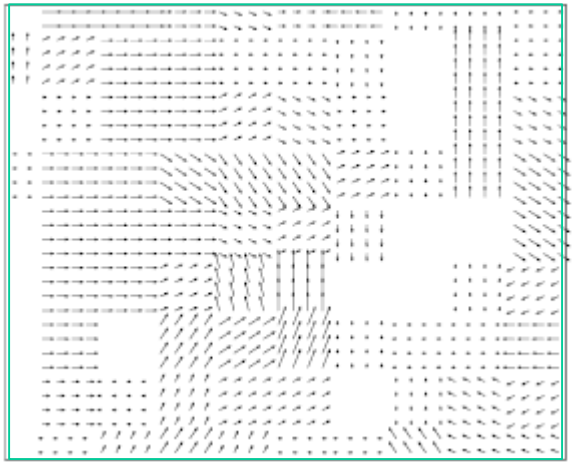
Frame 1

Block Matching

- **Advantages:**
  - Easy :)

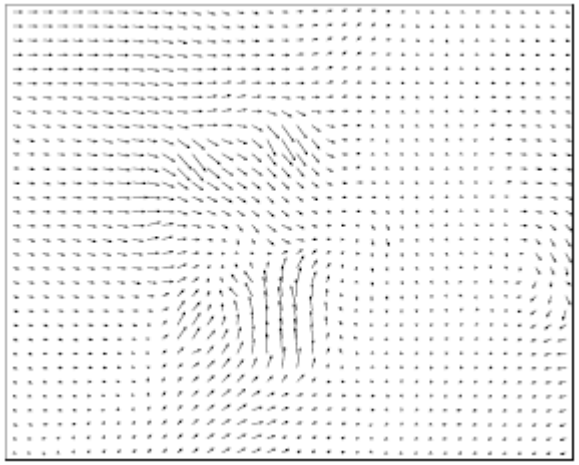
- **Disadvantages:**
  - Block artifacts
  - Edges inside a block are smoothed
  - Not accurate
  - Not efficient (define search windows, ...)
  - Dependent of block size
  - Problems caused by the brightness constancy assumption.

Block Matching



VS.

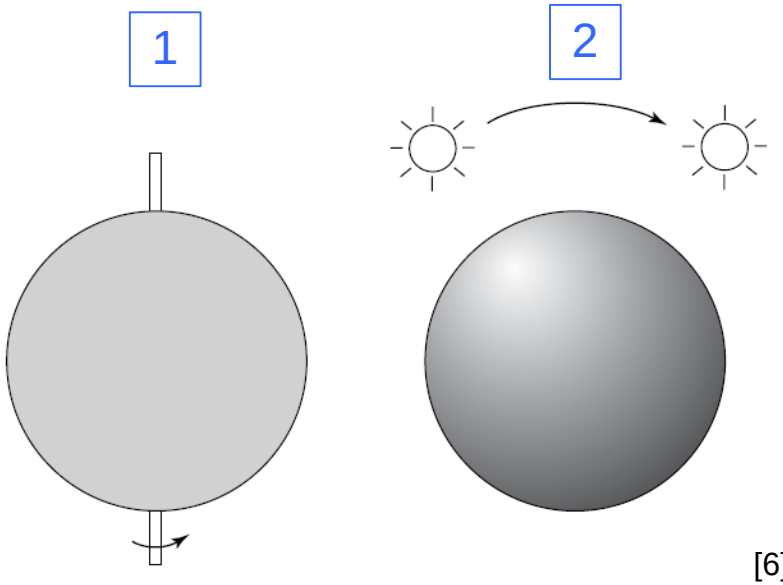
Gradient-based



Consequences of Brightness Constancy

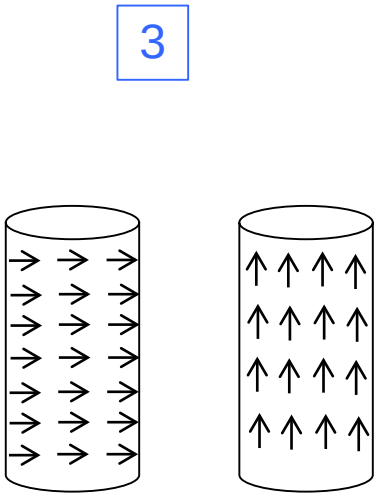
- **2D motion** is the projection of 3D motion. It depends on 3D object motion and projection operator.

- **Optical Flow** is the pattern of *apparent motion* of objects, surfaces and edges in a visual scene caused by the relative motion between a camera and the scene. It depends on illumination and object surface texture.



[6]

Barber‘s pole



motion field

optical flow

Gradient-Based Approaches

Optical Flow Equation (OFE)

Brightness (Color) Constancy: Pixel intensities (colors) don't change!

I … intensity
(x,y) … pixel
(u,v) .. flow vector
t … time

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

***u* and *v* are small, *I* is smooth -> linearize with Taylor expansion:**

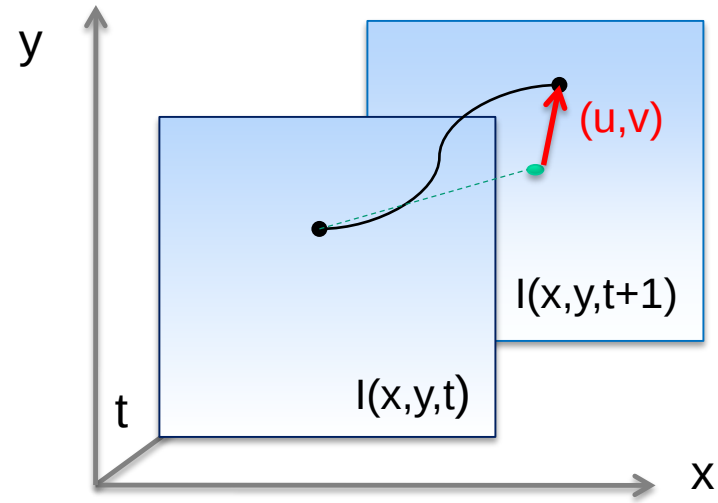
$$I(x+u, y+v, t+1) = I(x,y,t) + uI'_x(x,y,t) + vI'_y(x,y,t) + I'_t(x,y,t)$$

$$\cancel{I(x,y,t)} = \cancel{I(x,y,t)} + uI'_x(x,y,t) + vI'_y(x,y,t) + I'_t(x,y,t)$$

$$0 = uI'_x(x,y,t) + vI'_y(x,y,t) + I'_t(x,y,t)$$

gradient constraint equation

(a.k.a. brightness constancy constraint equation)
(a.k.a. image brightness constancy equation)
(a.k.a. intensity flow equation)

