

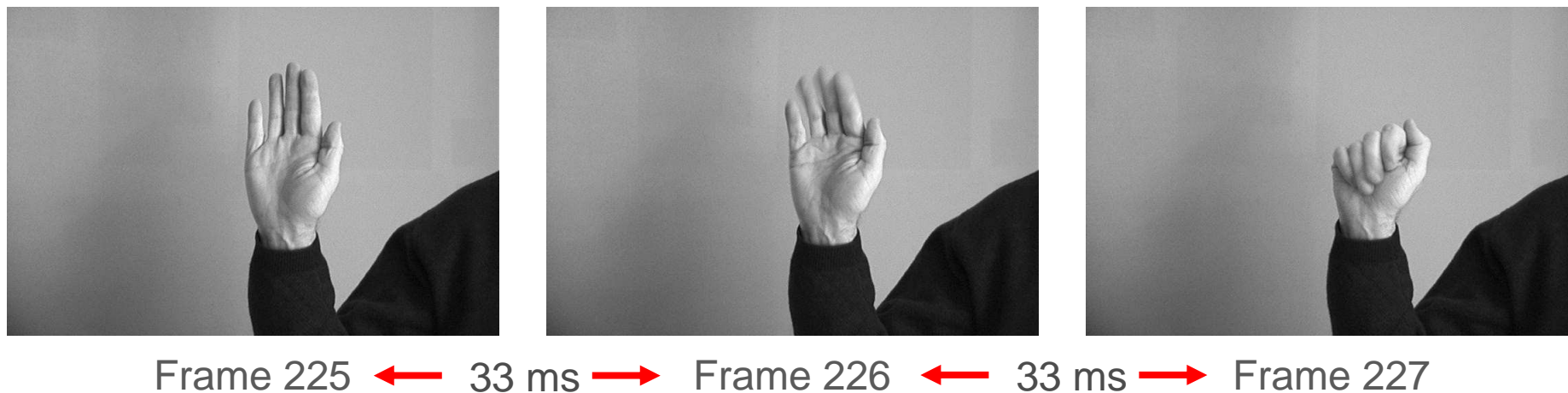
# 3D Tracking = Classification + Interpolation

Carlo Tomasi, Slav Petrov, and Arvind Sastry  
Department of Computer Science  
Duke University

[tomasi@cs.duke.edu](mailto:tomasi@cs.duke.edu)

