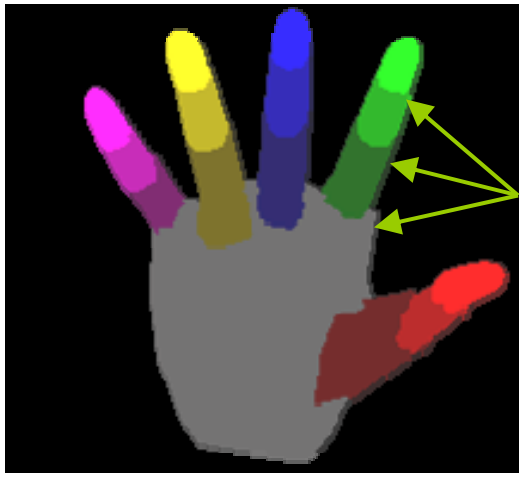
Database-Based Hand Pose Estimation

CSE 6367 – Computer Vision
Vassilis Athitsos
University of Texas at Arlington

Static Gestures (Hand Poses)

- Given a hand model, and a single image of a hand, estimate:
  - 3D hand shape (joint angles).
  - 3D hand orientation.

Input image



Articulated hand model

Static Gestures

- Given a hand model, and a single image of a hand, estimate:
  - 3D hand shape (joint angles).
  - 3D hand orientation.

Input image

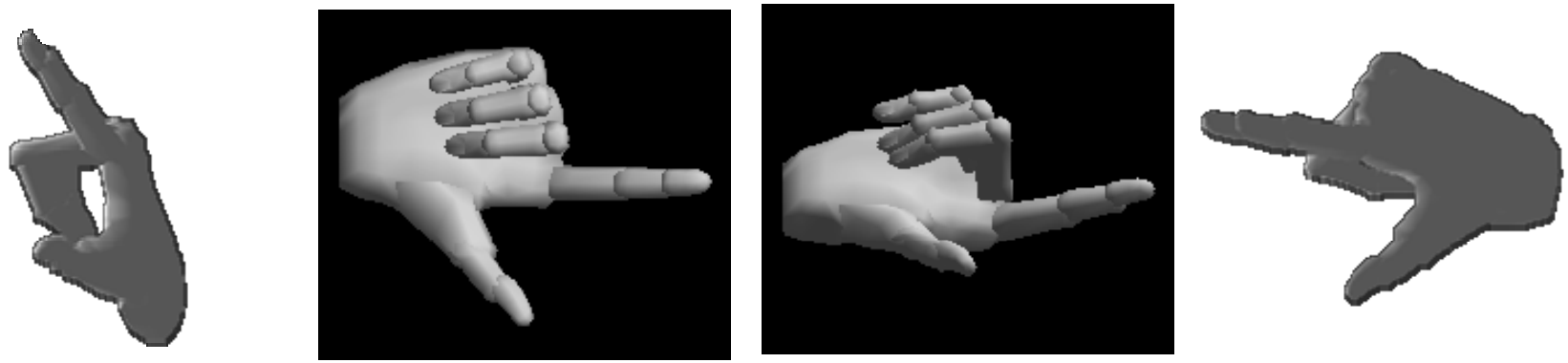
Articulated hand model

Goal: Hand Tracking Initialization

- Given the 3D hand pose in the previous frame, estimate it in the current frame.
  - Problem: no good way to automatically initialize a tracker.

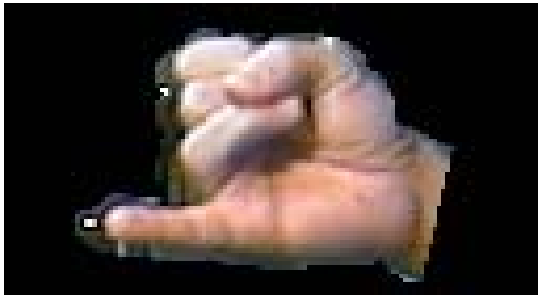
Rehg et al. (1995), Heap et al. (1996), Shimada et al. (2001), Wu et al. (2001), Stenger et al. (2001), Lu et al. (2003), …

Assumptions in Our Approach



- A few tens of distinct hand shapes.
  - All 3D orientations should be allowed.
  - Motivation: American Sign Language.

Assumptions in Our Approach



- A few tens of distinct hand shapes.
  - All 3D orientations should be allowed.
  - Motivation: American Sign Language.
- Input: single image, bounding box of hand.

