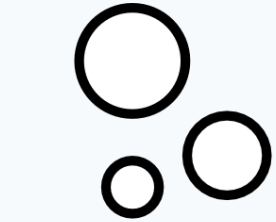
On-line Gesture Recognition



Sciling

Luis A. Leiva
name@sciling.com

Presentation Outline

Slides available at https://luis.leiva.name/lectures/gestures-famnit.pdf

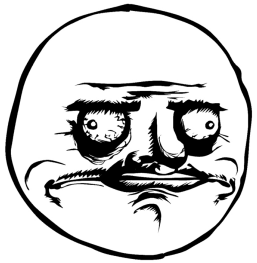
Introduction

Definitions

Gesture /ˈdʒstʃə/ *noun*
The use of motions of limbs or body as a means of expression.

Synonyms: signal, sign, motion, indication, gesticulation

"Gestures are hand-drawn strokes that do things."

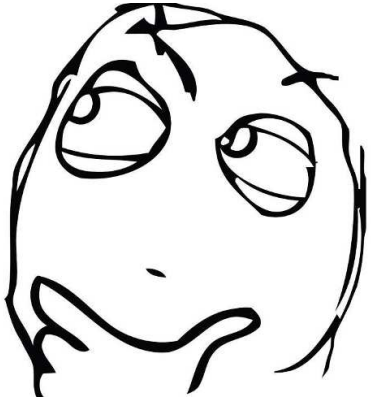


— Lipscomb (1991)

Definitions

1. **Off-line** gesture recognition:
   - *post-hoc*, processed after user interaction
   - static data, no temporal info available

2. **On-line** gesture recognition:
   - realtime, direct manipulation
   - sequential, time series data



Historical Precedents

Sketchpad

Sutherland (1963)

RAND tablet

Davis and Ellis (1964)

Gestures Today

Minority Report. Image by 20th Century Fox & DreamWorks

Input Devices

Wii. Image by Nintendo Co., Ltd.

Input Devices

T(ether). Image by Massachusetts Institute of Technology

Input Devices

Kinect for Xbox 360. Image by Microsoft Corporation

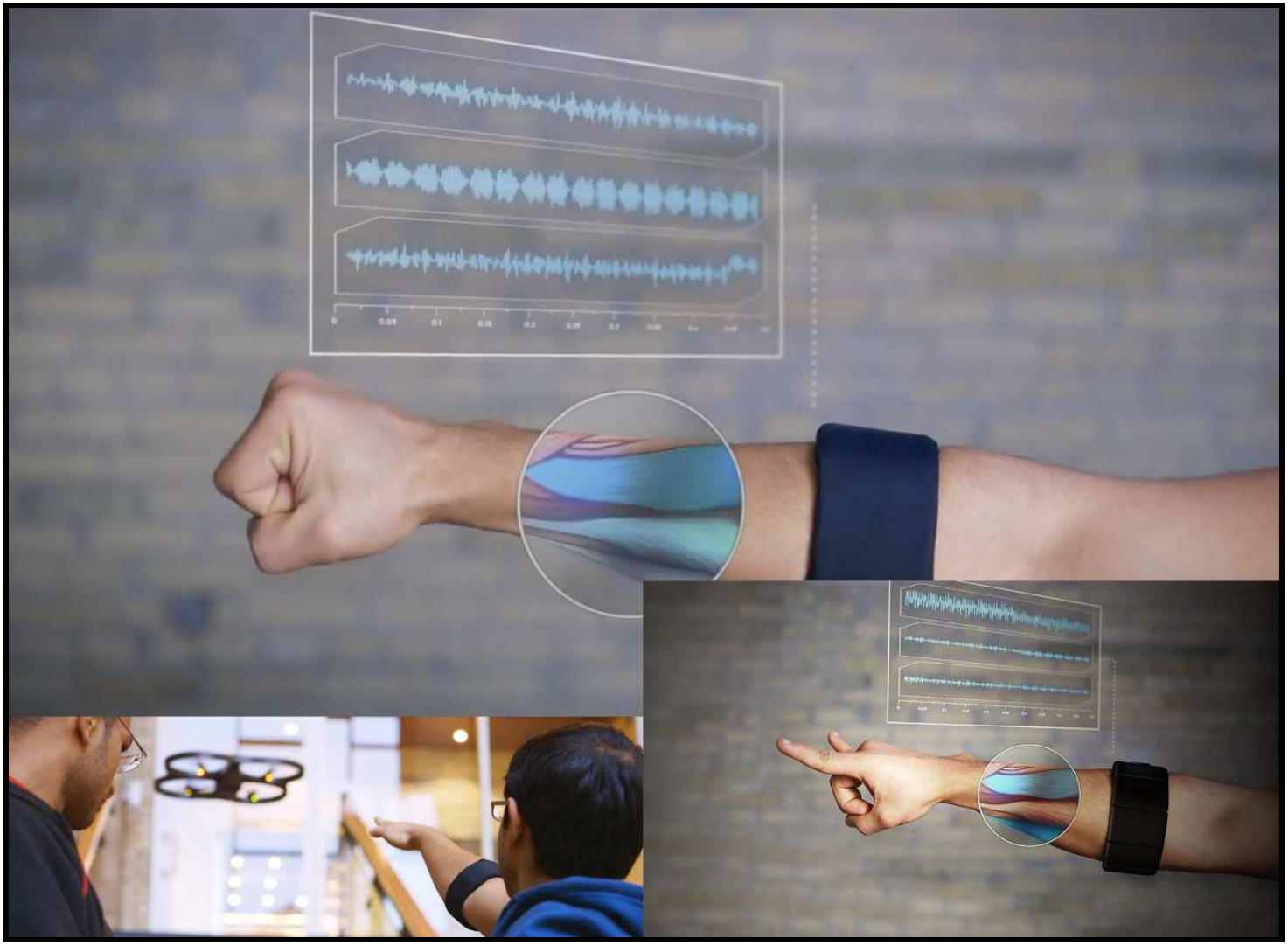
Input Devices

Humantenna. Image by Microsoft Research

Input Devices

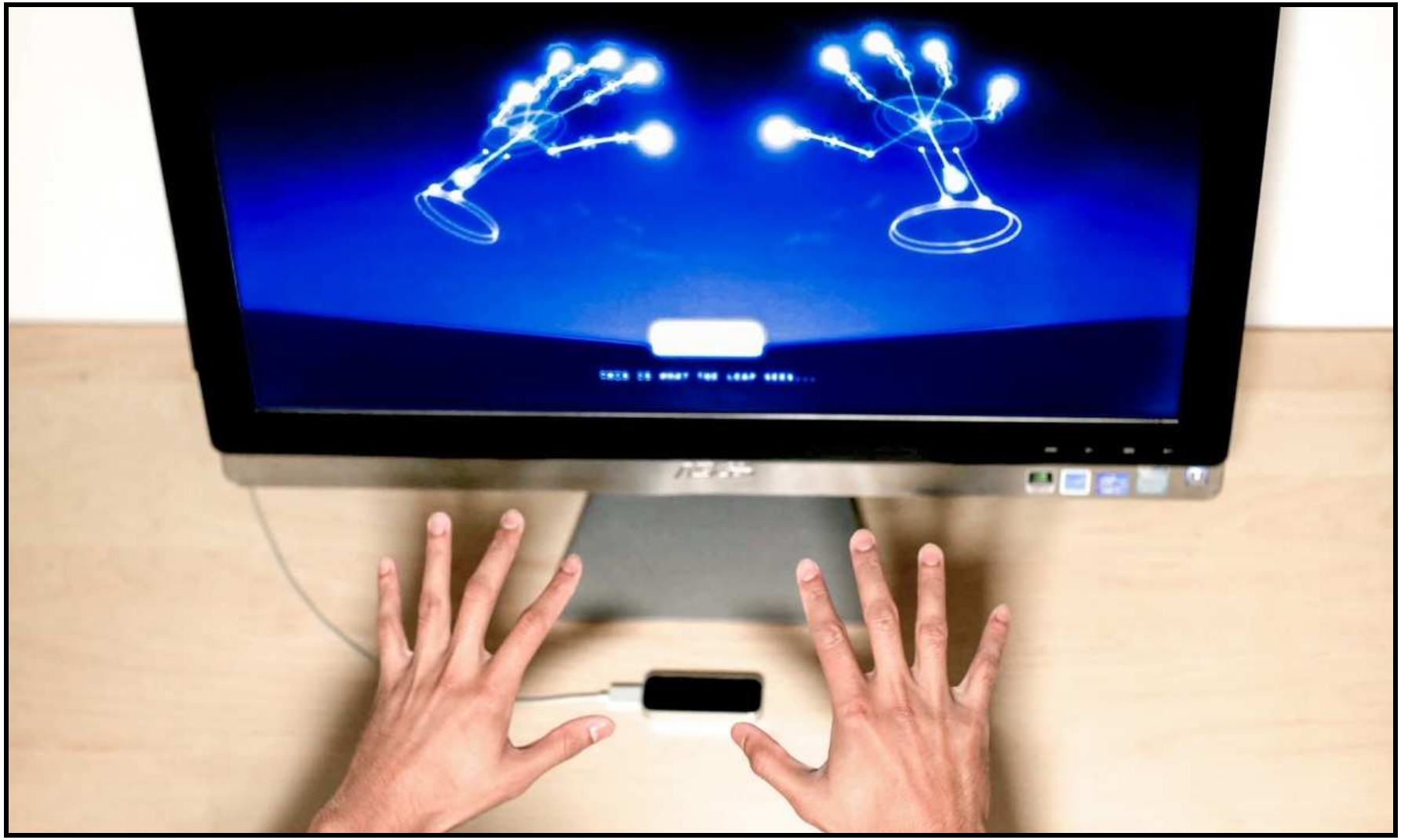
Skinput. Image by Carnegie Mellon University

