Interaction with an autonomous agent

Jim Little
Laboratory for
Computational Intelligence
Computer Science
University of British Columbia
Vancouver BC Canada

Observing people

Jim Little
Work with Jeff Boyd and Jesse Hoey

Activities

- Detection and recognition of action type from motion characteristics
- Biometrics: recognizing persons by
  - fingerprint, retinal scan, iris, motion
- Surveillance and monitoring
- Situation awareness

Action and Motion

- Recognizing people by their gait
- Identifying gestures and expressions
- Analyzing image sequences to identify activities of players from their trajectories – tracking and identification of context

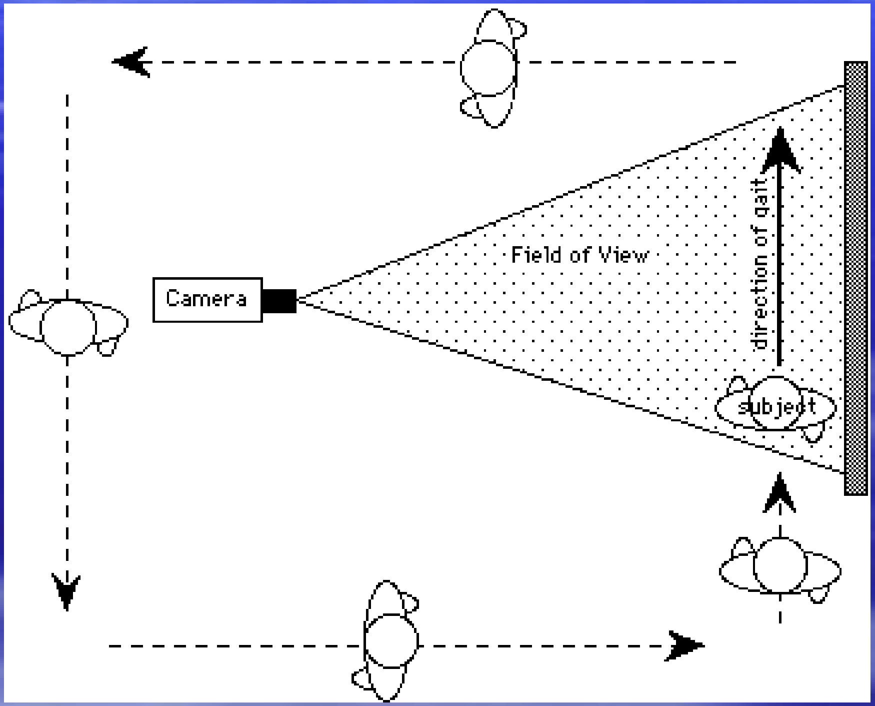
Recognizing People by Their Gait: The Shape of Motion

- James J. Little
- Jeffrey E. Boyd

Overview

We have developed a novel vision system that can recognize people by the way they walk. The system computes optical flow for an image sequence of a person walking, and then characterizes the shape of the motion with a set of sinusoidally-varying scalars. Feature vectors composed of the phases of the sinusoids are able to discriminate among people.

Apparatus



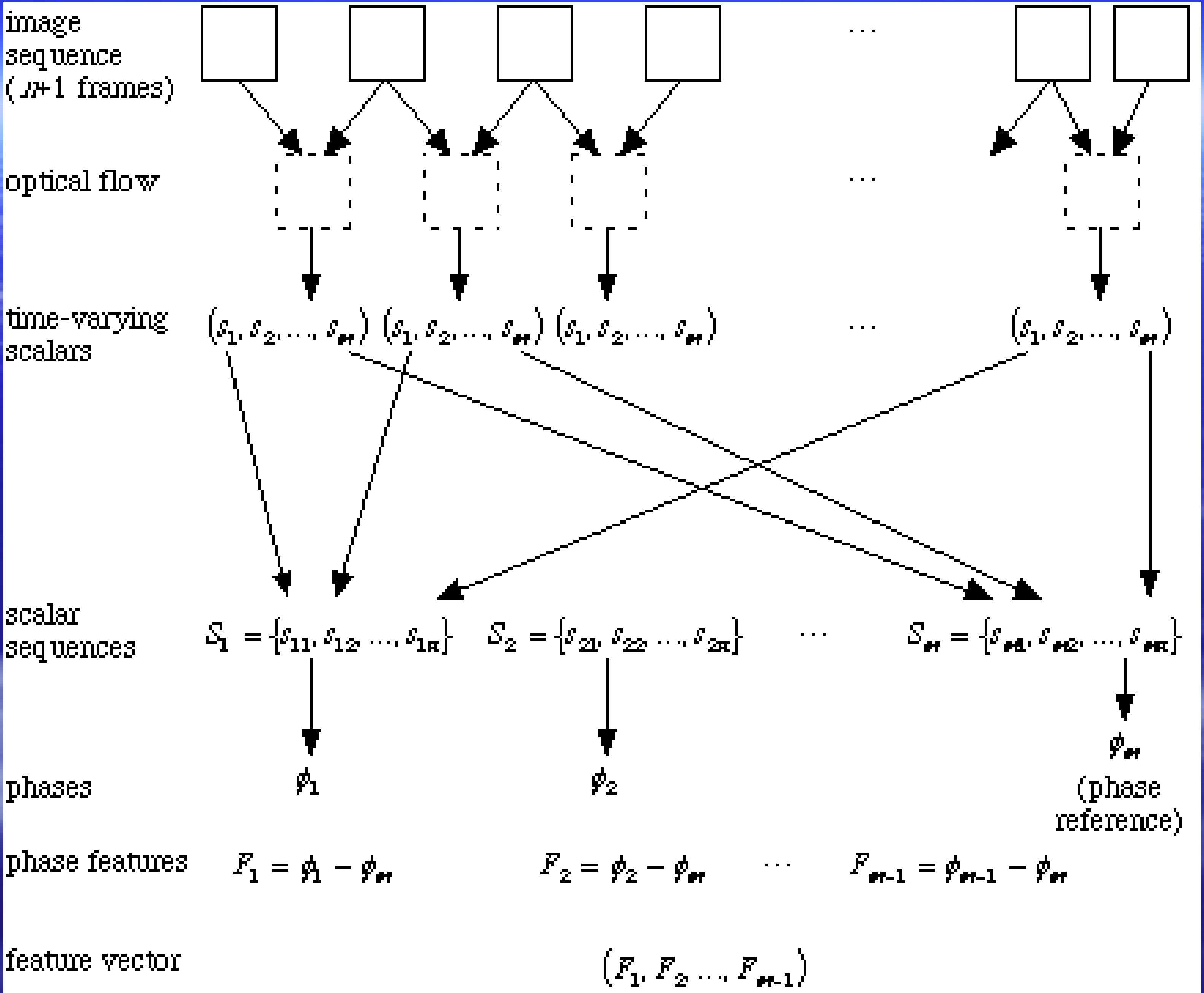


image sequence (n+1 frames)
optical flow
time-varying scalars
$(s_1, s_2, \ldots, s_m)$
scalar sequences
$S_1 = \{s_{11}, s_{12}, \ldots, s_{1n}\}$
$S_2 = \{s_{21}, s_{22}, \ldots, s_{2n}\}$
$S_m = \{s_{m1}, s_{m2}, \ldots, s_{mn}\}$
phases
ϕ_1
ϕ_2
ϕ_m
(phase reference)
phase features
$F_1 = \phi_1 - \phi_m$
$F_2 = \phi_2 - \phi_m$
$F_{m-1} = \phi_{m-1} - \phi_m$
feature vector
$(F_1, F_2, \ldots, F_{m-1})$