Assumptions in Our Approach

input image

skin detection

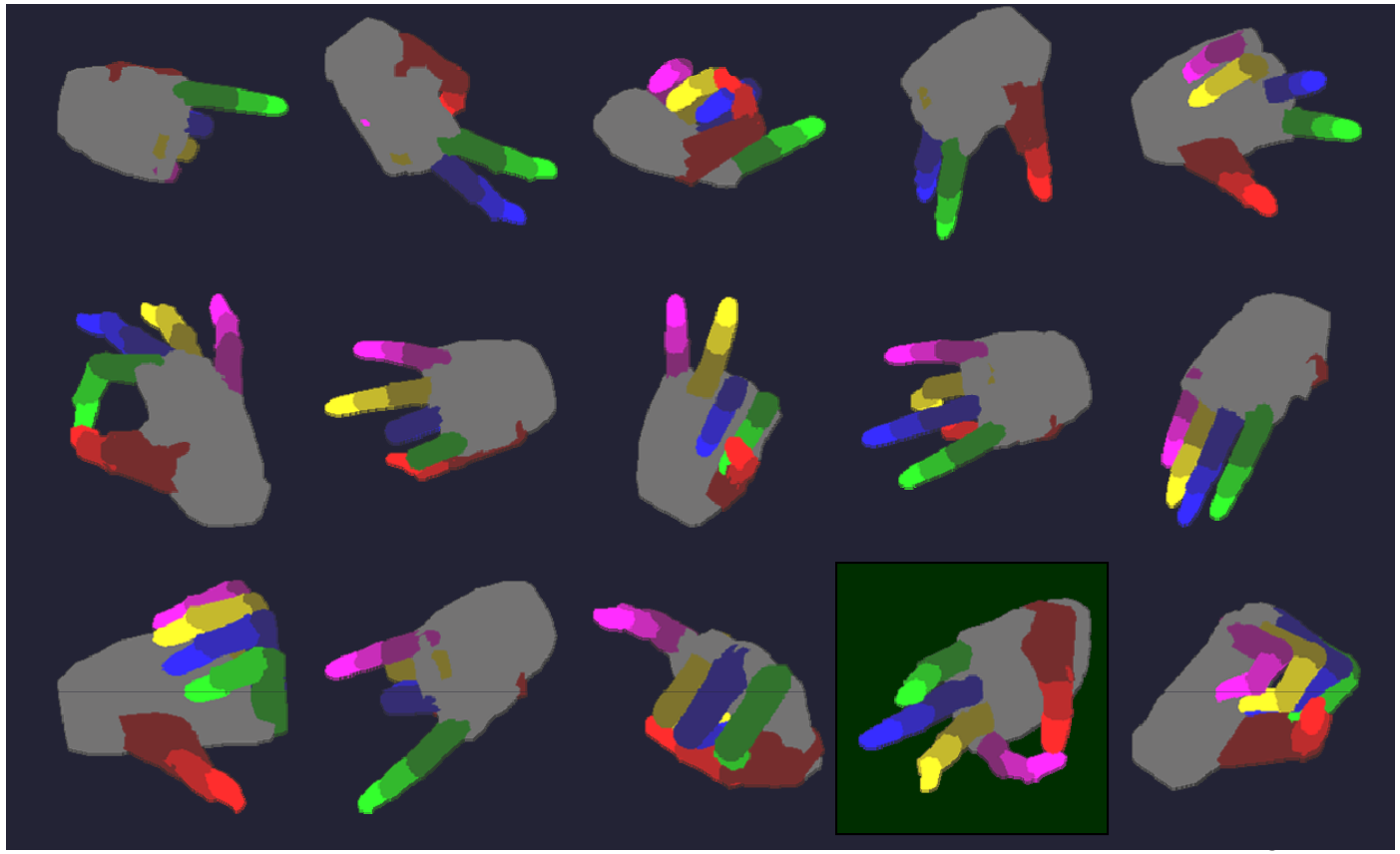
segmented hand

- We do not assume precise segmentation!
  - No clean contour extracted.

Approach: Database Search

- Over 100,000 computer-generated images.
  - Known hand pose.

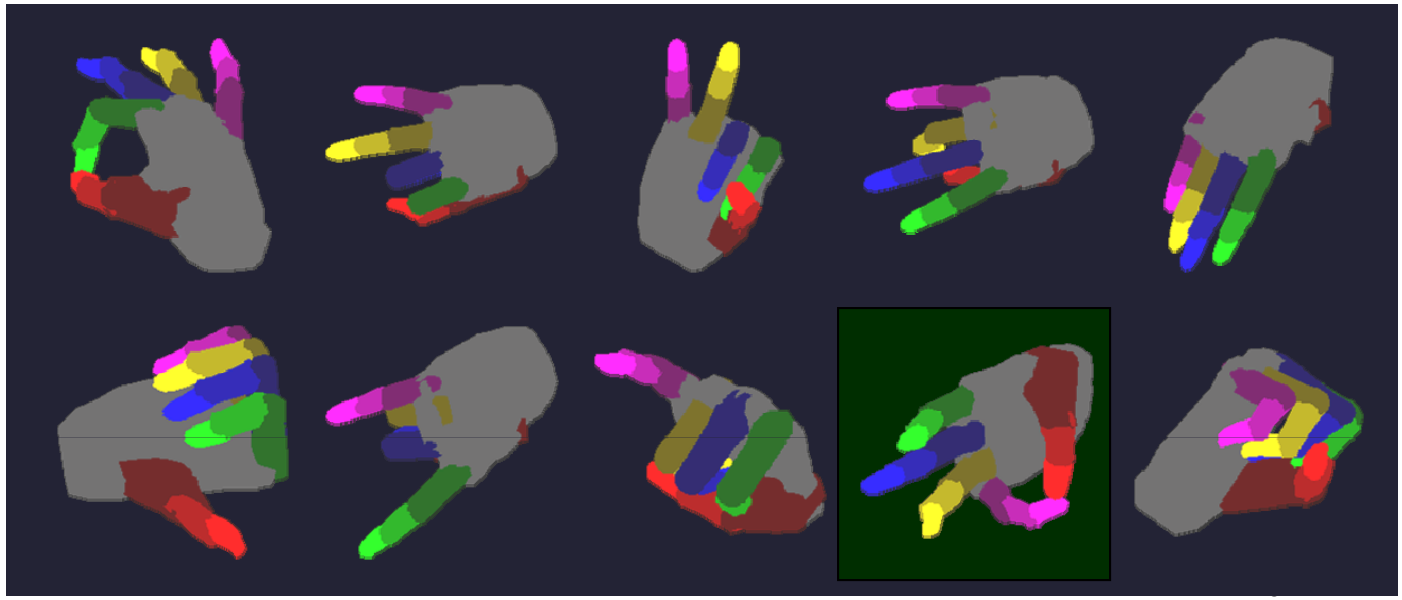
input



Why?