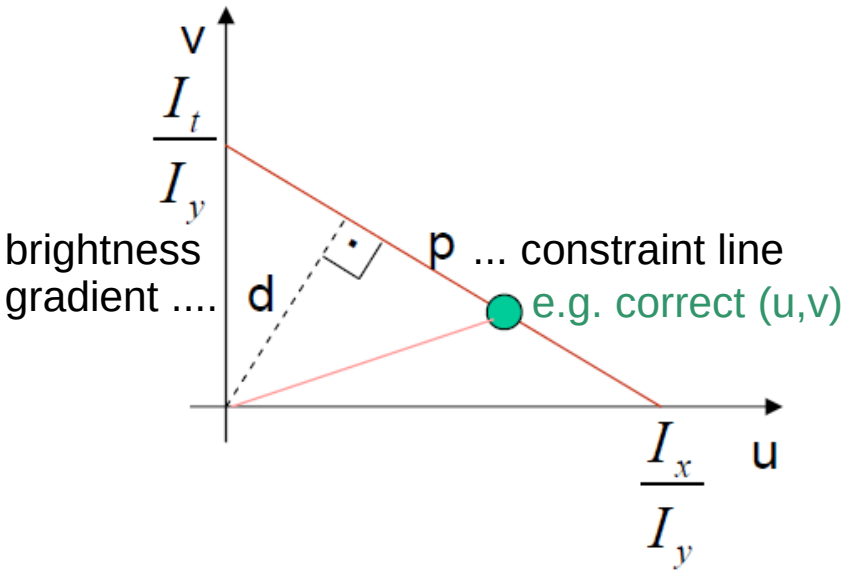


Aperture Problem

- For each pixel *(x,y)* we have one equation with two unknowns *d=(u,v).*

$$0 = uI'_x + vI'_y + I'_t$$

$$-v = u\frac{I'_x}{I'_y} + \frac{I'_t}{I'_y}$$



Correct solution lies on the constraint line p

- Regions with constant brightness -> flow is *indeterminate*!
- **To solve both unknowns we need additional constrains or have to include the neighbourhood!**

Lucas-Kanade

- **Basic Idea:** Constrain velocity locally using gradient constraints from nearby pixels

$$0 = ul'_{x1} + vl'_{y1} + l'_t$$
$$0 = ul'_{x} + vl'_{y} + l'_t$$
$$\dots$$
$$0 = ul'_{x2} + vl'_{y2} + l'_t$$

x1,y1	x,y1	x2,y1
x1,y	**x,y**	x2,y
x1,y2	x,y2	x2,y2

-> *overdetermined* system of equations,

can be solved by *Least-Squares Estimation* (LSE)

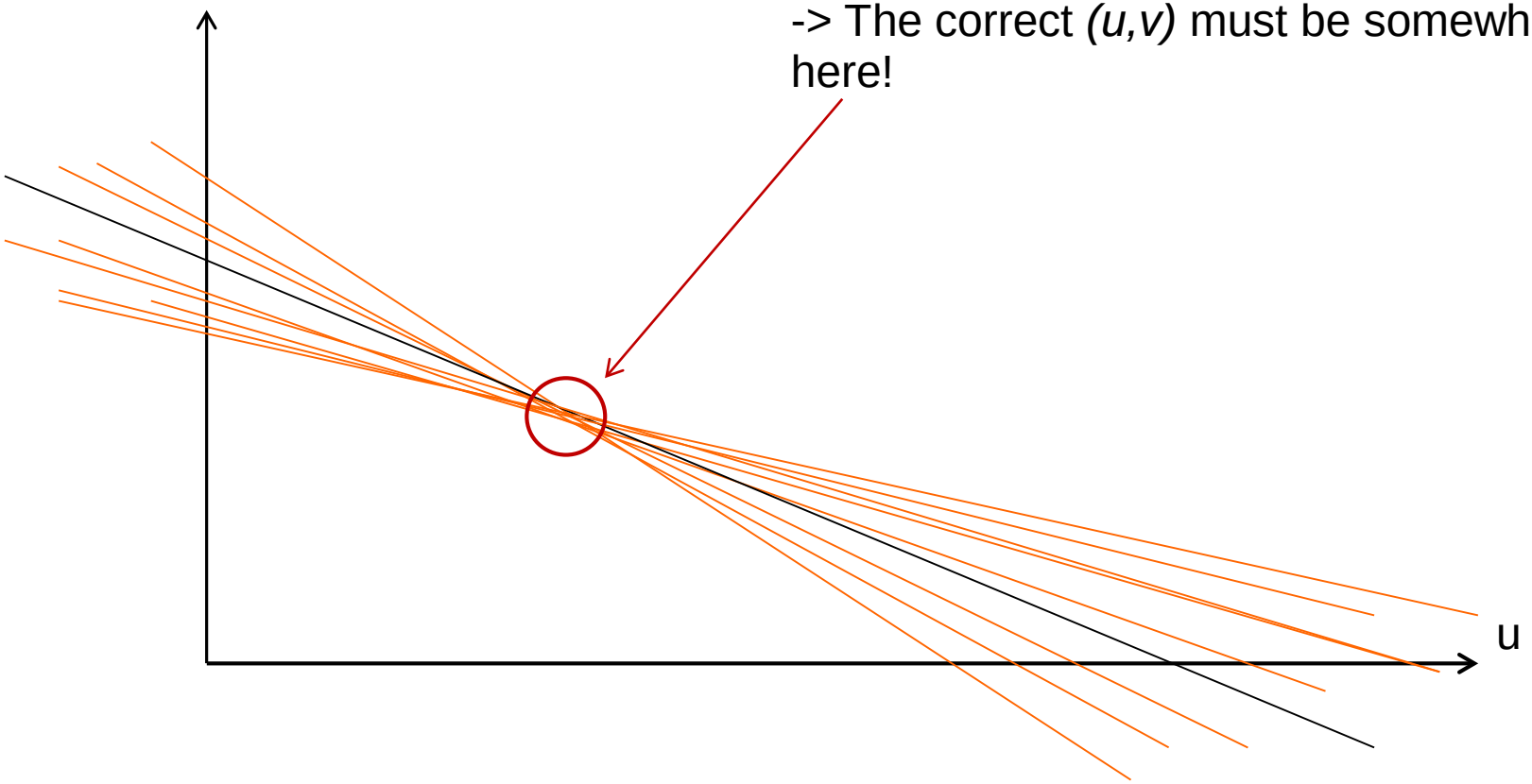
- **Energy function** which is minimized by using LSE:

$$E(u,v) = \sum (ul'_x + vl'_y + l'_t)^2$$

Lucas-Kanade

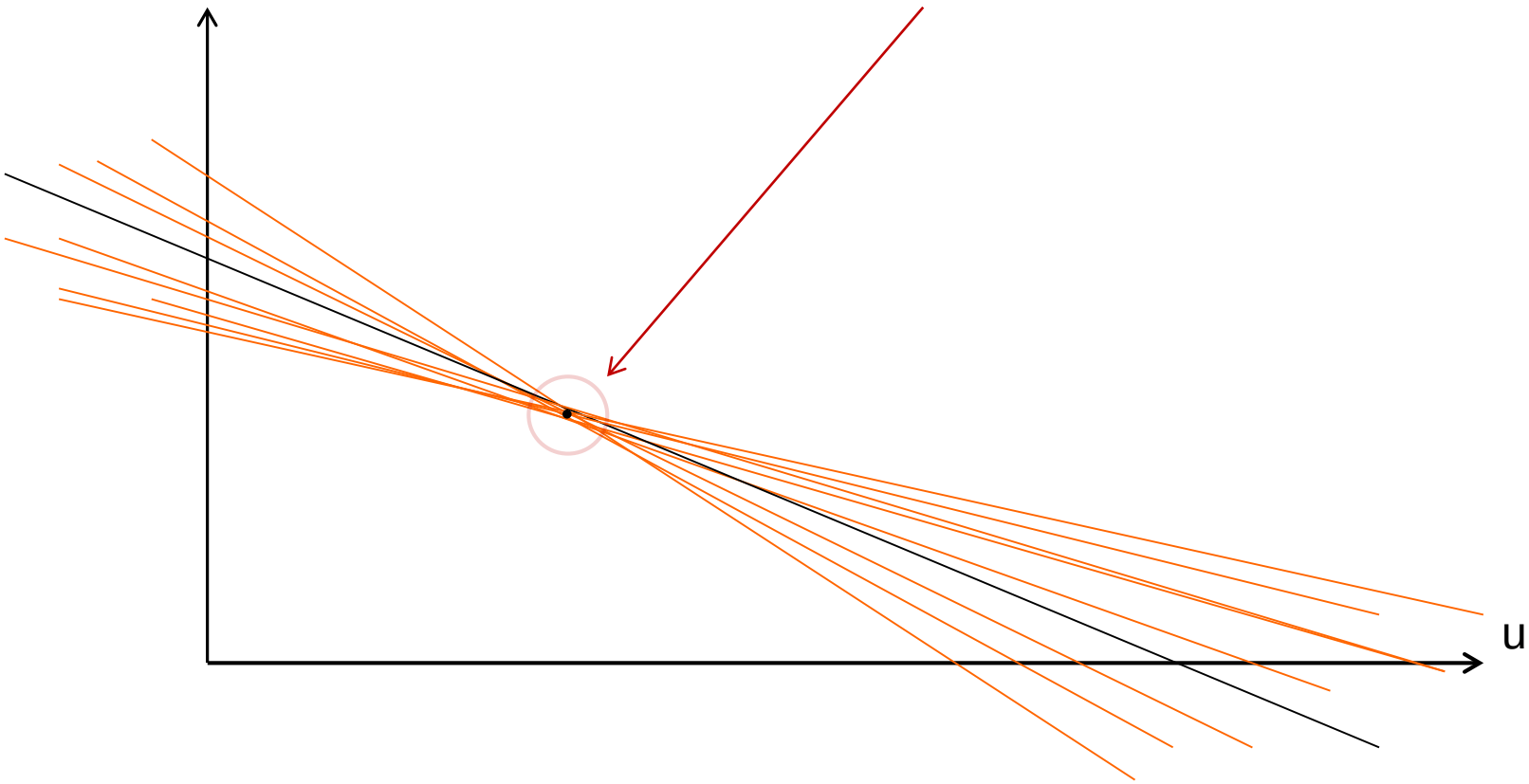
Constraint lines from nearby pixels cluster around a point in this area.

-> The correct *(u,v)* must be somewhere here!



Lucas-Kanade

LSE finds the ‚optimal‘ interception point, concerning the distances from all constraint lines.

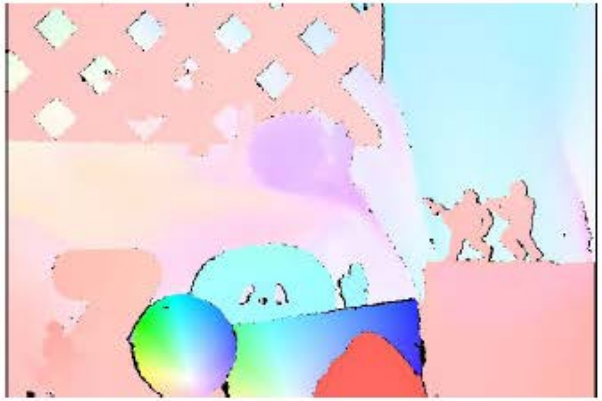


Lucas-Kanade

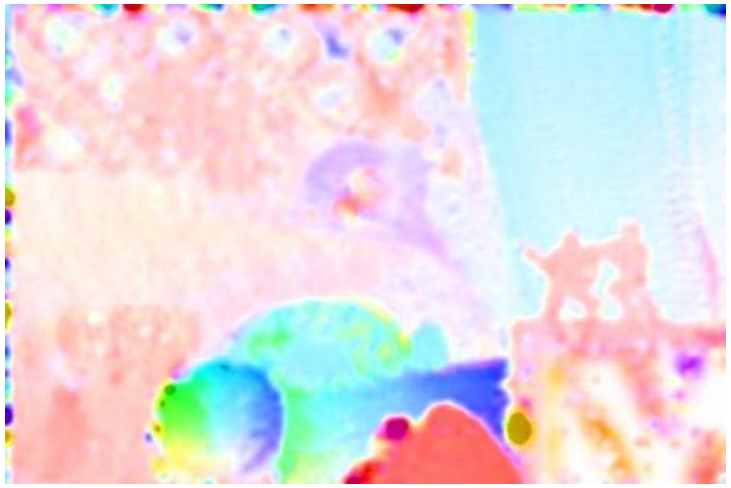
Frame 0

Frame 1

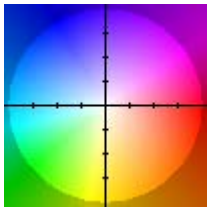
Correct Solution



Estimation Result



Error



OF-vectors are color coded

[1]

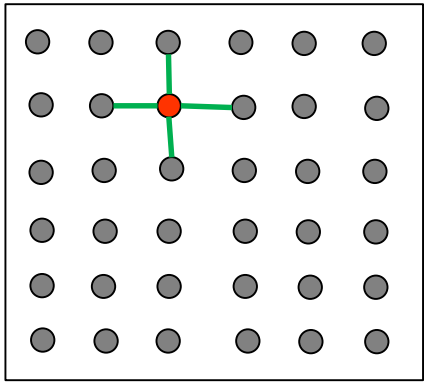
Lucas-Kanade

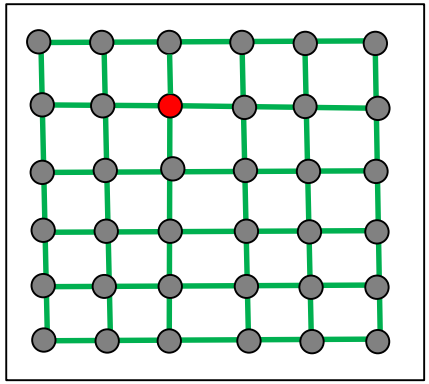
- **Advantages**
  - Dense vector field, no block artifacts
  - More efficient and reliable than block matching
  - Noise is smoothed, but still noisy result!
  - Good results in textured regions
  - Easy and fast calculation
- **Disadvantages**
  - Useless in textureless regions
  - Underestimation of large displacement
  - Errors at boundaries
  - Problems caused by the brightness constancy assumption

Horn-Schunck

- **Basic Idea:** Assume overall smoothness of the resulting flow field.

Local Constraint **vs.** **Global Constraint**





constrains the vector for a pixel by constrains of neighbors

e.g. Lucas-Kanade

constrains the vector for a pixel by average vector of all other pixels

e.g. Horn-Schunck

Horn-Schunck

quadric **penalty functions**

partial derivatives

regularization parameter for degree of smoothness

$$E(u,v) = \iint \left[(u I'_x + v I'_y + I'_t)^2 + \alpha \left(|\nabla u|^2 + |\nabla v|^2 \right) \right] dx\,dy$$

Data Term:

- Applies brightness constancy assumption
- Optical flow equation

- Large for large deviations from brightness constancy assumption
- Small if brightness constancy assumption holds

Smoothness Term:

- Applies smoothness assumption
- First derivatives

- Large if flow field is not smooth
- Small if flow field is smooth

By minimizing this function, solutions which are true for both constrains are favored above others. Wrong solutions are *penalized* with large values.