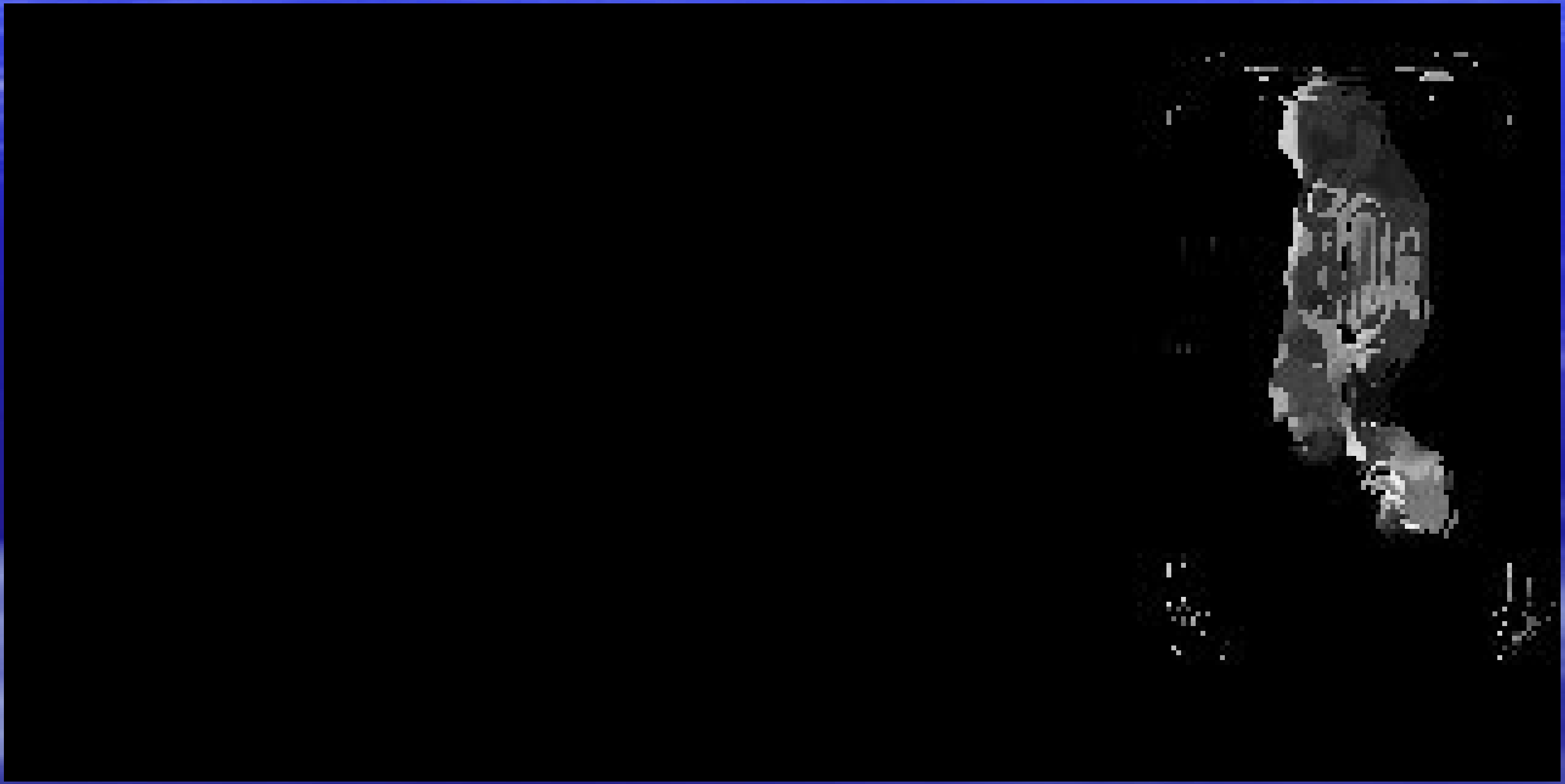
Input Sequence

Optical Flow

Optical flow (Little, Bulthoff and Poggio): n frames of (u,v) data, where u is the x flow and v is the y flow.

Points (white) where flow is non-zero.