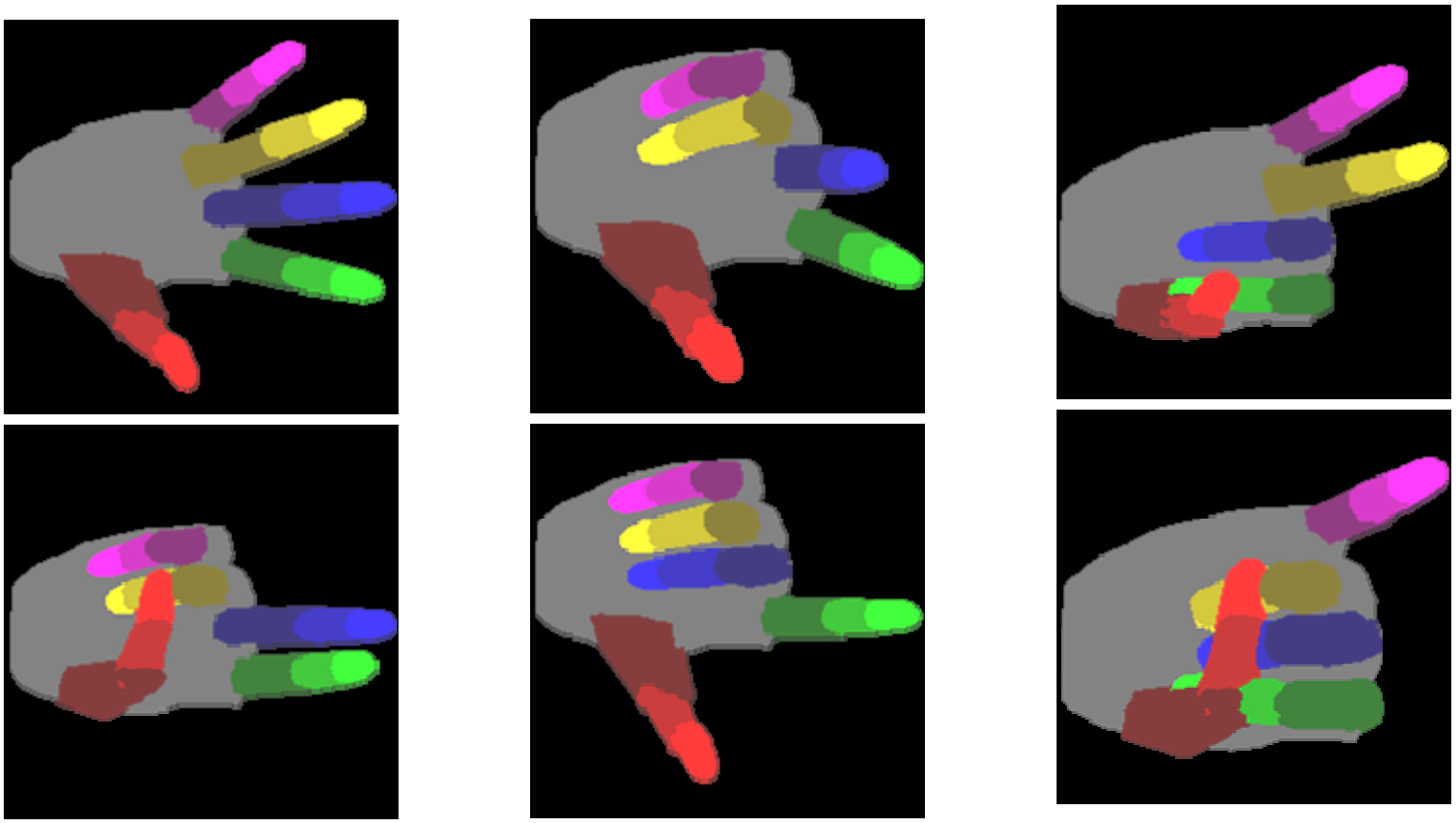
- We avoid direct estimation of 3D info.
  - With a database, we only match 2D to 2D.
- We can find all plausible estimates.
  - Hand pose is often ambiguous.

input



Building the Database

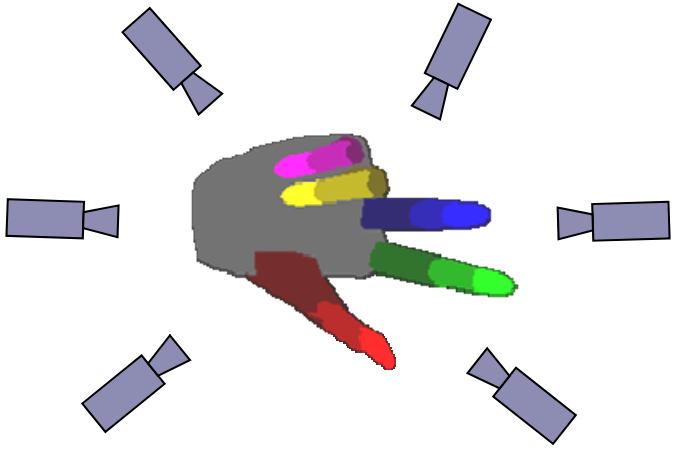
26 hand shapes

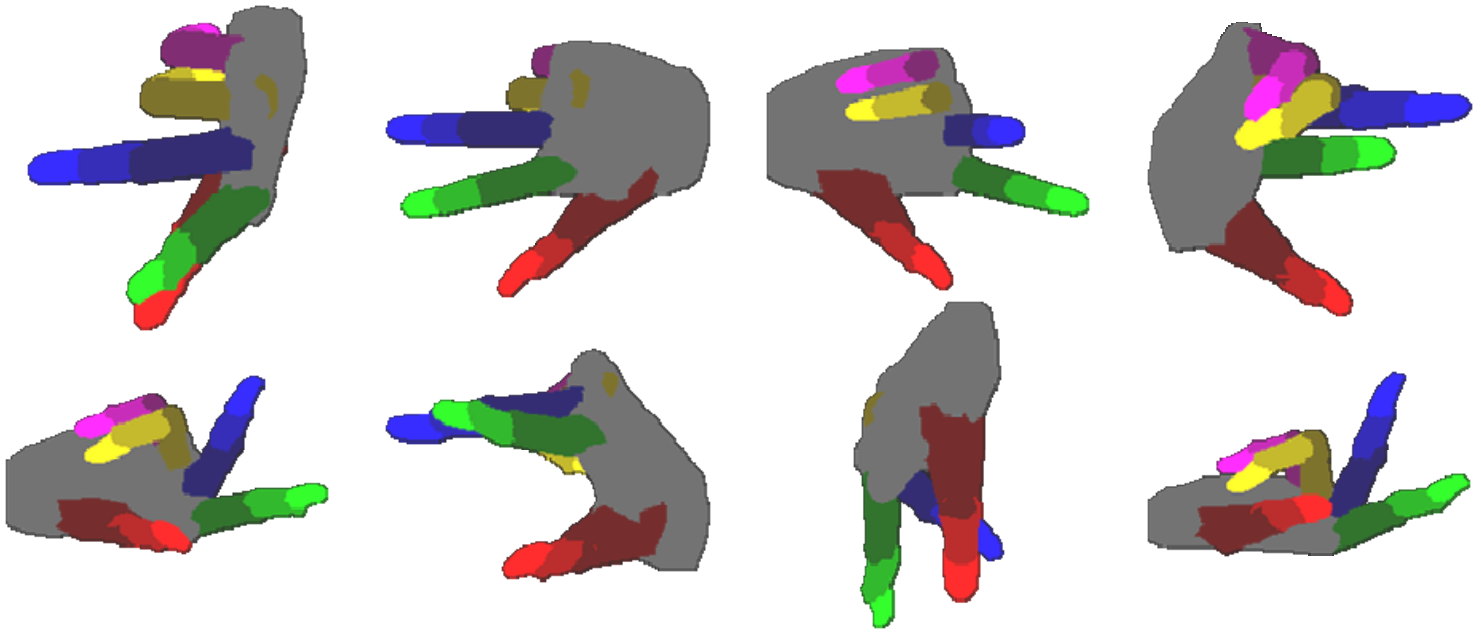


Building the Database

4128 images are generated for each hand shape.

