EECS 4422/5323 Computer Vision

Spatiotemporal Vision 1

Calden Wloka

13 November, 2019

Announcements

- Assignment 2 due date changed to Monday
- Project Demo Schedule Posted
- Course evaluations at end of class Nov. 20th, open through to Dec. 4th

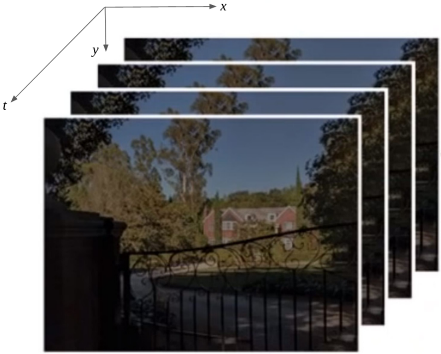
Outline

- Introduction to Spatiotemporal Processing
- Optical Flow
- Example Problems and Approaches

Video Processing adds an Extra Dimension

The shift to video adds an additional dimension to the problem.

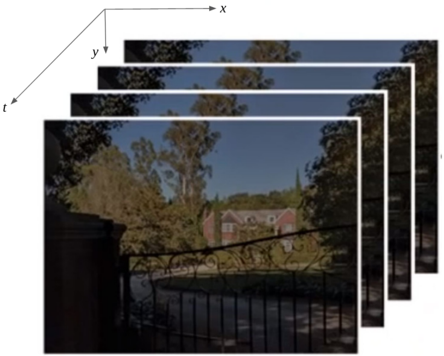
- A time axis is added to our data



Video Processing adds an Extra Dimension

The shift to video adds an additional dimension to the problem.

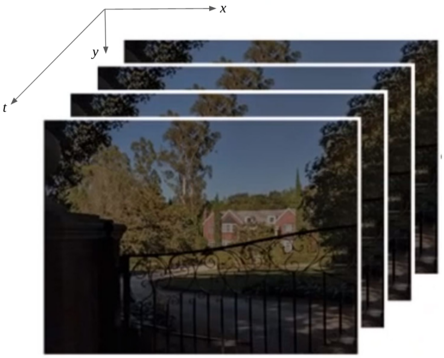
- A time axis is added to our data
- In addition to standard spatial features, we can explore features in the temporal domain



Video Processing adds an Extra Dimension

The shift to video adds an additional dimension to the problem.

- A time axis is added to our data
- In addition to standard spatial features, we can explore features in the temporal domain
- Usually we want to retain the power of our spatial features, so combine to form *spatiotemporal* features

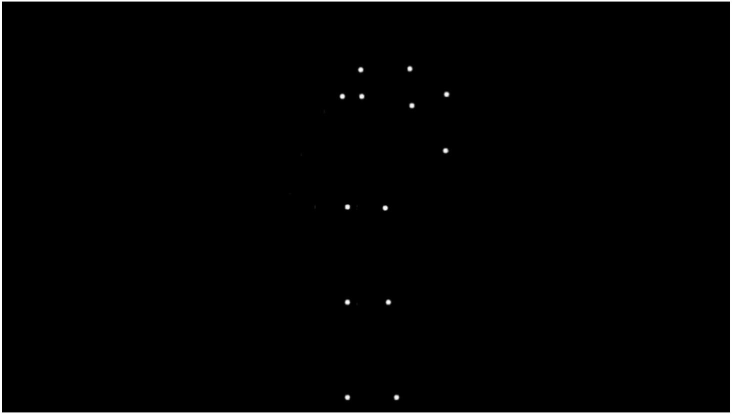


Motion is a Powerful Feature

Coherent motion can be a very powerful feature for identifying objects and initiating feature binding.

Sparse Motion

Even over a sparse field of dots, we can often extract highly meaningful information through motion.



A Primary Challenge in Video Processing: Speed

Spatiotemporal filtering is very challenging from a timing perspective. We often want *real-time* systems which operate at our video framerate, but the combinatorics of adding a temporal dimensions vastly increases our typical feature complexity.

