A stable skeletonization for tabletop gesture recognition

Andoni Beristain, Manuel Graña

Computational Intelligence Group,
Universidad del Pais Vasco

Contents

- Introduction
- Skeletonization algorithm
- Results
  - Real time implementation
  - Classification of gestures for tabletop interaction

Introduction

- Aim: gestual interaction with tabletop systems
  - Naturalness of the interaction
  - Basic gestures: grabbing, pointing, passing a page.
- Approach:
  - Skeleton as a key feature extraction
  - Classification based on skeleton features

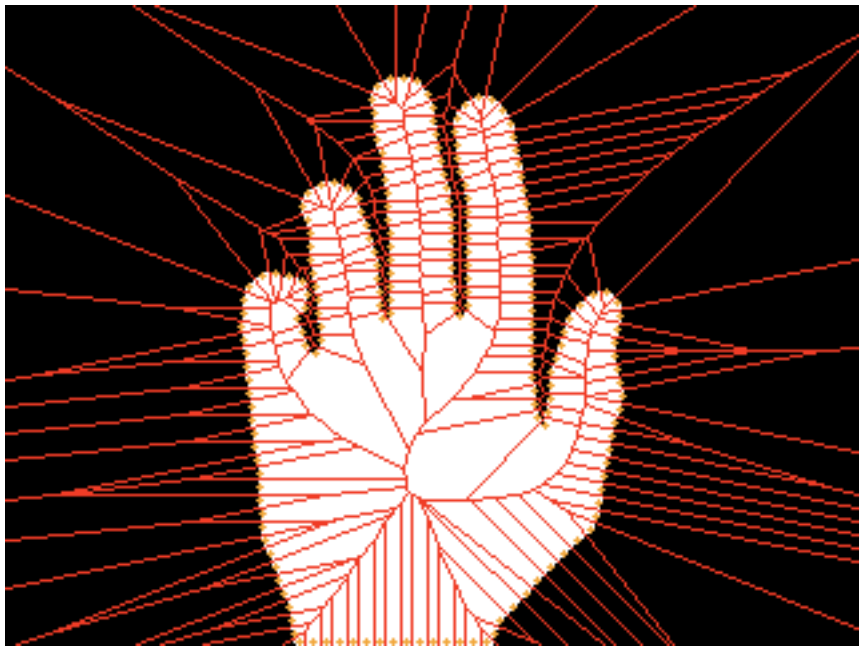
Introduccion

- Skeletonization procedure
  - Voronoi skeleton
  - Two Pruning steps
    - Removing edges crossing the object boundary
    - Discrete Curve Evolution based approach

