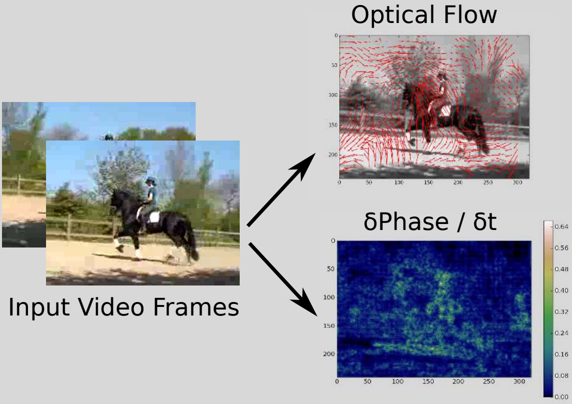
Making a Case for Learning Motion Representations with Phase

S. L. Pintea and J. C. van Gemert

Computer Vision Lab
Delft University of Technology, Netherlands

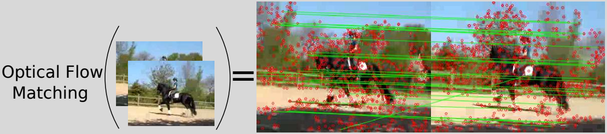
Eulerian versus Lagrangian Motion



- Lagrangian motion (optical flow) estimates the changes in position over time — match points/patches.
- Eulerian approach (phase variations over time) flux statistics over time — number of measurements stays constant.

Gains of Eulerian Motion

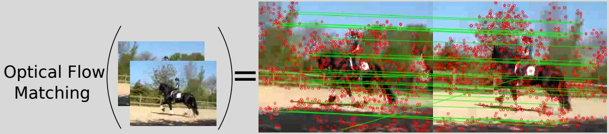
- **No feature extraction and matching** of patches between frames.
- **No missing correspondences**, rotating objects, occlusion.



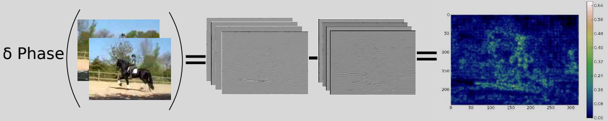
We focus on phase-based motion [Wadhwa et. al., SIGGRAPH, 2013].

Gains of Eulerian Motion

- **No feature extraction and matching** of patches between frames.
- **No missing correspondences**, rotating objects, occlusion.

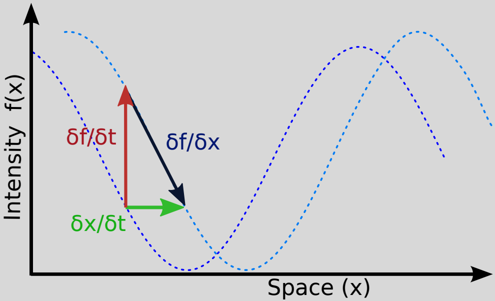


- **Constant number of measurements** over time.
- Phase information is **innate** to the image.



We focus on phase-based motion [Wadhwa et. al., SIGGRAPH, 2013].

Phase-based Motion



[Fleet et. al., IJCV, 1990] show that the temporal gradient of the phase over a spatially band-passed video over time directly relates to the motion field.

Phase-based Motion

- Local motion $\leftrightarrow$ local edges with different orientations.
- Using a steerable pyramid [Simoncelli, Trans. on Info. Theory, 1992] we decompose the image into localized subbands.
- The local-phase variations over time, give the local motion.

$$(G_\sigma^\theta + iH_\sigma^\theta) \otimes I(x,y) = A_\sigma^\theta(x,y)\,e^{i\phi_\sigma^\theta(x,y)}$$

Image Channels

Phase-based Motion

- Local motion $\leftrightarrow$ local edges with different orientations.
- Using a steerable pyramid [Simoncelli, Trans. on Info. Theory, 1992] we decompose the image into localized subbands.
- The local-phase variations over time, give the local motion.