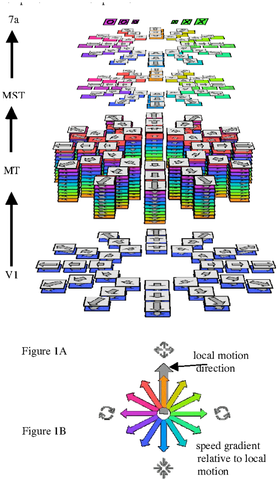
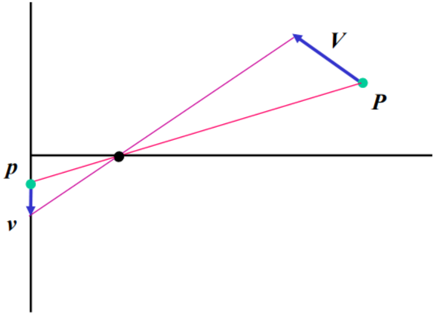


Image source: Tsotsos *et al.*, 2004

Motion vs. Optical Flow

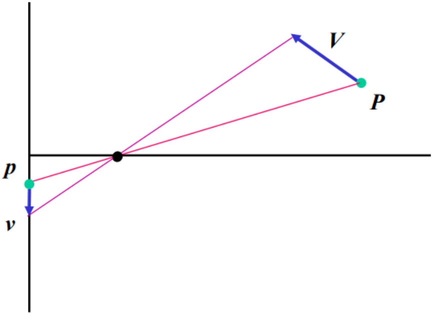
Motion typically refers to movement within the three dimensional world. A visual sensor only sees a projection of this movement, however.



Motion vs. Optical Flow

Motion typically refers to movement within the three dimensional world. A visual sensor only sees a projection of this movement, however.

The change in the image elicited from a given vector of motion is referred to as *optical flow*.

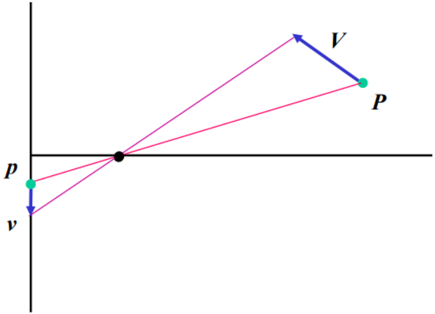


Motion vs. Optical Flow

Motion typically refers to movement within the three dimensional world. A visual sensor only sees a projection of this movement, however.

The change in the image elicited from a given vector of motion is referred to as *optical flow*.

Optical flow and motion will usually be strongly correlated, but there are possibly degenracies for which the correspondence is greatly diminished.



Optical Flow Consistency

In a similar vein to the challenge of correspondence in stereo scenes in which we can't know *a priori* how consistent a scene should be, optical flow can vary greatly across a scene depending on the movement of individual moving parts.

Optical Flow Consistency

In a similar vein to the challenge of correspondence in stereo scenes in which we can't know *a priori* how consistent a scene should be, optical flow can vary greatly across a scene depending on the movement of individual moving parts.

To add to the challenge, even within a single cohesive object the movement patterns may be highly variable depending on the type of motion and articulation or deformation of the object.

Optical Flow Consistency

In a similar vein to the challenge of correspondence in stereo scenes in which we can't know *a priori* how consistent a scene should be, optical flow can vary greatly across a scene depending on the movement of individual moving parts.

To add to the challenge, even within a single cohesive object the movement patterns may be highly variable depending on the type of motion and articulation or deformation of the object.

Example video

A Specific Challenge: Egomotion

Egomotion refers to the movement of the visual agent or camera with respect to the environment.

Ideally, we want to take this motion into account when computing optical flow and remove it from our estimates of external motion (and avoid blurring the image due to camera motion).

A Specific Challenge: Egomotion

Egomotion refers to the movement of the visual agent or camera with respect to the environment.

Ideally, we want to take this motion into account when computing optical flow and remove it from our estimates of external motion (and avoid blurring the image due to camera motion).