Input Devices



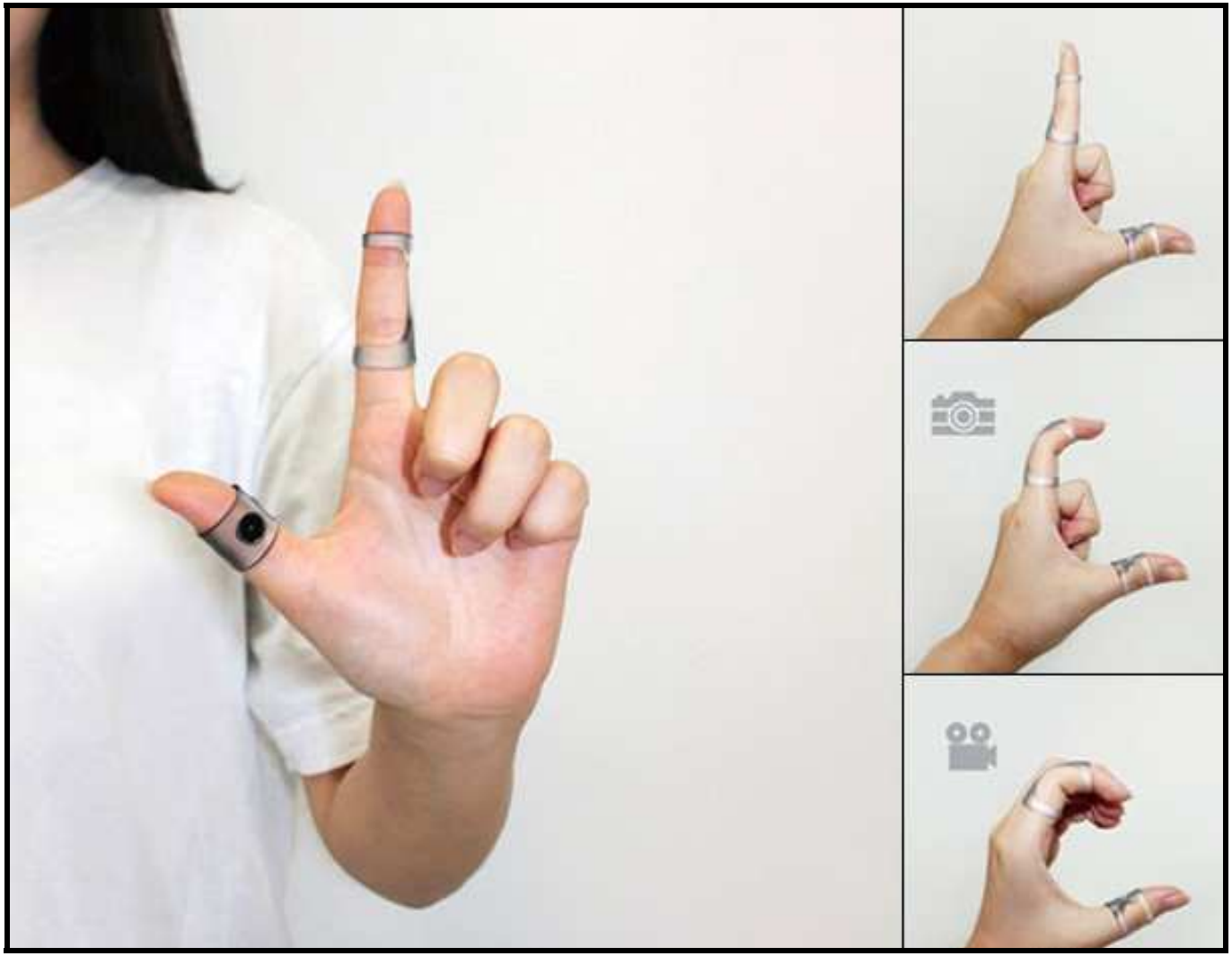
Myo. Image by Thalmic Labs

Input Devices



Leap motion. Image by Leap Motion, Inc.

Input Devices

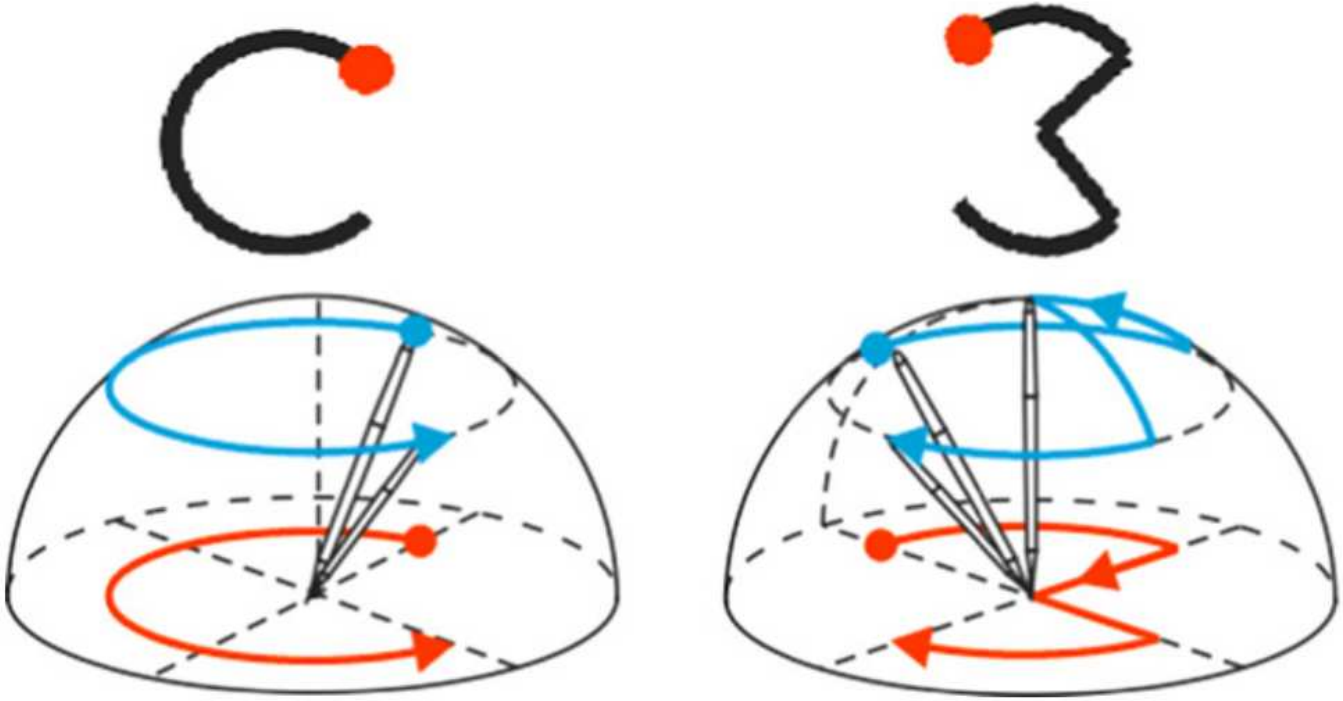


Air Clicker. Image by Yanko Design

Input Devices

TapTap. Image by Woodenshark LLC

Input Devices



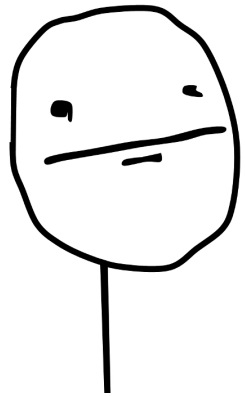
Pen Tail gestures, by Tian et al. (2012)

Advantages

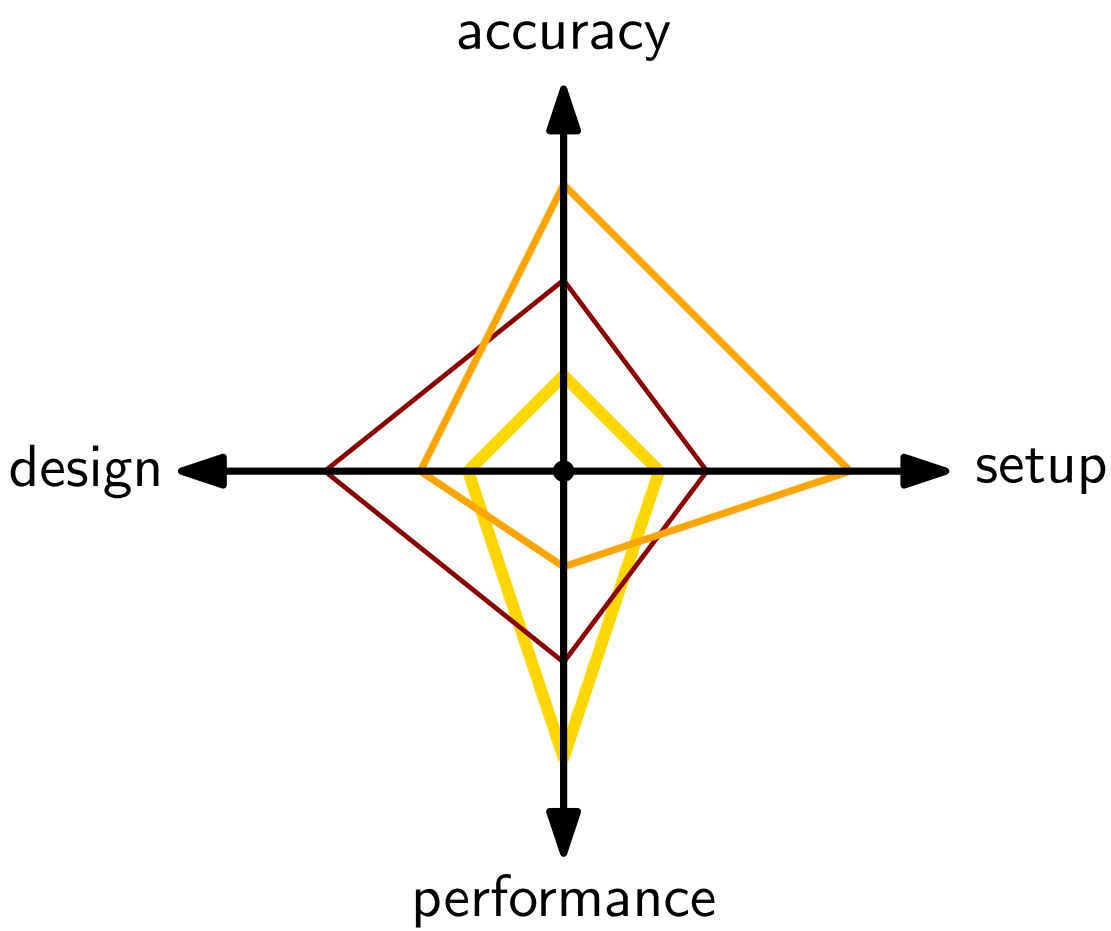
- Natural communication
- Expressiveness
- Ergonomics
- Usability
- Fun