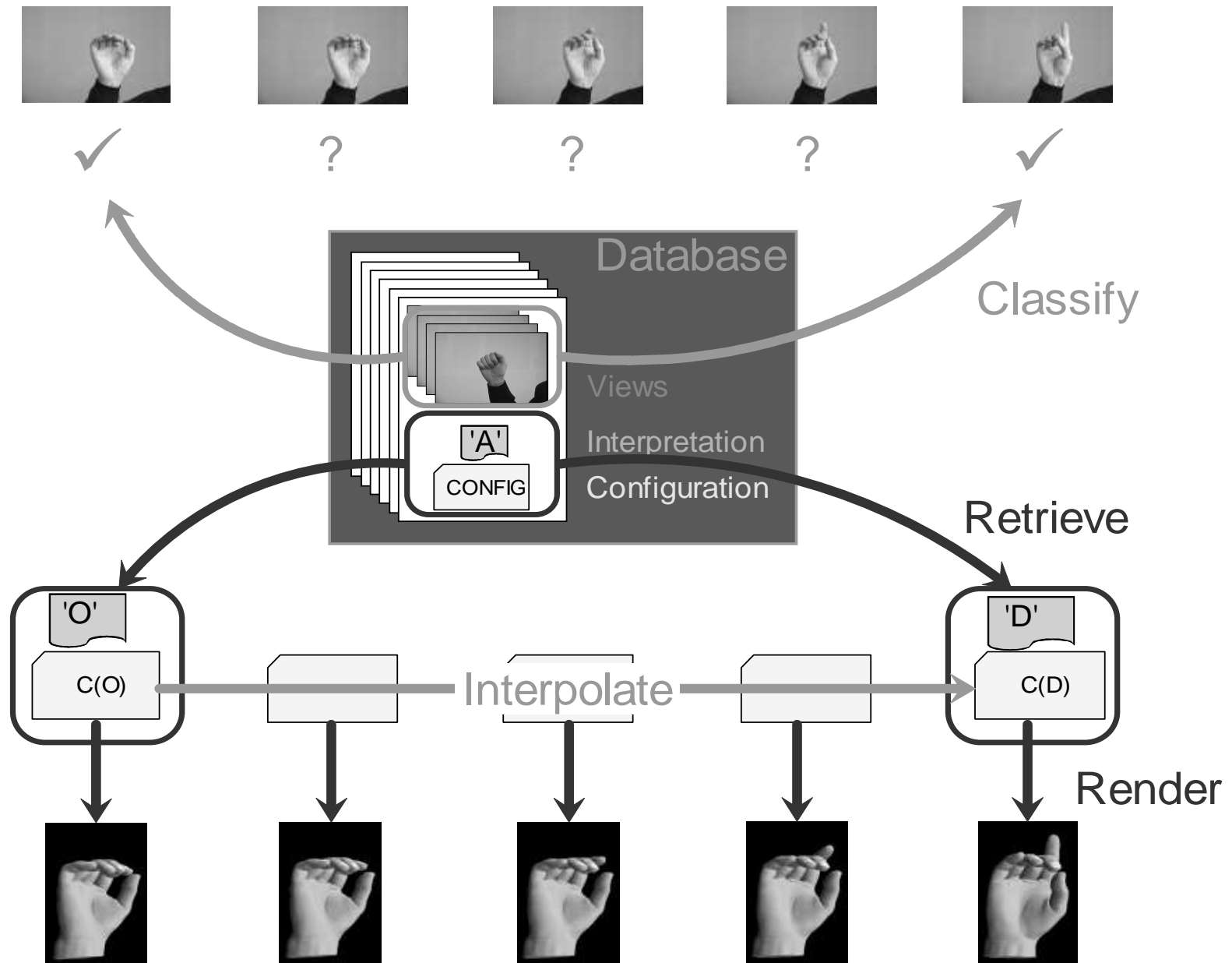
# Motivation of the Approach

- Applications: finger spelling: interpretation and transmission; musical tutoring; general UI; medicine
- Hand motions are too fast for tracking based on motion continuity:

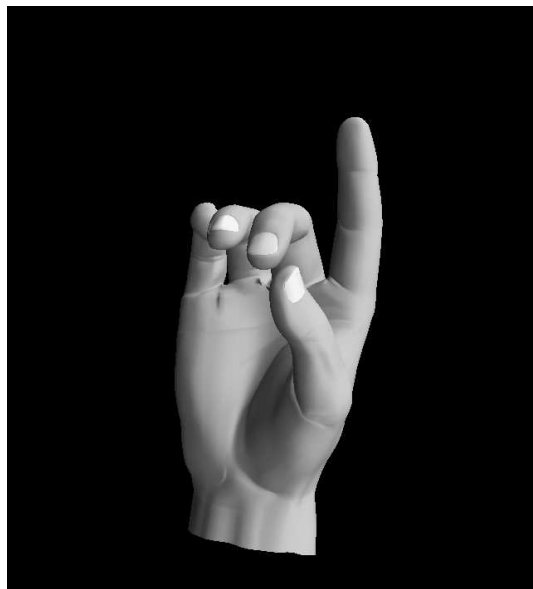


- Instead, classify recognizable frames, retrieve corresponding 3D models, and interpolate 3D motions in-between. Render to video if desired.