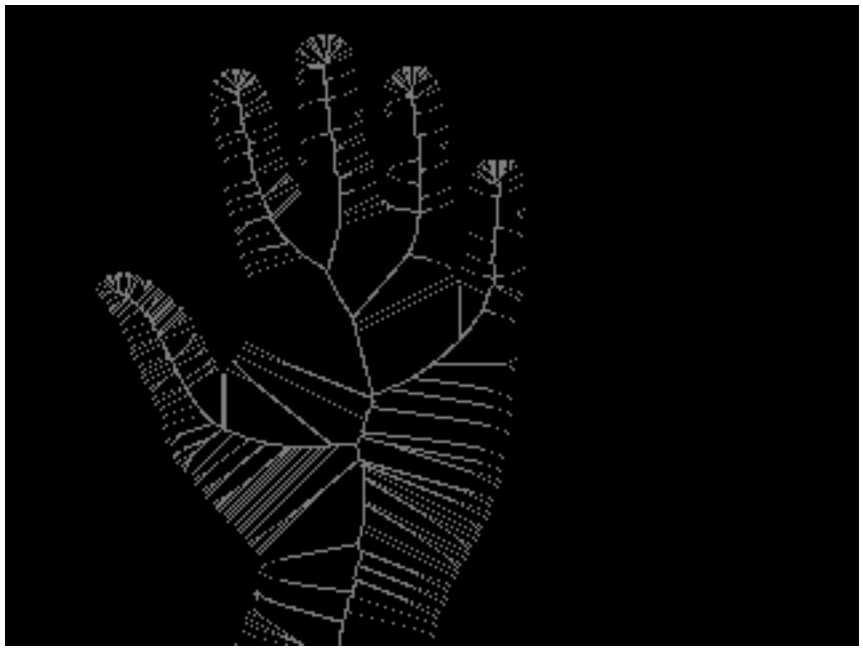
Skeletonization algorithm

- Starting from the boundary of the object
- Four steps
  - Shape boundary subsampling
  - Voronoi Tessellation computation:
    - the Voronoi edges inside the form are the skeleton branches
  - Discrete Curve Evolution computation (DCE) on the boundary
  - Pruning