Disadvantages

- May break fundamental interaction principles:
  - Discoverability, **Reliability**, Scalability, etc.
- Ambiguity: **non-deterministic decoding**
- Lack of standards
- Cultural issues



Trade-offs



Preliminaries

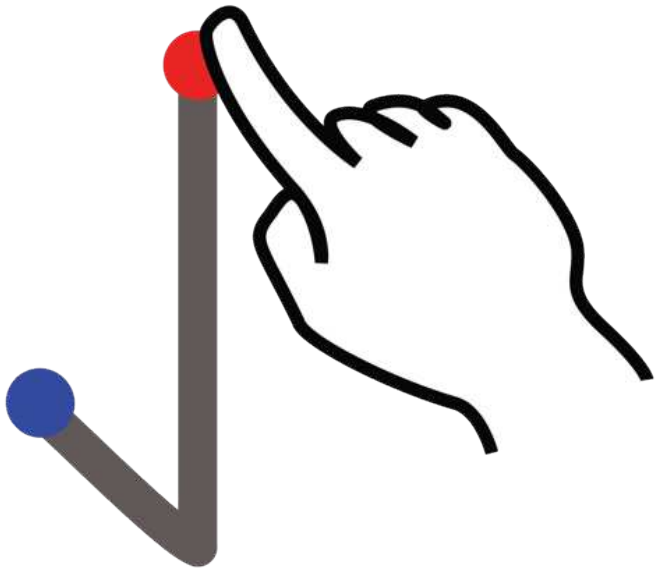
Interaction Paradigms

Mid-air

Onscreen

Definition

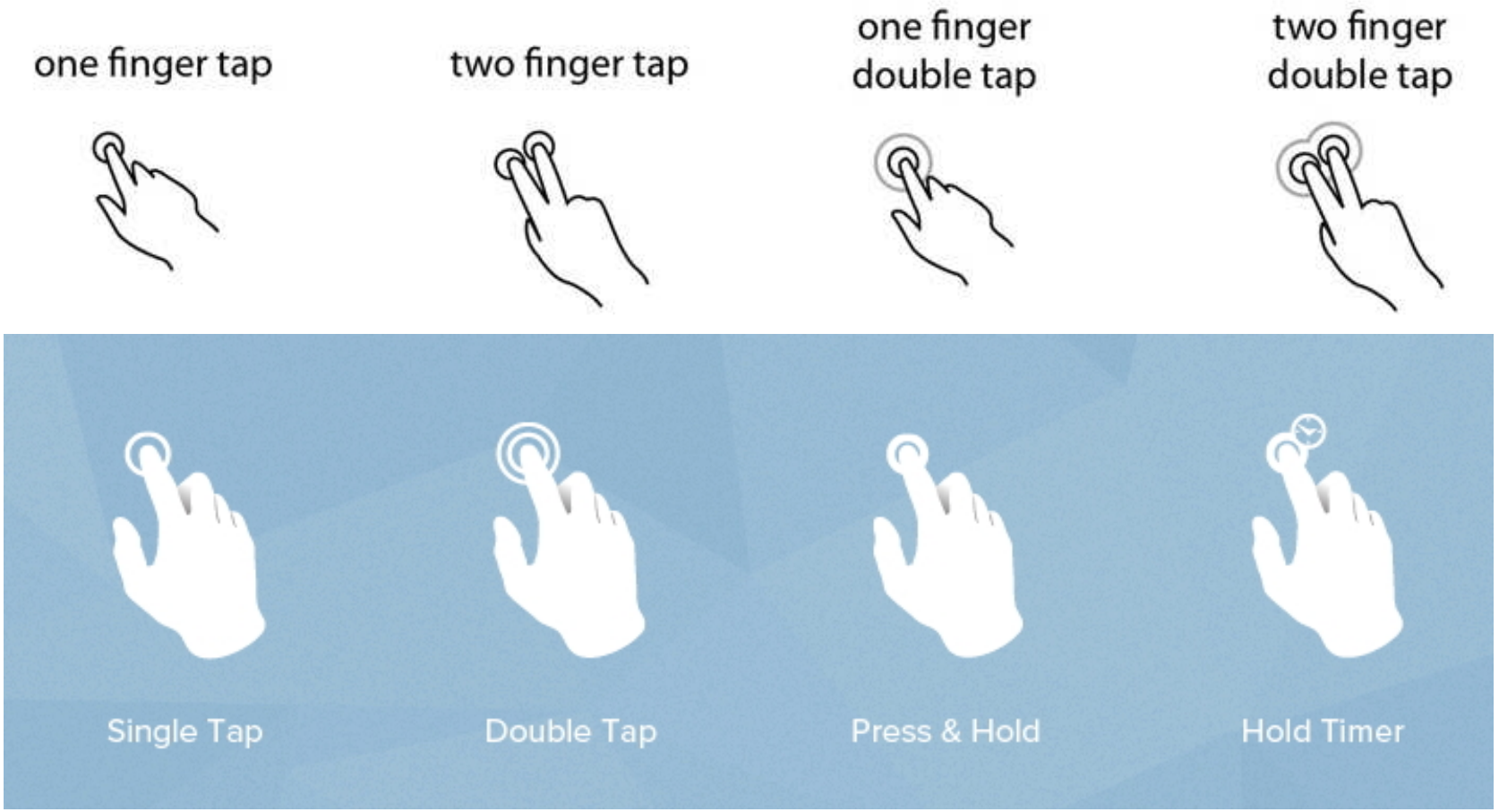
stroke $=$ pointer **down** $\rightarrow$ pointer **move** $\rightarrow$ pointer **up**



$$\boldsymbol{s} = \{(x_1, y_1, t_1) \cdots (x_j, y_j, t_j) \cdots (x_N, y_N, t_N)\}$$