Horn-Schunck / Energy Function

$$E(u,v) = \iint \left[(uI'_x + vI'_y + I'_t)^2 + \alpha(|\nabla u|^2 + |\nabla v|^2) \right] dx\,dy$$

- **Filling-in-Effect**:

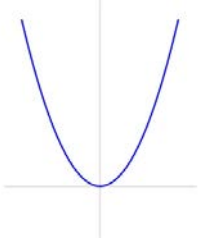
  If there is no information available (e.g. regions of nearly uniform intensity)

$$(uI'_x + vI'_y + I'_t)^2 \approx I'^2_t$$

the smoothness term takes over and „fills in“ the optical flow vectors from nearby pixels (adapt to local solution).

Horn-Schunck / Minimization

$$E(u,v) = \iint \left[(ul'_x + vl'_y + l'_t)^2 + \alpha(|\nabla u|^2 + |\nabla v|^2) \right] dx\,dy$$



E(.) is a **convex function.**

Euler-Lagrange Equations: The first variation of the energy functional (≈first derivate) must vanish.

L1: $(ul'_x + vl'_y + l'_t)l'_x - \alpha(\Delta u) = 0$

L2: $(ul'_x + vl'_y + l'_t)l'_y - \alpha(\Delta v) = 0$

L1: $(l'^2_x + \alpha)u + vl'_y l'_x = \alpha\bar{u} - l'_t l'_x$

L2: $ul'_y l'_x + (l'^2_y + \alpha)v = \alpha\bar{v} - l'_t l'_y$

Δ … Laplace Gradient can be approximated by nearby pixels:

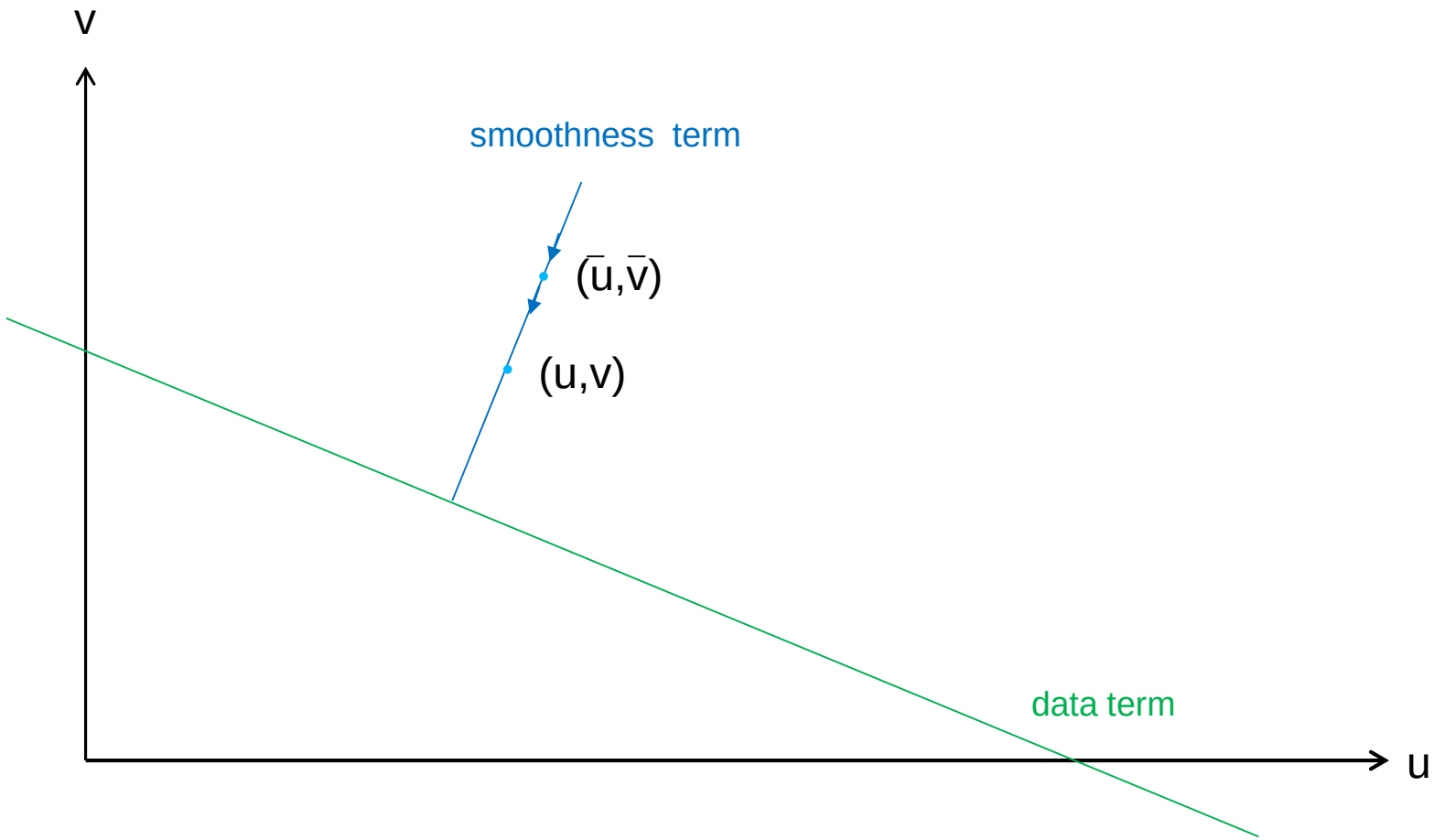
$$\Delta u = \bar{u} - u$$

Solved *iteratively*. Gets more precise with every iteration.