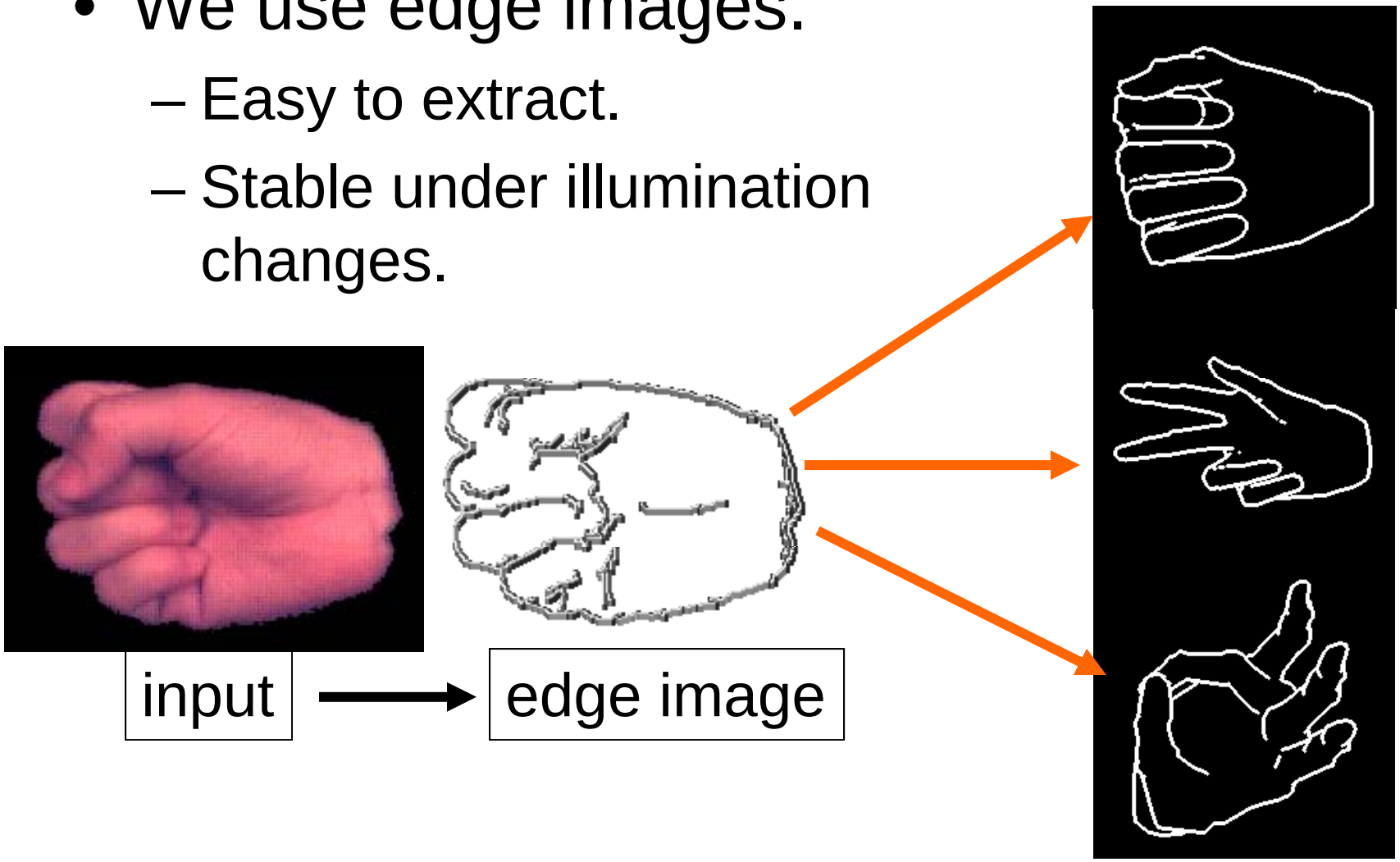
Total: 107,328 images.





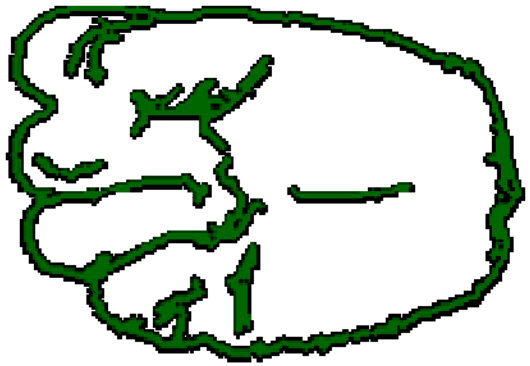
Features: Edge Pixels

- We use edge images.
  - Easy to extract.
  - Stable under illumination changes.