Magnitude of flow

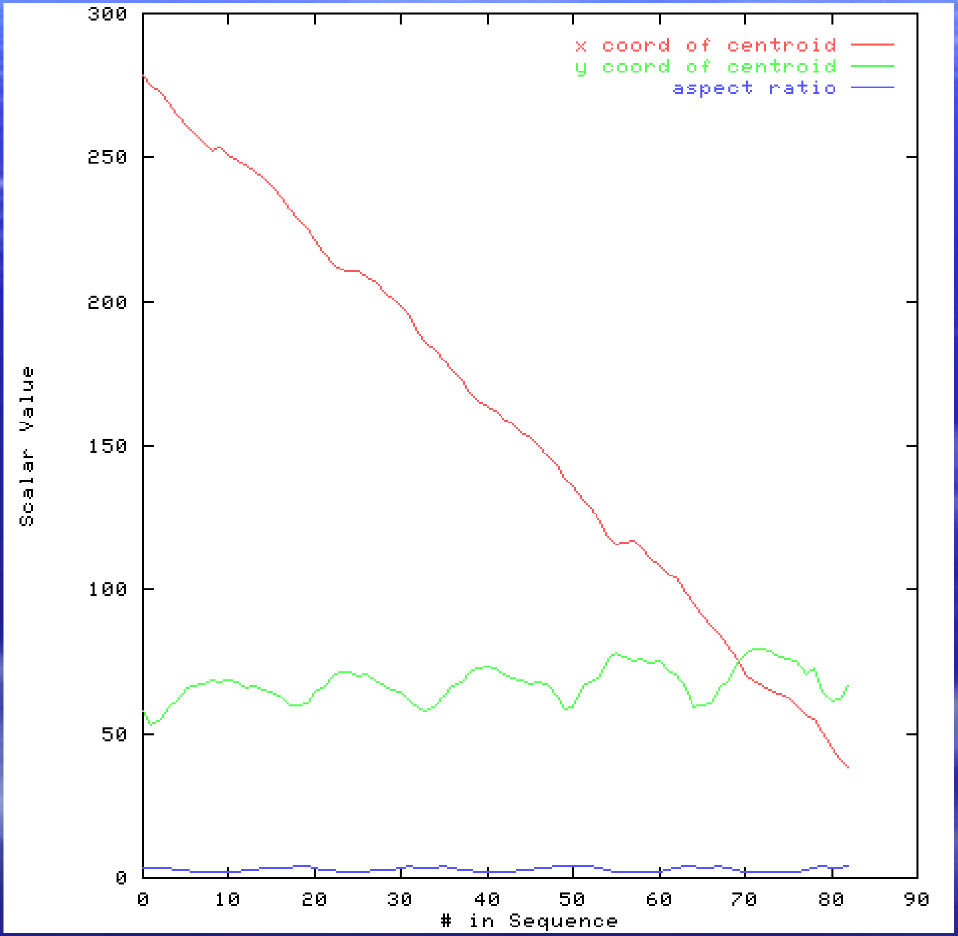


U component of flow (x-direction)

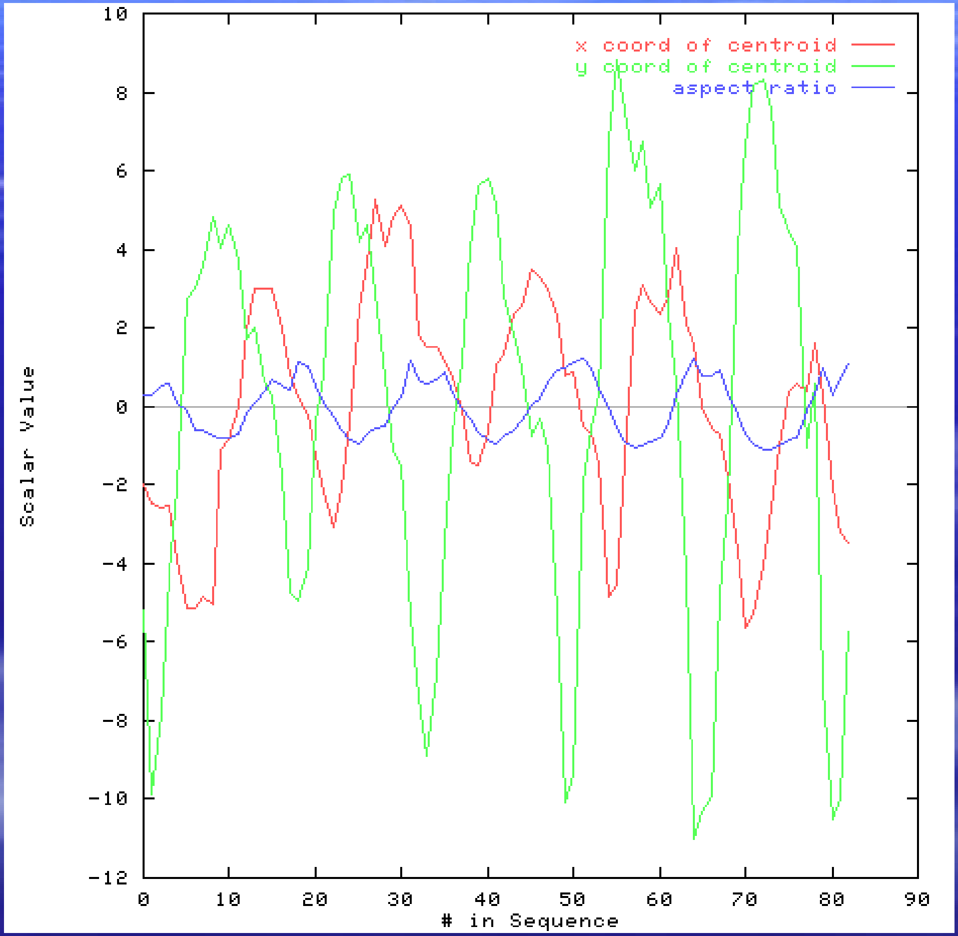
V component of flow (y-direction)