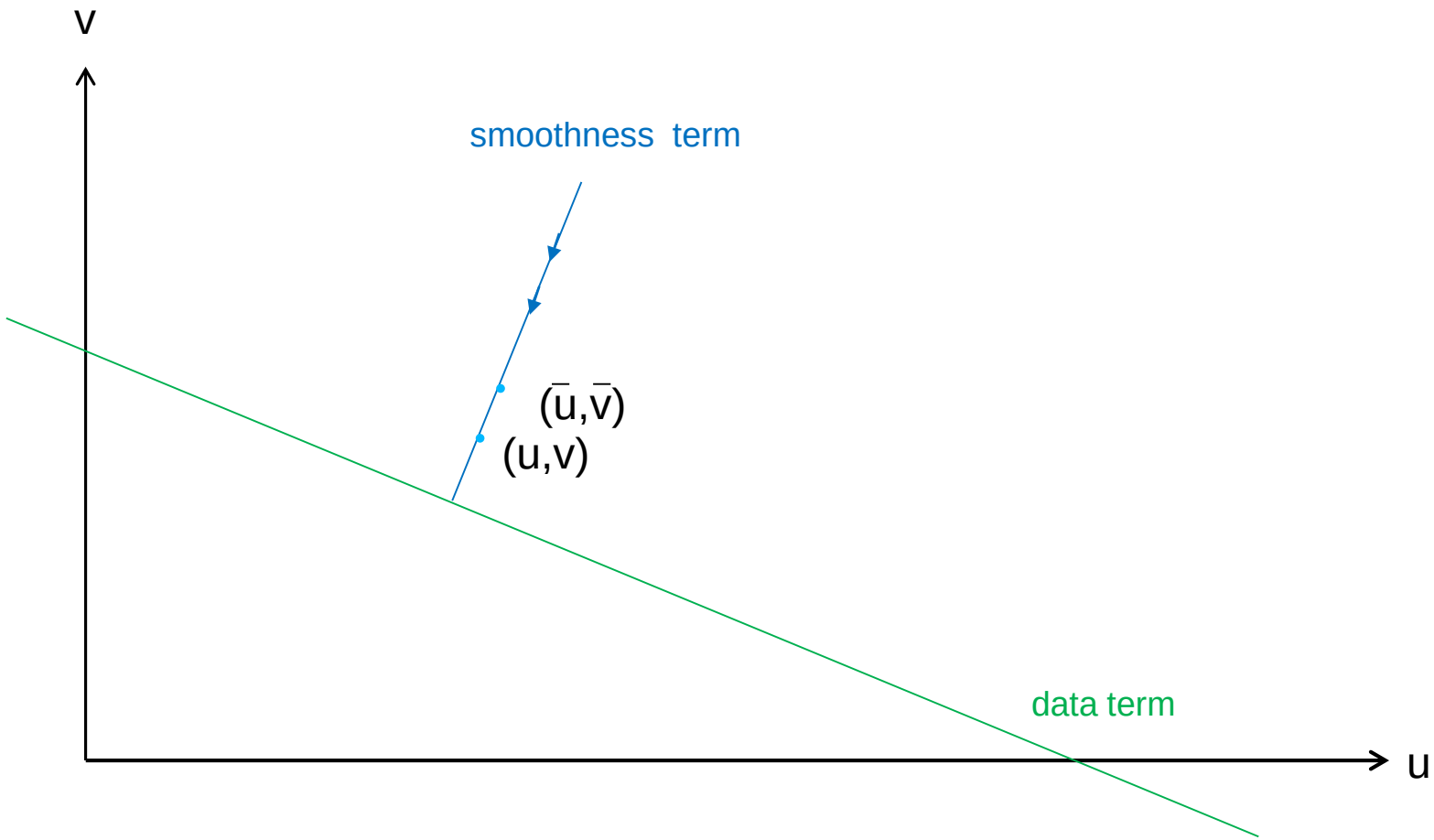
$$u = \bar{u} - \frac{(\bar{u}l'_x + \bar{v}l'_y + l'_t)l'_x}{l'^2_x + l'^2_y + \alpha}$$

$$v = \bar{v} - \frac{(\bar{u}l'_x + \bar{v}l'_y + l'_t)l'_y}{l'^2_x + l'^2_y + \alpha}$$

Horn-Schunck



Horn-Schunck

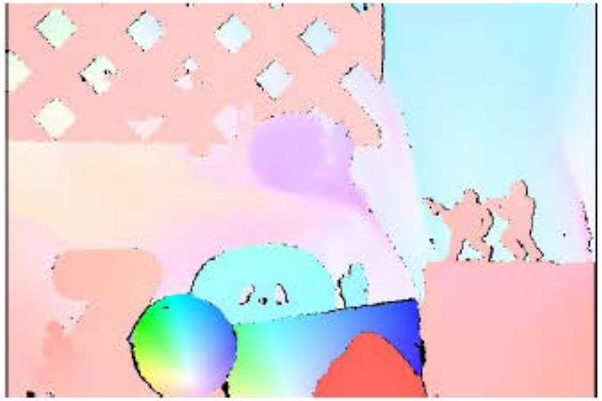


Horn-Schunck

Frame 0

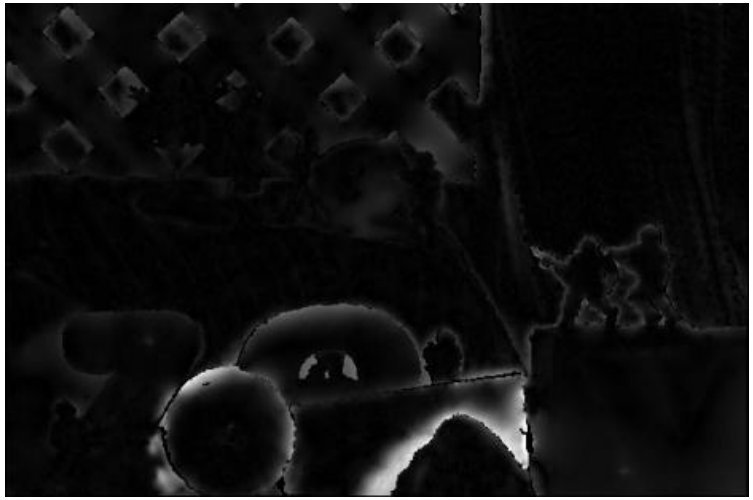
Frame 1

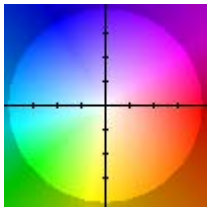
Correct Solution



Estimation Result

Error





OF-vectors are color coded

[1]

Horn-Schunck

- **Advantages**
  - Incorporates global information (“filling-in”)
  - Smooth flow
  - Good results for untextured regions