Note that this isn't just an average flow across the field, but is an average flow *with respect to the observer*.

A Specific Challenge: Egomotion

Egomotion refers to the movement of the visual agent or camera with respect to the environment.

Ideally, we want to take this motion into account when computing optical flow and remove it from our estimates of external motion (and avoid blurring the image due to camera motion).

Note that this isn't just an average flow across the field, but is an average flow *with respect to the observer*.

Some examples of flow fields giving the impression of rapid egomotion can be seen from 90's Sci Fi: Sliders, Stargate: SG1

Human Perception: The Vestibular System

Humans subconsciously take advantage of an additional sense when it comes to sensing egomotion: the *vestibular system*.

The vestibular system consists of a set of (nearly) orthogonal fluid filled tubes and chambers filled with hair cells, which can detect the motion of the fluid (the same sensors used in audition).

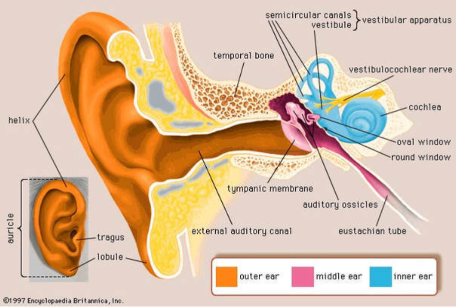


Image Source: Encyclopedia Britannica

Some Notes on the Vestibular System

- The vestibular detects acceleration, not velocity

Some Notes on the Vestibular System

- The vestibular detects acceleration, not velocity
- Continuous or intense acceleration can lead to intertial movement of the fluid, causing a residual sense of motion

Some Notes on the Vestibular System

- The vestibular detects acceleration, not velocity
- Continuous or intense acceleration can lead to intertial movement of the fluid, causing a residual sense of motion
- Our optical system and vestibular system are tightly connected by a number of important reflexes

Artificial “Vestibular” Systems: IMUs

Some artificial systems can use the same principles as our vestibular system to try and improve temporal processing through Intertial Measurement Units (IMUs).

For example, see this example video by Tsotsos *et al.*, 2015.

Classical Methods: Lucas-Kanade

The Lucas-Kanade method (Lucas & Kanade, 1981) is one of the earliest optical flow methods.

Classical Methods: Lucas-Kanade

The Lucas-Kanade method (Lucas & Kanade, 1981) is one of the earliest optical flow methods.

Assumptions:

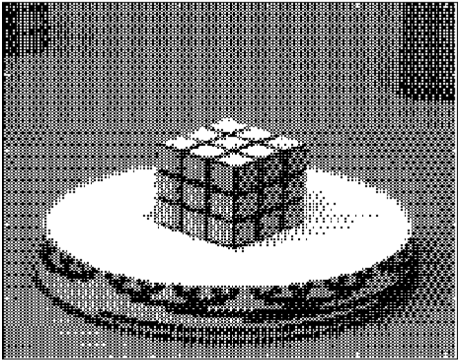
- Displacement of the image contents between frames is small
- Displacement is approximately the same for all pixels in a small neighbourhood

Lucas-Kanade Continued

For a given window W in image I, we solve for the velocity vector (v_x, v_y) via least squares over the system of equations:

$$\forall w_i \in W, \quad \frac{\partial I}{\partial x}(w_i)v_x + \frac{\partial I}{\partial y}(w_i)v_y = -\frac{\partial I}{\partial t}(w_i)$$

