

Gesture Recognition using Accumulated Optic Flow Fields

Ole Rehmsen
May 29, 2009



Why gestures?

Advantages

- ▶ no device (remote control), no batteries
- ▶ more natural / intuitive ?

Drawbacks

- ▶ straining over time
- ▶ limited set of commands

Conclusion

Use gestures for applications with

- ▶ few often used commands
- ▶ sporadical user interactions



Two types of gestures

1. Natural Gestures

Depend largely on the way hand is molded.

Advantage:

- ▶ do not have to be learned

Drawbacks:

- ▶ mapping unclear
- ▶ individual variability
- ▶ require complex models

2. Artificial Gestures

Can be designed to depend largely on motion.

Advantages:

- ▶ can be designed to be easily distinguishable
- ▶ inherent mapping from gesture to command

Drawback:

- ▶ have to be learned and practiced



What do we aim at?

Fair results despite

- ▶ simple models
- ▶ simple algorithms
- ▶ simple numerics



The basic technique

1. Acquire image sequence
2. Compute sequence of optic flow fields
3. Combine into *Accumulated Optic Flow Field*
4. Match Eigenspace Representations



Challenges

I. Detection

- i. when does a gesture start, when does it end?
- ii. unintentional movement

2. Recognition

- i. changing distance from and angle towards camera
- ii. robustness vs discrimination