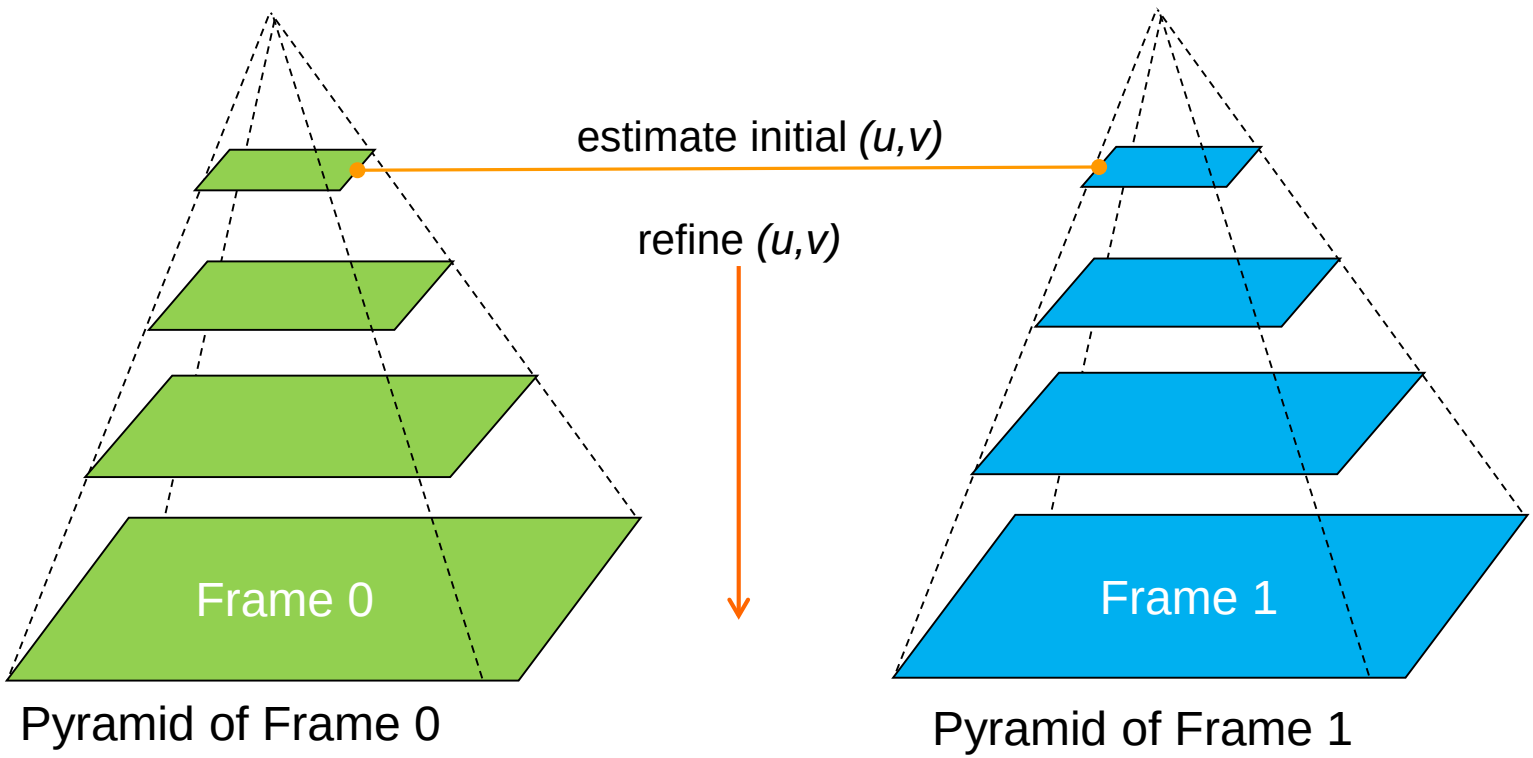
- **Disadvantages**
  - Oversmoothing of motion boundaries
  - Underestimation of large displacements
  - Iterative: slow

  - Problems caused by the brightness constancy assumption

Challenges

Large displacements

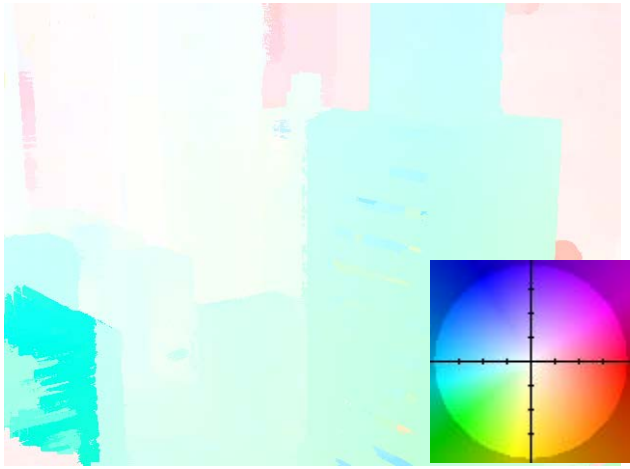
- **Problem:** Approximation of image derivatives does not work for large movement. HS and LK work only for small movements.
- **Solution**: *Coarse-to-fine estimation* using image pyramids



Large displacements

- **Problem:** Approximation of image derivatives does not work for large movement. HS and LK work only for small movements.
- **Solution**: Use a different optimization method, e.g.**,** ***Cost-Volume Filtering****.*
  - **Labels**: Vectors that represent the motion (i.e., *x,y* displacement).
  - **Costs-Volume**: Cost being moved according to a specific vector.
  - **Result**: Motion vectors.

Input: Frame 1 Frame 2 *Output*: Motion [7]



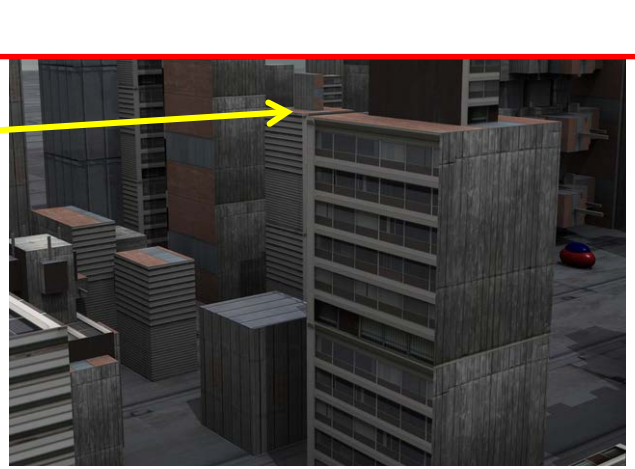
Large displacements

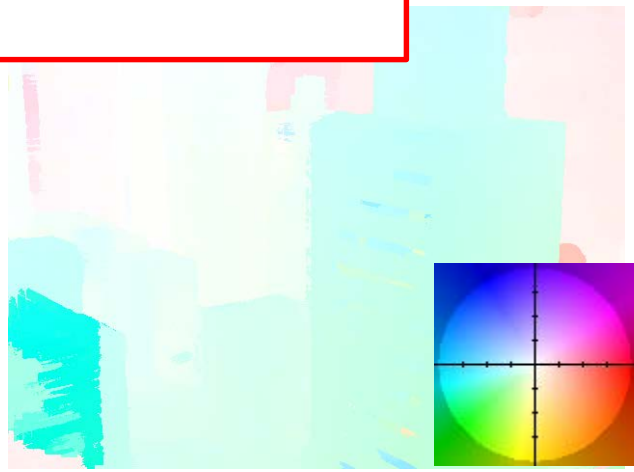
- **Problem:** Approximation of image derivatives does not work for large movement. HS and LK work only for small movements.
- **Solution**: Use a different optimization method, e.g.**,** ***Cost-Volume Filtering****.*
  - **Labels**: Vectors that represent the motion (i.e., *x,y* displacement).

New problem: Scalability of the optical flow estimation algorithm.

- Cost computation for a huge amount of labels (motion vectors).
- Filtering of many cost-volumes!

on [7]





Illumination Changes

- **Problem:** Illumination Changes, e.g., caused by:
  - Shadows
  - Different viewing angles
  - Light source flickering
  - Self-adaptive cameras

- **Solution 1:** Gradient Constancy Assumption
- **Solution 2:** Descriptive Features
- **Solution 3**: HSV Color Model

Illumination Changes: Gradient Constancy Assumption

- **Solution 1** : ***Gradient* constancy assumption**

Assumes that intensity changes are constant.

spatial gradient:

$$\nabla = \partial_x, \partial_y$$

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

Frame 0

Frame 1

The *gradient* constancy assumption is used the same way as the *brightness* constancy assumption.