- Voronoi tesselation & skeleton



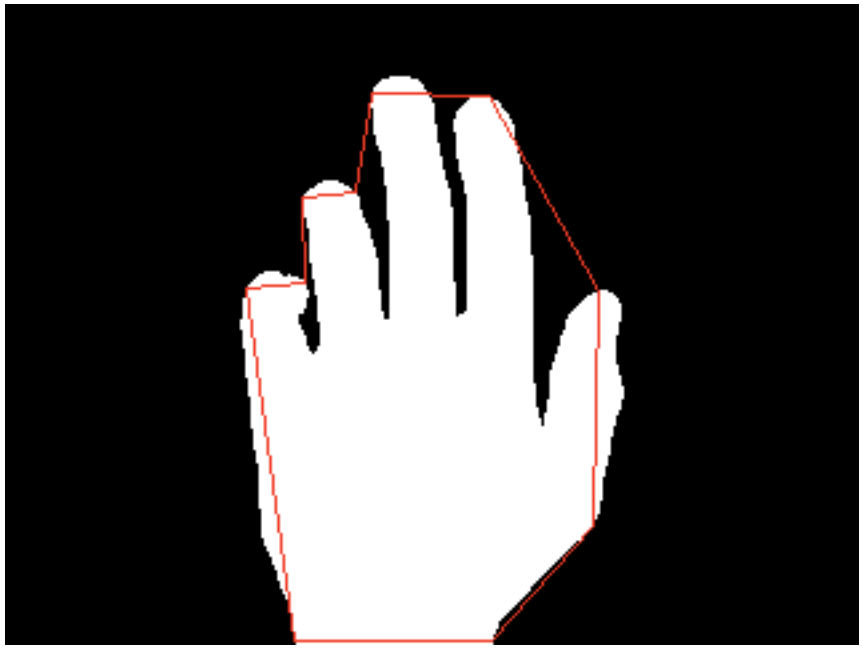
- Voronoi skeleton



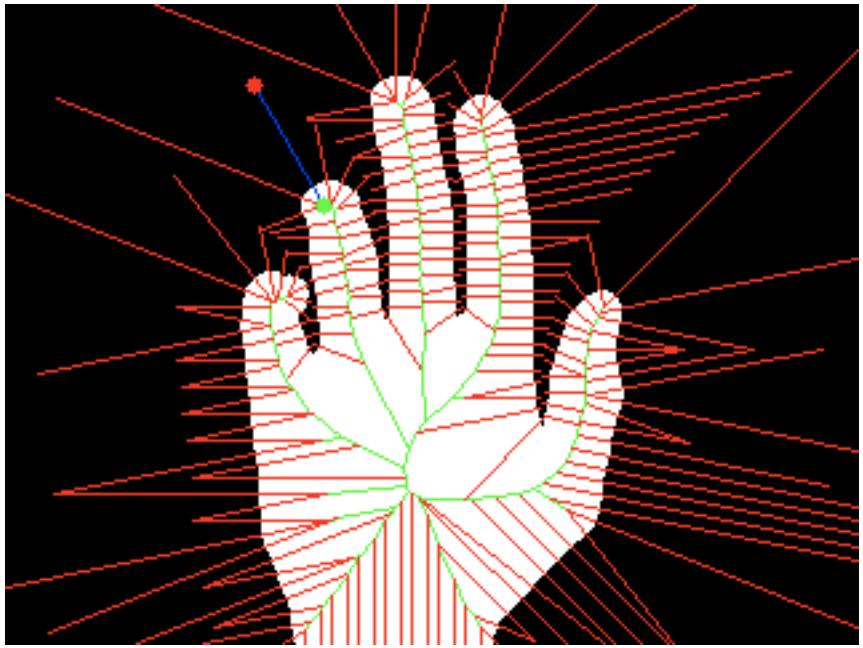
Skeletonization algorithm

- Discrete Curve Evolution
  - The contour is iteratively simplified removing the points with the lowest saliency curvature coefficient
  - The stopping criterion is the minimal number of points in the contour
  - Remaining vertices are joined by lines to form the simplified contour

- Result of DCE for the hand shape



- First pruning: removing Voronoi edges crossing the boundary

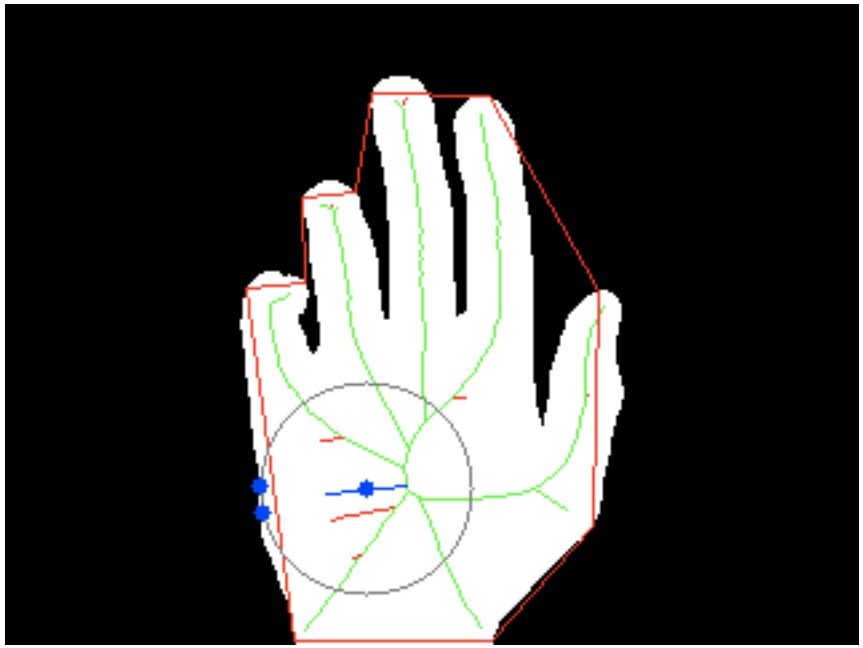


- Pruning speedup

  – Only the two extreme points of the Voronoi segment need to be tested to determine if the whole segment belongs to the object

- Second pruning step
  - Compute the convex hull on the DCE contour
  - Voronoi edges with generative points belonging to the same convex hull segment are removed