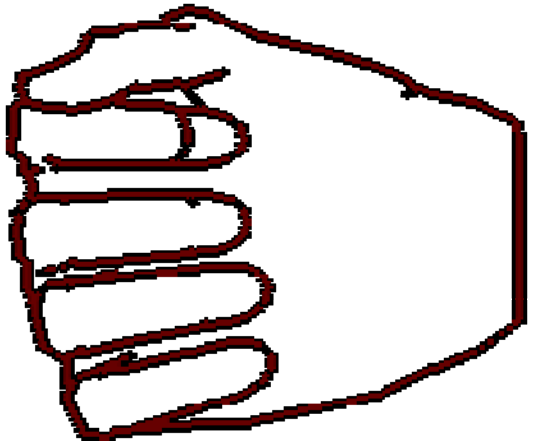


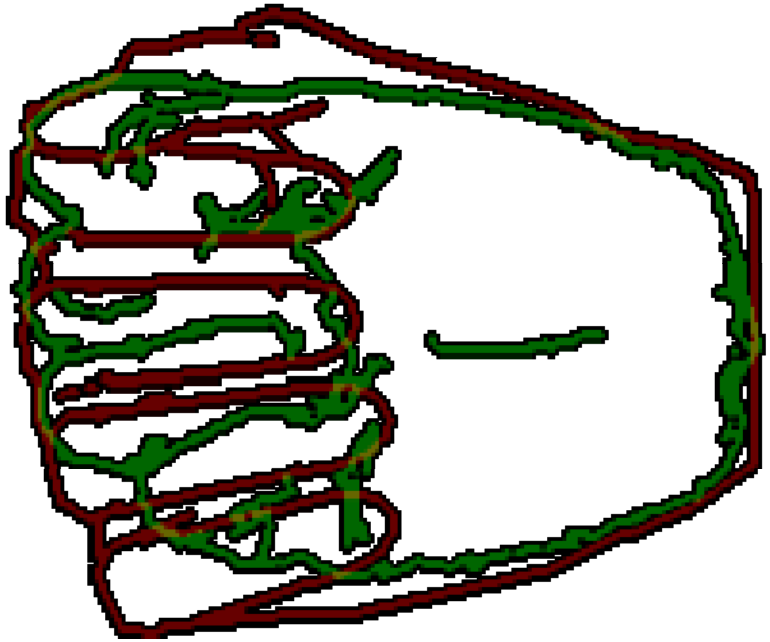
Chamfer Distance

input



model



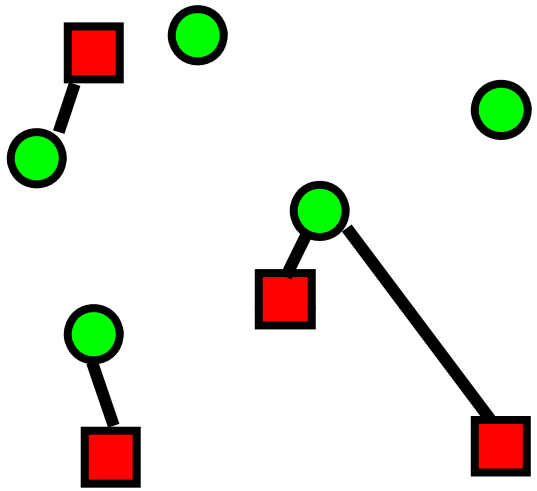


Overlaying input and model

How far apart are they?

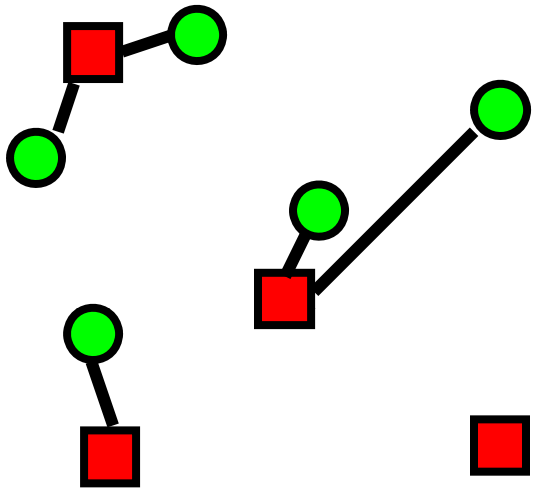
Directed Chamfer Distance

- Input: two sets of points.
  - red, green.
- c(red, green):
  - Average distance from each red point to nearest green point.



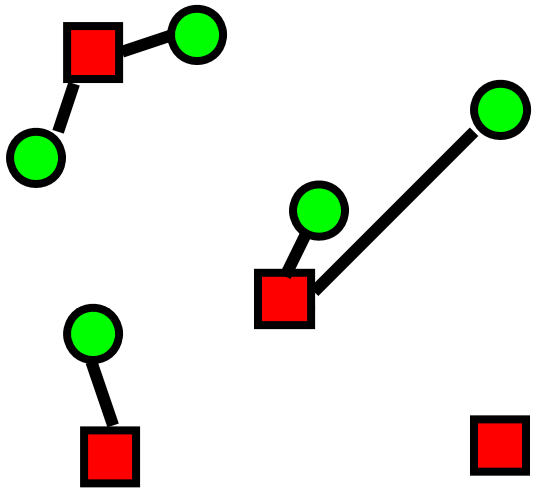
Directed Chamfer Distance

- Input: two sets of points.
  - red, green.
- c(red, green):
  - Average distance from each red point to nearest green point.
- c(green, red):
  - Average distance from each red point to nearest green point.



Chamfer Distance

- Input: two sets of points.
  - red, green.
- c(red, green):
  - Average distance from each red point to nearest green point.
- c(green, red):
  - Average distance from each red point to nearest green point.

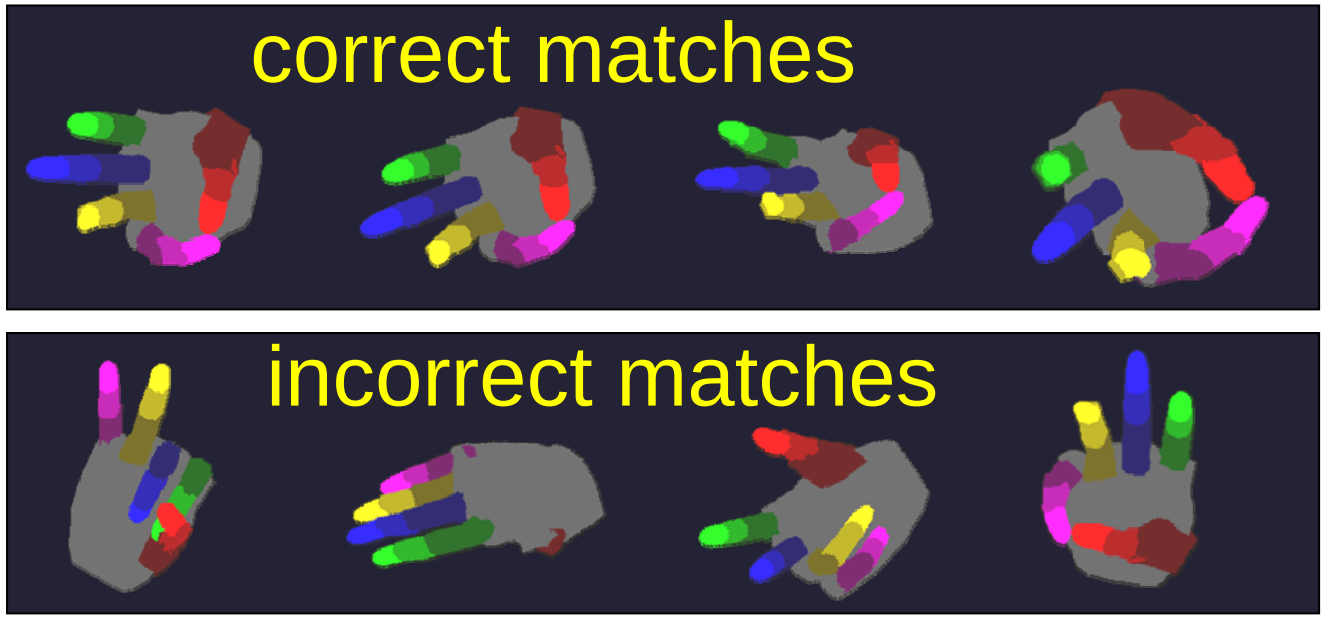


Chamfer distance:
C(red, green) = c(red, green) + c(green, red)

Evaluating Retrieval Accuracy

- A database image is a *correct match* for the input if:
  - the hand shapes are the same,
  - 3D hand orientations differ by at most 30 degrees.

