Computer Vision

Lecture 7: Optical Flow

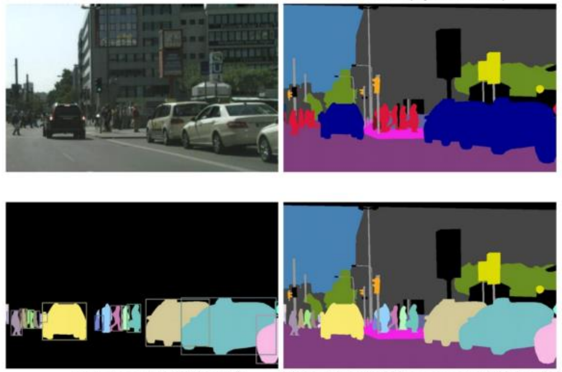
Oliver Bimber

JOHANNES KEPLER UNIVERSITY LINZ

Last Week: Segmentation

Types of Segmentation

Semantic (by Classes)



Instance (by Objects) Panoptic (by Classes and Objects)

JKU JOHANNES KEPLER UNIVERSITY LINZ

Example: Clustering by Graph Eigenvectors

affinity between element i and j
X
association of elements j with cluster n

$$\boxed{w_n^T}\boxed{Aw_n} = \lambda \qquad Aw_n = \lambda w_n$$

association of elements j with cluster n X

$$\begin{bmatrix} w_{n,0} \dots w_{n,j} \end{bmatrix} \begin{bmatrix} a_{0,0} & \dots & a_{i,0} \\ \dots & \dots & \dots \\ a_{0,j} & \dots & a_{i,j} \end{bmatrix} \begin{bmatrix} w_{n,0} \\ \dots \\ w_{n,j} \end{bmatrix}$$

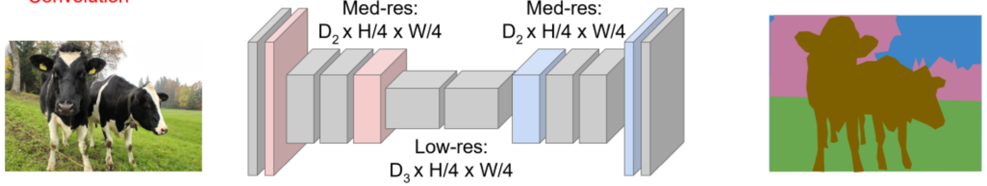
JKU JOHANNES KEPLER UNIVERSITY LINZ

Segmentation using CNNs

Downsampling: Pooling, Strided Convolution

Design Network as a Bunch of Convolutional Layers, with **Downsampling** and **Upsampling** inside the Network!

Upsampling: Strided Transposed Convolution (or Unpooling)



Input:
3 x H x W

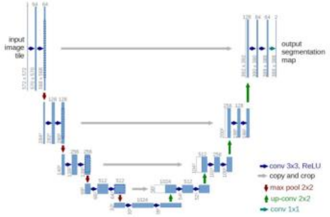
High-res:
D_1 x H/2 x W/2

High-res:
D_1 x H/2 x W/2

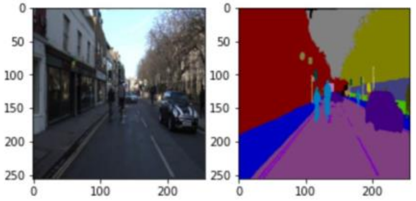
Predictions:
H x W

JKU JOHANNES KEPLER UNIVERSITY LINZ

Connected Autoencoders (U-Nets)



Connected Autoencoder (U-Net)



Predicted Segmentation

- U-Nets overcome this problem by connecting corresponding encoder-decoder layers with skip connections:
  - the output of an encoder level is skip-connected (concatenation) with the input of the corresponding decoder level

JKU JOHANNES KEPLER UNIVERSITY LINZ

JKU JOHANNES KEPLER UNIVERSITY LINZ

Course Overview

CW	Topic	Date	Place	Lab
41	Introduction and Course Overview	08.10.2024	Zoom	Lab 1
42	Capturing Digital Images	15.10.2024	Zoom	Lab 2
43	Digital Image Processing	22.10.2024	Zoom	Assignment 1
44	Machine Learning	29.10.2024	Zoom	
45	Feature Extraction	05.11.2024	Zoom	Open Lab 1
46	Segmentation	12.11.2024	Zoom	Assignment 2
47	Optical Flow	19.11.2024	Zoom	Open Lab 2
48	Object Detection	26.11.2024	Zoom	Assignment 3
49	Multi-View Geometry	03.12.2024	Zoom	Open Lab 3
50	3D Vision	10.12.2024	Zoom	Assignment 4
3	Trends in Computer Vision	14.01.2025	Zoom	Open Lab 4
4	Q&A	21.01.2025	Zoom	
5	Exam	28.01.2025	HS1 (Linz), S1/S3 (Vienna), S5 (Bregenz)	
9	Retry Exam	25.02.2025	HS1 (Linz), S1/S3 (Vienna), S5 (Bregenz)	

Optional Research Challenge

Given: are two image stacks, each stored in a .tif file, and each having 300 layers (individual images)

You can view these stacks, for example, in ImageJ (drag&drop): https://imagej.net/ij/ , or simply implement a python script that loads and display them (you can use ChatGPT & co. to generate one)

Both images stacks encode focus / defocus of a scenery (plants). Focus changes from top (first layer in the stacks) to bottom (last layers in the stacks). See example layers in the following slides.

It appears, that out-of-focus regions always have a similar noise pattern (i.e. the standard deviation in such regions is similar) while in-focus regions have very different noise patterns (either extremely high at edge boundaries, or extreme low at surfaces) with corresponding standard deviations.

Task: Can you distinguish (segment) between out-of-focus (with common noise pattern for out-of-focus regions) and in-focus (with noise patterns that are very different from the out-of-focus patterns) regions?

Hint: Models, such as computing and comparing std. deviation, lead to noisy results. They are too simple to describe the noise statistics. Approach an ML-solution that learns how to describe the noise statistics better (e.g., via more than one (learned) operations, such as kernels in multiple conv. layers – as std. deviation is only one modeled (not learned) operation).

Hint 2: The noise statistics of out-of-focus and in-focus regions are "similar" within the same stack. They might be slightly different in other stacks.

Submission: Submit only (OK) working solutions (no solutions that don't work at all or deliver bad results). You need to submit your code, the segmented results (images), and a short description (you chose the format and length) on how you solved the problem. Any help (including LLMs) is allowed. Deadline: **10.12.2024, 12 noon**, to oliver.bimber@jku.at (3 weeks).

General: At max. you can **gain 25% of your lab points** (which is **equivalent to one full assignment**) in addition. These points can be used to compensate for missing assignment points (to improve your grade), or to skip any upcoming assignments (and compensate for it with the points of the research challenge). Note, that points of the research challenge are on top the points of the regular assignments. Thus, the research challenge is optional, and all students doing only the regular assignments have a change to achieve 100% of their lab grade with them only. This research challenge is intended for motivated students that are underchallenged with the regular assignments and are interested in solving more research oriented tasks. You can approach this research challenge **alone or as team of two** (each getting the same points).

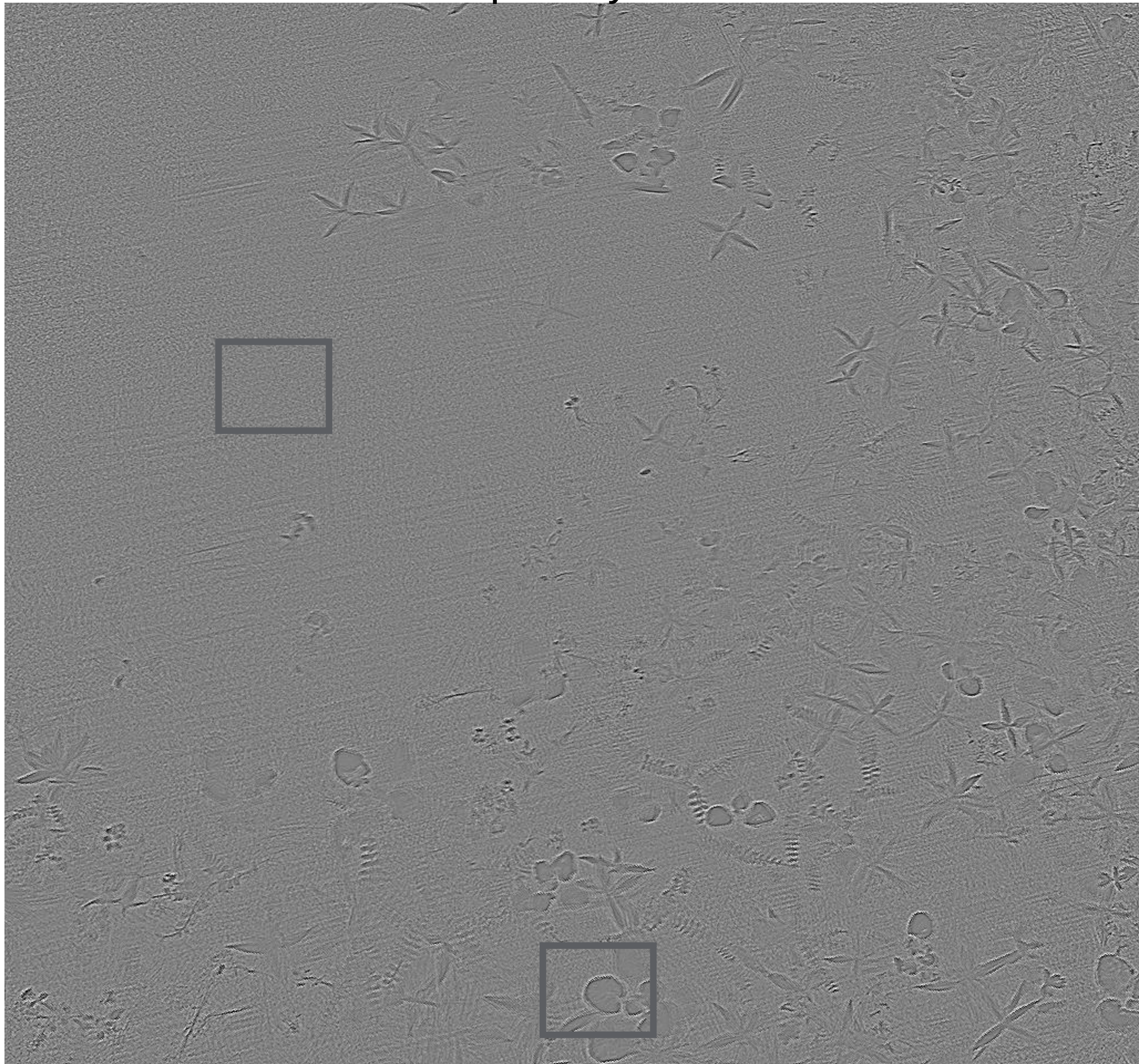
Final: You can contact me any time with questions: oliver.bimber@jku.at

Data: https://drive.google.com/file/d/1KwbC89ks9_h6PpRQjID5M5OviVrMSPue/view?usp=drive_link (600 MB), 7Zip: https://7-zip.de/download.html