# Architecture



# 3D Tracking: Configuration and Pose



Different Pose  
→



→  
Different Configuration

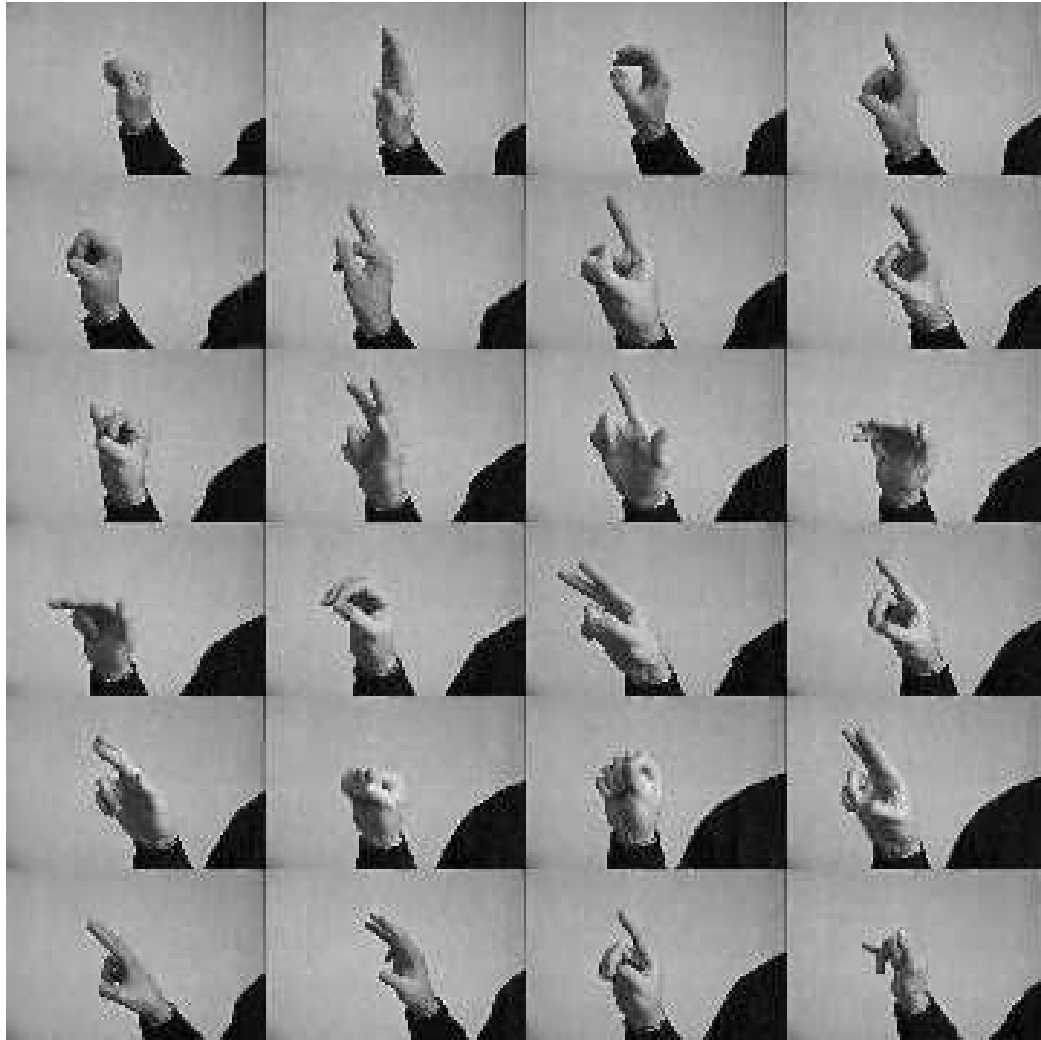


# View and Configuration Databases



Several real-video views + one graphics-based 3D model for each configuration

# Configuration, Out-of-Plane Rotation



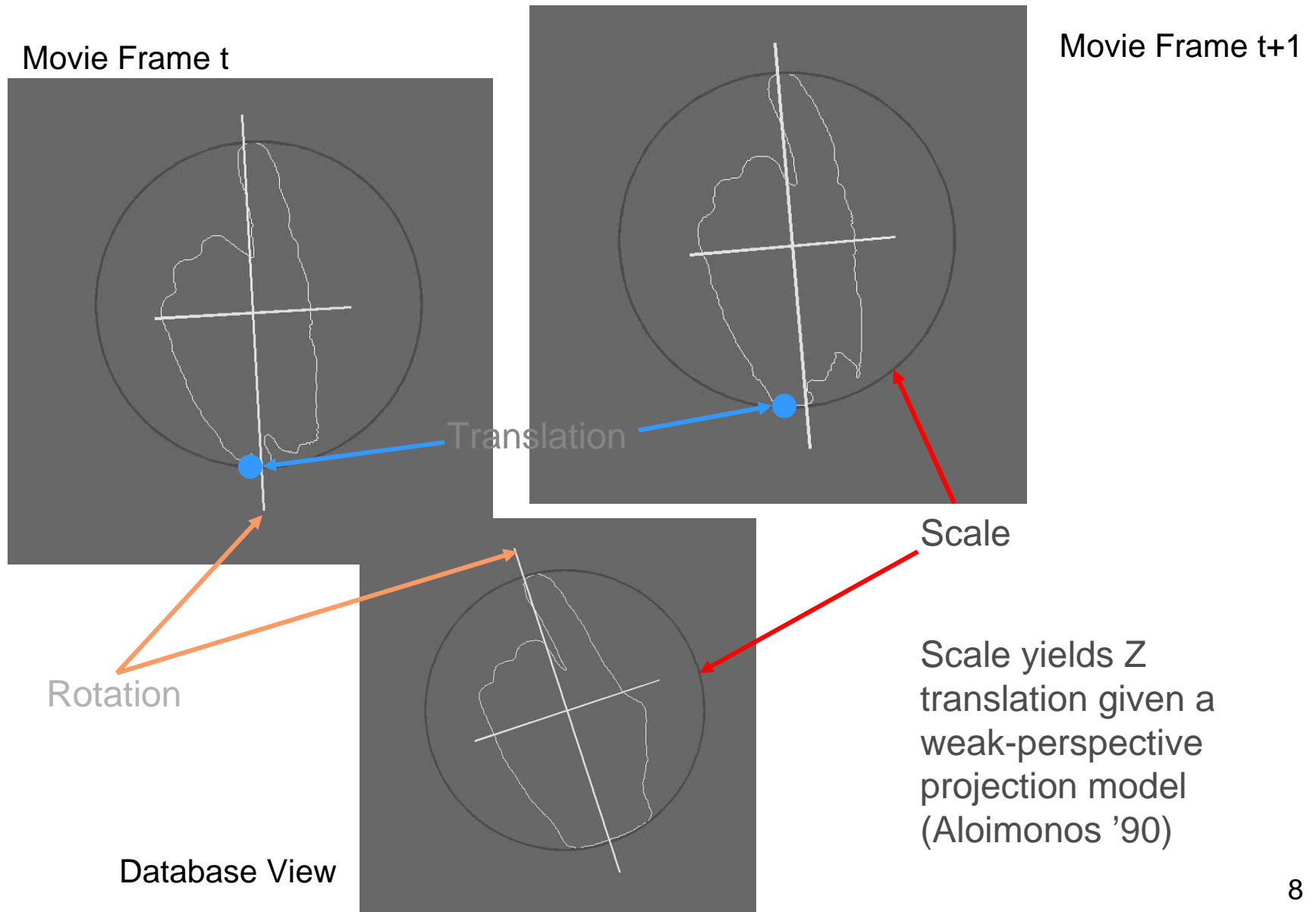
Known configurations and their out-of-plane rotations are recognized by matching with database views

2 rotation DOF in general, only one for finger spelling

# Sources of Pose Information

- Out-of-plane rotation from the database
  - 2 DOF, only 1 for finger spelling
- In-plane rotation:
  - Compare frame and view moments when well defined
  - Compare rotated images otherwise
  - Key frames only. Interpolate for intermediate frames
- In-plane translation:
  - Anchor point in consecutive frames if available
  - Circum-circle bottoms otherwise
  - All frames
- Out-of-plane translation:
  - Frame and view circum-circle radii
  - Requires an estimate of camera geometry
  - Key frames only. Interpolate for intermediate frames