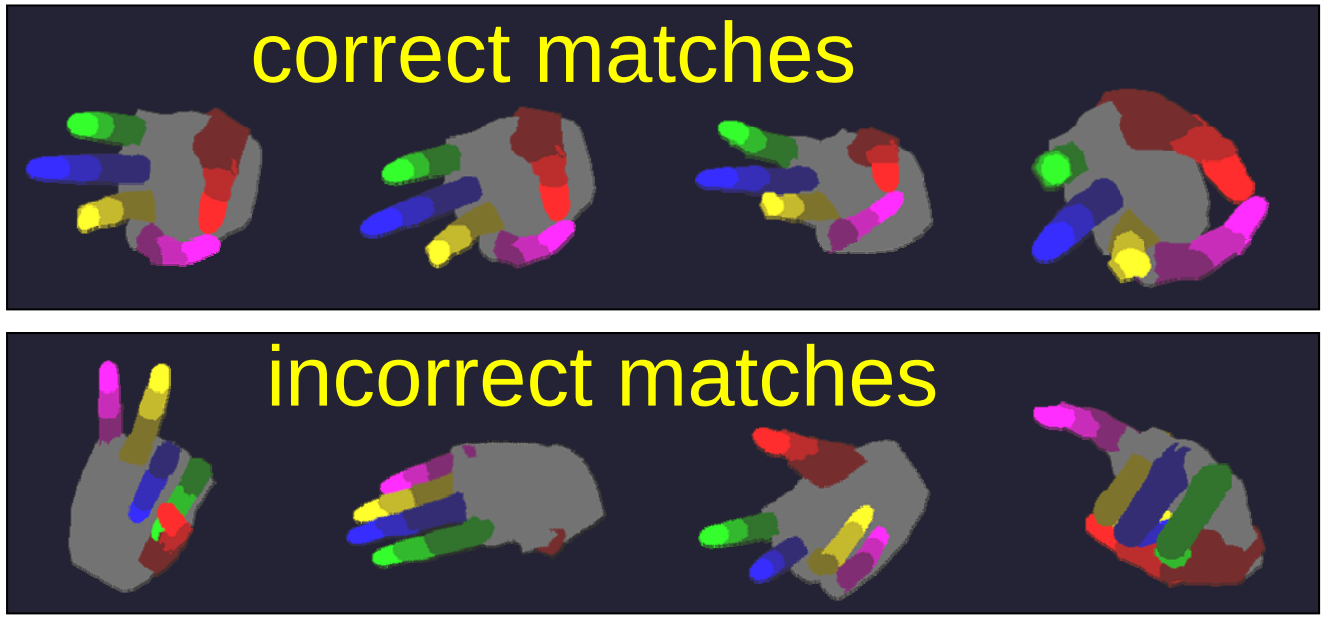
input



Evaluating Retrieval Accuracy

- An input image has 25-35 correct matches among the 107,328 database images.
  - Ground truth for input images is estimated by humans.

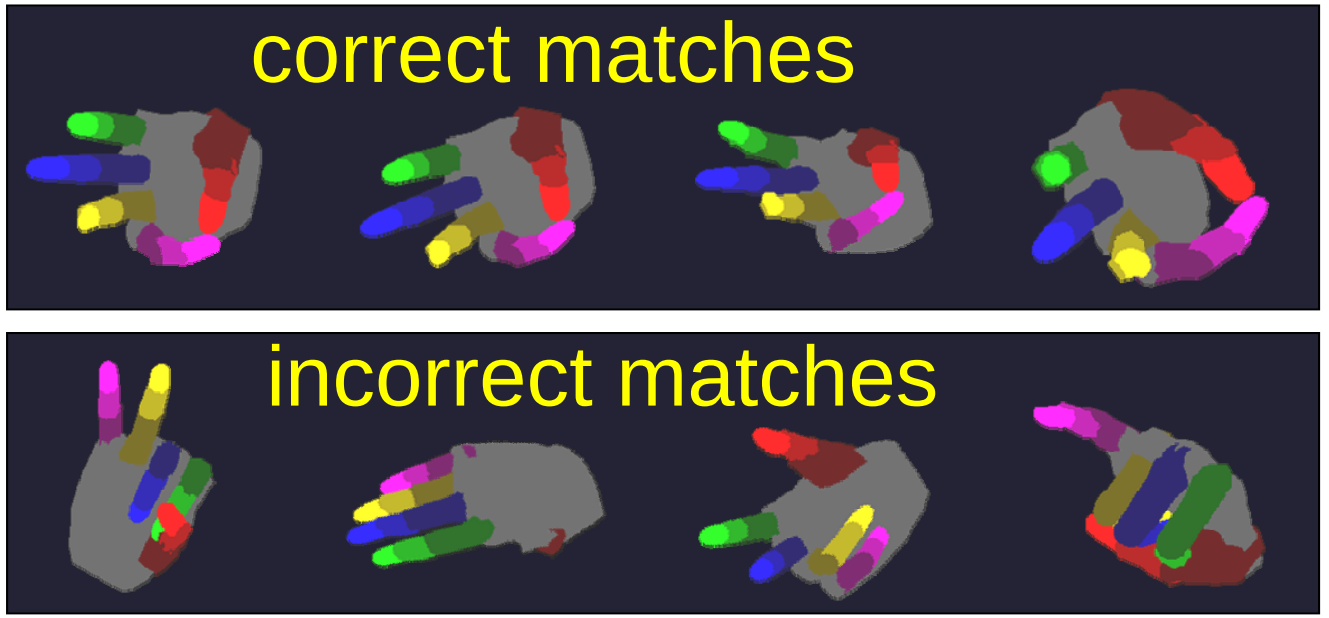
input



Evaluating Retrieval Accuracy

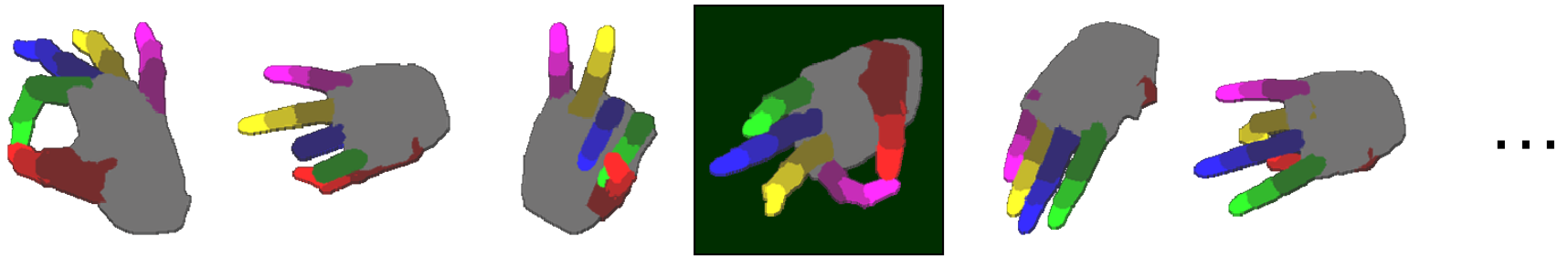
- Retrieval accuracy measure: what is the rank of the highest ranking correct match?