Optional Research Challenge

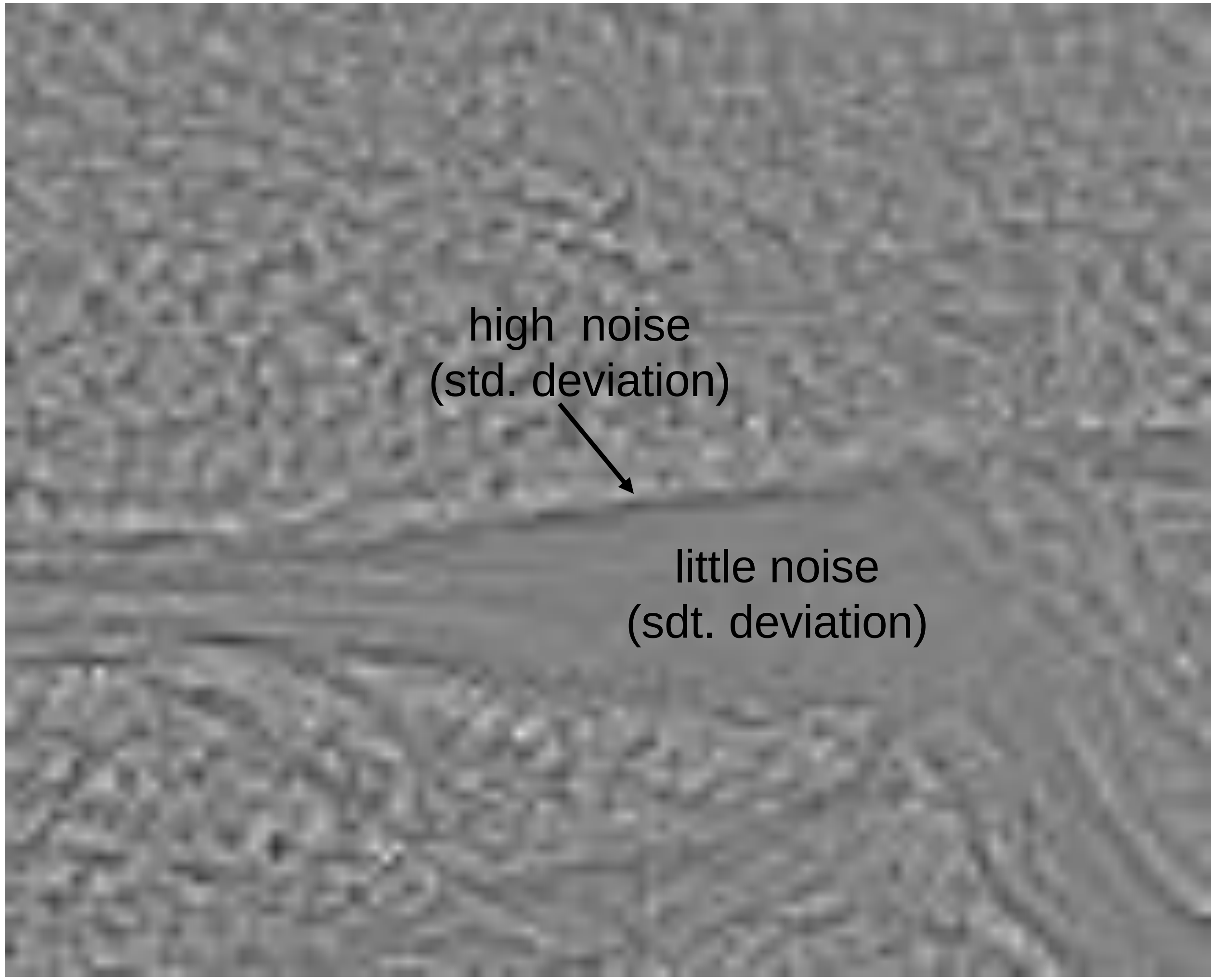
Example Layer of Stack 1

Example Layer of Stack 2

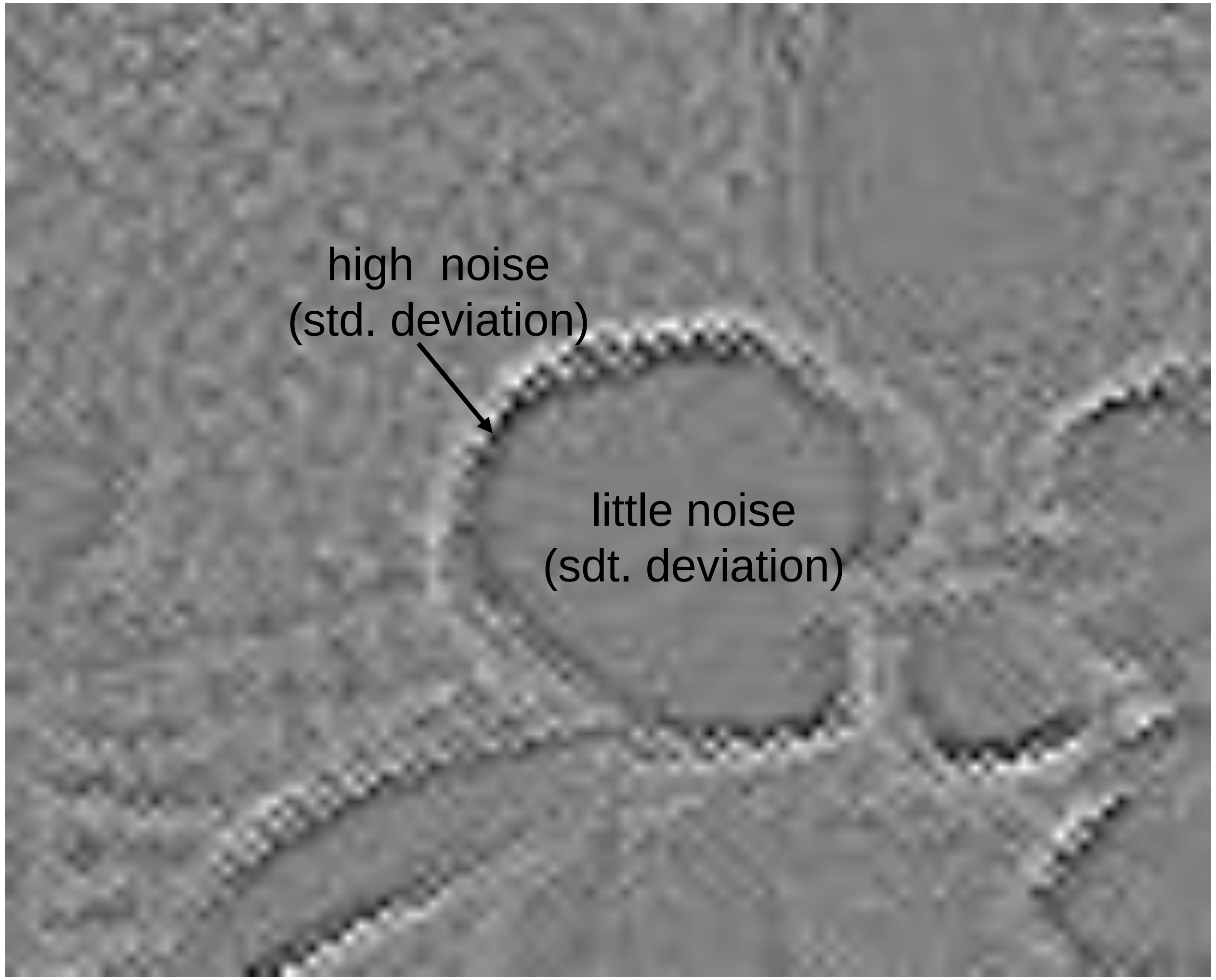


Optional Research Challenge

Close-Up for In-Focus Regions
Example Layer of Stack 1

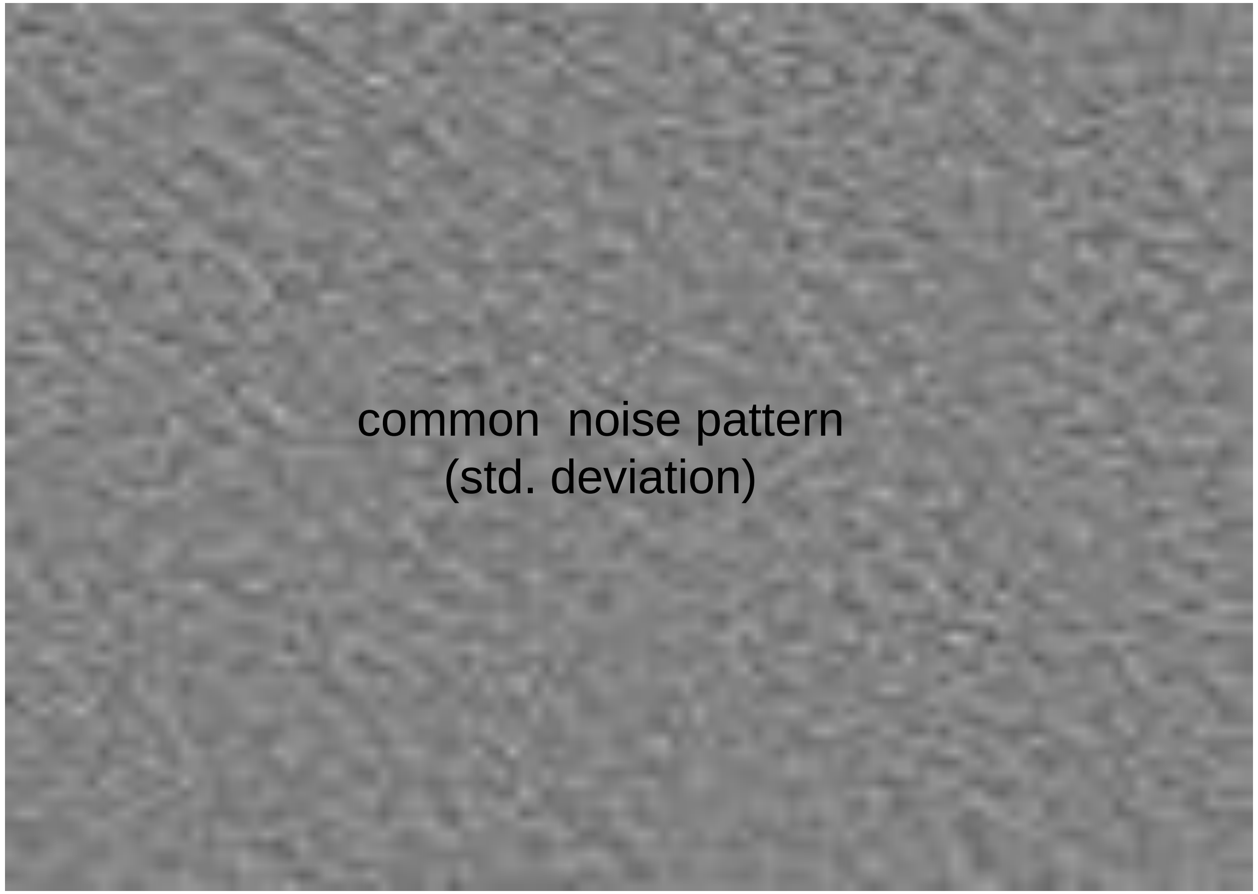


Close-Up for In-Focus Regions
Example Layer of Stack 2