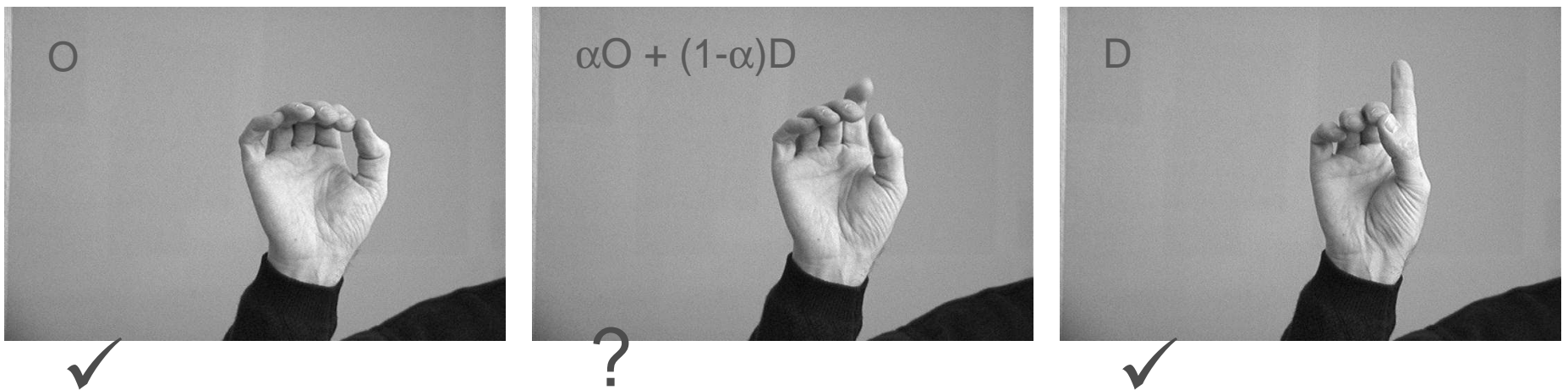
# In-Plane Pose Parameters



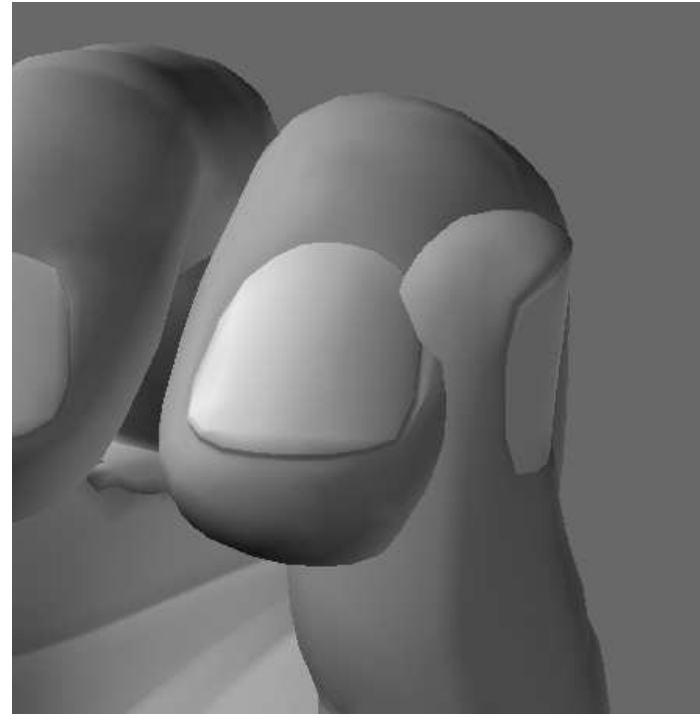


# 3D Motion Interpolation

- Intermediate frames are interpolated
- Input is buffered until the next known pose is seen
- Finger collision avoidance requires path planning in hand configuration space



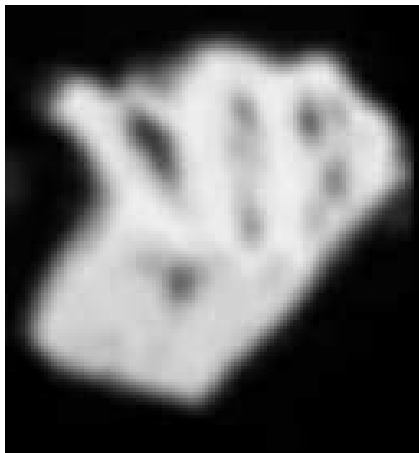
# Collision Avoidance



- General solution: path planning in hand configuration space
- Reality is easier: transition poses near poses that violate the natural order of the fingers
- Transition poses are inserted by a parser monitoring the interpretation stream