- Segments removed in the second pruning step



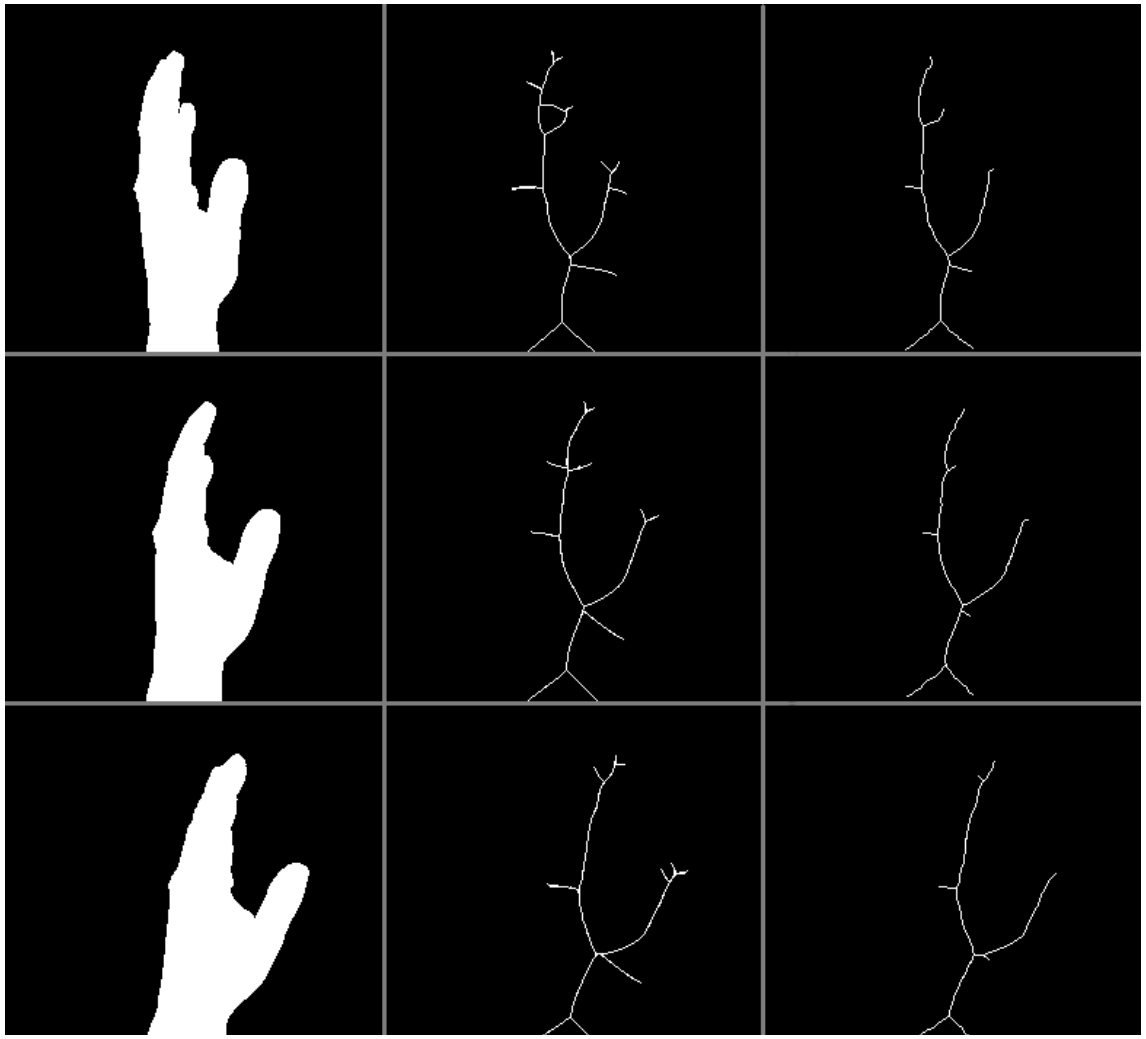
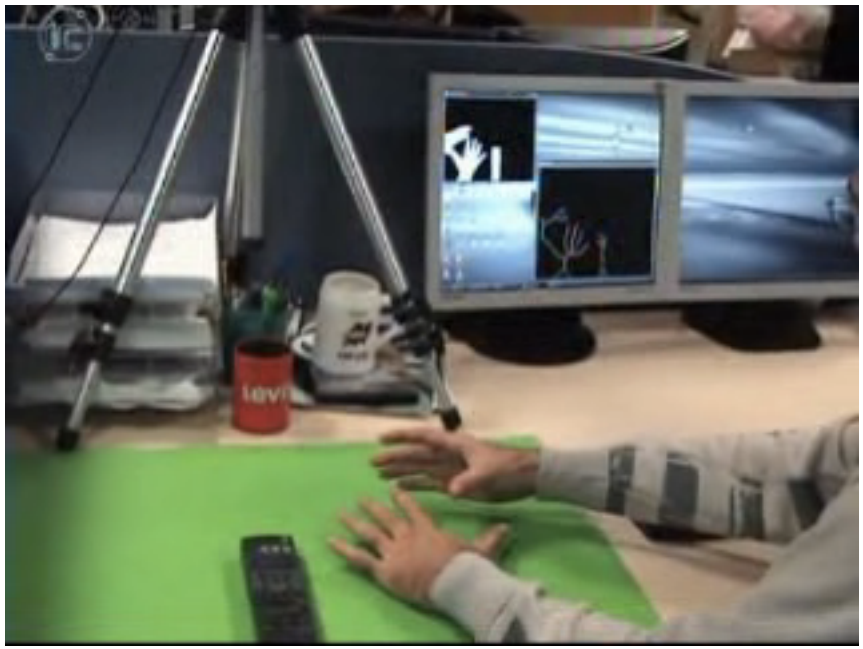