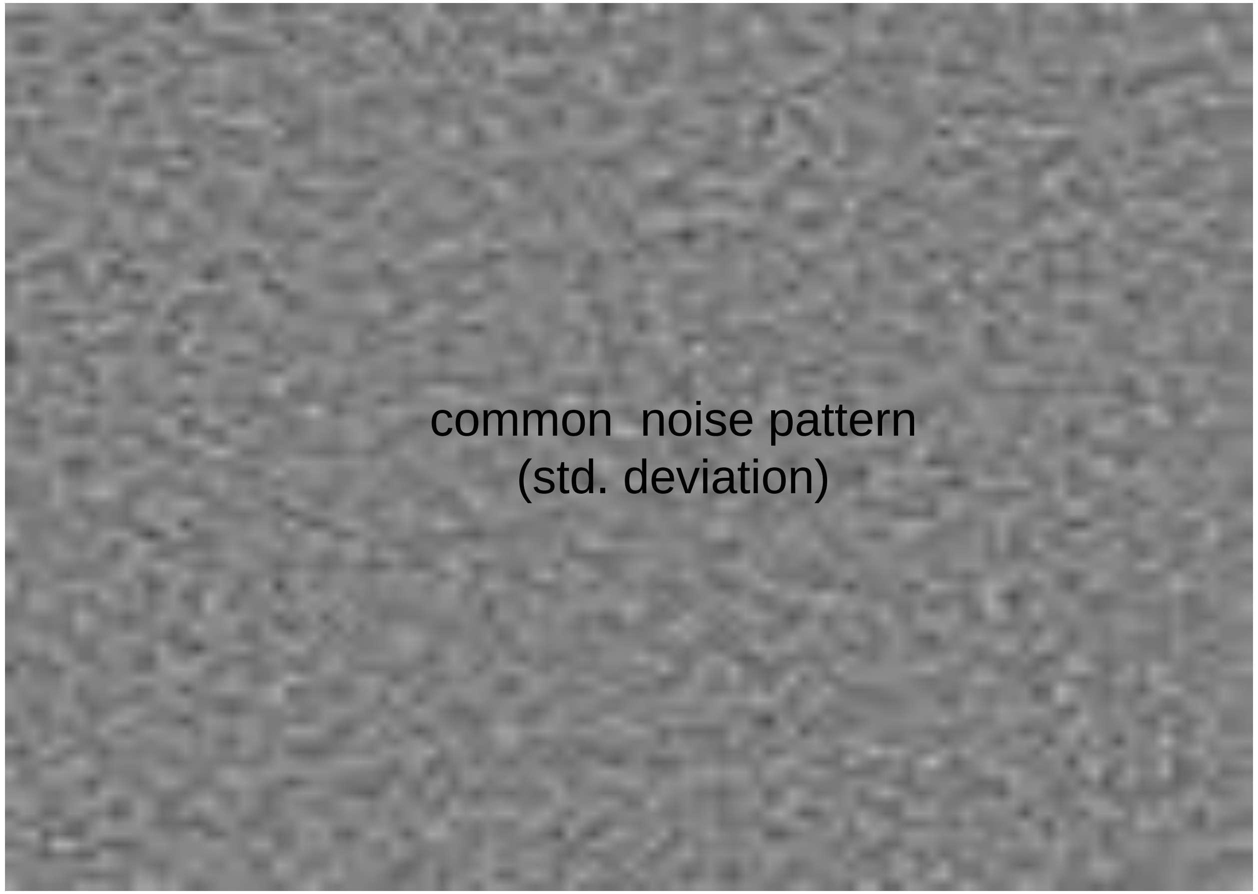


Optional Research Challenge

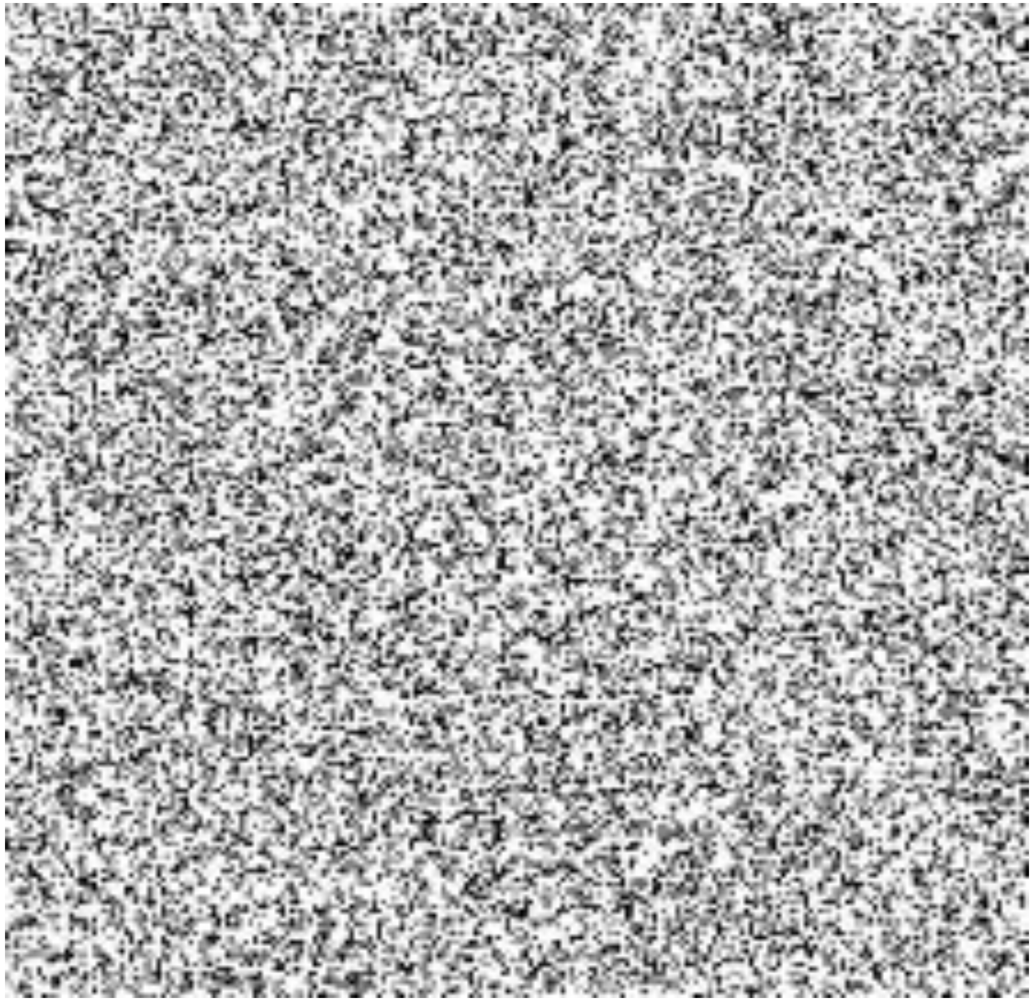
Close-Up for Out-of-Focus Regions
Example Layer of Stack 1



Close-Up for Out-of-Focus Regions
Example Layer of Stack 2

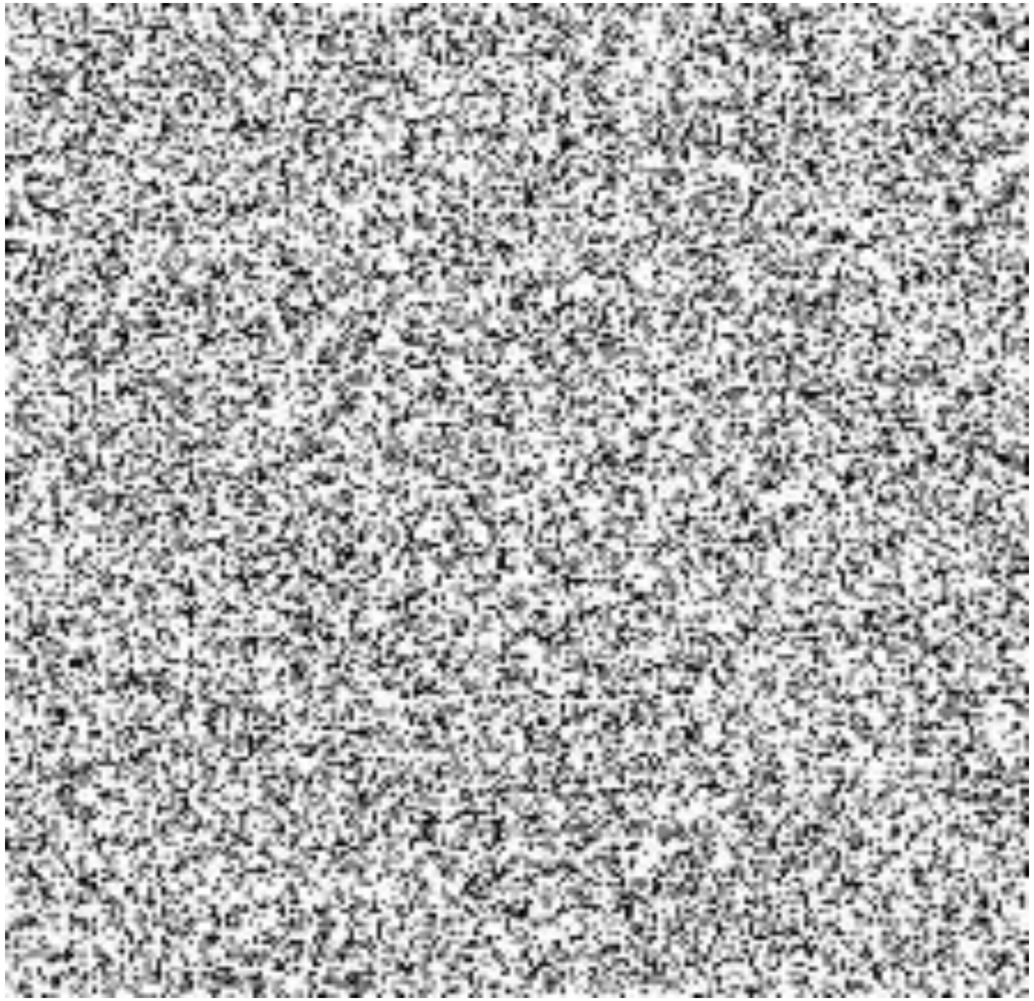


Perception of Motion

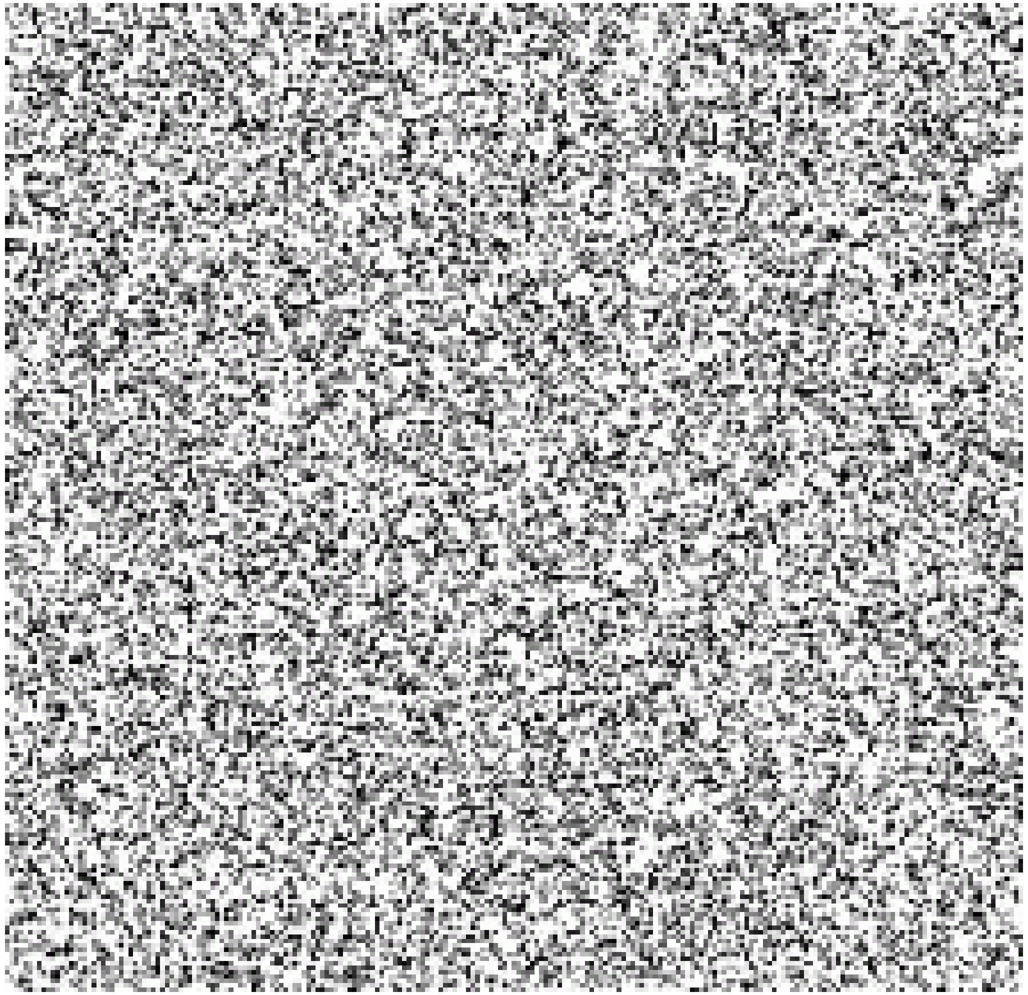


Can you see a Patter?