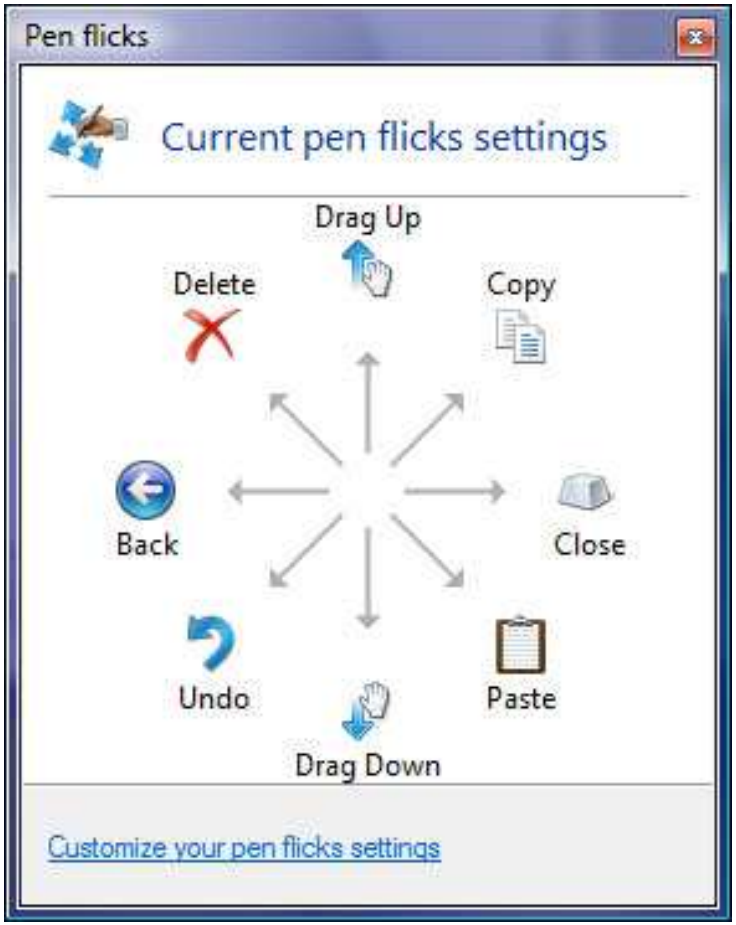
A Taxonomy

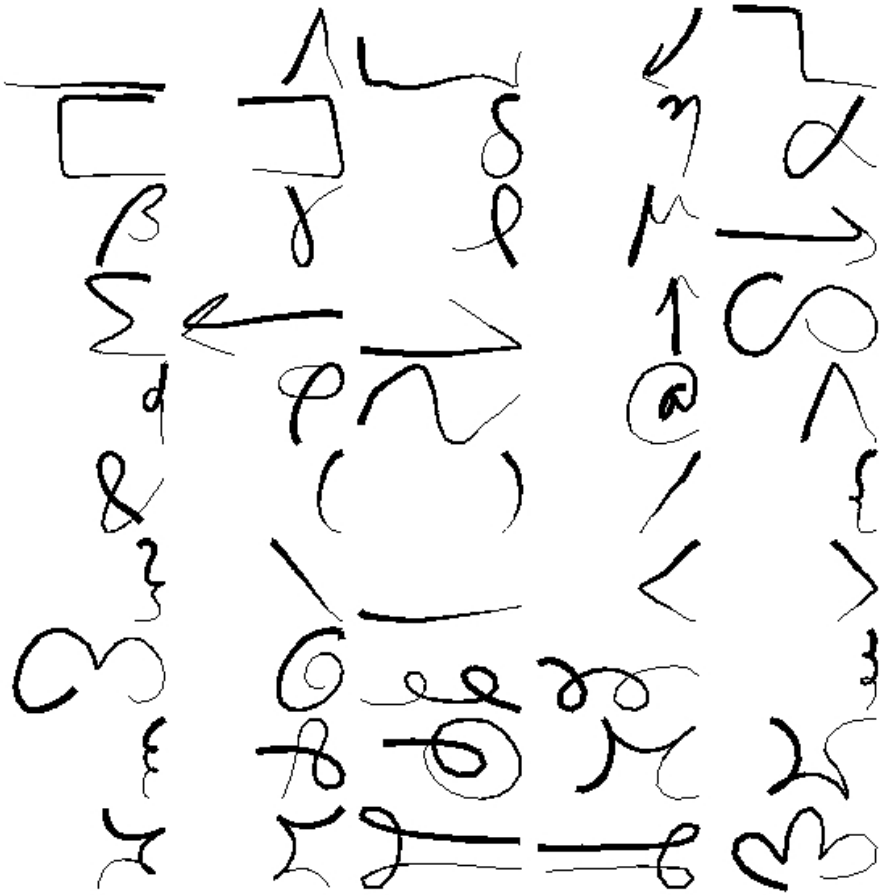
1. **zero-order** gestures



A Taxonomy

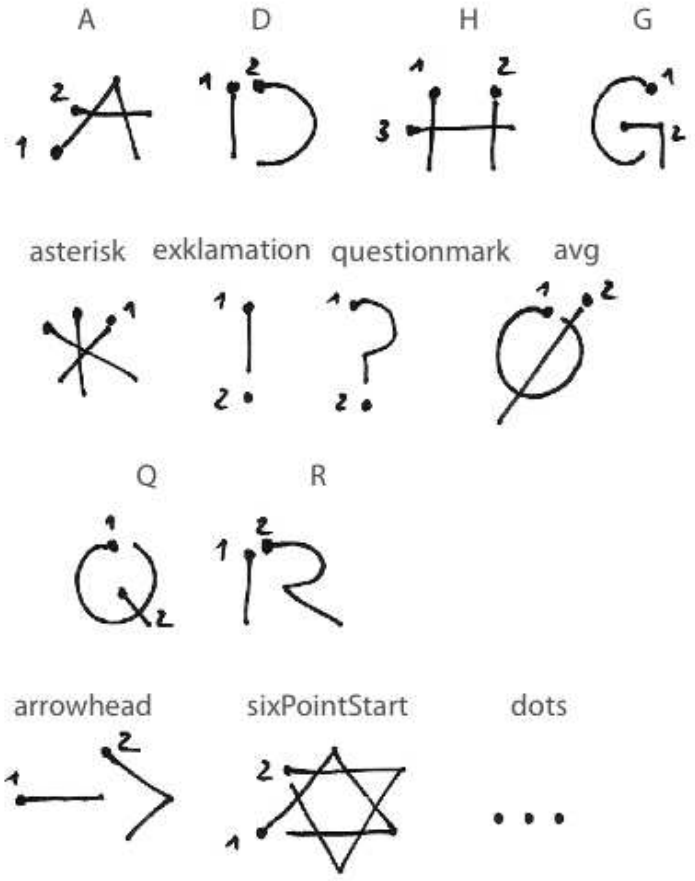
2. **first-order** gestures (unistrokes)



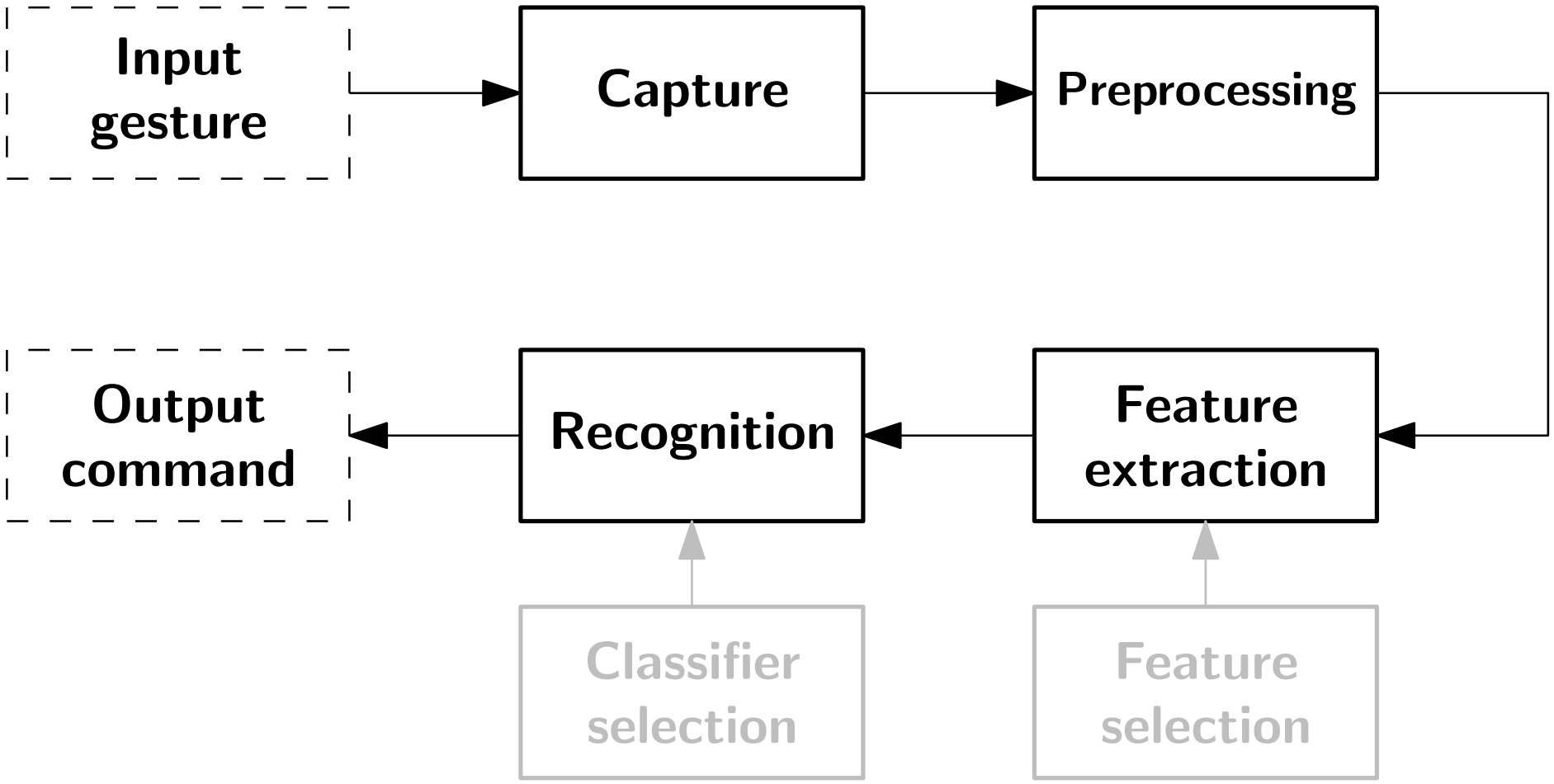


A Taxonomy

3. **higher-order** gestures (multistrokes)

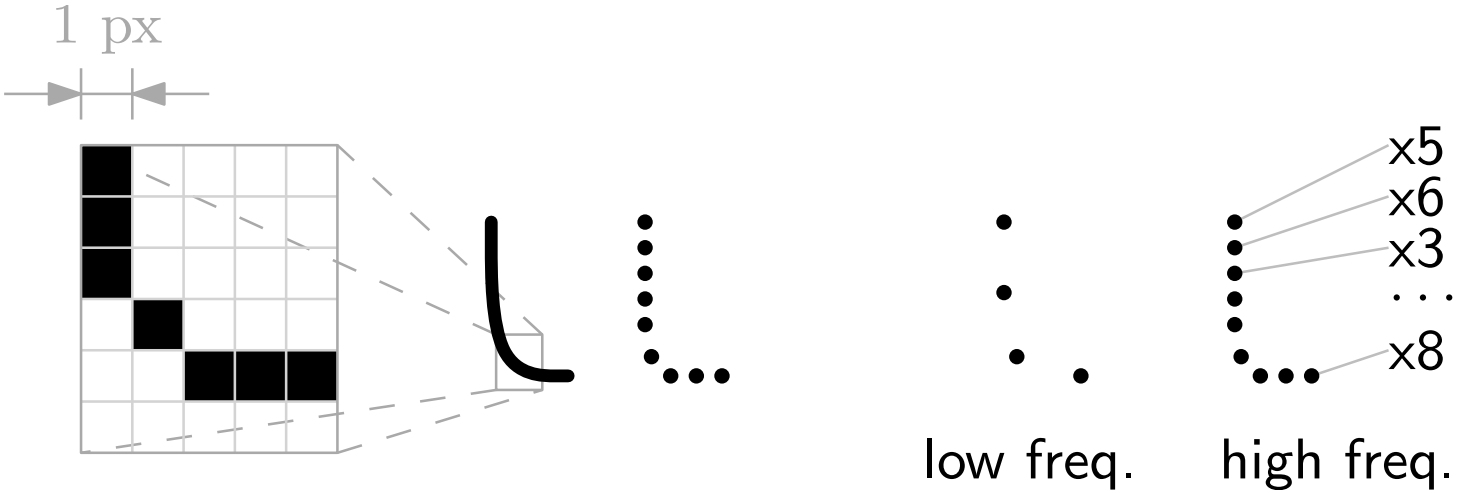


Processing Pipeline



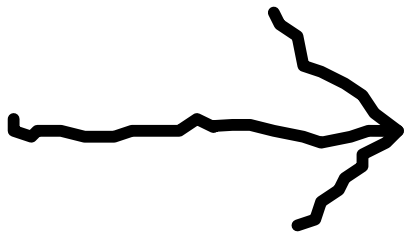
Capture

- Event-based
- Polling (constant frequency)

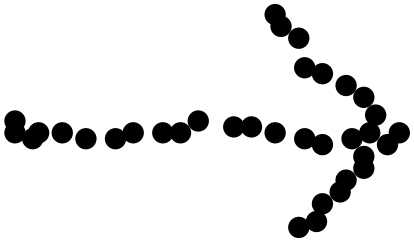


Sampling rate matters!