Detecting a gesture

Advantages

Drawbacks

Multimodal

new problems arise

No movement

easy to implement

not robust

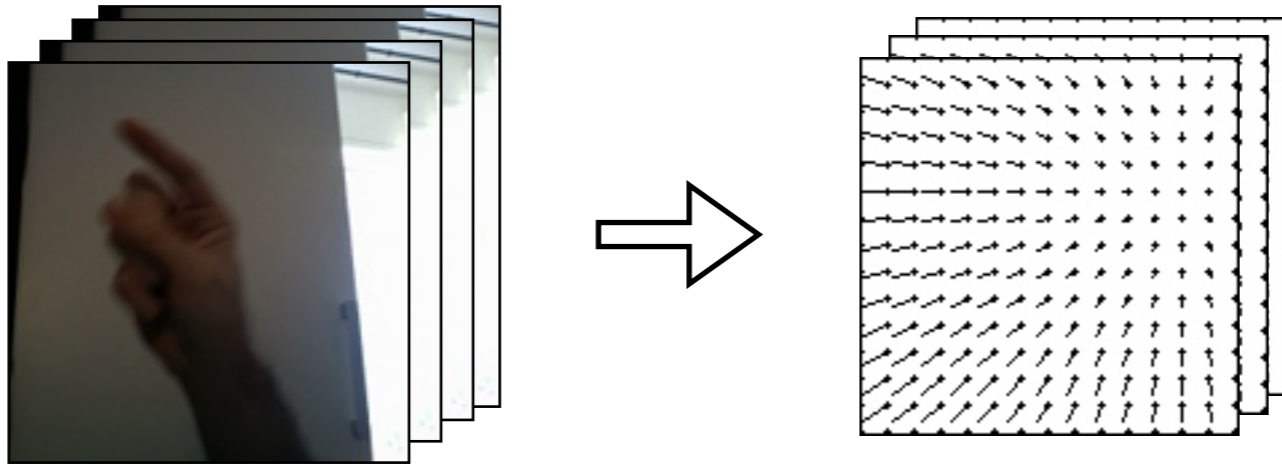
Opening gesture

can be chosen to be easily distinguishable

less intuitive



Sequence of Optic Flow Fields



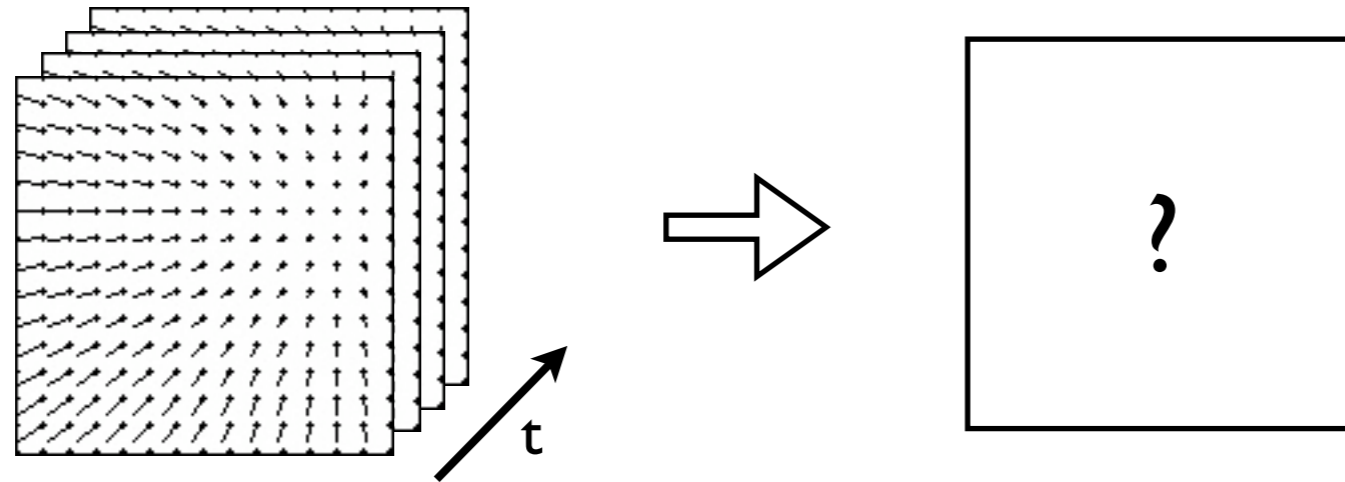
Requirements for the optic flow algorithm:

- ▶ can be efficiently computed
- ▶ yields dense flow fields
- ▶ works with large displacements?

Start off simple: Horn and Schunck



The Accumulated Optic Flow Field



Each frame is uninformative by itself

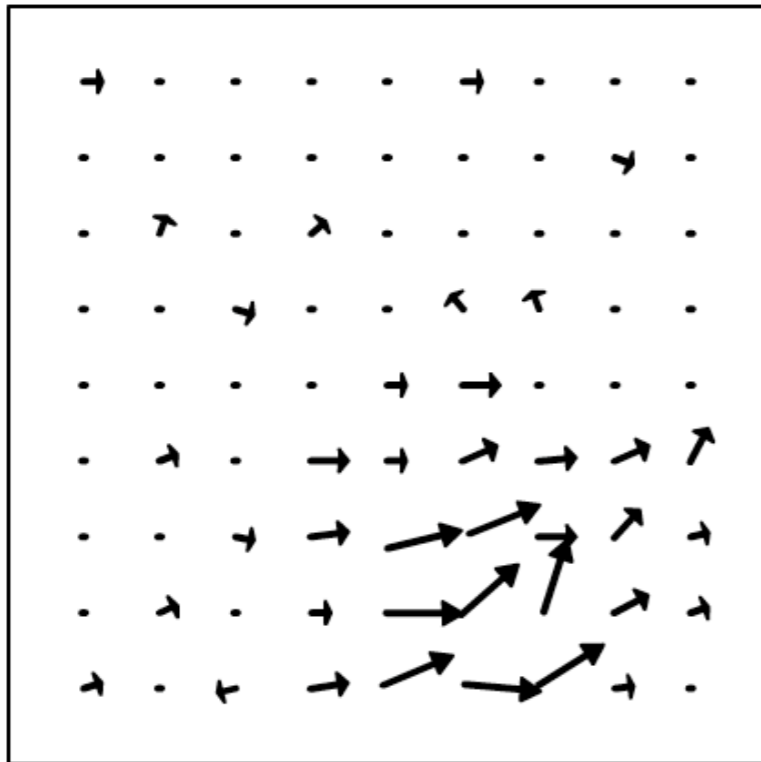
Appearance of flow field at time depends on