Perception of Motion

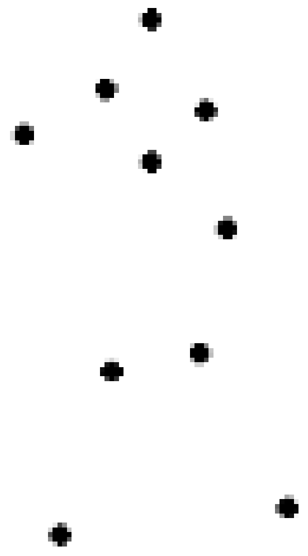


Can you see a Patter?



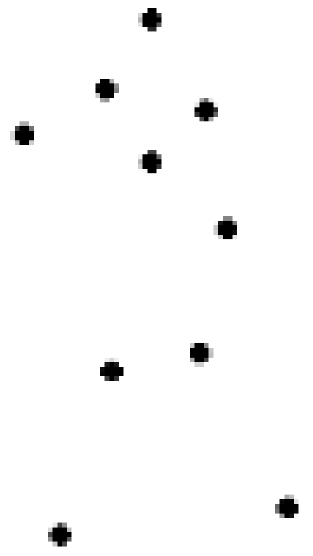
Sometimes Motion is the only Cue

Perception of Motion

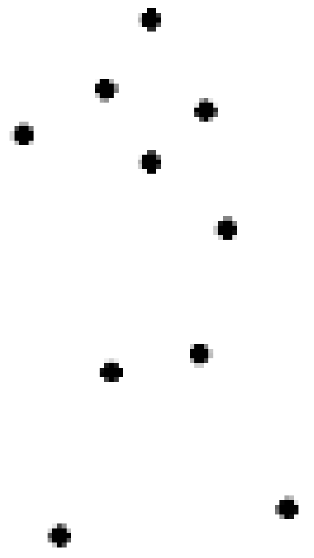


Can you see a Patter?

Perception of Motion



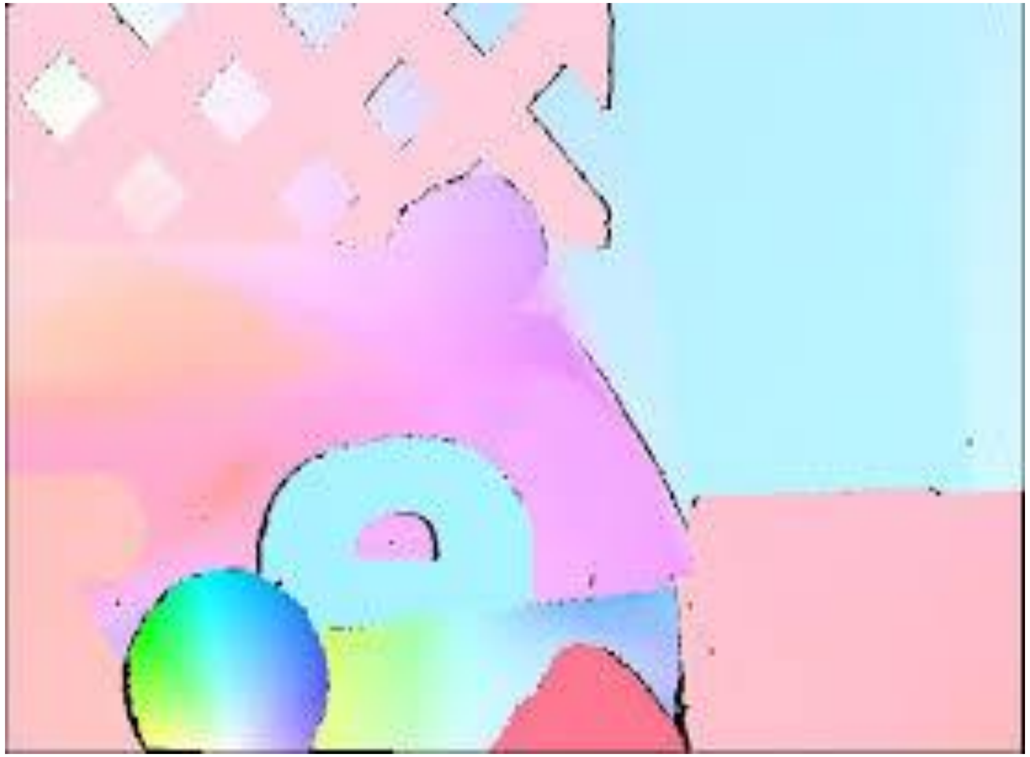
Can you see a Patter?



Sometimes Motion is the only Cue
(even for sparse Data)

What is Optical Flow?

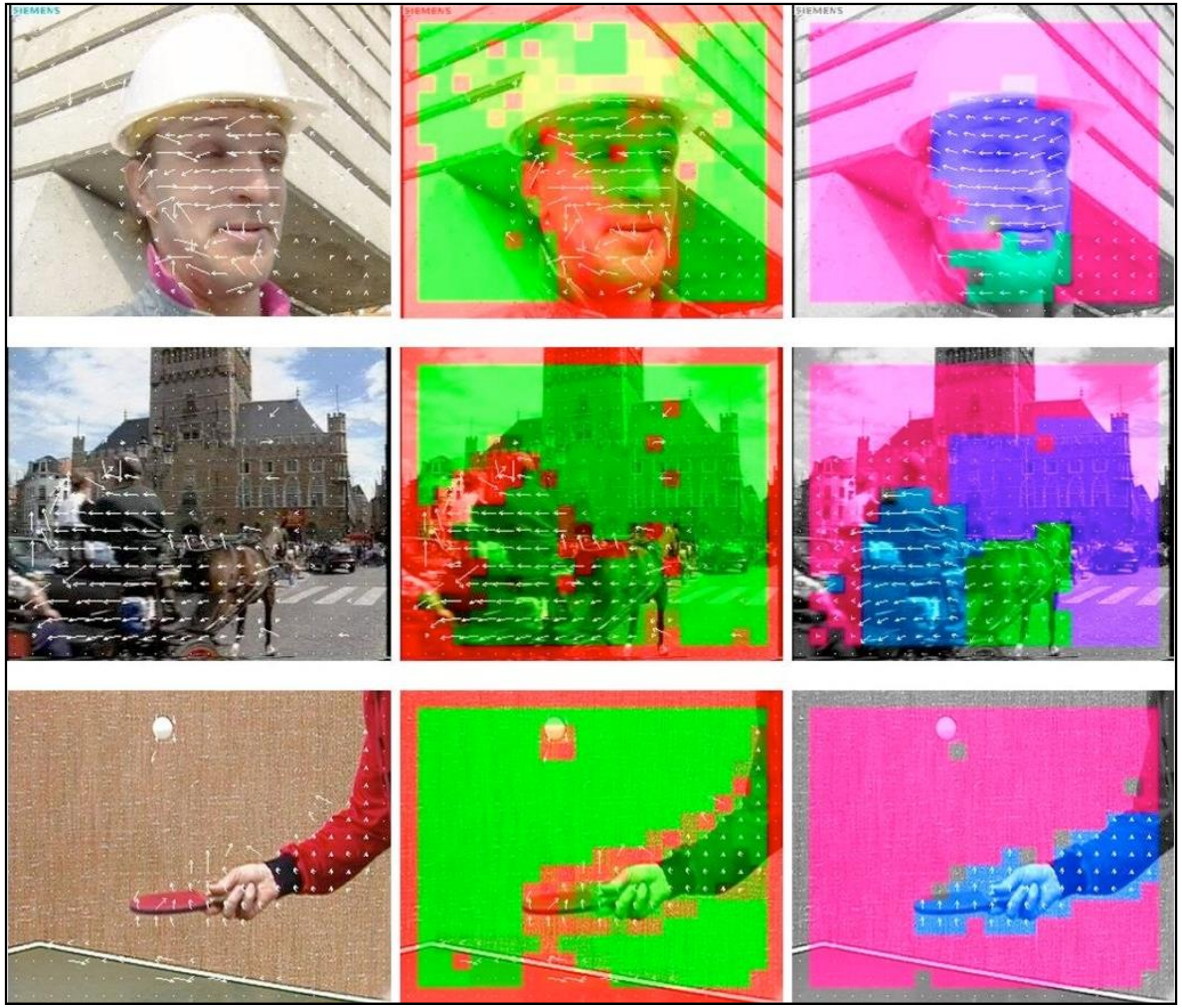
Sparse Optical Flow (Flow Vector per Region)



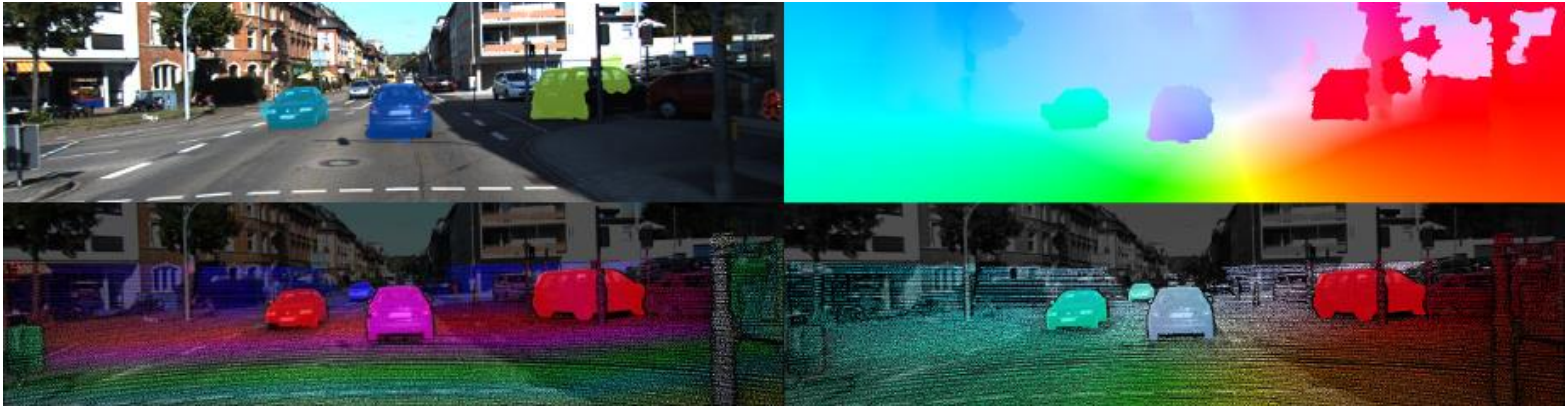
Dense Optical Flow (Flow Vector per Pixel)

Motion Vector of Pixel in Time Series (two consecutive Video Frames at Times t and t+1)