Resultant Flowfield



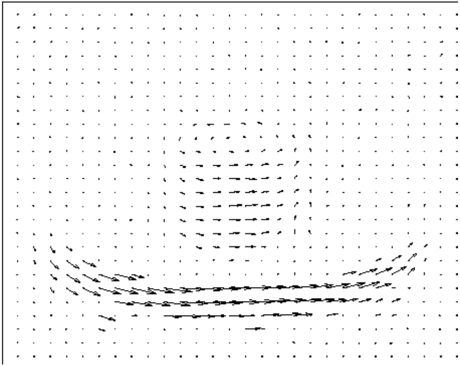
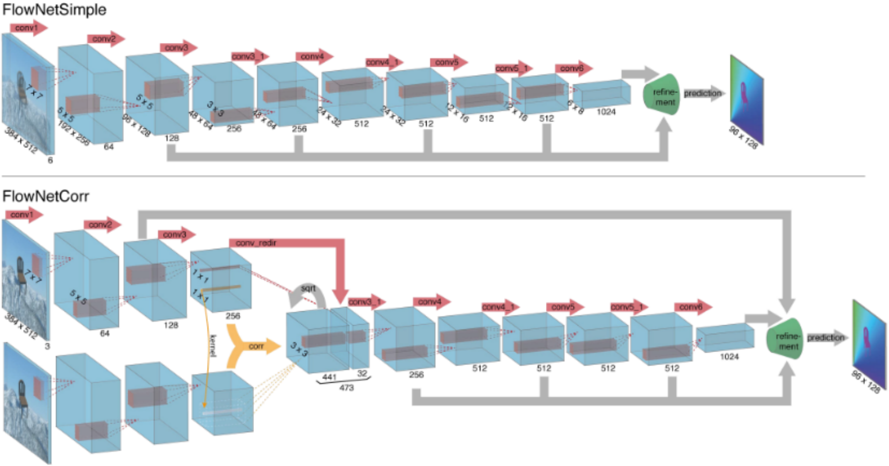


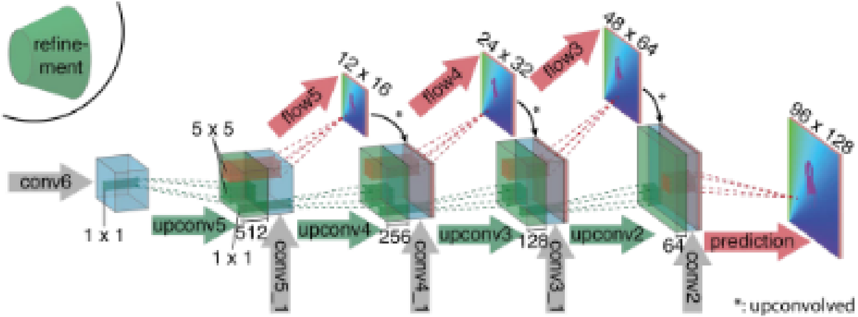
Image Source: Bauer *et al.*, 2006

Deep Learning Example: FlowNet



Two different versions of FlowNet
Image Source: Dosovitskiy *et al.*, 2015

FlowNet Architecture Continued



Expansion of the refinement part of the network
Image Source: Dosovitskiy *et al.*, 2015

Flying Chairs

In addition to demonstrating the efficacy of convolutional networks to optical flow computations, one of the outputs of the FlowNet paper was the introduction of the “Flying Chairs” dataset, an interesting example of synthetic data proving more effective for training than natural data.

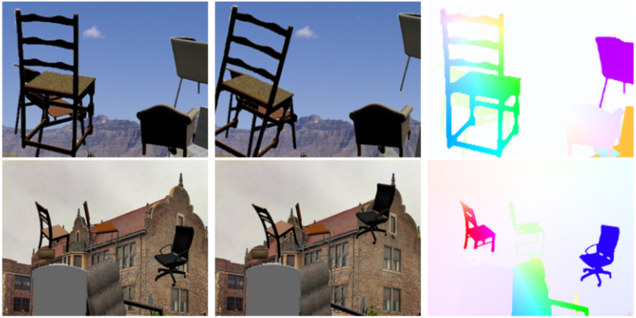


Image Source: Dosovitskiy *et al.*, 2015

Adversarial Optical Flow

Ranjan *et al.*, 2019 demonstrate that optical flow networks are also vulnerable to small patch adversarial attacks, including when printed and displayed in real-world scenes.

Image Source: Ranjan *et al.*, 2019