Fitting Ellipses to motion

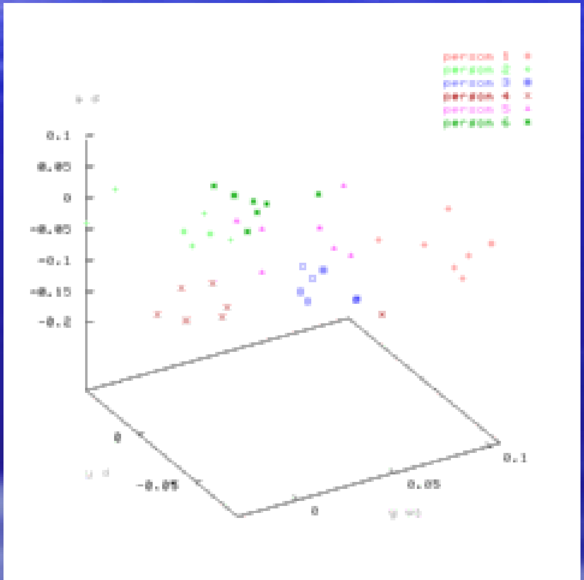
X and Y position over time

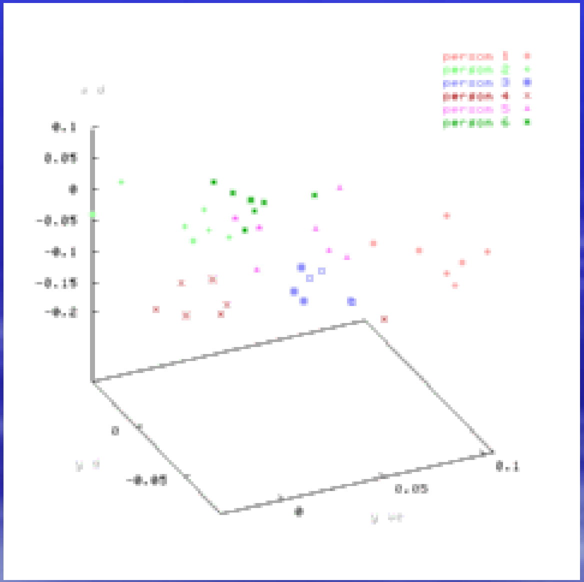


Scalar Signals

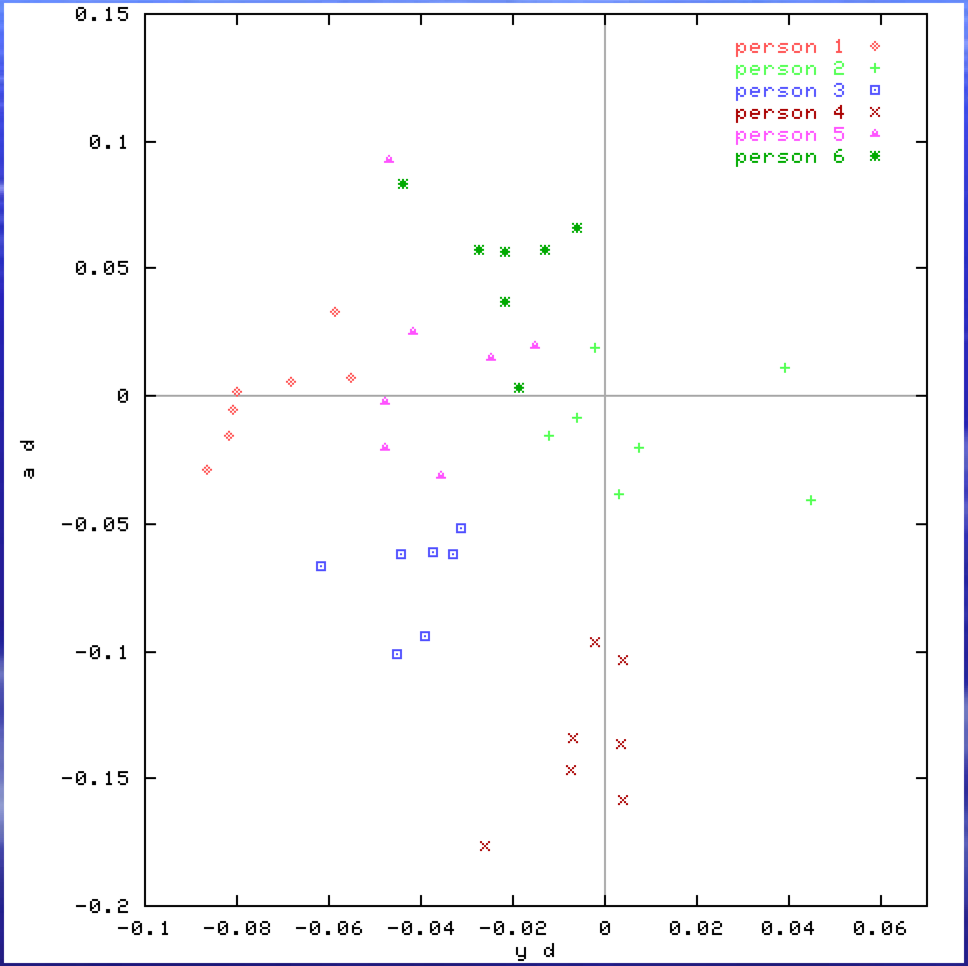


Stereo Features



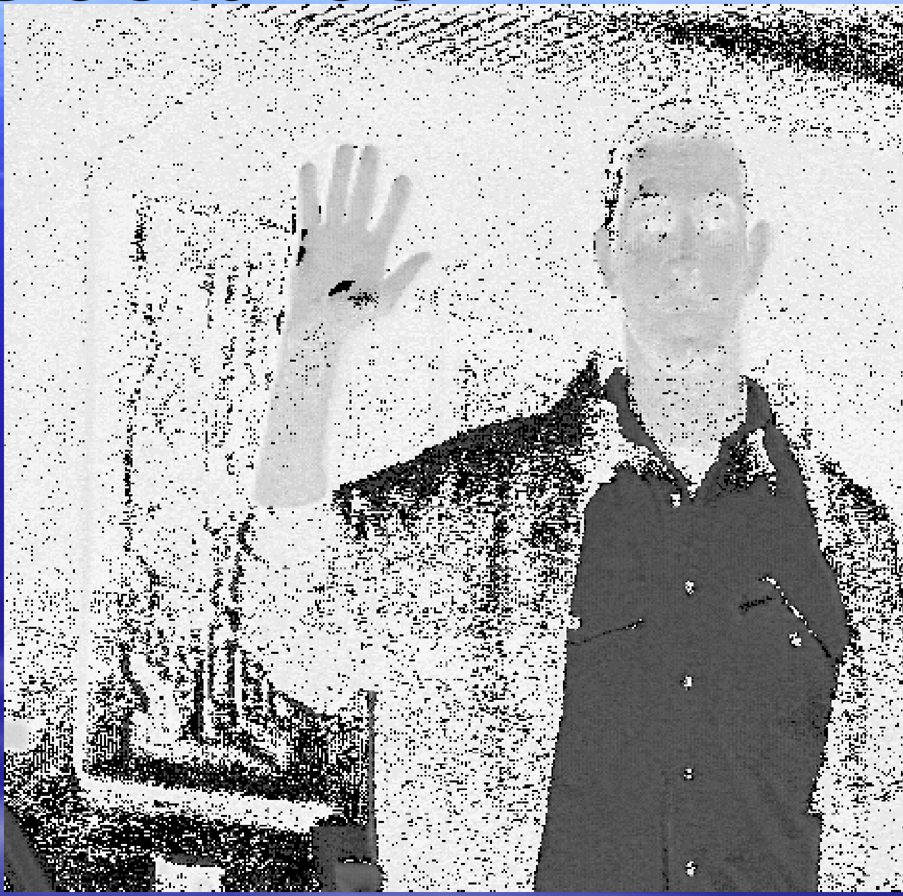


Scatterplot



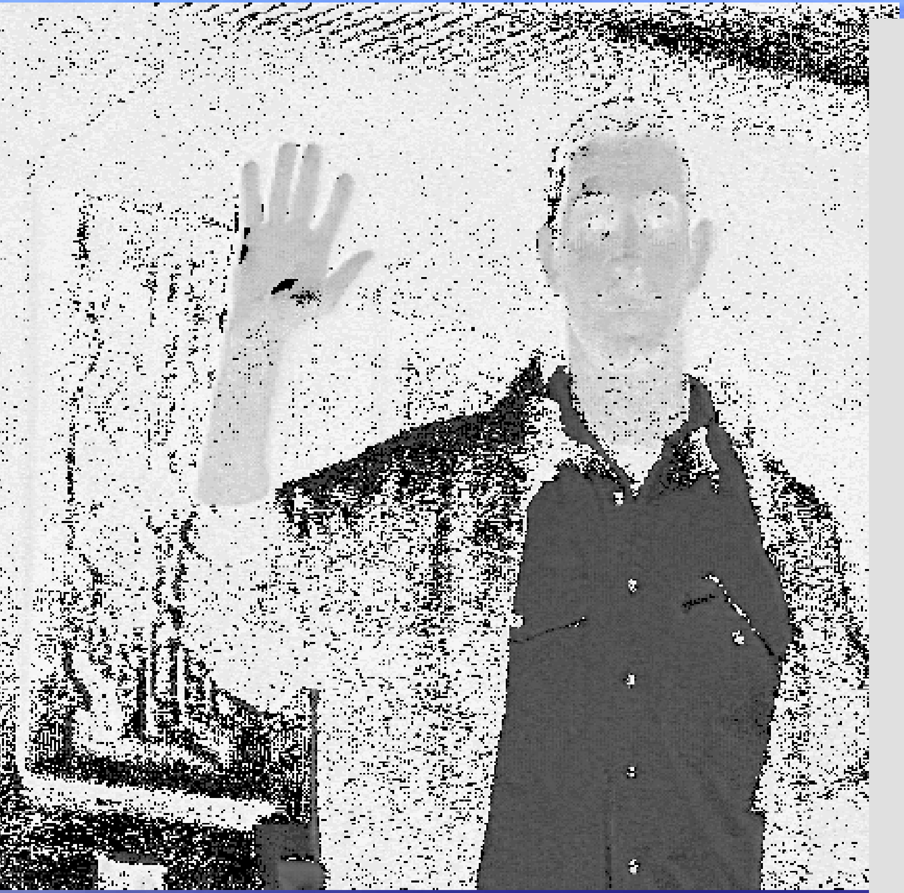
Simple Gestures

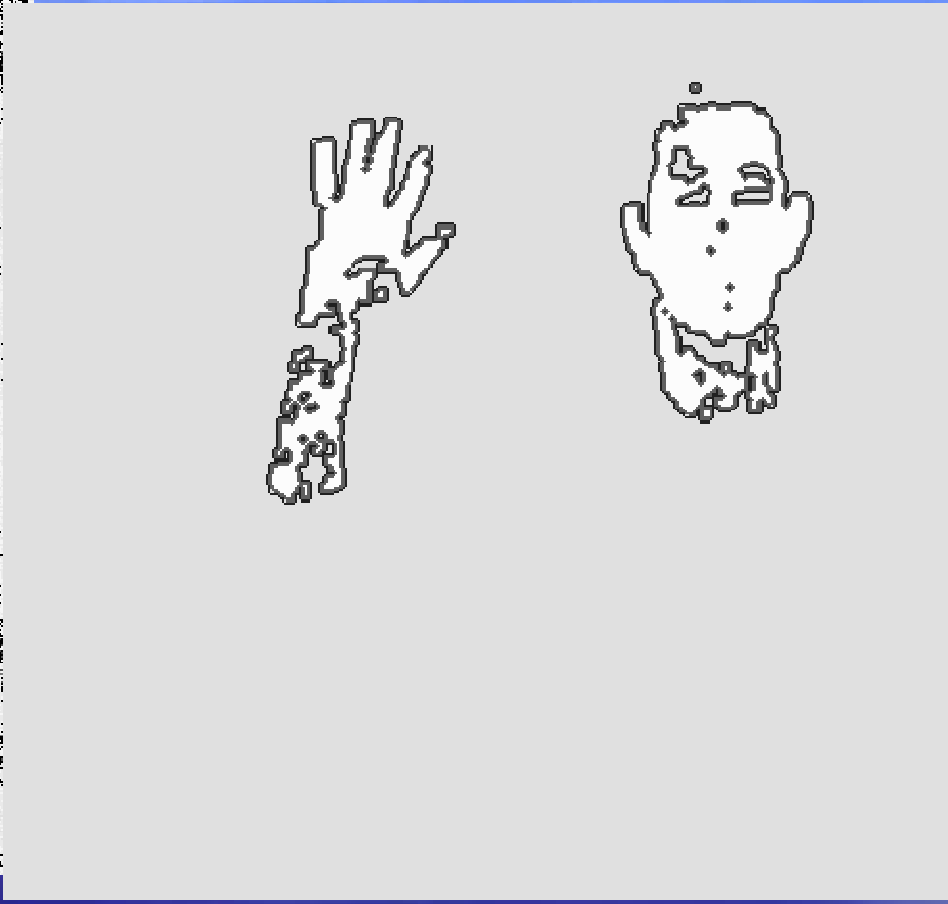
B/W image



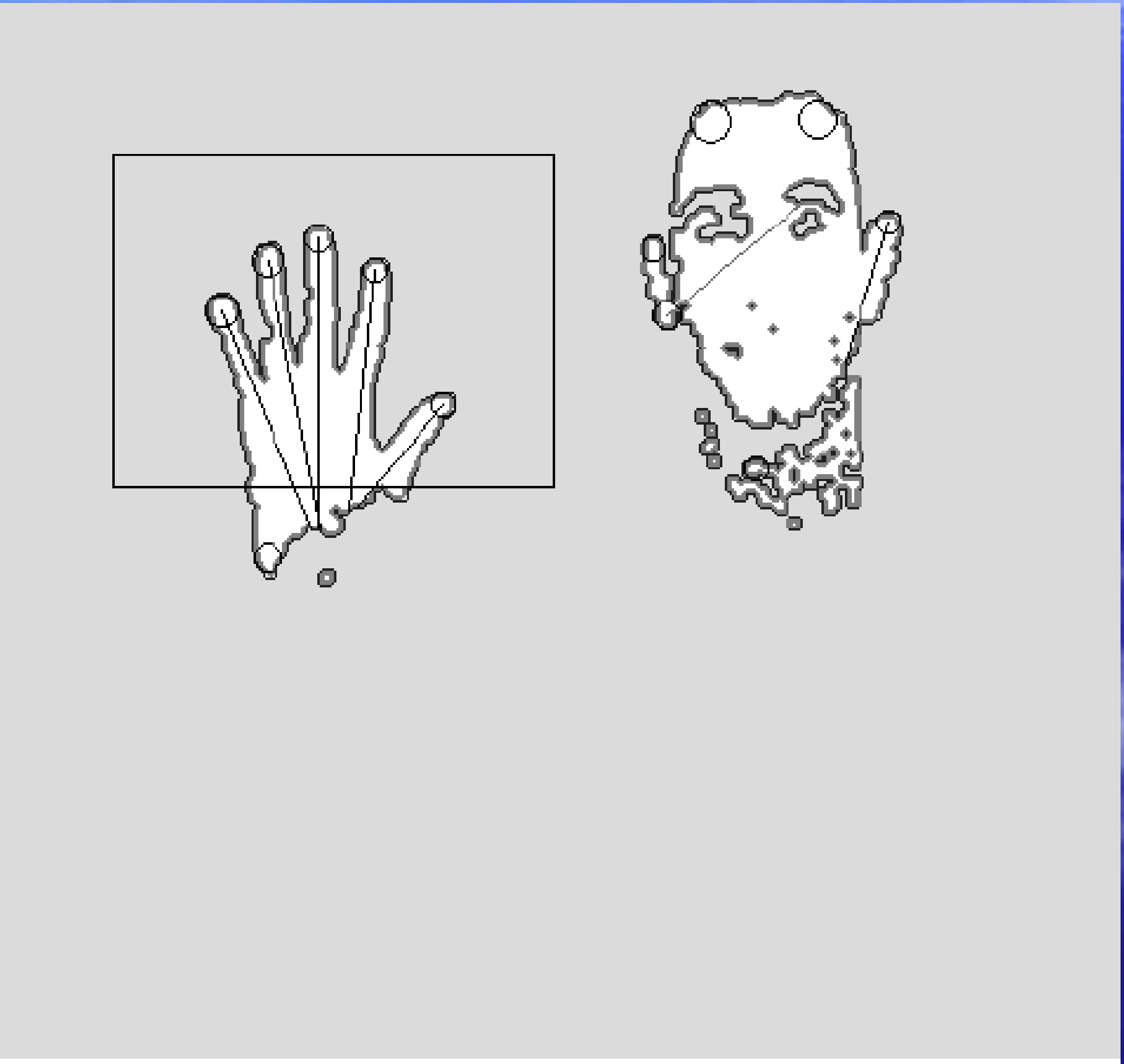
Hue: from RGB

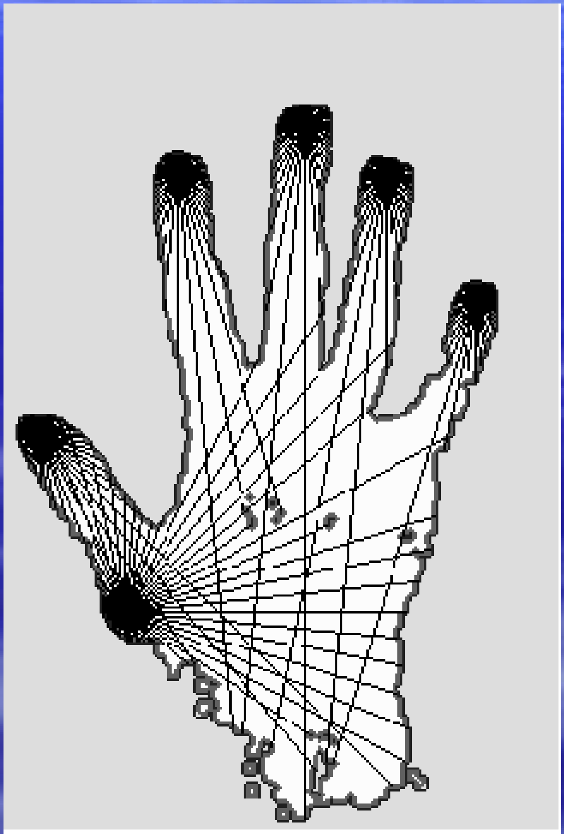
Binarized





Local analysis of shape





Jose: the Robot Waiter