Example: MPEG Compression

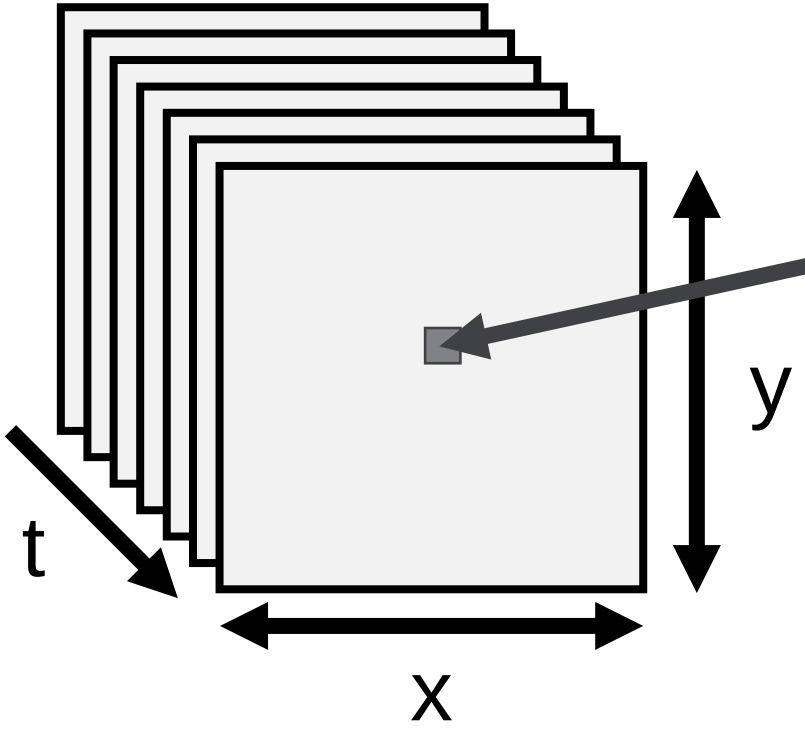


Example: Autonomous Driving



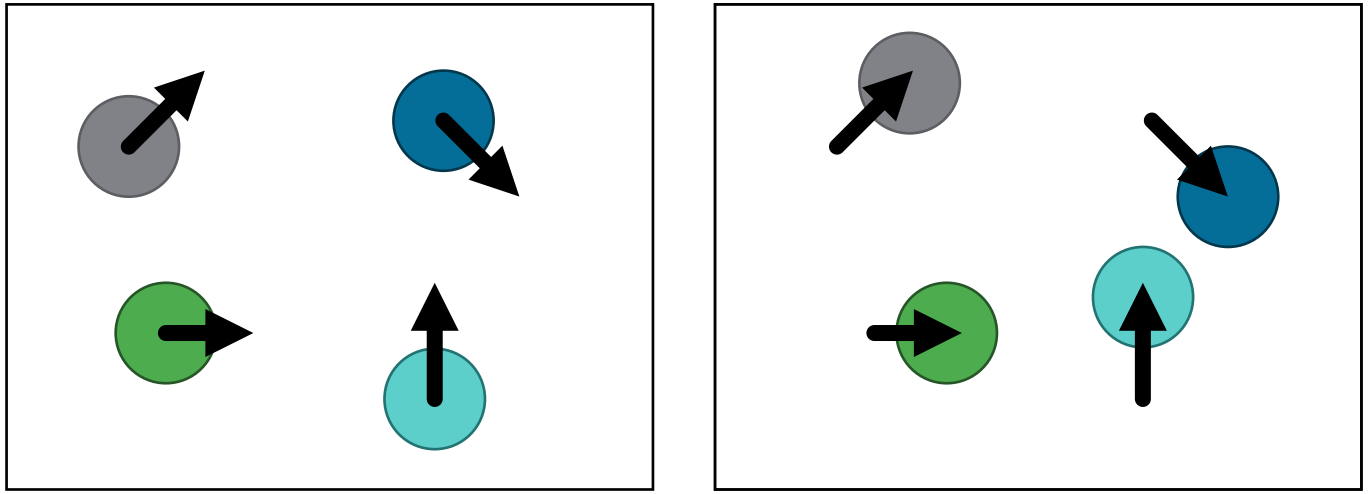
Example: Traffic Monitoring

Video: A Sequence of Frames over Time



I(x,y,t)
x,y – Location
t – Time

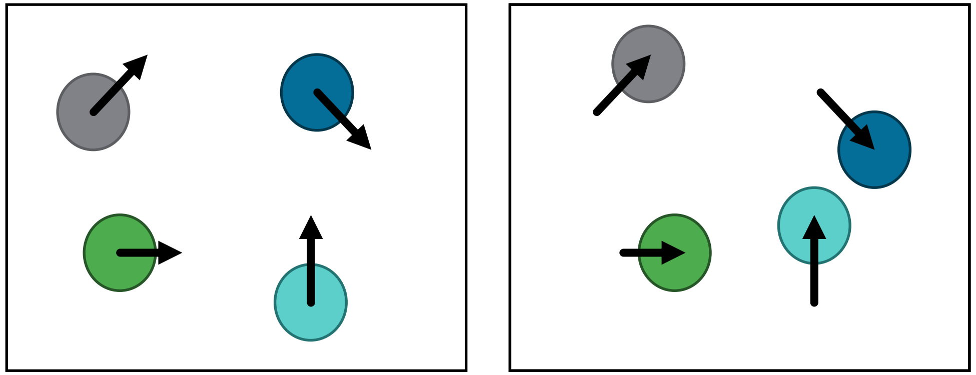
Optical Flow



I(x,y,t) I(x,y,t+1)

Want to estimate Pixel Motion from Image I(x,y,t) to Image I(x,y,t+1)

Assumption



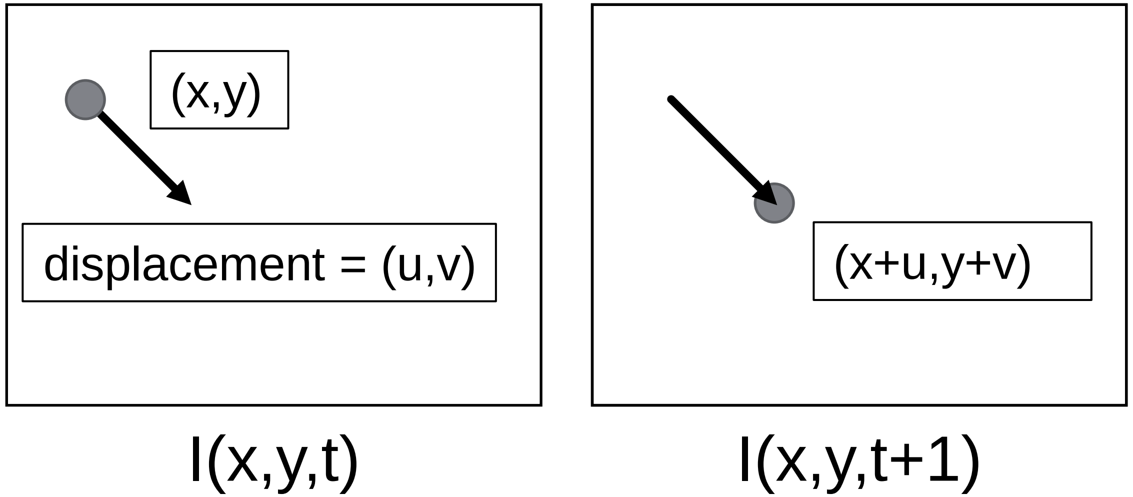
I(x,y,t) I(x,y,t+1)

Solve correspondence Problem: given Pixel at Time t, find **nearby** Pixels of the **same Color** at Time t+1

Key assumptions:

- **Color/Brightness Constancy**: Point at Time t looks same at Time t+1
- **Small Motion**: Points do not move very far

The Optical Flow Equation

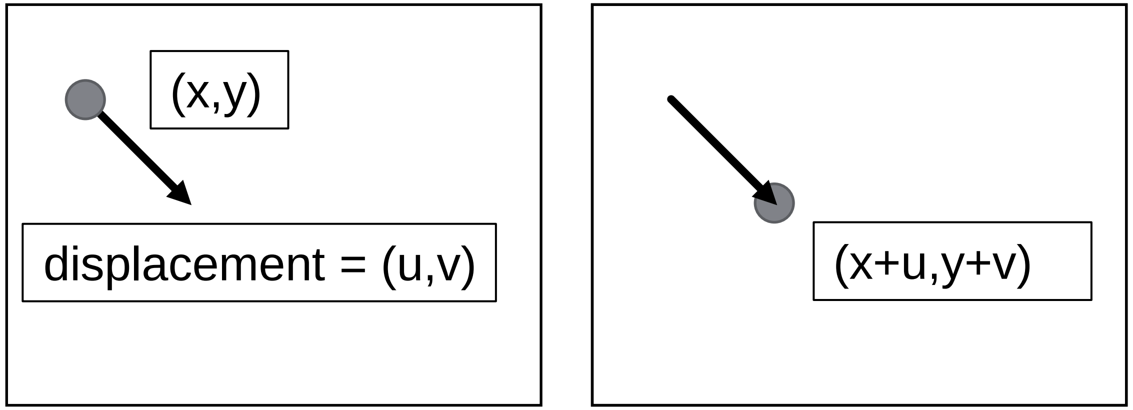


Brightness Constancy:

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

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

The Optical Flow Equation



I(x,y,t) I(x,y,t+1)

Brightness Constancy:

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

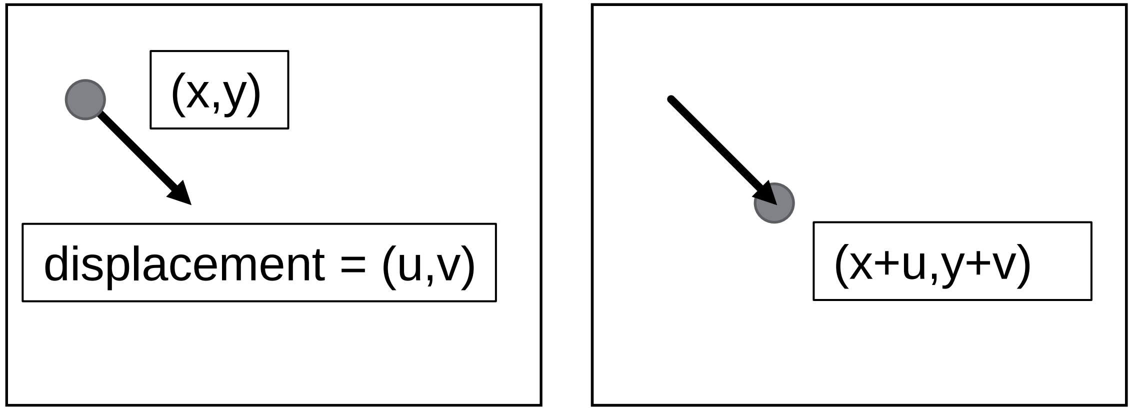
$$0 \approx I(x+u, y+v, t+1) - I(x, y, t)$$

Taylor Expansion:

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

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

The Optical Flow Equation



I(x,y,t) I(x,y,t+1)

Brightness Constancy:

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

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

Taylor Expansion:

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

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

Optical Flow Equation:

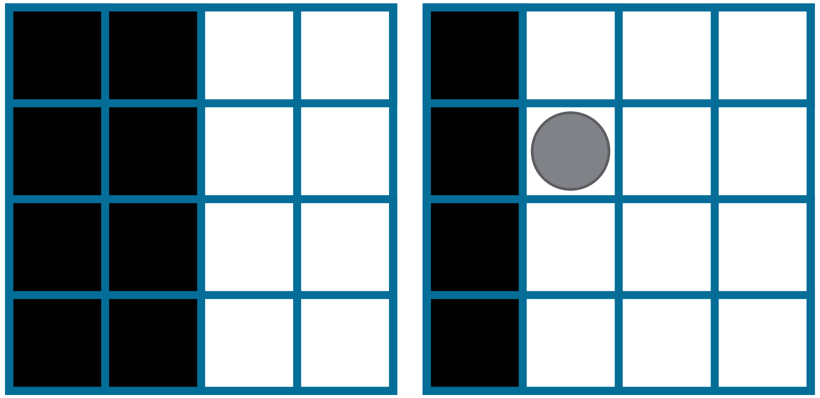
$$0 = I_t + I_x u + I_y v$$

$$= I_t + \nabla I \cdot [u, v]$$

I_t, I_x, I_y are partial derivatives of image intensity (gradients) in *t,x,y*

Example

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



t   t+1

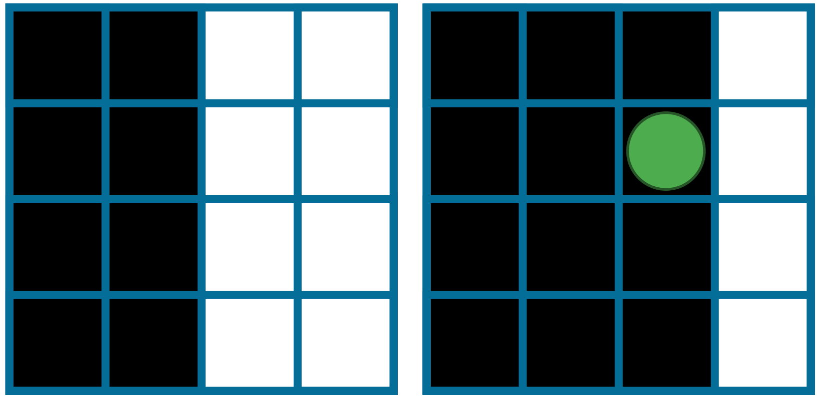
@

I_t = 1-0 = 1

I_y = 0

I_x = 1-0 = 1

What's u?