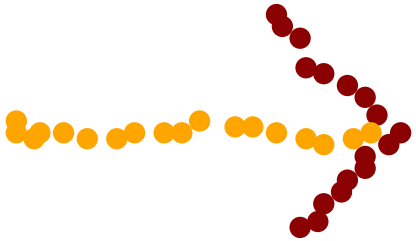
Preprocessing



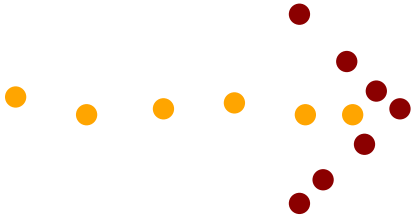
input



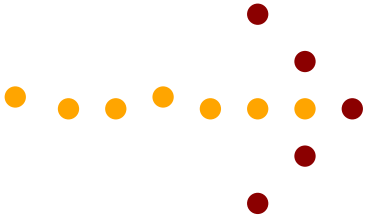
capture



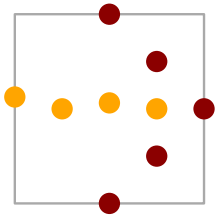
segmentation



noise removal*



resampling*



normalization*

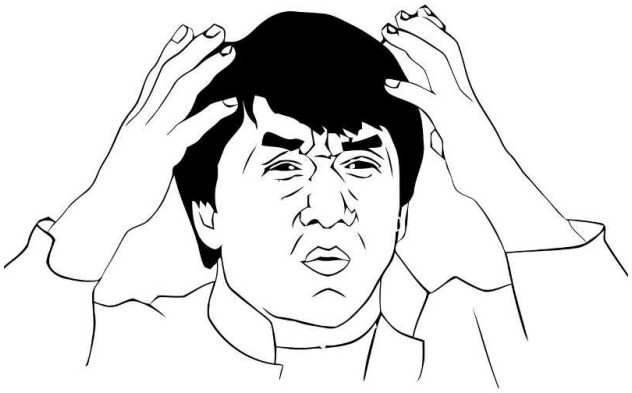
*optional steps

Features

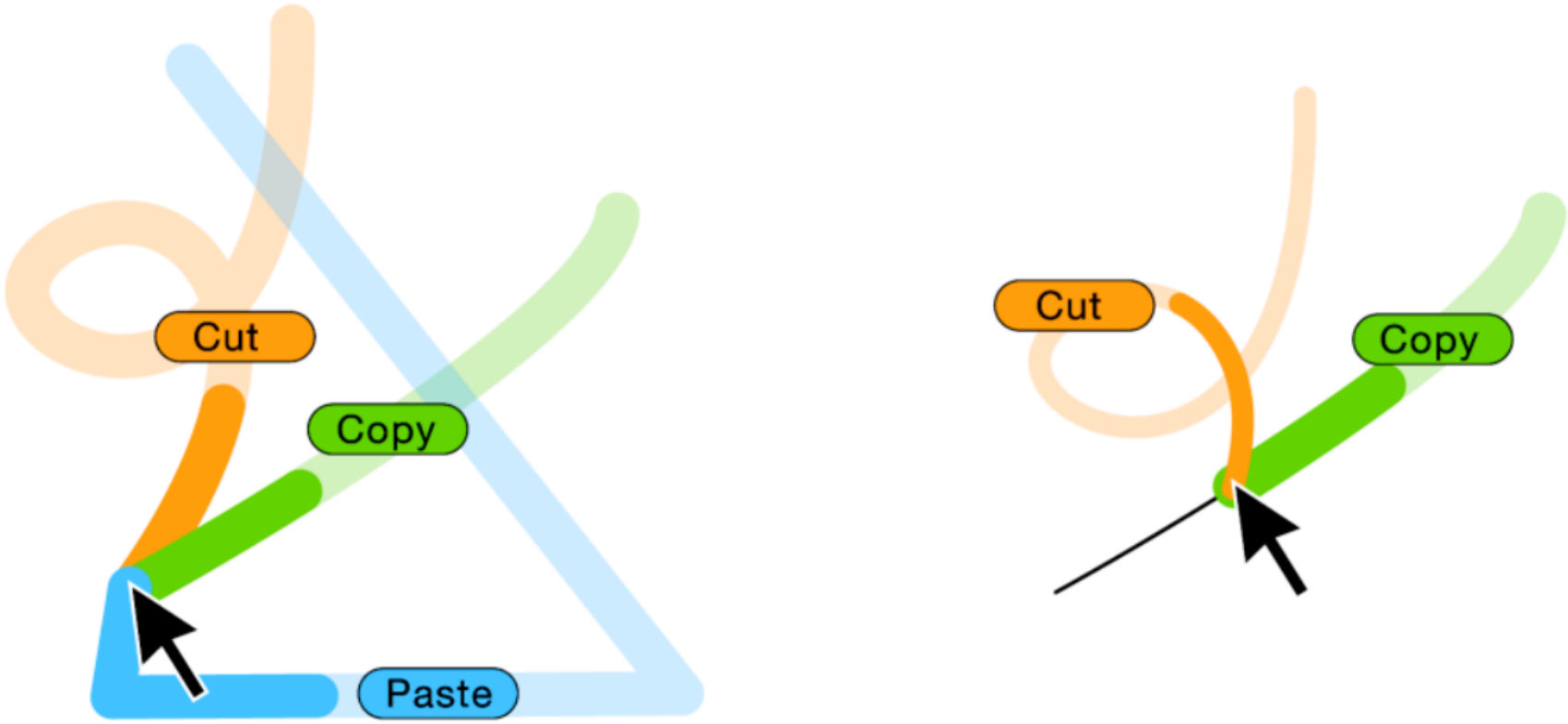
Feature Engineering is an art!

Recognition Techniques

- **Hashing**: dictionary lookup, zone coding, chain codes
- **Parametric**: linear fitting, corner detection
- **Matching**: DTW, k-NN, “dollar family”
- **Statistical**: GLM, RF, ANN, HMM, CRFs
- **Ad-hoc**: knowledge-based, decision trees, FSM



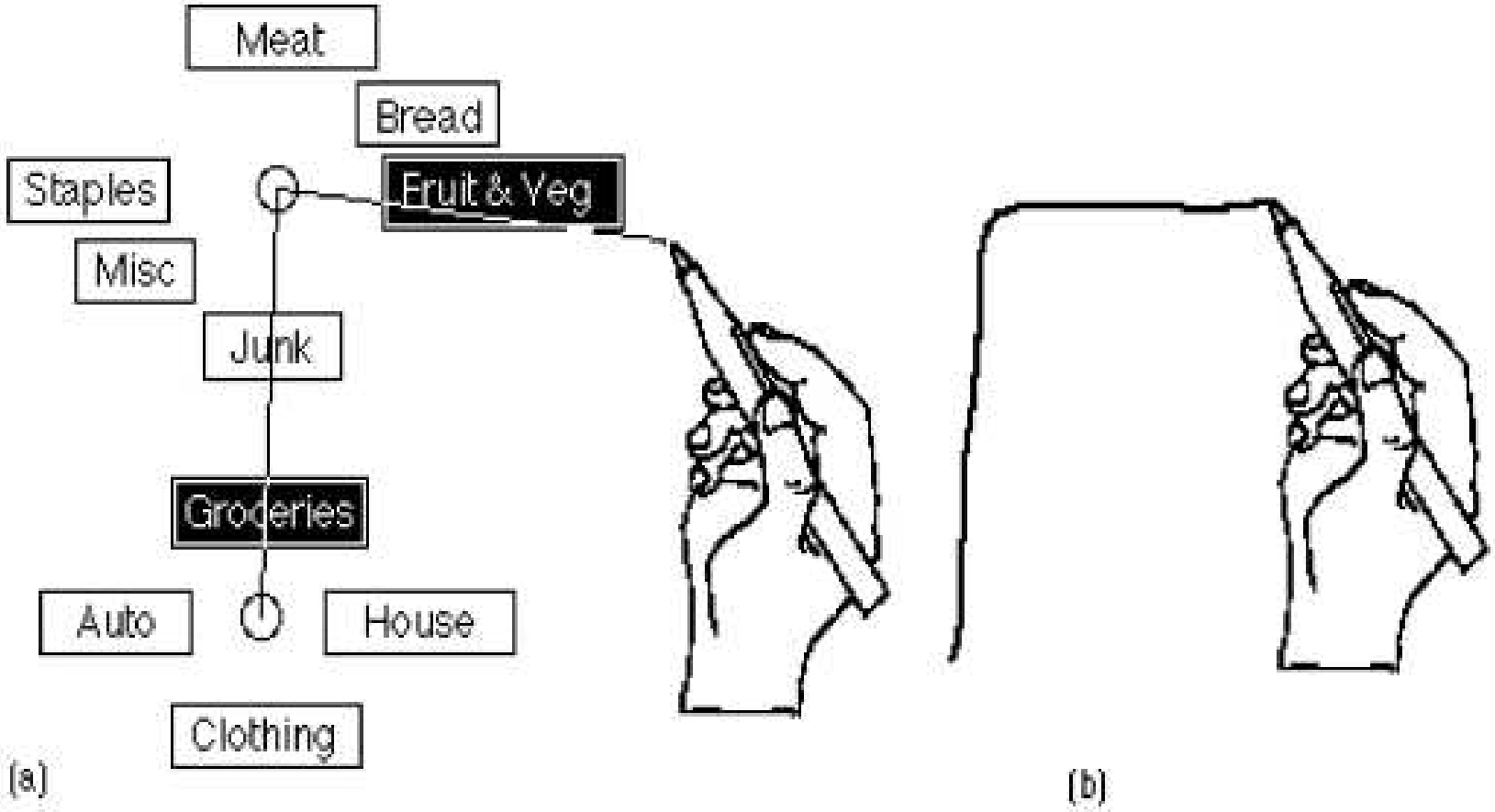
Bonus: Continuous Recognition



OctoPocus, by Bau and Mackay (2008)