Fig. 1. Left column: Two binary images of the turning page gesture (see section 3.2 below). Middle Column: Skeletons obtained using the implementation of algorithm [2] provided in [1]. Right column: Skeletons obtained using the approach in this paper.

Results

- Real time implementation



- http://www.ehu.es/ccwintco/index.php/2009-10-24-video-skeleton

Results

- Tabletop gesture interaction database
- Generated in-house
- Made available to the public

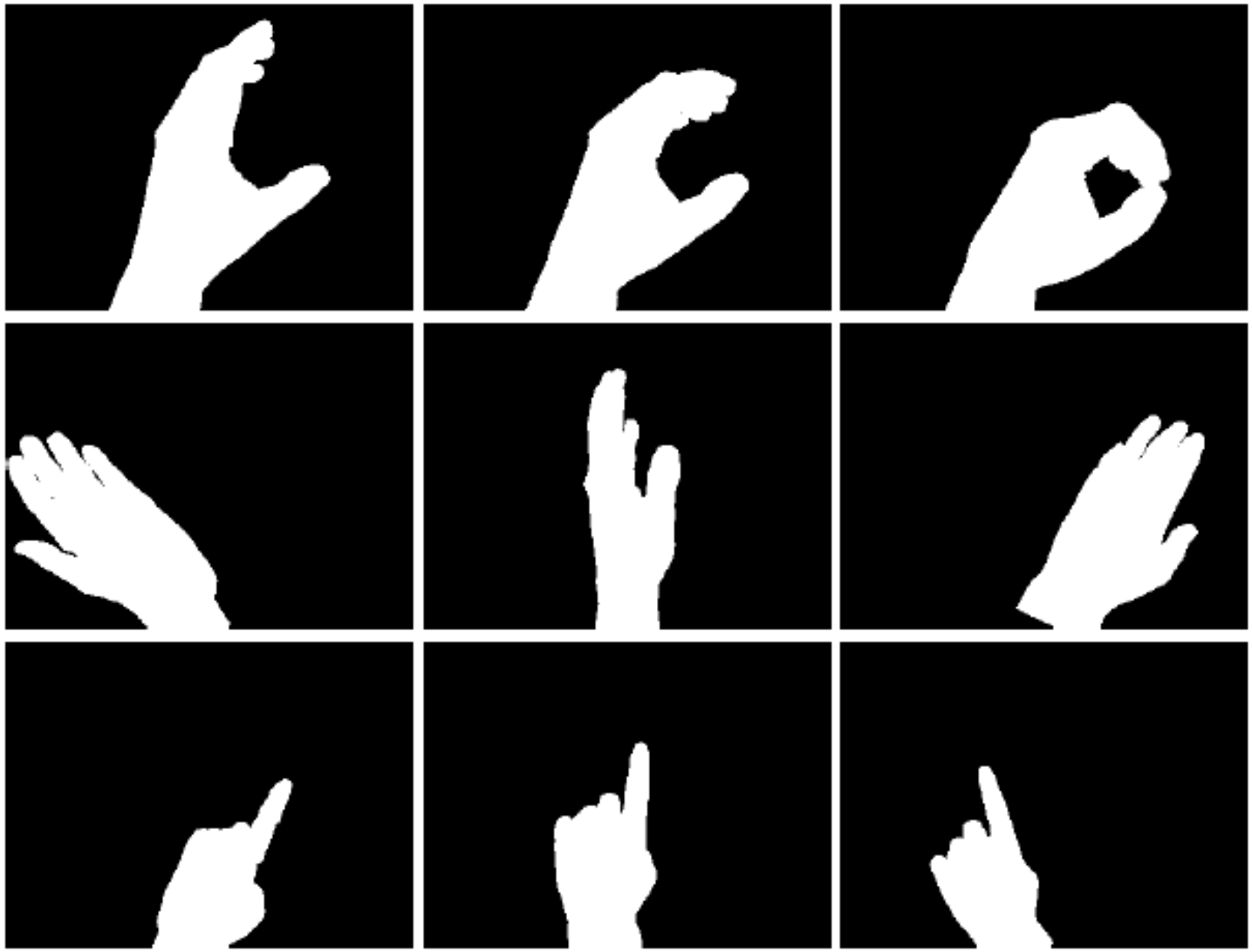


Fig. 3. Hand gestures for natural tabletop interaction, from top to bottom, the sequences correspond to grab,

- Distance between skeletons computed by a greedy matching algorithm
  - Nodes are extreme vertices of the Voronoi edges
  - Associate nodes in pairs following the order from the closest to the image window boundary to its center

$$|A,B|_{ND} = \alpha\,|A_{dt} - B_{dt}| + (1-\alpha)\,\sqrt{A_r^2 + B_r^2 - 2A_rB_r\cos\left(A_\theta - B_\theta\right)} \qquad (5)$$

$$|A,B|_{ND} \in [0,1].$$

$$GS\left(T_i,S\right) = \frac{2\sum_k\left(1 - |t_k,s_k|_{ND}\right)}{\#T_i + \#S} \qquad (6)$$

- Patter recognition:
  - K-NN algorithms based on the greedy graph distance
- Validation over the tabletop gesture database
  - Perform ten-fold cross validation
  - Compare with the Bai skeletonization algorithm

ϕ_E	α	Classifier	Bai	Beris