t   t+1

@

I_t = 0-1 = -1

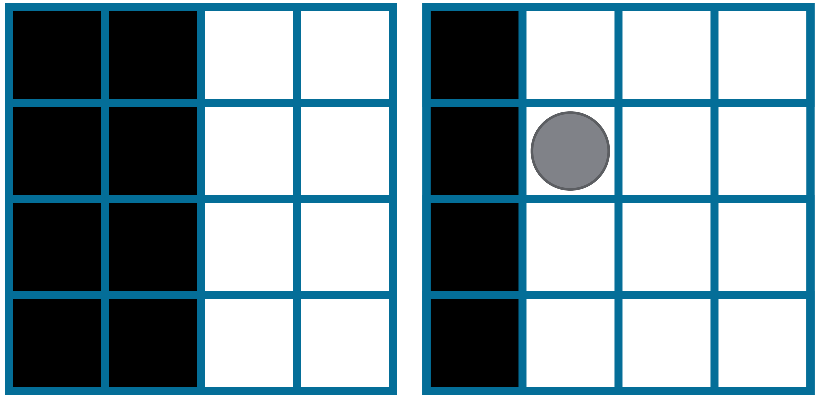
I_y = 0

I_x = 1-0 = 1

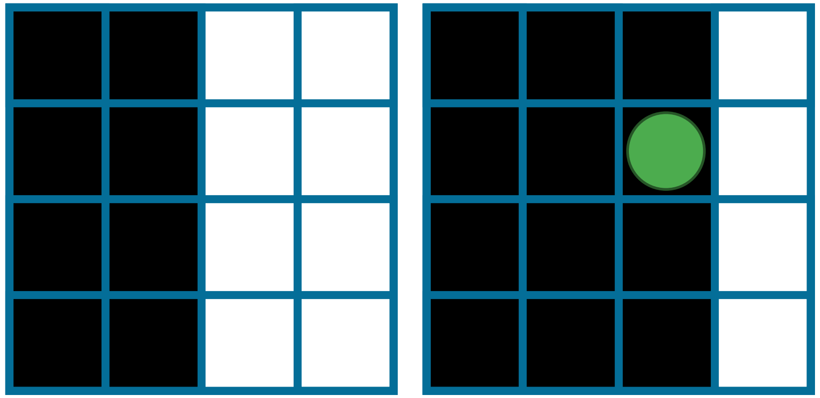
What's u?

Example

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



t t+1



t t+1

@

I_t = 1-0 = 1

I_y = 0

I_x = 1-0 = 1

What's u?

@

I_t = 0-1 = -1

I_y = 0

I_x = 1-0 = 1

What's u?

Can only identify motion along gradient, but not along arbitrary directions!

How to overcome this Problem?

Lucas-Kanade 1981

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

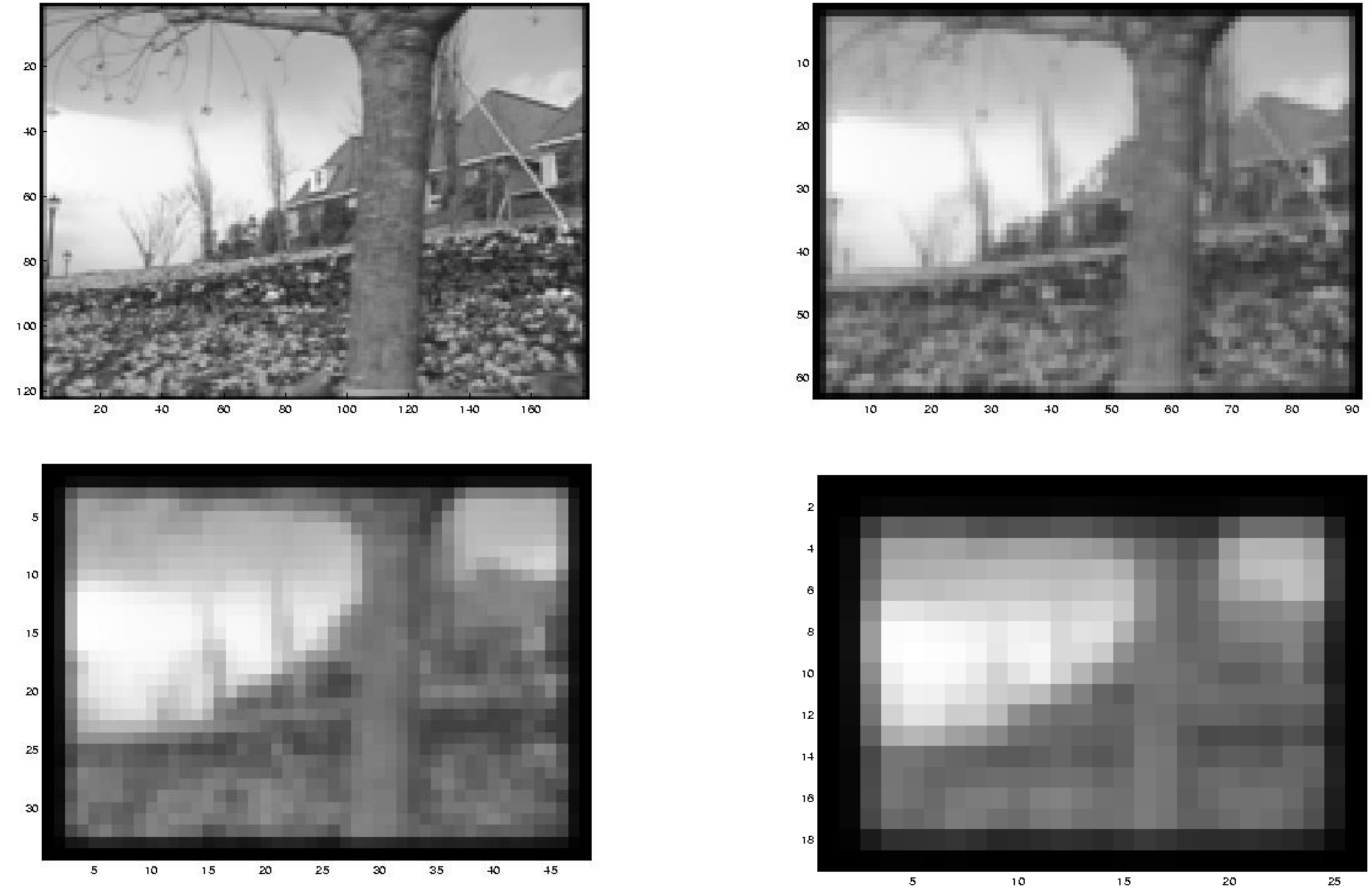
$$\downarrow$$

$$\begin{bmatrix} I_x(p_1) & I_y(p_1) \\ \vdots & \vdots \\ I_x(p_{25}) & I_y(p_{25}) \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = -\begin{bmatrix} I_t(p_1) \\ \vdots \\ I_t(p_{25}) \end{bmatrix}$$

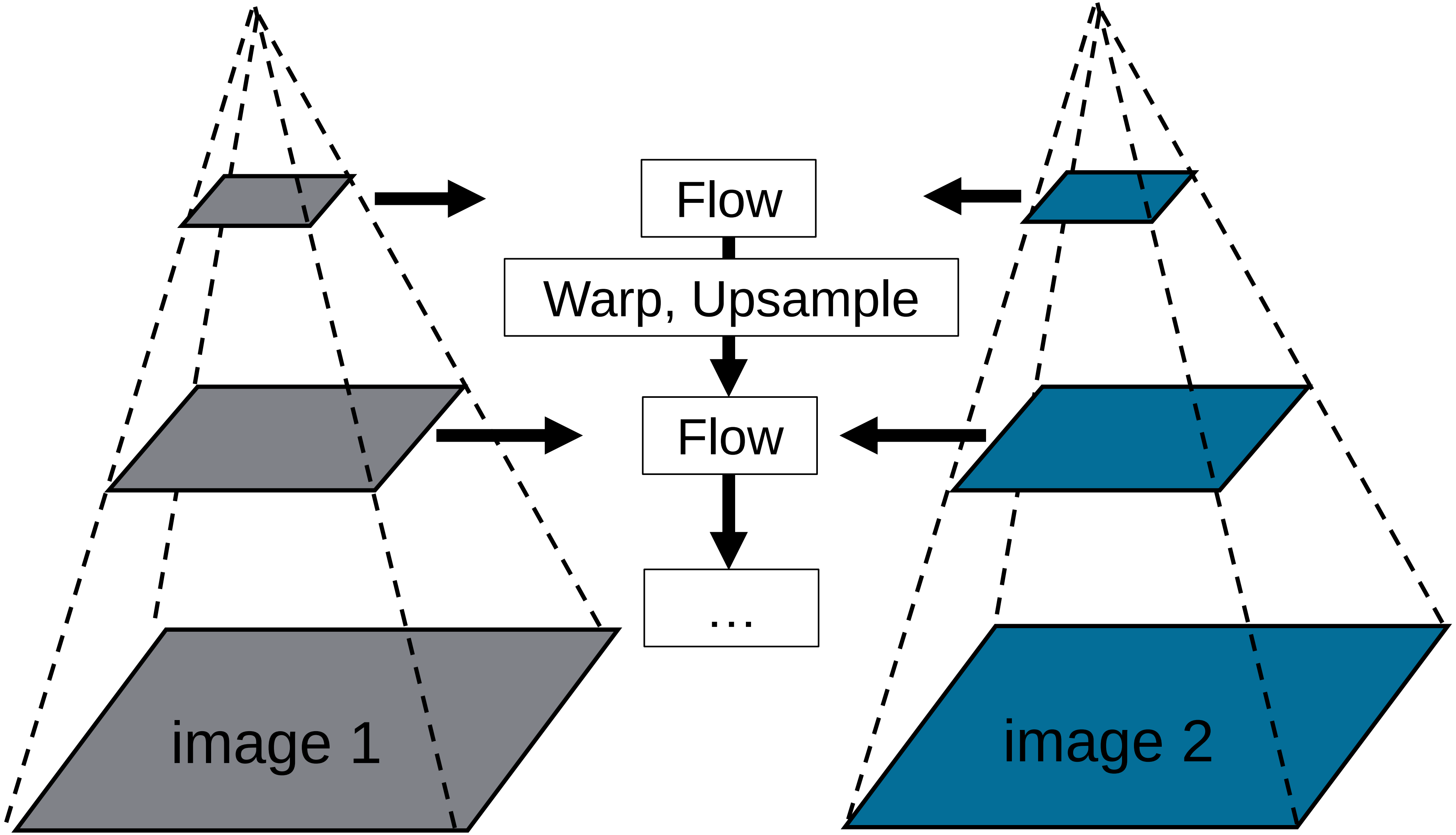
- 2 Unknowns [u,v], 1 Equation per Pixel
- How do we get more Equations?
- Assume ***Spatial Coherence***: Pixel's Neighbors have same [u,v] (same Optical Flow on local Region)
- Example: 5x5 Window gives 25 Equations

What if Motion is larger than one Pixel?

Reduce Resolution!