$$(G_\sigma^\theta + iH_\sigma^\theta) \otimes I(x, y) = A_\sigma^\theta(x, y)e^{i\phi_\sigma^\theta(x,y)}$$

Complex Filters
scale - σ
orientation - θ

Image
Channels

Phase-based Motion

- Local motion $\leftrightarrow$ local edges with different orientations.
- Using a steerable pyramid [Simoncelli, Trans. on Info. Theory, 1992] we decompose the image into localized subbands.
- The local-phase variations over time, give the local motion.

$$(G_\sigma^\theta + iH_\sigma^\theta) \otimes I(x,y) = A_\sigma^\theta(x,y)e^{i\phi_\sigma^\theta(x,y)}$$

Complex Filters
scale - σ
orientation - θ

Image
Channels

Amplitude

Phase-based Motion

- Local motion $\leftrightarrow$ local edges with different orientations.
- Using a steerable pyramid [Simoncelli, Trans. on Info. Theory, 1992] we decompose the image into localized subbands.
- The local-phase variations over time, give the local motion.

$$(G_\sigma^\theta + iH_\sigma^\theta) \otimes I(x,y) = A_\sigma^\theta(x,y)\, e^{i\phi_\sigma^\theta(x,y)}$$

Complex Filters
scale - σ
orientation - θ

Image
Channels

Amplitude

Phase

4 Phase Applications

1. Action Recognition
   — Starting point: Two stream network [Simonyan et. al., CVPR, 2014].

2. Motion Prediction in Static Images
   — Starting point: RCNNs [Liang et. al., CVPR 2015], [Pinheiro et.al., ICML 2014].

3. *Motion Transfer in Static Images
   — Starting point: Artistic style transfer [Gatys et. al., 2015].

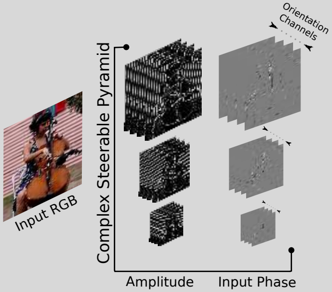
4. *Motion Transfer in Videos
   — Starting point: Artistic style transfer in videos [Ruder et. al., 2016]

4 Phase Applications

- Convolve the input frame/image with the steerable complex filters.

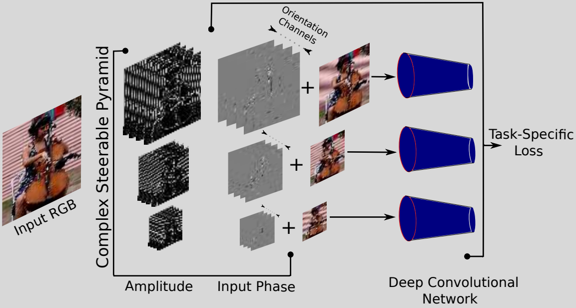
4 Phase Applications

- Get oriented amplitude and phase at different scales.

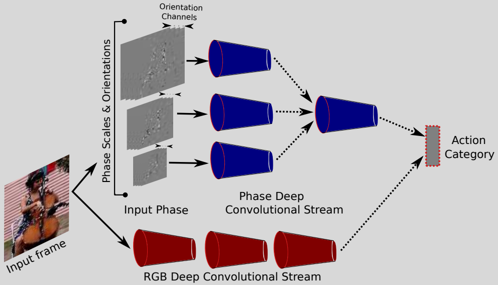


4 Phase Applications

- Add the image appearance info into a deep net formulation.



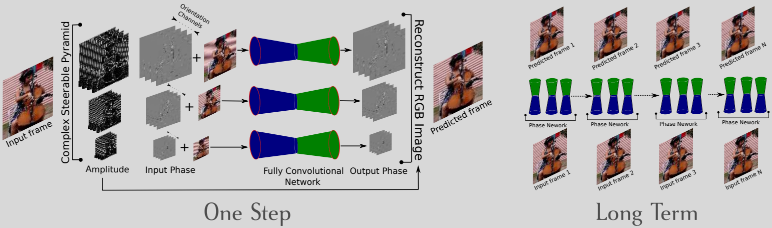
Phase-based Action Recognition



Experimental Setup:

- Compare against [Simonyan et. al., CVPR, 2014].
- Action Recognition datasets: HMDB51 and UCF101.