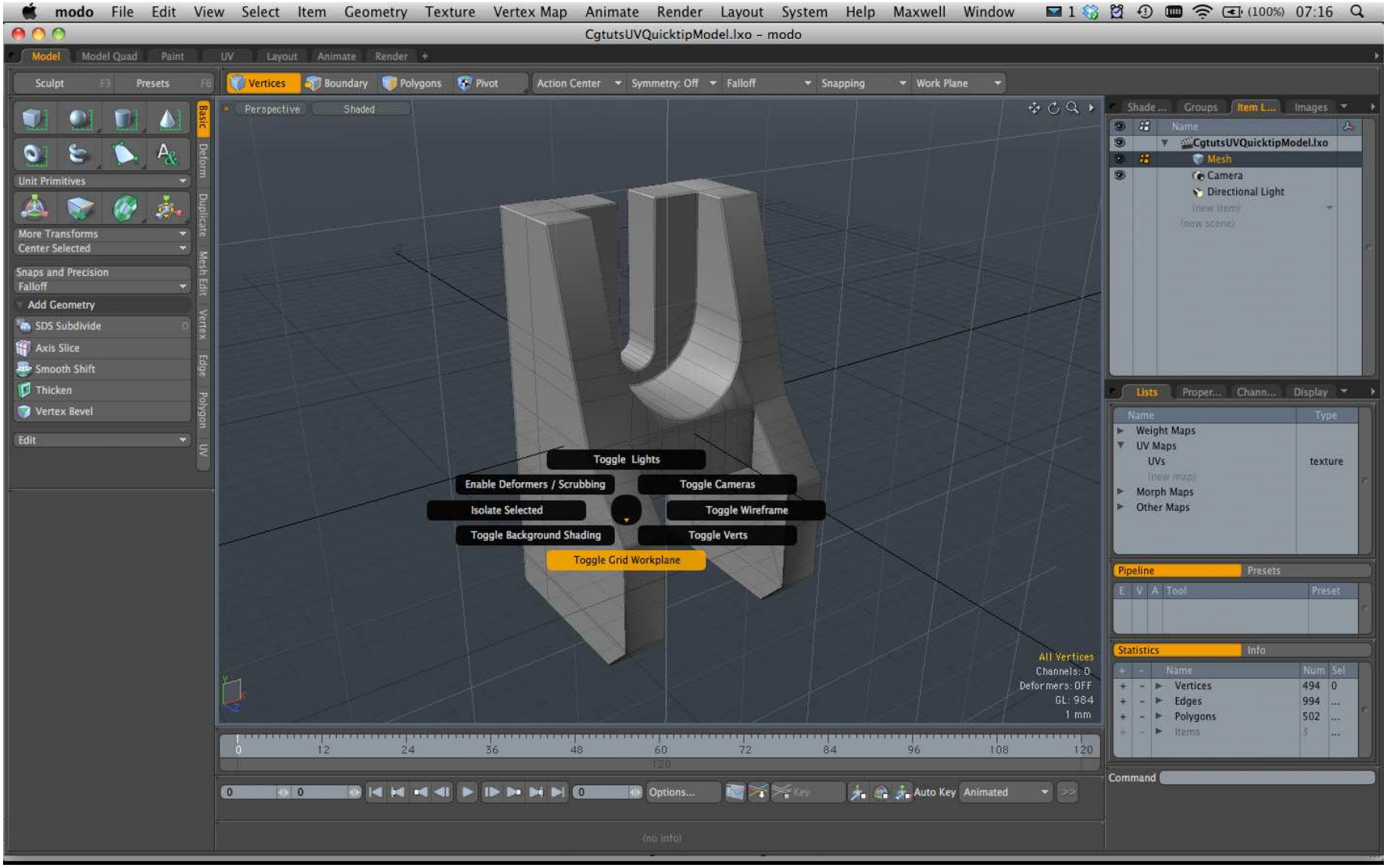
Some Techniques

Marking Menus



by Kurtenbach (1991)

Marking Menus



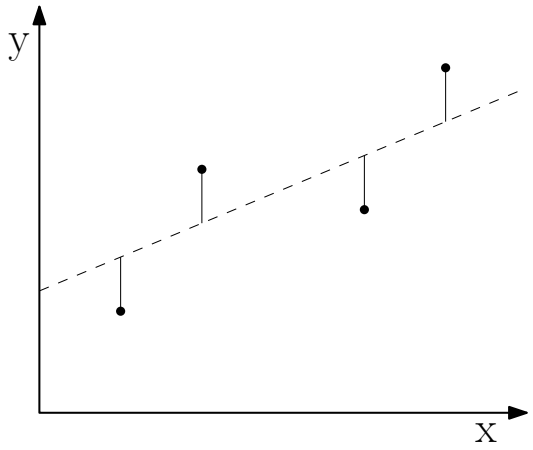
Blender. Image by Blender Foundation

Linear Fitting

$$\hat{\boldsymbol{y}} = a + b\boldsymbol{x}$$

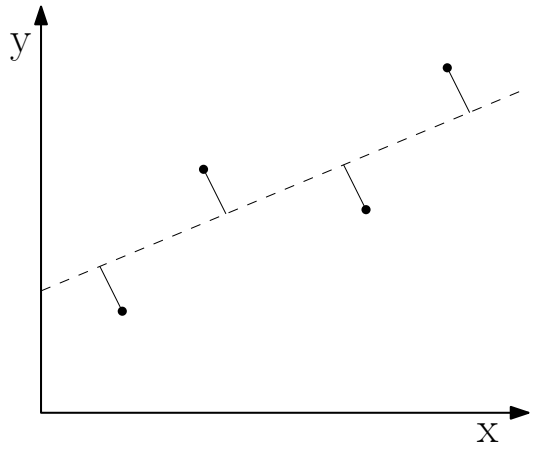
minimize $R^2 = \sum_{i=1}^{N} {r_i}^2$

vertical offsets



$$r_i = y_i - (a + bx_i)$$

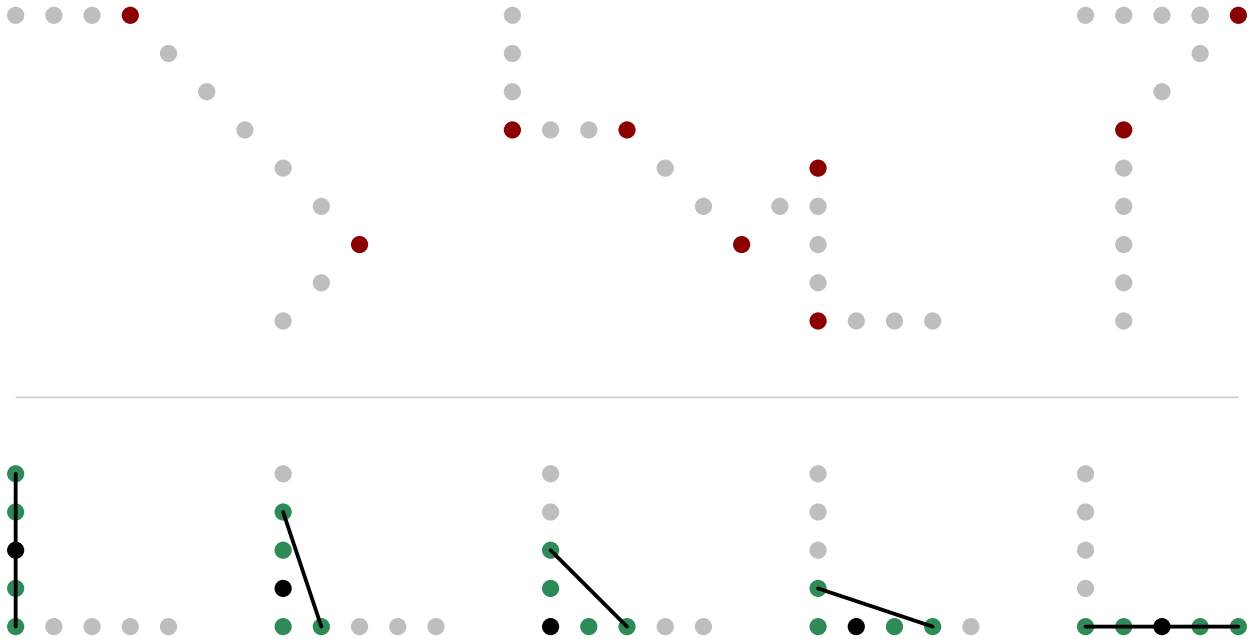
perpendicular offsets



$$r_i = \frac{|y_i - (a + bx_i)|}{\sqrt{1 + b^2}}$$

Corner Detection

PDL, ShortStraw, Firefox's QuickGestures, etc.