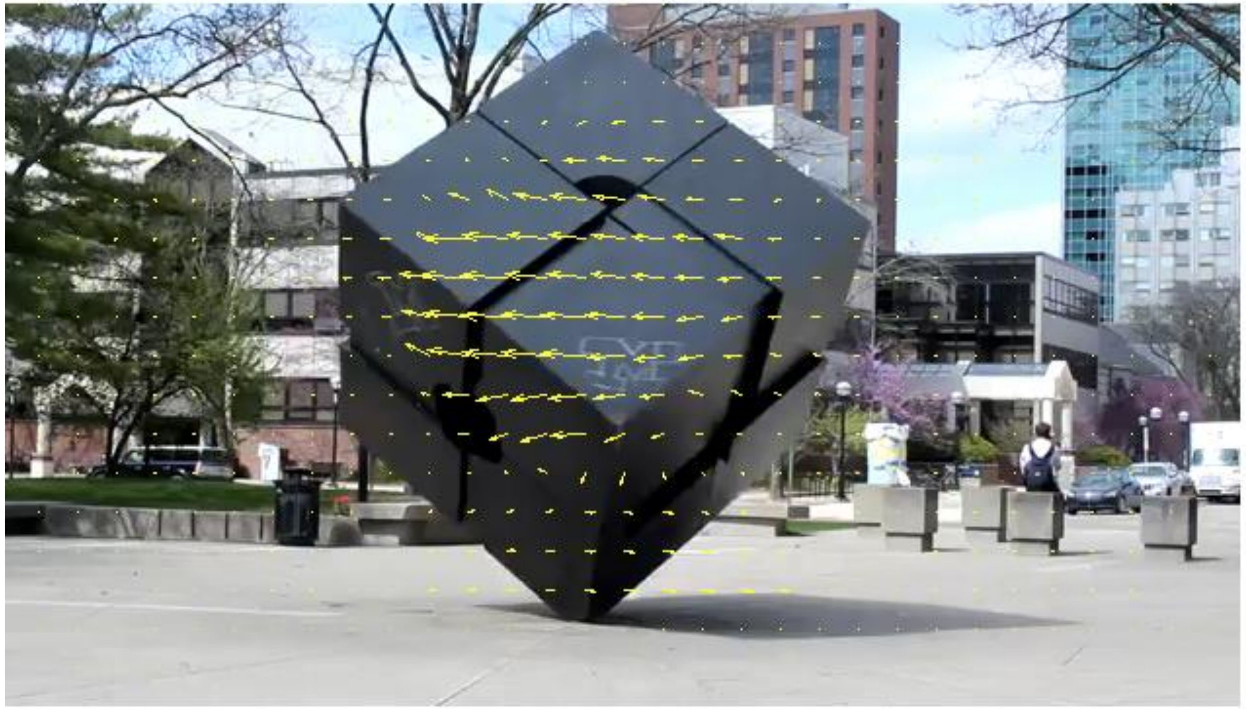
Using Gaussian Image Pyramids



Lucas-Kanade Optical Flow

Input Frames

Output



Optical Flow Field is sparse

Dense Optical Flow

Key Assumption:

- Most Objects in the World are rigid or deform elastically and move together coherently
- We expect the flow fields to be **SMOOTH**

Basic Idea:

- Enforce Brightness Constancy for every Pixel
- Enforce Smoothness Constraint for Flow Vectors

Enforce Brightness Constancy

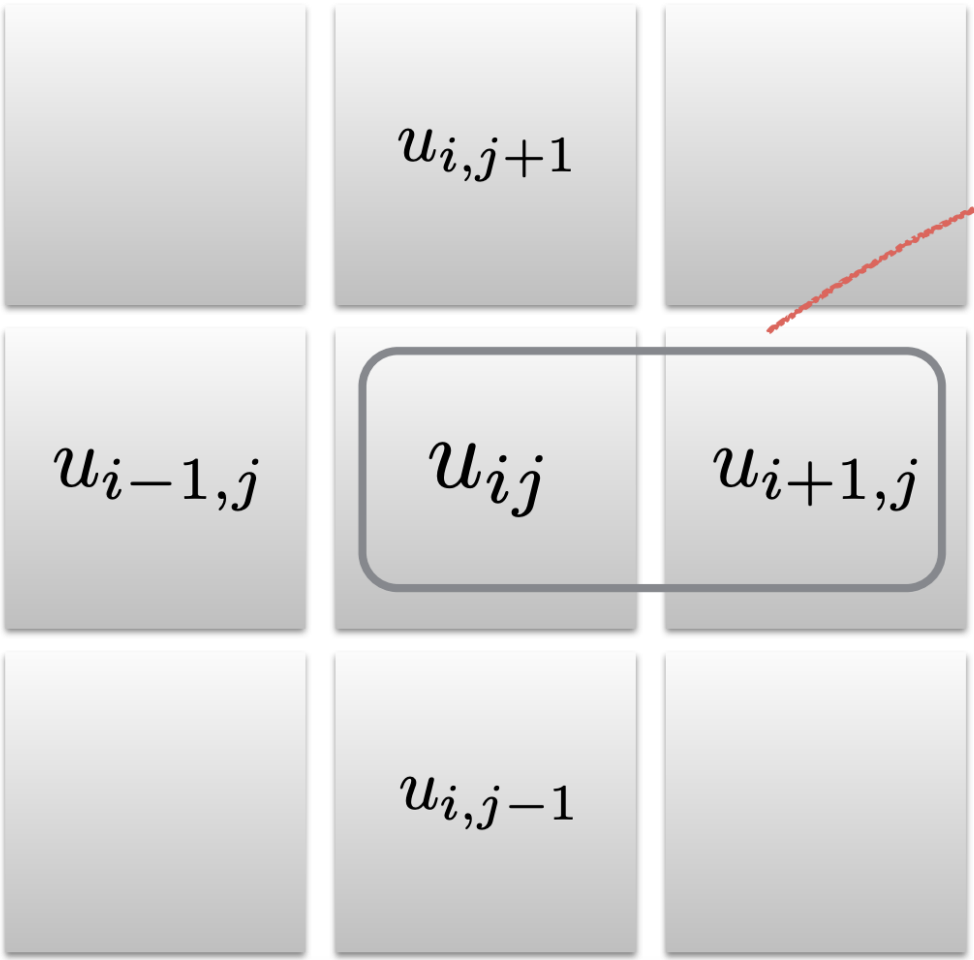
$$I_x u + I_y v + I_t = 0$$

$$\downarrow$$

For every Pixel (i,j):

$$\min_{u,v}\left[I_x(i,j)u_{ij} + I_y(i,j)v_{ij} + I_t(i,j)\right]^2$$

Enforce Smoothness Constraint



$$\min_{\boldsymbol{u}} (u_{i,j} - u_{i+1,j})^2$$

Same for ***v***

Objective Function

Horn-Schunck 1981

Brightness Constraint

Smoothness Constancy

$$\min_{\boldsymbol{u},\boldsymbol{v}} \sum_{i,j} \{E_d(i,j) + \lambda E_s(i,j)\}$$

Weight

λ = regularization constant (larger → more smooth optical flow)

Brightness Constancy

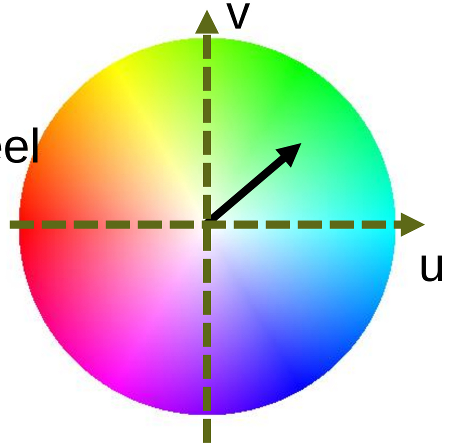
$$E_d(i,j) = \left[I_x(i,j)u_{ij} + I_y(i,j)v_{ij} + I_t(i,j)\right]^2$$

Smoothness Constraint

$$E_s(i,j) = \frac{1}{4}\left[\left(u_{ij} - u_{i+1,j}\right)^2 + \left(u_{ij} - u_{i,j+1}\right)^2 + \left(v_{ij} - v_{i+1,j}\right)^2 + \left(v_{ij} - v_{i,j+1}\right)^2\right]$$