input



Evaluating Retrieval Accuracy

input



rank 1 rank 2 rank 3 rank 4 rank 5 rank 6 …

highest ranking
correct match

Results on 703 Real Hand Images

Rank of highest ranking correct match	Percentage of test images
1	**15%**
1-10	**40%**
1-100	**73%**

Results on 703 Real Hand Images

Rank of highest ranking correct match	Percentage of test images
1	**15%**
1-10	**40%**
1-100	**73%**

- Results are better on “nicer” images:
  - Dark background.
  - Frontal view.
  - For half the images, top match was correct.

Examples

initial image

segmented hand

edge image

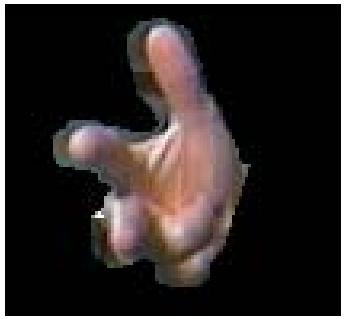
correct match

rank: 1

Examples

initial image

segmented hand



edge image

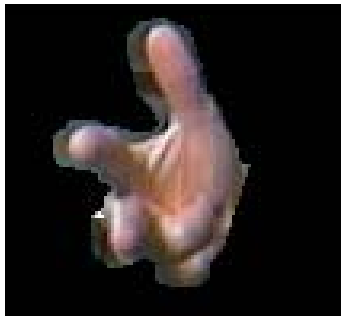
correct match

rank: 644

Examples

initial image

segmented hand



edge image

incorrect match

rank: 1

Examples

initial image

segmented hand

edge image

correct match

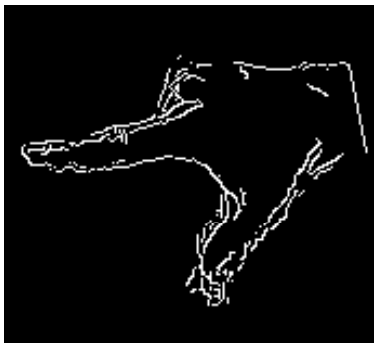
rank: 1

Examples

initial image

segmented hand

edge image



correct match

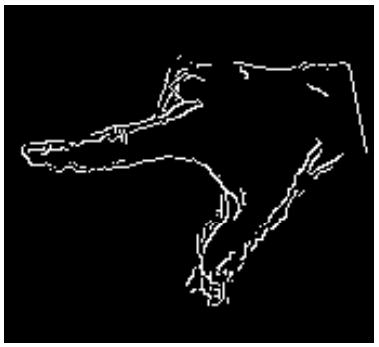
rank: 33

Examples

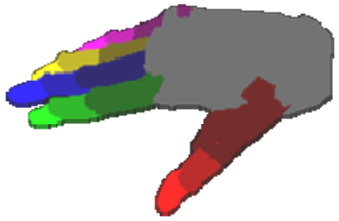
initial image

segmented hand

edge image



incorrect match

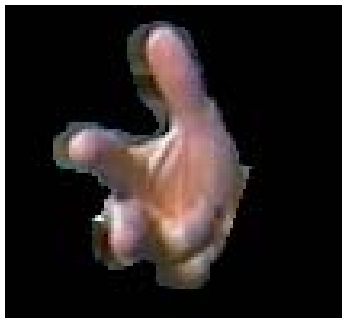


rank: 1

Examples

"hard" case

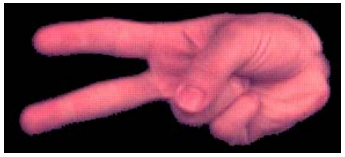
segmented hand



edge image

"easy" case

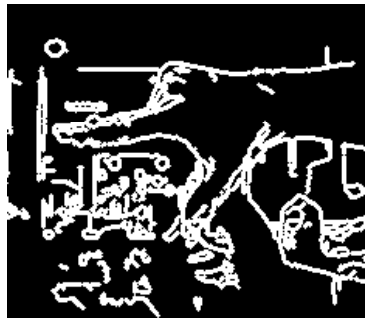
segmented hand



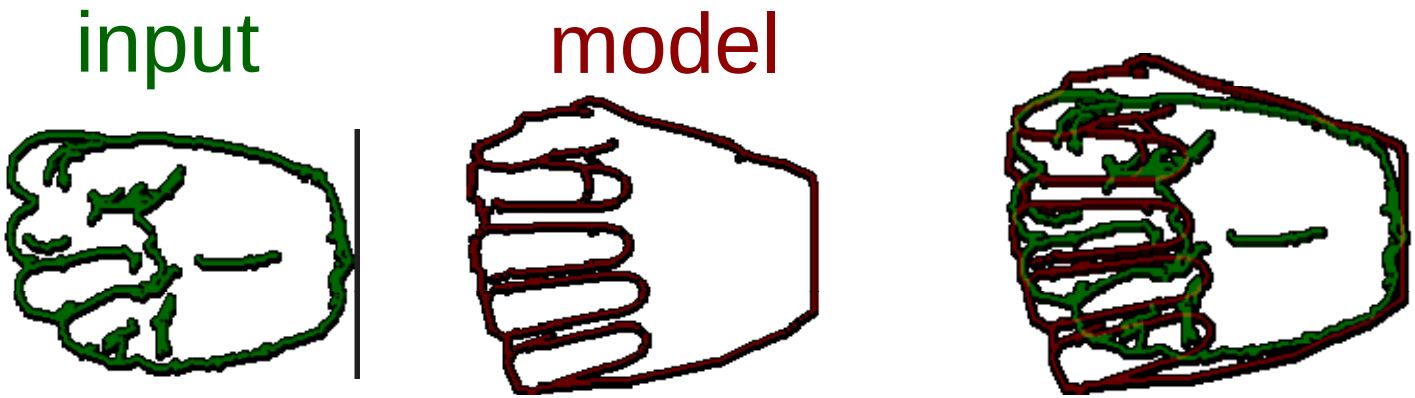
edge image

Research Directions

- More accurate similarity measures.
- Better tolerance to segmentation errors.
  - Clutter.
  - Incorrect scale and translation.
- Verifying top matches.
- Registration.