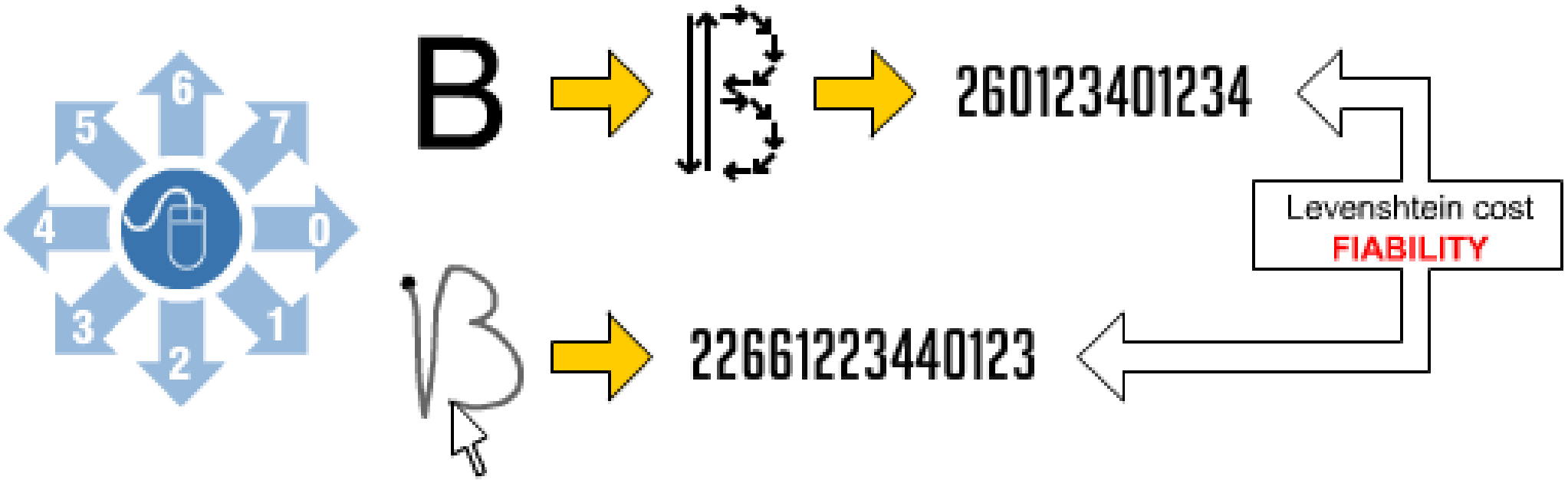


$$G = s_1, \ldots, s_n, \ldots, s_N \mid s_n \in \{L, R, U, D\}$$

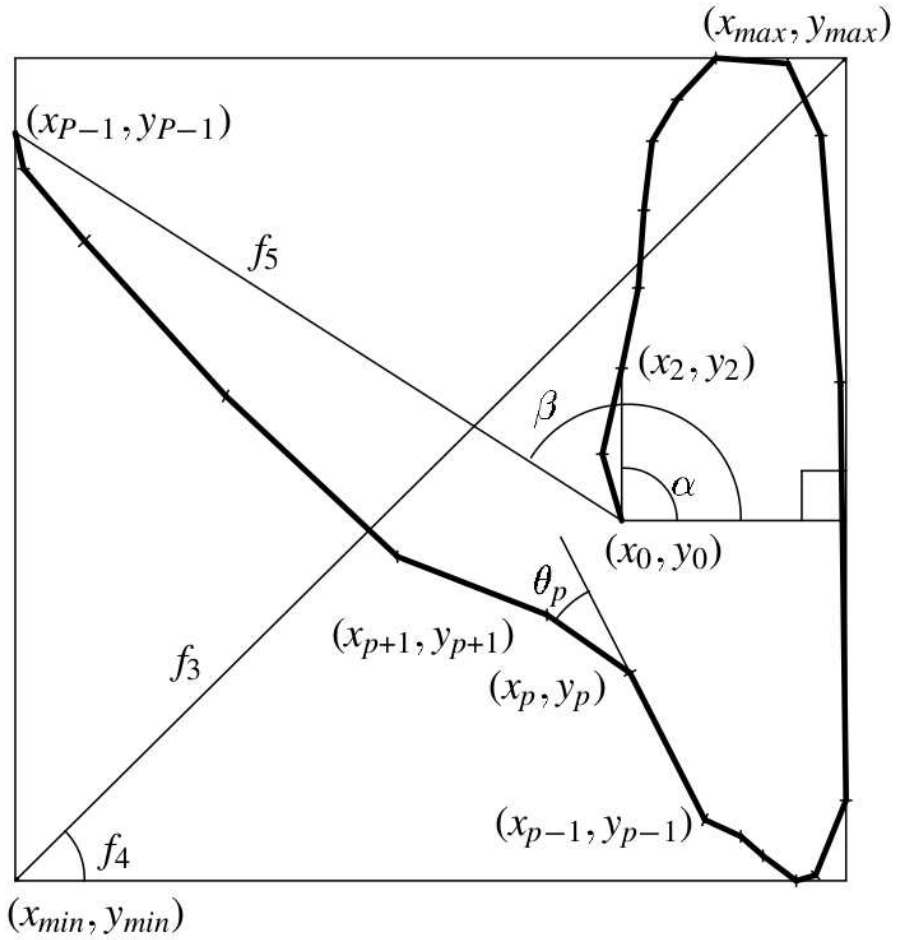
Graffiti & Unistrokes

Comparison by Castellucci and MacKenzie (2008)

Graffiti & Unistrokes



Rubine recognizer



by Rubine (1991)

Rubine recognizer

Linear classifier using $N = 13$ stroke features

$$f\left(\boldsymbol{w}^T \boldsymbol{g}\right) = w_o + \sum_{i=1}^{N} \Sigma^{-1} \mu_i$$

$$w_0 = -\frac{1}{2} \sum_{i=1}^{N} w_i \mu_i$$

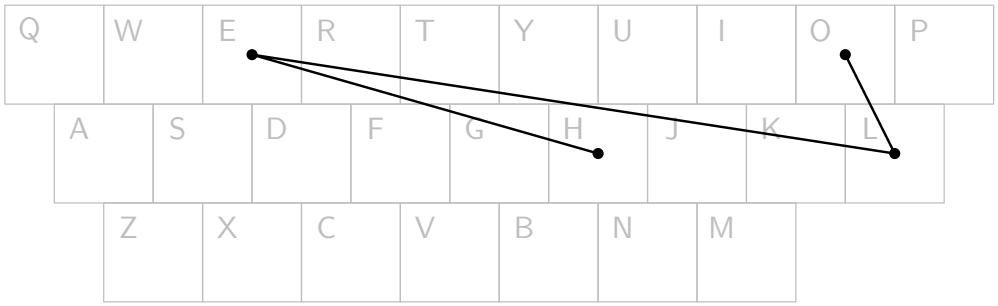
Weight estimation:
perceptron, LSBF, LDA, SVM, logistic regression, etc.

Shapewriting

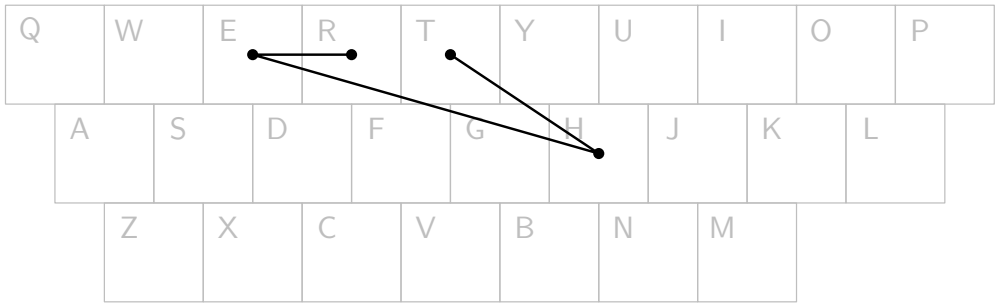
SHARK^2, by Kristensson and Zhai (2004)

Shapewriting

sokgraph of “hello”



sokgraph of “there”

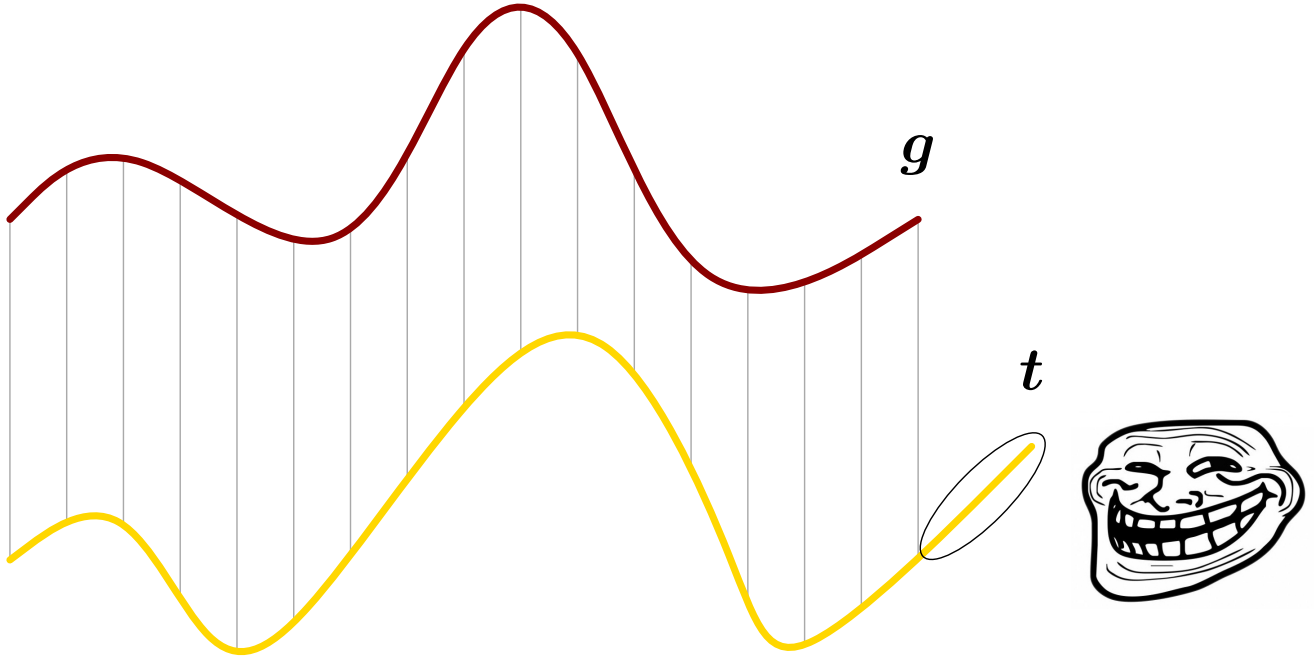


$$\hat{W} = \arg\max_{W} P(W|\boldsymbol{g})$$

$$\hat{W} = \arg\max_{W} \frac{P(\boldsymbol{g}|W)P(W)}{P(\boldsymbol{g})} = \arg\max_{W} P(\boldsymbol{g}|W)P(W)$$