Phase-based Motion Prediction in Static

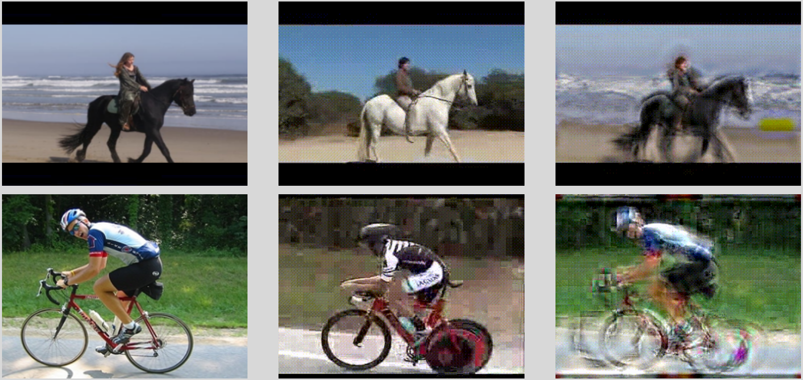


Experimental Setup:

- Compare against optical-flow motion prediction [Walke, ICCV, 2015].
- Motion prediction datasets: HMDB51 and UCF101.

Phase-based Motion Transfer in Static

Given an input **static image** transfer the style of a video motion.



Static   Video   Transferred

Phase-based Motion Transfer

In static:

- Correctly aligning the parts of the objects that have similar motion is essential for the motion transfer task.
- We propose to add a wight pixels-wise correlation to the "style loss" of [Gatys et. al., 2015].

In video:

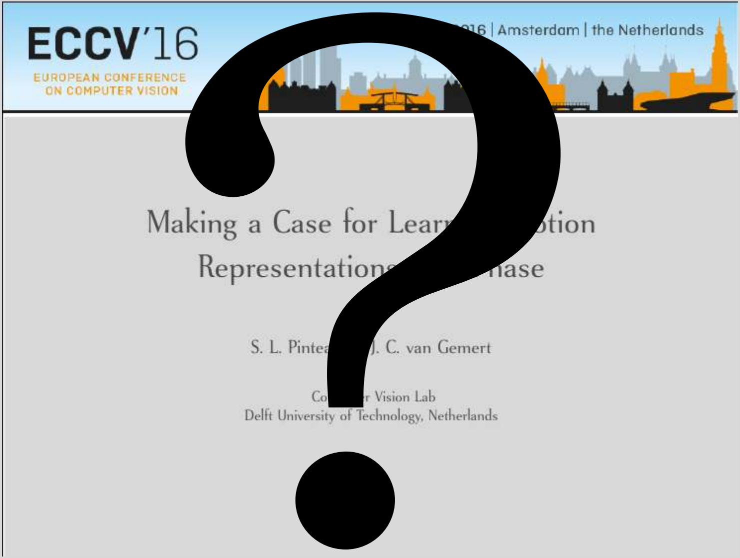
- Follow [Ruder et. al., 2016] and add a temporal constrain between the transferred frame phases.

Conclusions

- We propose an Eulerian — phase-based — approach to motion representation learning.
- We argue for the intrinsic stability of the phase-based motion description.
- We explore a set of motion learning tasks in an Eulerian setting:
    - (a) action recognition,
    - (b) motion prediction in static images,
    - (c) motion transfer from a video to a static image and
    - (d) motion transfer in videos.

  And we propose a phase-based approach.
- We provide a small proof of concept.

Thank you


ECCV'16
EUROPEAN CONFERENCE ON COMPUTER VISION
16 | Amsterdam | the Netherlands
Making a Case for Lear ption
Representation hase
S. L. Pinte J. C. van Gemert
Co r Vision Lab
Delft University of Technology, Netherlands