Efficiency of the Chamfer Distance



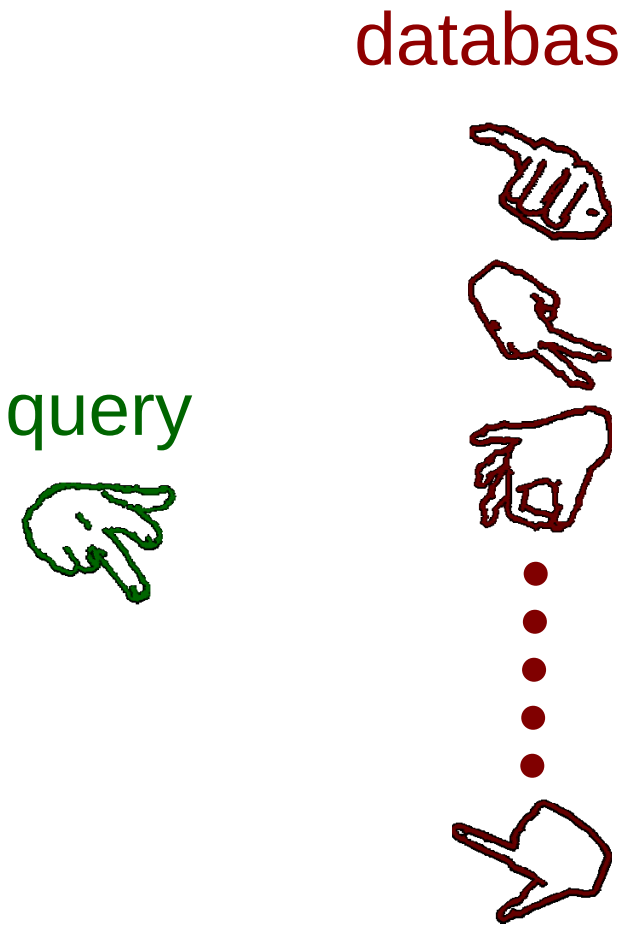
- Computing chamfer distances is slow.
  - For images with d edge pixels, O(d log d) time.
  - Comparing input to entire database takes over 4 minutes.
    - Must measure 107,328 distances.

The Nearest Neighbor Problem

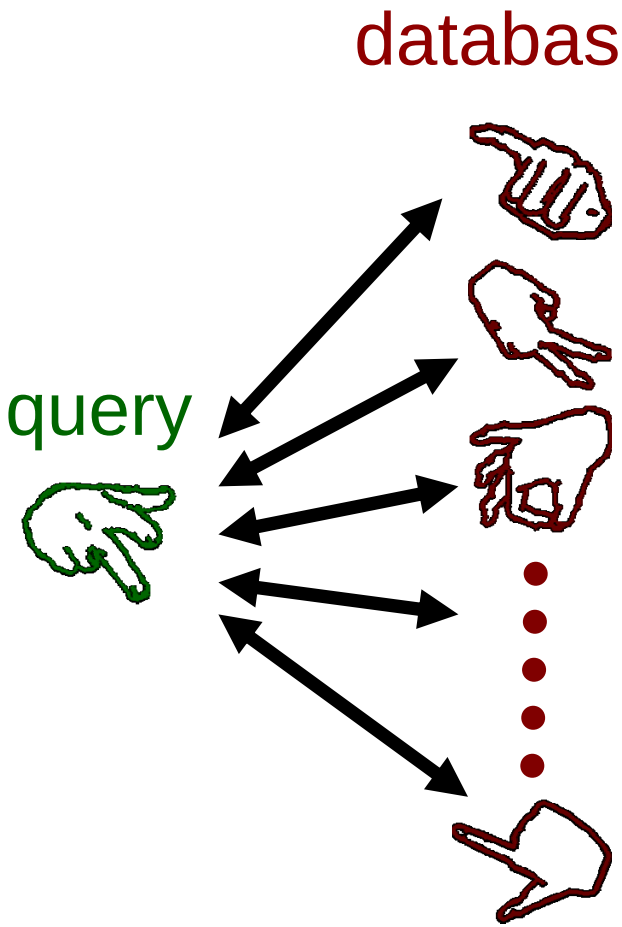
database

The Nearest Neighbor Problem

- Goal:
  - find the k nearest neighbors of query q.

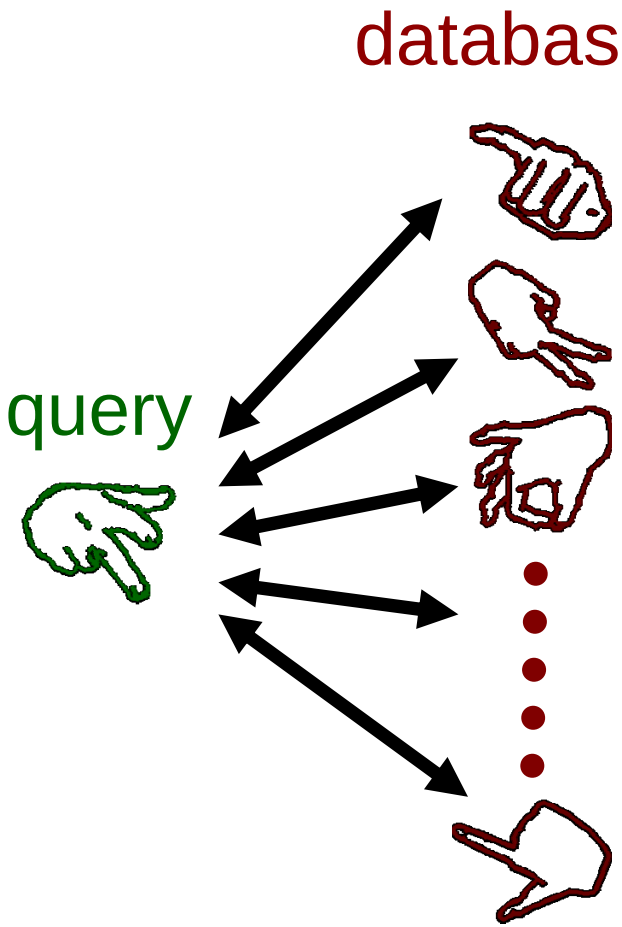


The Nearest Neighbor Problem



- Goal:
  - find the k nearest neighbors of query q.
- Brute force time is linear to:
  - n (size of database).
  - time it takes to measure a *single distance*.

The Nearest Neighbor Problem



- Goal:
  - find the k nearest neighbors of query q.
- Brute force time is linear to:
  - n (size of database).
  - time it takes to measure a *single distance*.