Representation and Recognition of Complex Human Motion

Jesse Hoey
Jim Little
Computer Science
University of British Columbia

Why human motion?

- Human-Computer interaction
- Psychological research
- Video coding, search

It's hard because:

- Many types of motion
- Articulated, non-rigid motion
- Different scales

Our goals

- Find a general representation for any type of motion at any scale: Zernike polynomials
- Identify actions in motion sequences by extracting a low-dimensional feature vector for high-level processing

Zernike Polynomials

$$U_n^m(\rho,\theta) = R_n^m(\rho,\theta)e^{im\theta}$$

$$R_n^m(\rho) = \sum_{l=0}^{(n-m)/2} \frac{(-1)^l (n-l)!}{l![\frac{1}{2}(n+|m|)-1]![\frac{1}{2}(n-|m|)-l]!}\rho^{n-2l}$$

Examples:

n	m	ZP	
0	0	1	translation
1	1	$\rho\cos(\theta)$	affine
3	1	$(3\rho^3-2\rho)\cos(\theta)$	n: radial spatial frequency
3	3	$(3\rho^3-2\rho)\cos(3\theta)$	m:angular spatial frequency

Optical Flow Method

- Gradient based $I_x u + I_y v + It = 0$
- Graduated non-convexity with robust error norm
- Locally smooth
- Preserves discontinuities

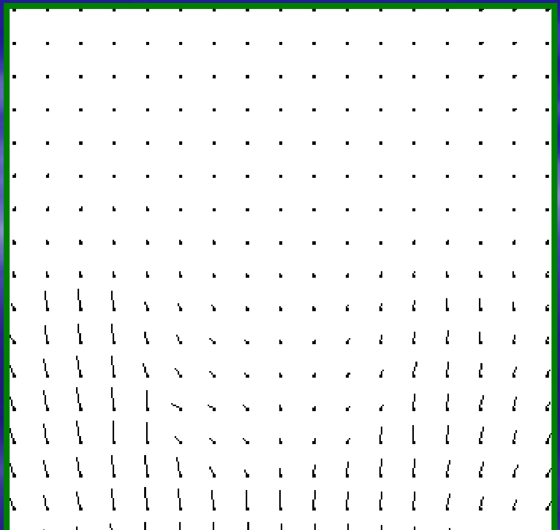
(Black & Anandan, *CVIU* 63(1) 1996)