Illumination Changes: Gradient Constancy Assumption

- **Advantages:**
  - Two more equations
  - Invariant to additive brightness changes

- **Disadvantages:**
  - Not rotation invariant
  - Very noise sensitive

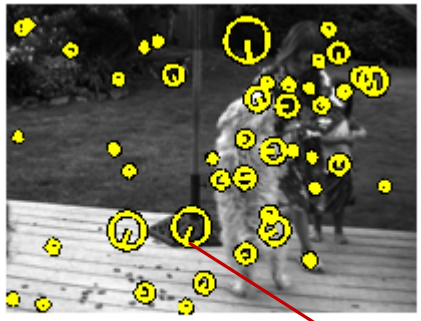
Brightness Constancy and Gradient Constancy both have desired and not wanted effects, which complement each other -> Combine them!

$$E = E_{brightness} + E_{gradient}$$

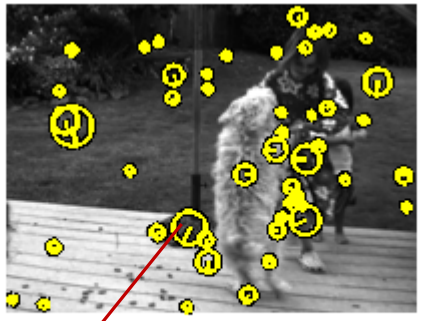
- **Solution 2** : **Descriptive Features**
  - e.g. **SIFT-Features**: Scale Invariant Features

    Points, that can be extracted to provide a description of an object.

    This description can be used to locate the object in another image.



Frame 0



Frame 1

e.g., SIFT-Matching: These are corresponding SIFT keypoints.

-> compute (u,v) from their locations in Frame 0 and Frame 1.

Illumination Changes: Descriptive Features

- **Advantages:**
  - Less sensitive to image deformations
  - Large motion is possible
  - Parts of the object can be occluded

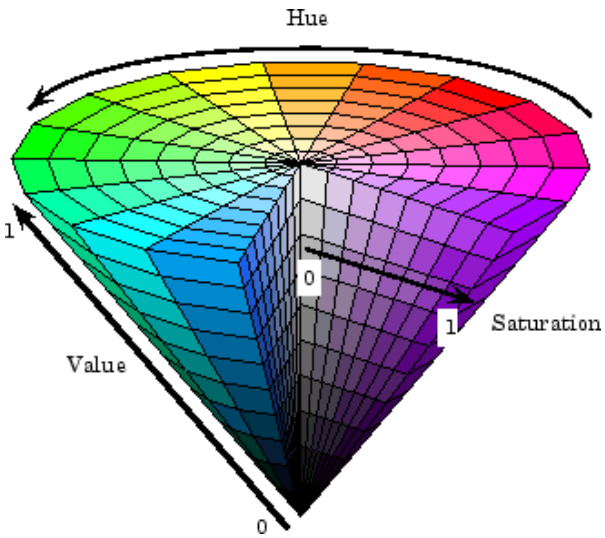
- **Disadvantages:**
  - We don‘t get a dense flow field!

    One Flow vector per SIFT point. If we interpolate these vectors over the whole image -> low accuracy, especially at motion boundaries.
  - Combination with common methods: Descriptive features would need special treatment in optimization process.

Illumination Changes: HSV Color Model

- **Solution 3**: **Spherical Color Model** e.g. HSV
  - **Hue** (H): Color
  - **Saturation** (S)

  Invariant to multiplicative illumination changes (e.g. shadows, shading)
  - **Value** (V): Brightness



$$E = E(H) + E(S) + E(V)$$

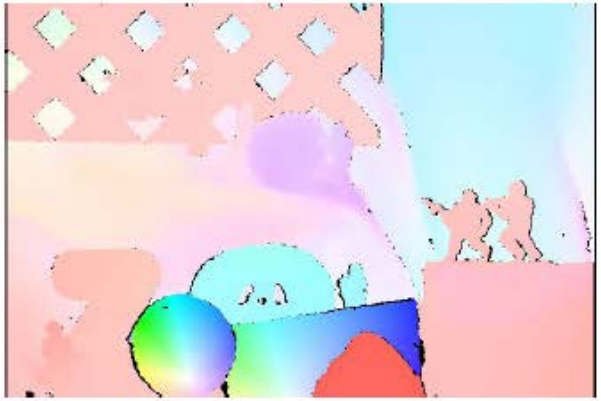
-> picks invariant components when needed

Evaluation

- **Ground Truth (GT)**

Frame 0

Frame 1



GT

Ground Truth: The correct solution (e.g., flow field) for a given problem

- The absence of GT data has represented a major problem in computer vision
- Computer-generated GT images do often not reflect the challenges of real data recorded with a camera.
- It is difficult to measure the progress in a field if there is no commonly agreed data set with GT solution.

[1]

Middlebury Data Set

- **Ground truth (GT) for 12 image sequences**
  - 3 computer generated sequences
  - 9 recorded sequences
- **Quality Evaluation in the Middlebury Benchmark**
  - **Error metrics:**
    - **Angular error** (AE): angle between GT-vector and result
    - **Endpoint error** (EE): Euclidean distance between vectors – more reliable!
    - **Interpolation error** (IE): difference GT and intermediate image
    - **Normalized interpolation error** (NE) : normalized IE
  - **Region Masks**
    - All not occluded pixels
    - Untextured regions
    - Pixels near motion discontinuities
  - **Statistics (accuracy, robustness, …)**

Middlebury Data Set / Online Evaluation

http://vision.middlebury.edu/flow/

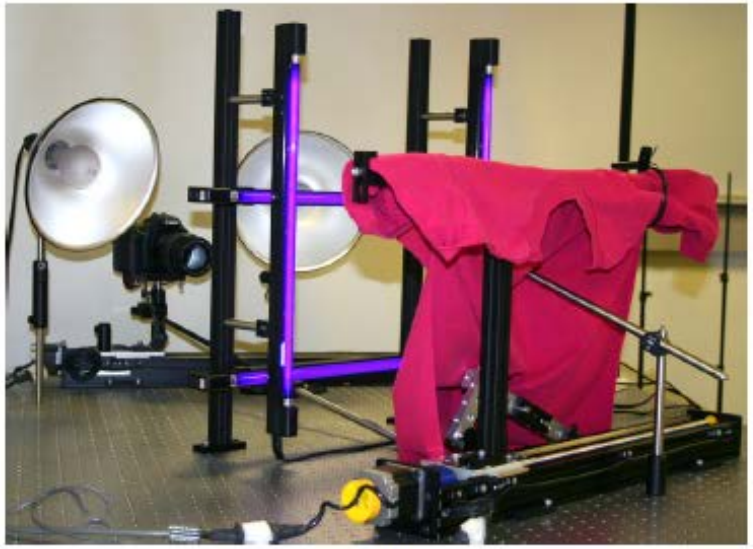
Optical flow evaluation results

Statistics: Average SD R0.5 R1.0 R2.0 A50 A75 A95
Error type: endpoint angle interpolation normalized interpolation