# Finding Hands

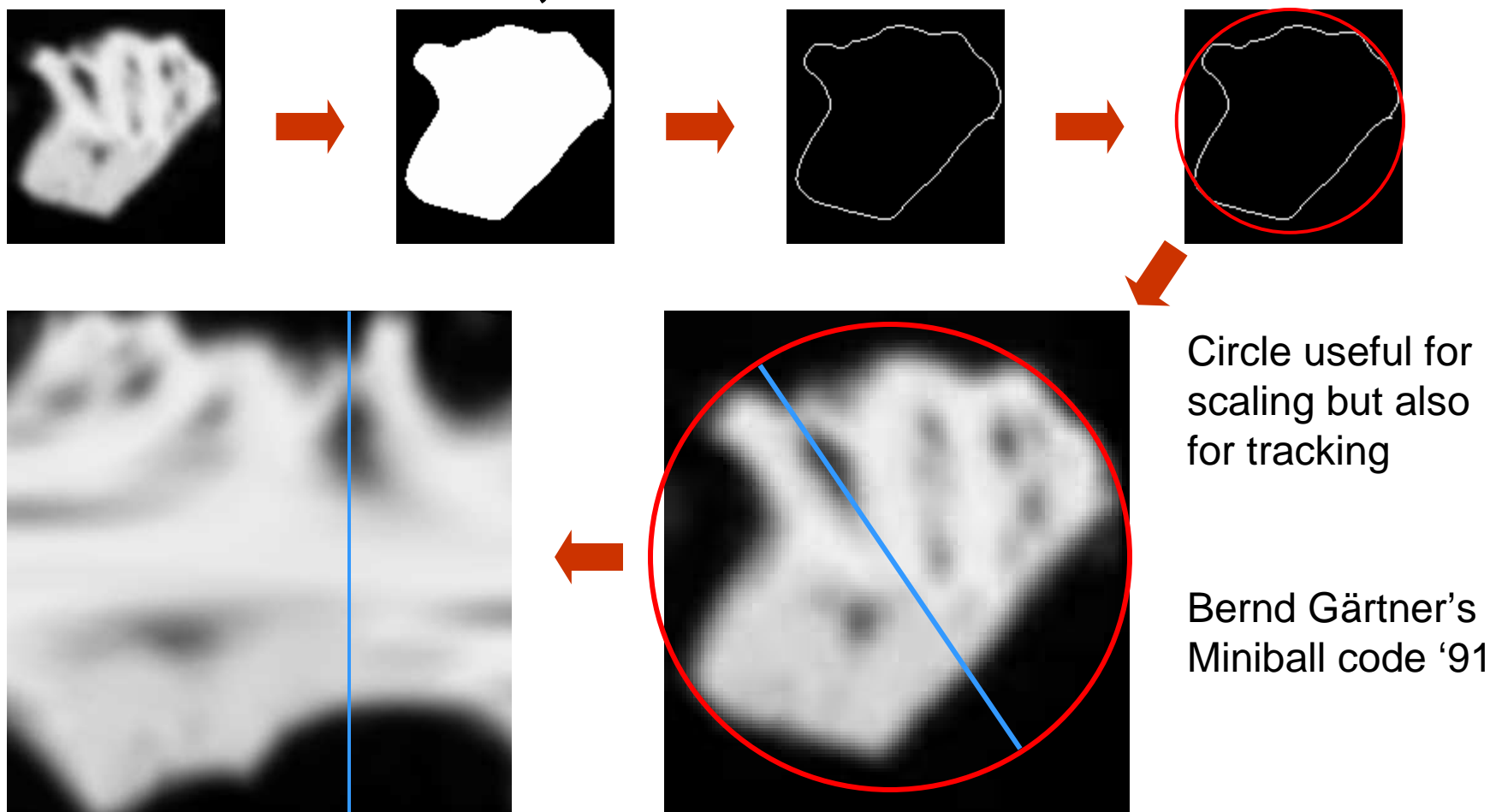
- Skin detection by Jones and Rehg '98
- Faces, wood, ...
- Largest connected component