- ▶ frames per second
- ▶ speed at which gesture is performed

➔ sum up pixel-wise

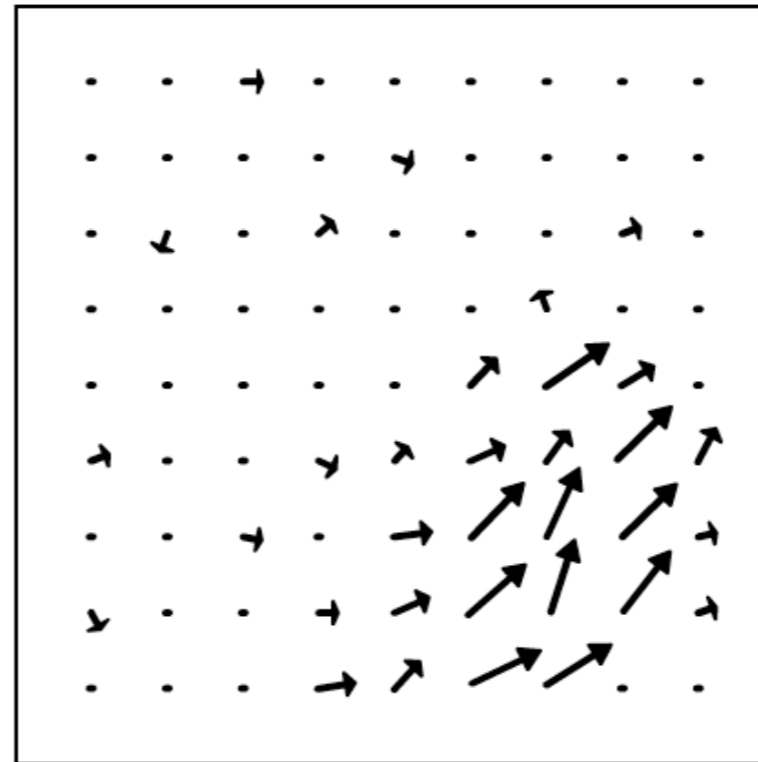


What could it look like?