Horn-Schunck Optical Flow

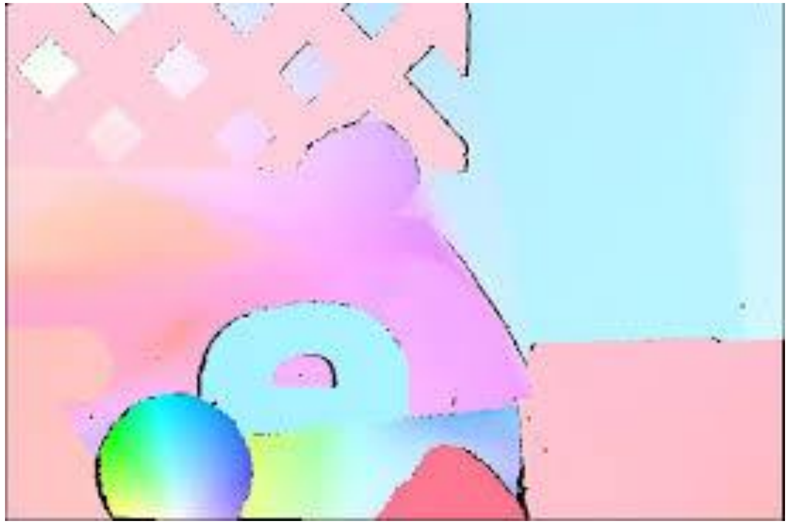
Color Wheel



t

t+1

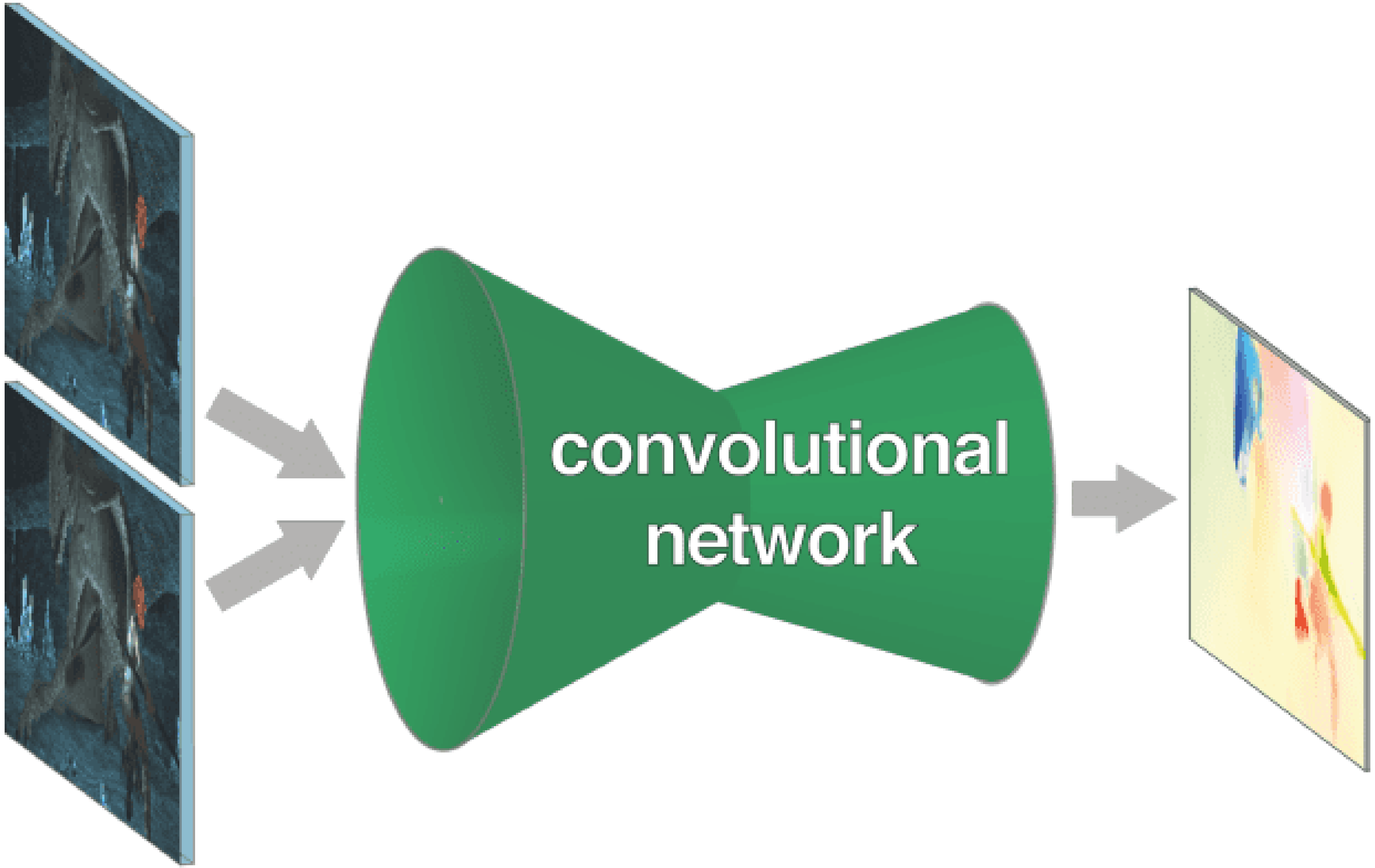
Flow Vectors visualized with Colors



Optical Flow Field is dense

JOHANNES KEPLER
UNIVERSITY LINZ

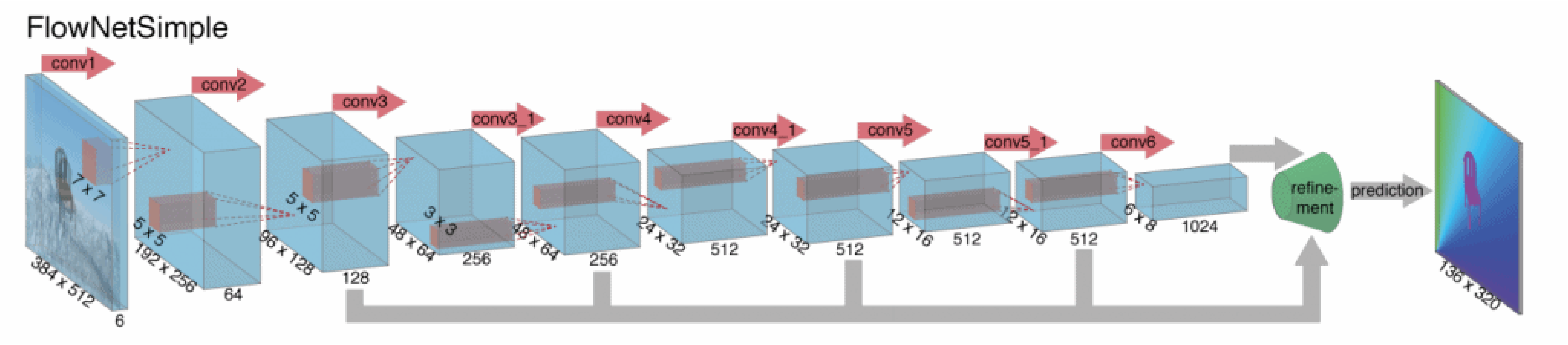
Optical Flow and Machine Learning



Encoder+Decoder Architectures (e.g. U-Nets)

Example: FlowNetS (Simple)

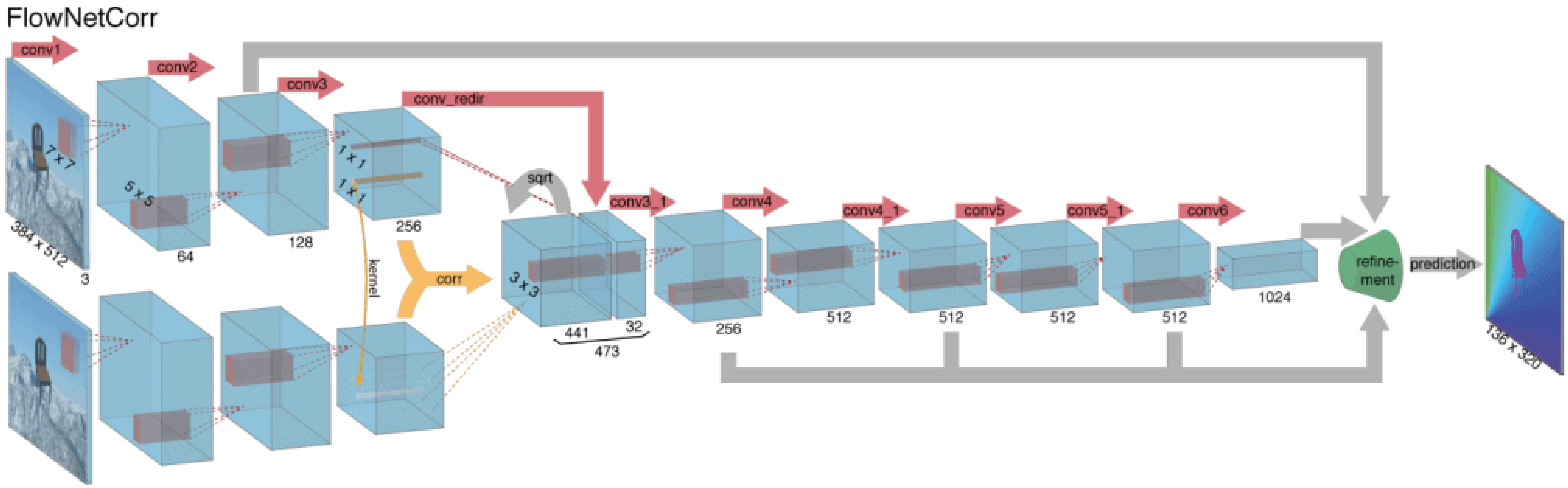
https://lmb.informatik.uni-freiburg.de/Publications/2015/DFIB15/flownet.pdf



Input: Tensor of 2 RGB Images

Example: FlowNetCorr (Correlation)

https://lmb.informatik.uni-freiburg.de/Publications/2015/DFIB15/flownet.pdf



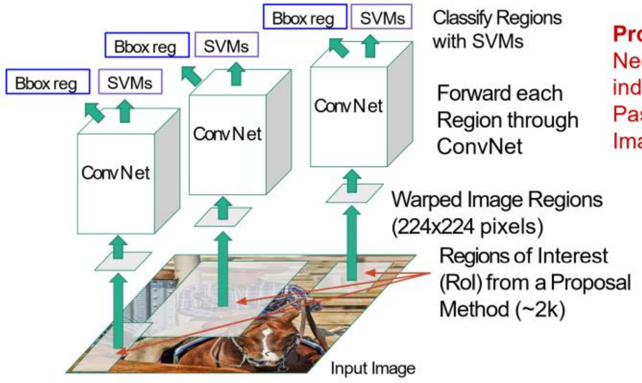
Input: 2 Tensors of individual RGB Images (Feature Maps are computed later → Correlation Layer)

Course Overview

CW	Topic	Date	Place	Lab
41	Introduction and Course Overview	08.10.2024	Zoom	Lab 1
42	Capturing Digital Images	15.10.2024	Zoom	Lab 2
43	Digital Image Processing	22.10.2024	Zoom	Assignment 1
44	Machine Learning	29.10.2024	Zoom	
45	Feature Extraction	05.11.2024	Zoom	Open Lab 1
46	Segmentation	12.11.2024	Zoom	Assignment 2
47	Optical Flow	19.11.2024	Zoom	Open Lab 2
48	Object Detection	26.11.2024	Zoom	Assignment 3
49	Multi-View Geometry	03.12.2024	Zoom	Open Lab 3
50	3D Vision	10.12.2024	Zoom	Assignment 4
3	Trends in Computer Vision	14.01.2025	Zoom	Open Lab 4
4	Q&A	21.01.2025	Zoom	
5	Exam	28.01.2025	HS1 (Linz), S1/S3 (Vienna), S5 (Bregenz)	
9	Retry Exam	25.02.2025	HS1 (Linz), S1/S3 (Vienna), S5 (Bregenz)	

Next Week: Object Detection

Regional-Based (R)-CNN

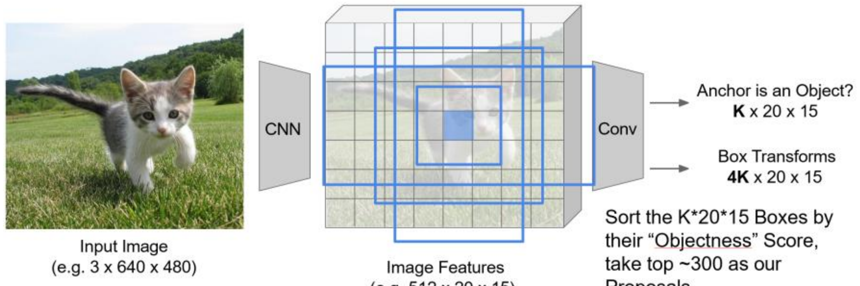


Problem: Very slow! Need to do ~2k independent forward Passes for each Image!

Predict "Corrections" to the RoI: 4 Numbers: (dx, dy, dw, dh)

Region Proposal Network

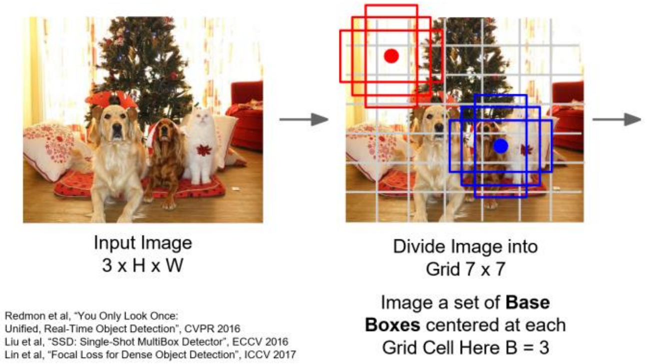
In practice use **K** different Anchor Boxes of different Size / Scale at each Point



Sort the K*20*15 Boxes by their "Objectness" Score, take top ~300 as our Proposals

JKU JOHANNES KEPLER UNIVERSITY LINZ

Single-Shot Object Detectors: YOLO/SSD/RetinaNet



Within each Grid Cell:

- Regress from each of the B Base Boxes to a final Box with 5 Numbers: (dx, dy, dh, dw, confidence)
- Predict Scores for each of C Classes (including Background as a Class)
- Looks a lot like RPN, but category-specific!
- Only a single Stage needed (Classification and prediction of BB in the same step)

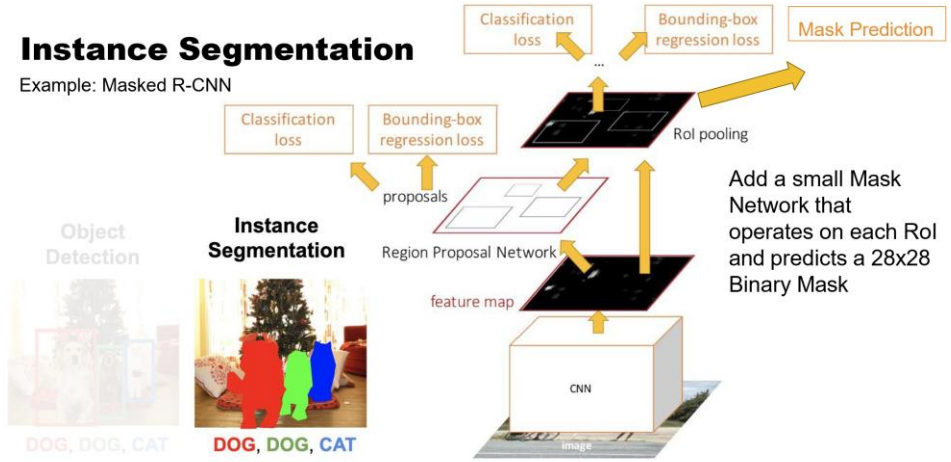
Output:
7 x 7 x (5 * B + C)

Redmon et al, "You Only Look Once: Unified, Real-Time Object Detection", CVPR 2016
Liu et al, "SSD: Single-Shot MultiBox Detector", ECCV 2016
Lin et al, "Focal Loss for Dense Object Detection", ICCV 2017

JKU JOHANNES KEPLER UNIVERSITY LINZ

Instance Segmentation

Example: Masked R-CNN



Add a small Mask Network that operates on each RoI and predicts a 28x28 Binary Mask

He et al, "Mask R-CNN", ICCV 2017

JKU JOHANNES KEPLER UNIVERSITY LINZ

JKU JOHANNES KEPLER UNIVERSITY LINZ

Thank You