15	0.25	1NN	72.06	70.74
15	0.25	3NN	80.53	80.21
15	0.25	5NN	83.67	83.81
15	0.50	1NN	78.30	80.73
15	0.50	3NN	87.01	89.17
15	0.50	5NN	89.84	91.96
15	0.75	1NN	82.30	84.78
15	0.75	3NN	90.71	92.27
15	0.75	5NN	**93.25**	**94.57**

Table 1. Recognition success of the 3 hand gesture classes.

	Grab	Point	Turn Page
Grab	**266725**	603	32672
Point	82	**298783**	1135
Turn Page	10607	3731	**285662**

Table 2. Total recognition Confusion Matrix for the best case of our algorithm (Beris) with 3 classes: pruning 15, alpha 0.75 and 5NN

Conclusions

- Our algorithm (Beris) outperforms the Bai algorithm recognition in most of cases.
- Higher values of the α parameter in the Node Distance produce better results.
  - This results mean that the geometrical position of nodes is more important for matching than the DT value.

- Against our expectations, less DCE pruning does not produce better recognition results.
  - it seems that higher DCE pruning parameter values add clutter to the representation, instead of enriching it.
- The confusion between the grab and page turn classes can be explained as follows:
    - We are classifying all the instantaneous poses of a gesture as this gesture.
    - Some of the intermediate poses of these classes are nearly indistigu- ishable.
    - The classification of these poses is at the surce of confussion.