Skin Likelihood

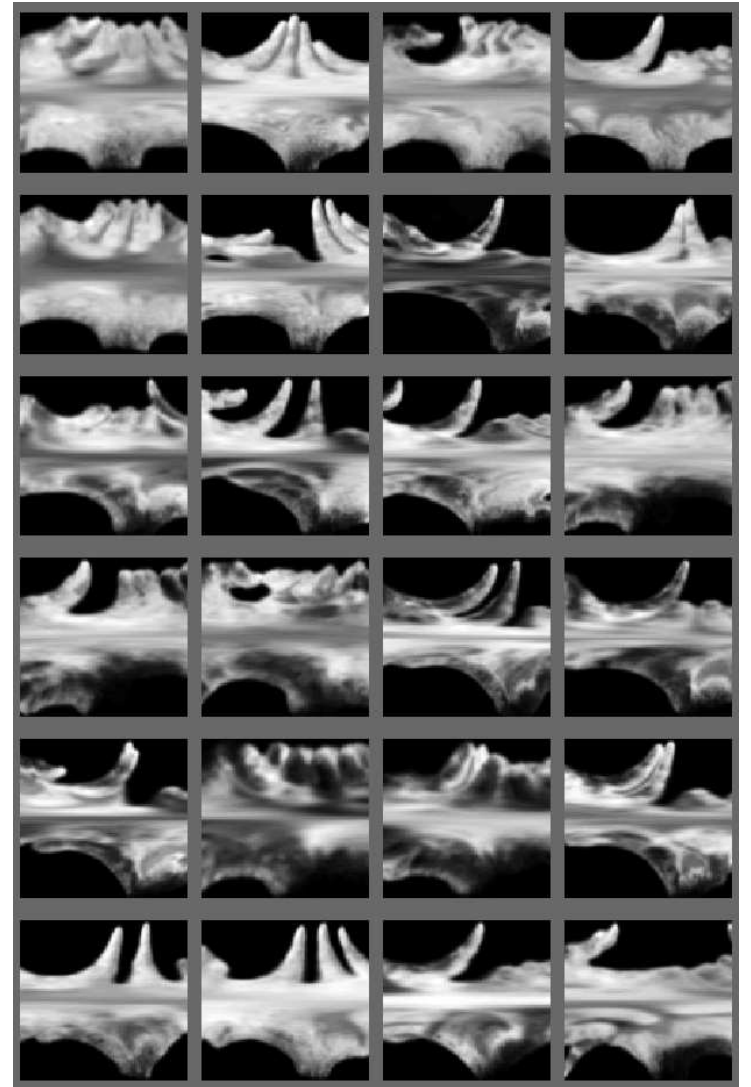
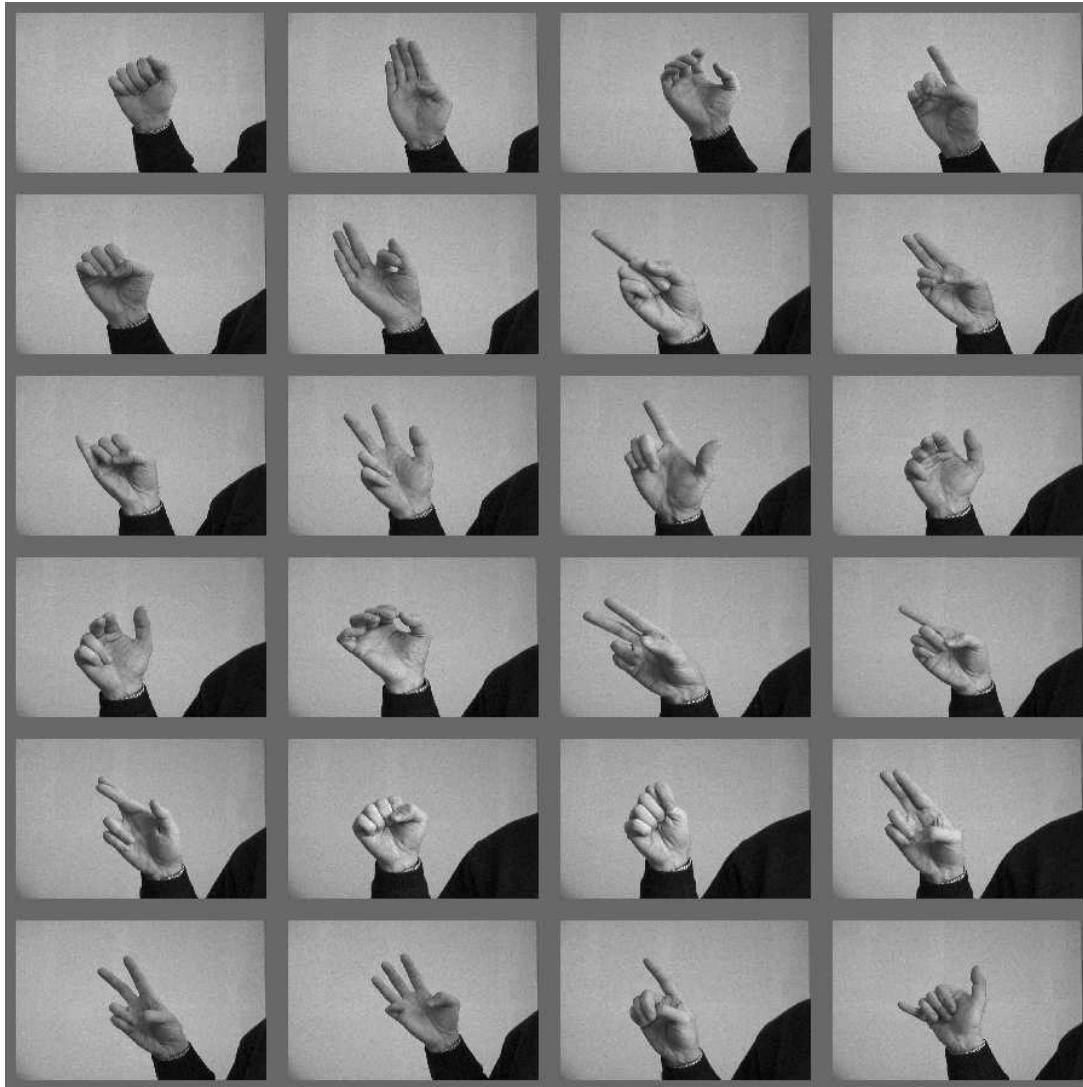


