



# 3D Hand Pose Reconstruction With ISOSOM

Haiying Guan, Rogerio S. Feris, and Matthew Turk

Four Eyes Lab, Department of Computer Science  
University of California, Santa Barbara, USA  
{ haiying, rferis, mturk }@cs.ucsb.edu



## 1. Abstract

We propose an Isometric Self-Organizing Map (ISOSOM) method for nonlinear dimensionality reduction, which integrates a Self-Organizing Map model and an ISOMAP dimension reduction algorithm, organizing the high dimension data in a low dimension lattice structure. We apply the proposed method to the problem of appearance-based 3D hand posture estimation. As a learning stage, we use a realistic 3D hand model to generate data encoding the mapping between the hand pose space and the image feature space. The intrinsic dimension of such nonlinear mapping is learned by ISOSOM, which clusters the data into a lattice map. We perform 3D hand posture estimation on this map, showing that the ISOSOM algorithm performs better than traditional image retrieval algorithms for pose estimation. We also show that a 2.5D feature representation based on depth edges is superior to intensity edge features commonly used in previous methods.

## 2. Motivation

Objective:

Non-intrusive 3D hand pose estimation given a 2D unmarked hand image in any viewpoint

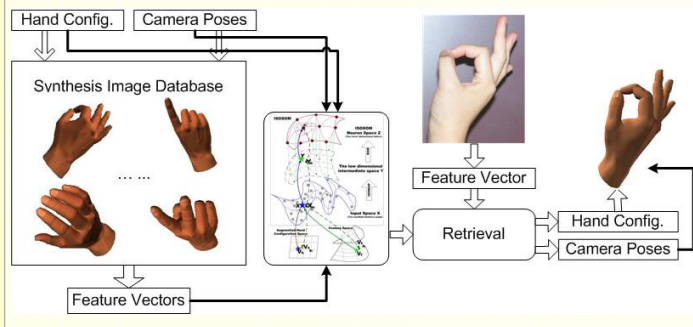
What Can We Do With Vision-Based Hand Pose Recognition?

Next generation user interfaces:

- Gesture control of true 3D display
- Large screen display
- Sign language recognition
- Wearable device interface



## 3. System Overview



## 4. ISOSOM

SOM

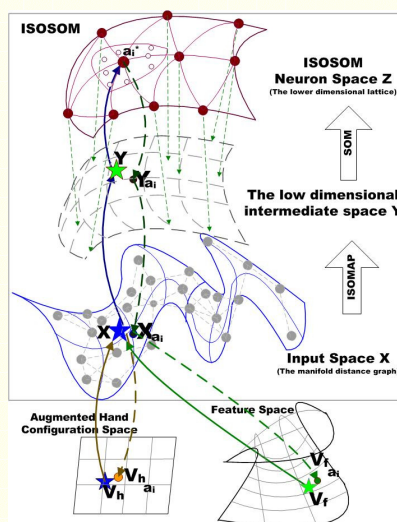
- clustering in a low dimensional grid

ISOMAP

- nonlinear dimension reduction preserving the dataset's global geometry

ISOSOM

- clustering to reduce the redundancy
- following the better topology of the dataset
- preserving the geometry relationships in high dimensional input space to the low dimension ISOSOM lattice map

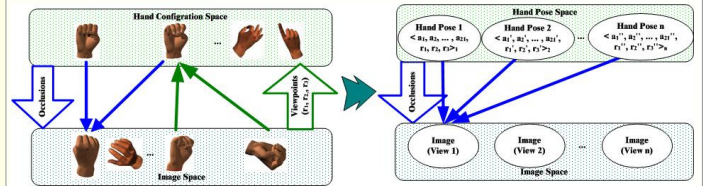


## 5. ISOSOM for Hand Pose Estimation

3D hand pose estimation from a 2D image is a highly nonlinear, many-to-many mapping problem in high dimension space.

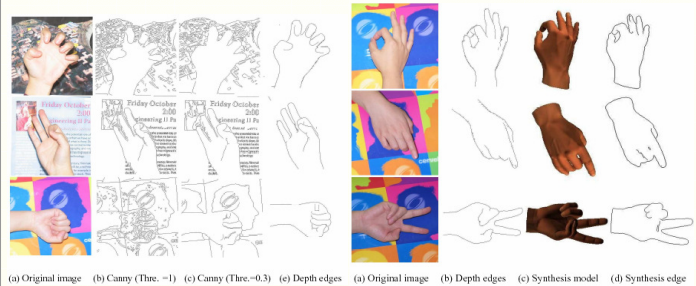
Feature vectors of different hand configurations are highly mixed up in feature space.

Augment hand configuration with 3 global camera rotation parameters to avoid feature mix-up.



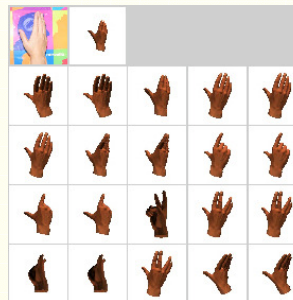
## 6. Depth Edge by Multi-flash Camera

A multi-flash camera with flashes strategically positioned to cast shadows along depth discontinuities is used to bypass scene reconstruction and detect depth edges directly. This method allows reliable hand shape acquisition (including internal finger edges), while considerably reducing background clutter.

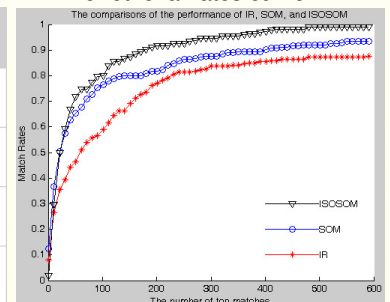


## 7. Results

1. The retrieval result



2. The retrieval rates curve



3. The correct retrieval rates

Number	IR	SOM	ISOSOM
Top 40	44.25%	62.39%	65.93%
Top 80	55.75%	72.12%	77.43%
Top 120	64.60%	78.76%	85.40%
Top 160	70.80%	80.09%	88.50%
Top 200	76.99%	81.86%	91.59%
Top 240	81.42%	85.84%	92.48%
Top 280	82.30%	87.17%	94.69%

In order to achieve more than 85% successful rate, basic image retrieval (IR) provides more than 280 hand pose possibilities. ISOSOM just requires less than half of that number with the same precision rate.

## 8. Conclusion

Contributions

- ISOSOM for nonlinear dimension reduction
- Formulation of 3D hand pose estimation from a 2D image to a nonlinear mapping approach in high dimension space
- Depth edges by multi-flash camera