| Average endpoint error | avg. rank | Army (Hidden texture) GT all | Army im0 all | Army im1 all | Army GT disc | Army im0 disc | Army im1 disc | Army GT untext | Army im0 untext | Army im1 untext | Mequon (Hidden texture) GT all | Mequon im0 all | Mequon im1 all | Mequon GT disc | Mequon im0 disc | Mequon im1 disc | Mequon GT untext | Mequon im0 untext | Mequon im1 untext | Schefflera (Hidden texture) GT all | Schefflera im0 all | Schefflera im1 all | Schefflera GT disc | Schefflera im0 disc | Schefflera im1 disc | Schefflera GT untext | Schefflera im0 untext | Schefflera im1 untext | Wooden (Hidden texture) GT all | Wooden im0 all | Wooden im1 all | Wooden GT disc | Wooden im0 disc | Wooden im1 disc | Wooden GT untext | Wooden im0 untext | Wooden im1 untext | Grove (Synthetic) GT all | Grove im0 all | Grove im1 all | Grove GT disc | Grove im0 disc | Grove im1 disc | Grove GT untext | Grove im0 untext | Grove im1 untext | Urban (Synthetic) GT all | Urban im0 all | Urban im1 all | Urban GT disc | Urban im0 disc | Urban im1 disc | Urban GT untext | Urban im0 untext | Urban im1 untext | Yosemite (Synthetic) GT all | Yosemite im0 all | Yosemite im1 all | Yosemite GT disc | Yosemite im0 disc | Yosemite im1 disc | Yosemite GT untext | Yosemite im0 untext | Yosemite im1 untext | Teddy (Stereo) GT all | Teddy im0 all | Teddy im1 all | Teddy GT disc | Teddy im0 disc | Teddy im1 disc | Teddy GT untext | Teddy im0 untext | Teddy im1 untext |
|---|---|---

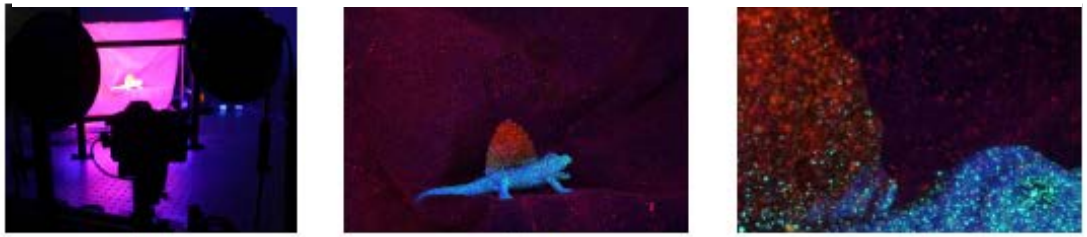
Middlebury Data Set / GT

- **How do they generate GT data?**
  - **Hidden fluorescent texture / splatter pattern in scene**
    - High resolution images in UV + ambient light, motion < 2 pixels
    - *Test sequences*: Every 40th ambient light-frame, downsample images
    - *GT*: Tracking of small windows in original UV sequence



Setup for obtaining GT flow

Setup under visible illumination



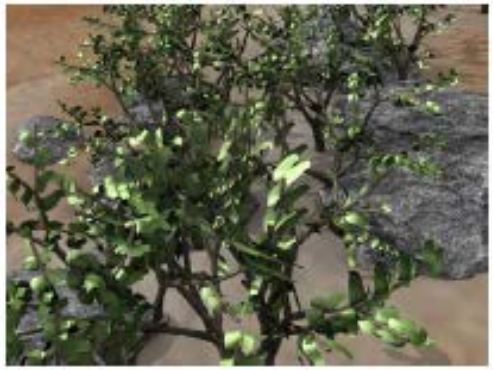
Setup under UV illumination [1]

Middlebury Data Set / GT

- **How do they generate GT data?**
  - **Realistic Synthetic Imagery**
    - Generated with Renderman
    - *Test sequence*: Downsampled images
    - *GT*: Project to 3D motion of the scene to 2D

Frame 0

Frame 1



GT, flow field

[1]

Middlebury Data Set / GT

- **How do they generate GT data?**
  - **Frame interpolation imagery**
    - Record scene with a high-speed camera: 60 frames / second
    - *Test sequences*: skip some frames from recorded sequence
    - *GT*: consists of image frames, use skipped frames as GT
    - Own frame interpolation algorithm to compare performance at same conditions

Frame 0

Frame 1

GT, interpolated frame

[1]

MPI-Sintel Flow Dataset

The **Sintel Flow Dataset** provides naturalistic synthetic videos that are *challenging for current methods*. It is designed to encourage research on *long-range motion*, *motion blur*, *multi-frame analysis*, *non-rigid motion*.

http://sintel.is.tue.mpg.de/

Summary

- **Optical Flow** is the movement of intensity patterns in the image.
- **Techniques:** There are various approaches to estimate optical flow, which differ in how they handle the three key problems of motion estimation.

  e.g. Block Matching

  e.g. Feature Matching

  e.g. Gradient-based Methods

  e.g. Local Methods

  e.g. Global Methods
- **Challenges:** Optical Flow estimation poses many challenges, which several approaches try to overcome.

  e.g. Illumination changes

  e.g. Noise and occlusions

  e.g. Oversmoothing

  e.g. Large/fast motion

  e.g. Untextured regions
- **Evaluation:** The Middlebury Dataset offers test sequences and their GT, which can be used to evaluate Optical Flow Estimation algorithms.

Sample Exam Questions

- How are the key problems solved for e.g., the HS method?
- What is the brightness constancy? What is the Optical Flow Equation?
- Explain the relationship of 2D motion and OF. Give examples.
- Compare the HS and LK method.
- Explain a challenge and a possible solution for it.
- Where and how can you evaluate an OF algorithm?
- What is the aperture problem?
- List three applications of optical flow.

References

[1] Backer S., Scharstein D., Lewis J., Roth S., Black M.J., Szeliski R., “A database and evaluation methodology for optical flow. In ICCV 2007: International Conference on Computer Vision, pages 1–8, 2007.

[2] Horn B. K. P., Schunck G., “Determining optical flow. Artificial Intelligence, Vol. 17, Nr. 1–3, pages 185–203, 1981.

[3] Lucas B. D., Kanade T., “An iterative image registration technique with an application to stereo vision. In IJCAI 1981: International Joint Conferences on Artificial Intelligence, pages 674–679, 1981.

[4] Bruhn A., Brox T., “Tutorial: Variational Optical Flow Estimation”, ICCV 2009.

[5] Fleet D. J., Weiss Y., “Optical Flow Estimation”, In Handbook of Mathematical Models in Computer Vision, pages 239-257, 2005.

[6] Wang Y., Ostermann J., Zhang Y., “Video Processing and Communications”, pages 142-145, 2001.

[7] Low, Toby and Wyeth, Gordon. “*Obstacle detection using optical flow”.* In Proceedings of Australasian Conference on Robotics and Automation 2005, 5 - 7 2005.

[8] Wang, J., Bhat, P., Colburn, R. A., Agrawala, M., and Cohen, M. F. “Interactive video cutout”. In *ACM SIGGRAPH 2005*, 585-594, 2005

[9] Lempitsky V., Roth S., Rother C., “FusionFlow: Discrete-continuous optimization for optical flow estimation”. In CVPR 2008, pages 1–8, 2008.

[10] Brox, T.; Bregler, C.; Malik, J.; , "Large displacement optical flow," *CVPR 2009.*, vol., no., pp.41-48, 20-25, 2009.

[11] Zimmer, H., Bruhn, A., Weickert, J., Valgaerts, L., Salgado, A., Rosenhahn, B., and Seidel, H. “Complementary Optic Flow”. In *Proceedings of the 7th international Conference on Energy Minimization Methods in Computer Vision and Pattern Recognition* , 2009.

[12] M. Guttmann, L. Wolf, D. Choen-Or. *Semi-automatic Stereo Extraction from Video Footage*, ICCV 2009.