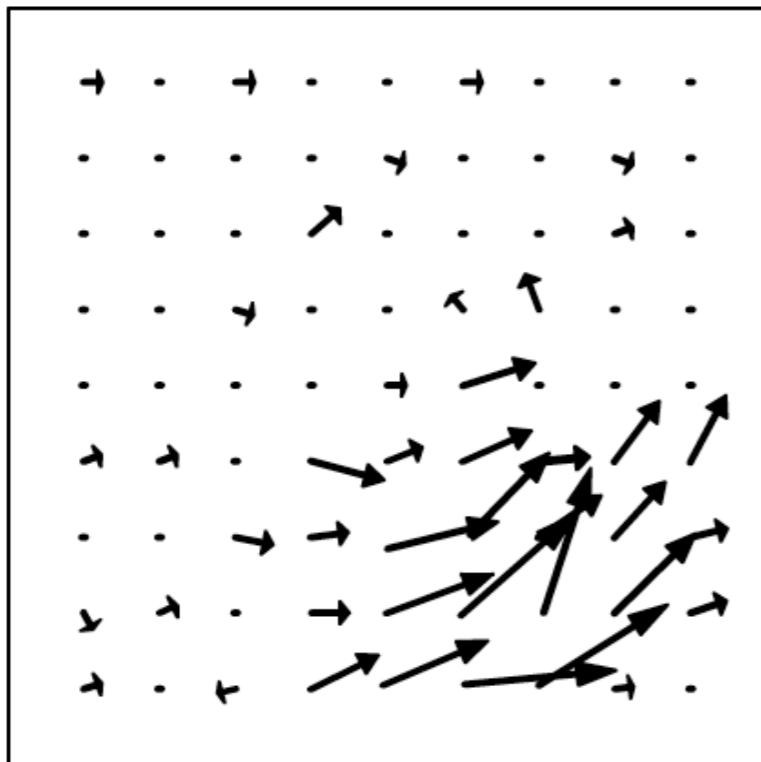


Optic Flow field, frame t

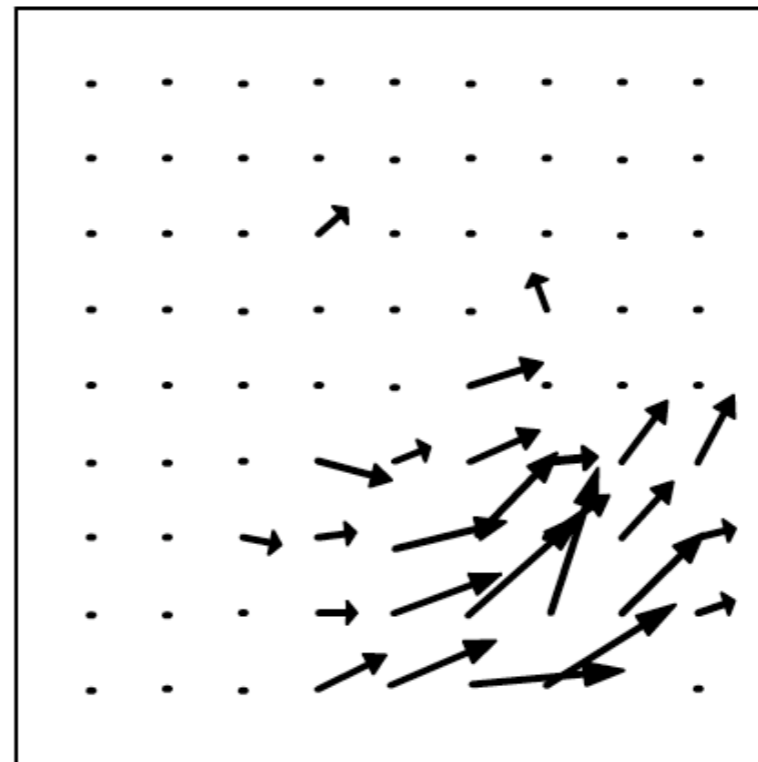
+



Optic Flow field, frame t+1



Accumulated Optic Flow field, frames t and t+1



Accumulated Optic Flow field with thresholding



AOFF - Details

How many frames to combine?

- A. all frames of one gesture
- B. all frames of fixed interval in time
- C. adaptively find discontinuities in time in flow field sequence

Normalization:

1. (double) threshold to cancel out noise
2. find smallest square that fully contains non-zero flow field
3. sample to some globally fixed size, i.e. 128 x 128 pixels
4. set mean 0 and squared intensity to 1



Eigenspace Method

Learning Phase:

1. record several appearances of each gesture and compute their AOFFs
2. compute the Eigenspace of each object and the representation of each appearance in it

Recognition Phase:

1. compute the representation of the AOFF of the new gesture in the Eigenspace of each object in the database
2. the gesture with the representations that has the smallest Euclidean distance to the new gesture is recognized.
3. if this distance exceeds a certain predefined threshold, no gesture is recognized



Eigenspace Method - Details

- ▶ we have flow components u and v at position x and y , possibly even t (depending on earlier choice)
- ▶ but: Eigenspace Method is defined for 1D Signals

Options:

- A. concatenate all into one long 1D signal
- B. treat u and v separately (yields improved results if average displacements in u and v differ strongly)



Challenges, revisited

I. Detection

- i. when does a gesture start, when does it end?
- ii. unintentional movement

2. Recognition

- i. changing distance from and angle towards camera
- ii. robustness vs discrimination



Literature

B. Horn, B. Schunck: Determining optical flow. In *Artificial Intelligence*, Vol. 17, pp 185-203, 1981.

E. Trucco, A. Veri: *Introductory Techniques for 3-D Computer Vision*. Prentice-Hall, Upper Saddle River, 1998.

T. F. Cootes, C. J. Taylor: *Statistical Models of Appearance for Computer Vision*. Technical Report, Wolfson Image Analysis Unit, University of Manchester, England, 2001 7 8