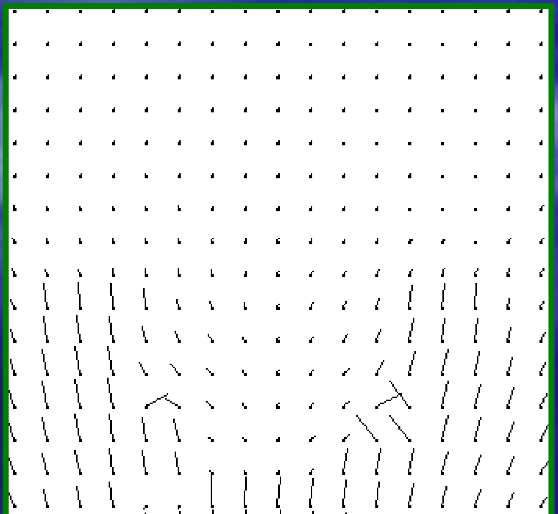
Optical Flow

t=0

t=1

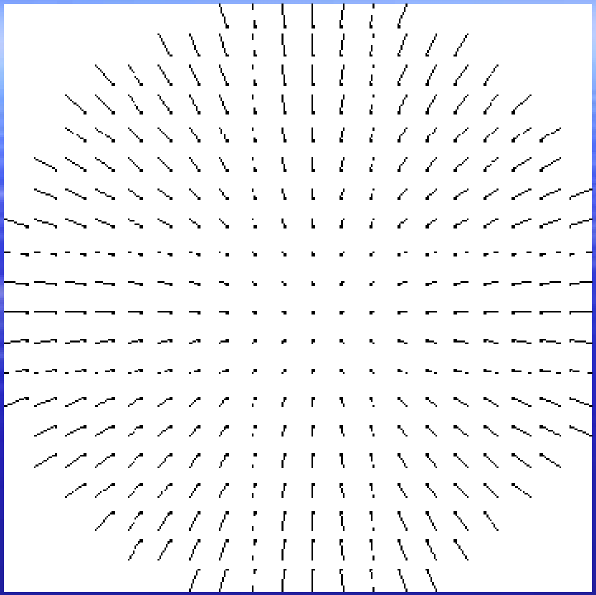
t=2



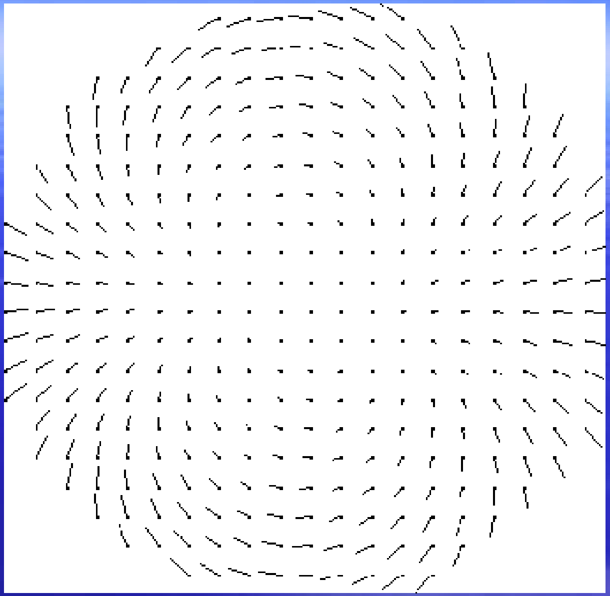


Example Flows

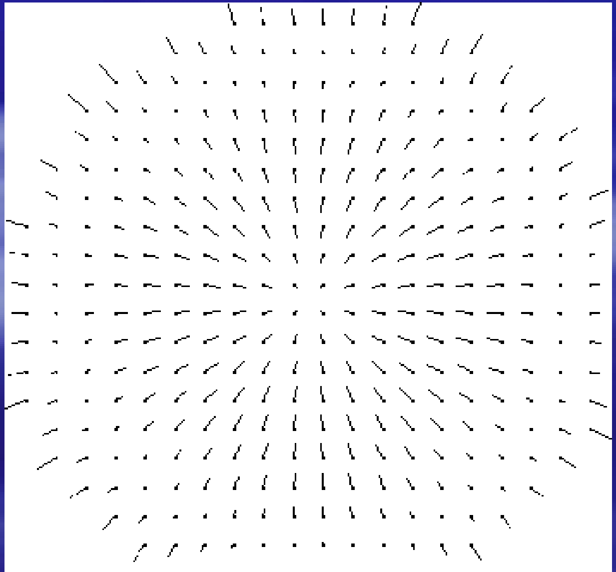
n=1 m=1 (affine)



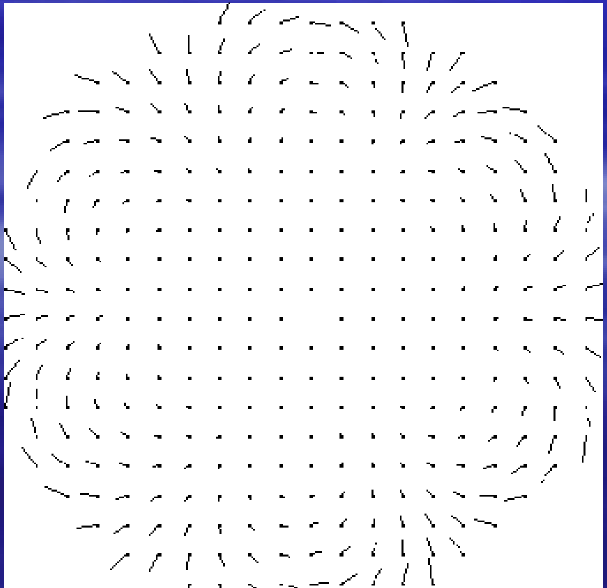
n=2 m=2



n=4 m=0 (radial only)

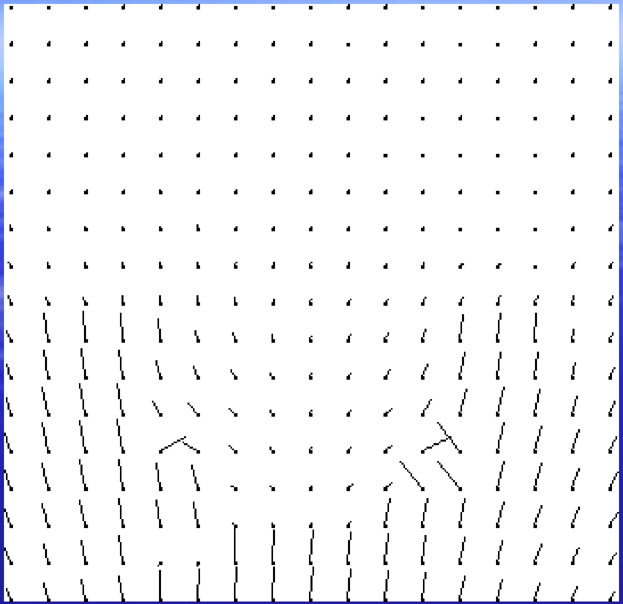


n=4 m=2

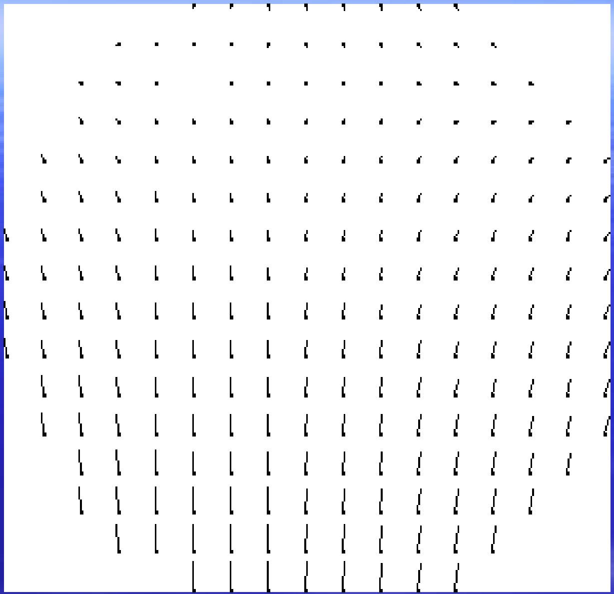


Reconstructed Flows:

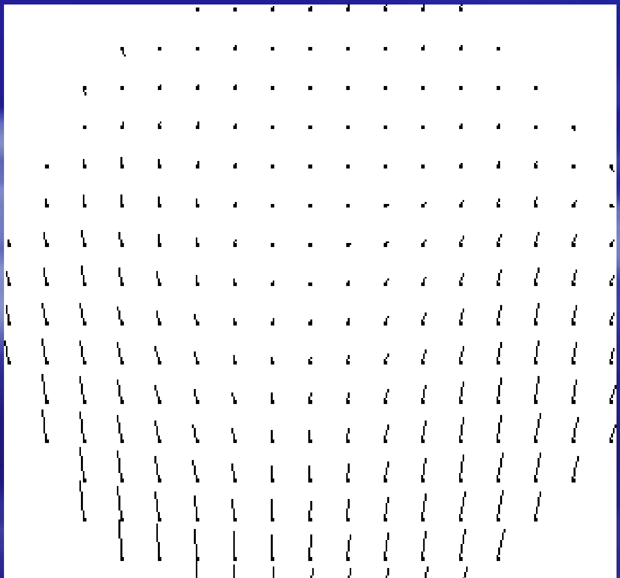
original flow field



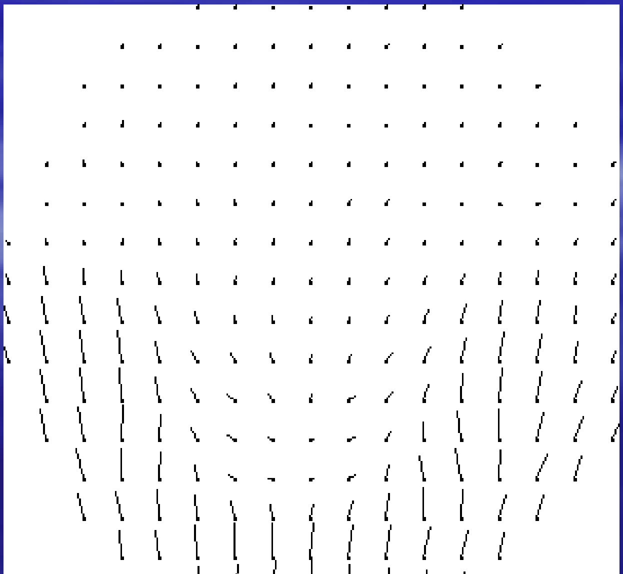
affine flow



first 7 ZPs



first 49 ZPs

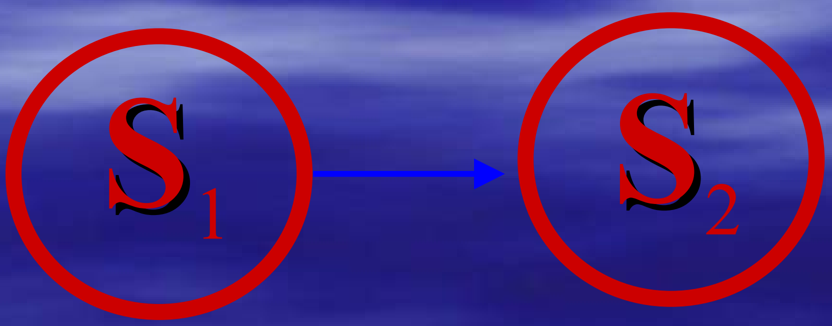


Zernike Basis

- Complete, orthogonal, basis of ***Zernike polynomials*** (ZPs) defined on the unit disk
- Can write any sufficiently smooth 2D function as a sum of Zernike polynomials
- dot product of flows with ZPs $\rightarrow$vector
- Simple flows (affine) $\rightarrow$ 6D vector (2 ZPs)
- Complex flows $\rightarrow$ higher dimensional vectors

Project onto Zernike Basis

- Create Z_1 and Z_2, Zernike features
- For optical flow shown above
- Map into temporal model
- Each state S_I depends on Z_i



Temporal Model

- Constant flow in temporal slices
- Model temporal progression with continuous density hidden Markov model
- Build Hidden Markov Model (HMM) for each motion type
- Classify new sequences as maximum likelihood HMM

Results

- Facial expressions with no rigid motion, only deformation
- Facial expressions with head motion
- Facial expression database
- Lip-reading

Facial expression

- with significant rigid head motion
- 1 subject - 5 expressions

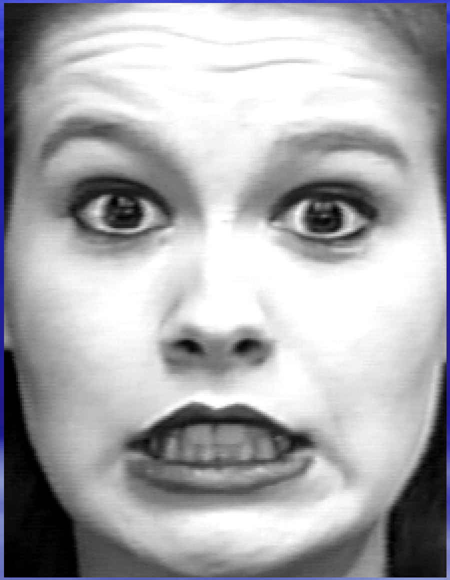


Translation (1 ZP):	**40%**
Affine (2 ZPs):	**69%**
First 7 ZPs	**94%**
PCA (7 components: 2-9)	**91%**

Facial expression

-no rigid head motion
-72 subjects - 6 expressions*

*Cohn-Kanade Facial Expression Database



Affine (2 ZPs): **71%**

(267 sequences, 4604 frames)

First 7 ZPs **90%**

Lip-Reading

- Tulips1 database
- 12 subjects - 4 words

Affine (2 ZPs):	**66%**
First 7 ZPs	**76%**
2,4,8,9,10,14,22:	**79%**

(96 sequences, 835 frames)

Facial expression

- no rigid head motion
-1 subject - 5 expressions



Translation (1 ZP):	**66%**
Affine (2 ZPs):	**98%**
First 7 ZPs	**100%**

Summary

- Zernike polynomials are an effective model-free basis for representing optical flow
- Typical flows – faces, lips – can be well represented in the Zernike basis
- Learning HMM models of flow leads to recognition rates exceeding affine bases
- Outperforms PCA (model-based) analysis