# Star Maps: Translation and Scale Invariance, Attention to Interior

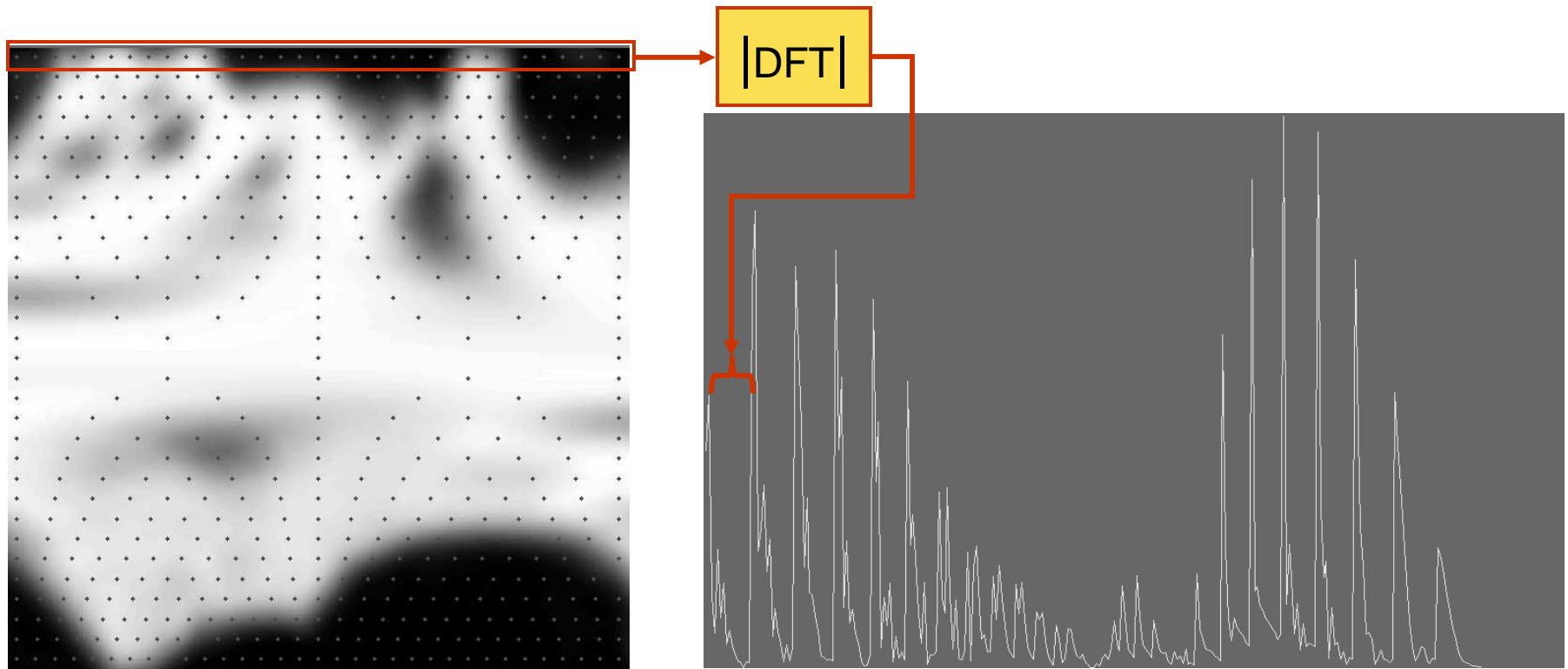


Related to Sandini, Tagliasco '80 (biological relevance of log-polar); Yamamoto, Yeshurun, Levine '96; Wiebe, Basu '97

# Star Map Examples



# Star Maps: Rotation Invariance



Related to Mellin transform (Casasent, Psaltis '76; Zwicke, Kiss '83; Sheng, Duvernoy '86), but handles scale and translation separately, and avoids numerical problems.

Actual feature algorithm only computes the required samples

# Feature Computation Summary



# Classification

- One sample per class → nearest-neighbor
- Raw features: not compact!
- Compactify:
  - PCA: project onto space with greatest data variation (remove commonalities)
  - VQPCA, Kambhatla & Leen '97: describe  $k$  different clusters by different PCAs
  - Hierarchical VQPCA, Tomasi & Petrov '03:  $k=2$ , but keep splitting until residual is small enough





# Results



Thank you for your attention.