Euclidean Matching

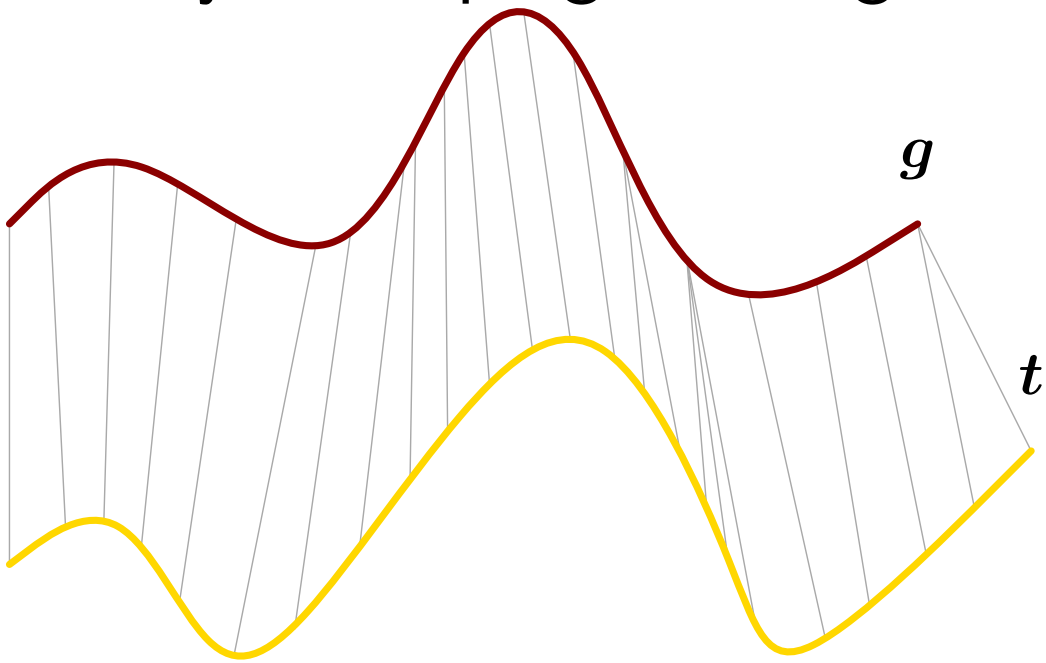
point-wise distances



$$D(\boldsymbol{g}, \boldsymbol{t}) = \frac{1}{|\boldsymbol{g}|} \sum_{i=1}^{|\boldsymbol{g}|} \|g_i - t_i\|$$

Elastic Matching

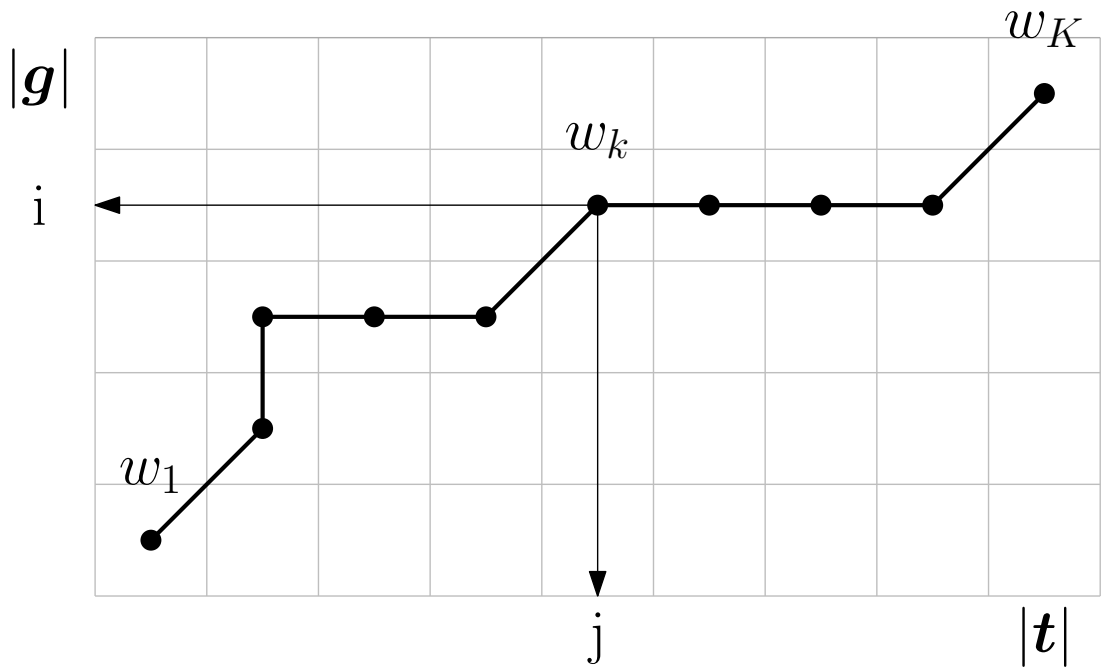
dynamic programming



$$D(\boldsymbol{g}, \boldsymbol{t}) = \min_{W \in \mathcal{W}} \frac{1}{|W|} \sum_{k=1}^{|W|} w_k$$

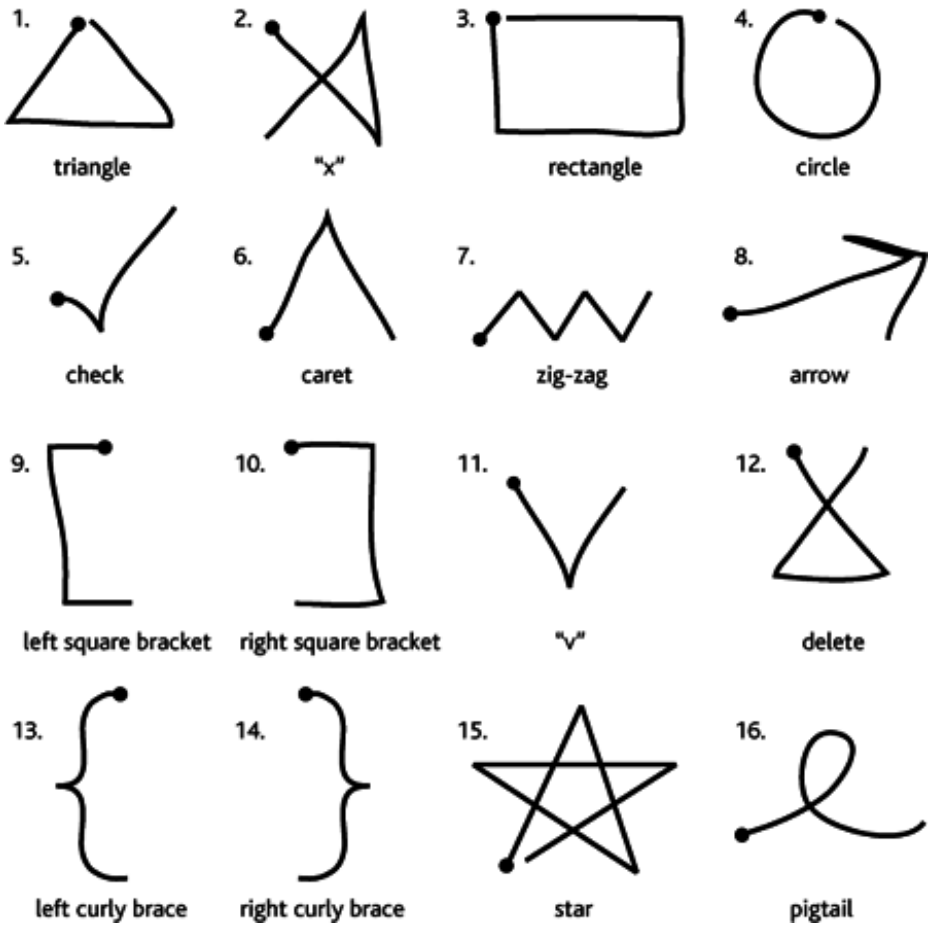
Elastic Matching

$$W = w_1, \ldots, w_k, \ldots, w_K$$



$$\gamma(i,j) = d(i,j) + \min\{\gamma(i-1,j), \gamma(i,j-1), \gamma(i-1,j-1)\}$$

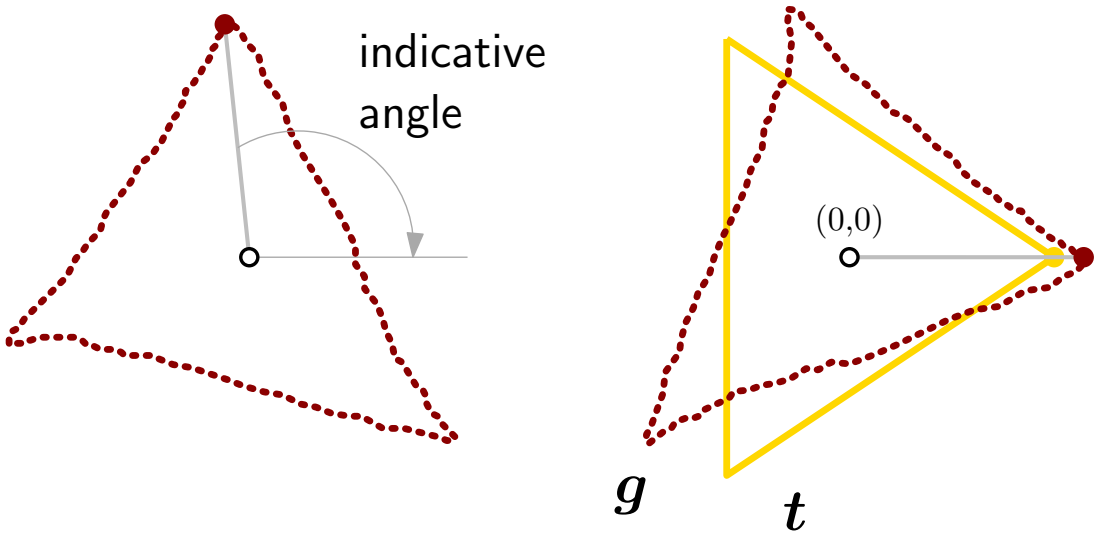
The Dollar Family: $1 recognizer



by Wobbrock et al. (2007)

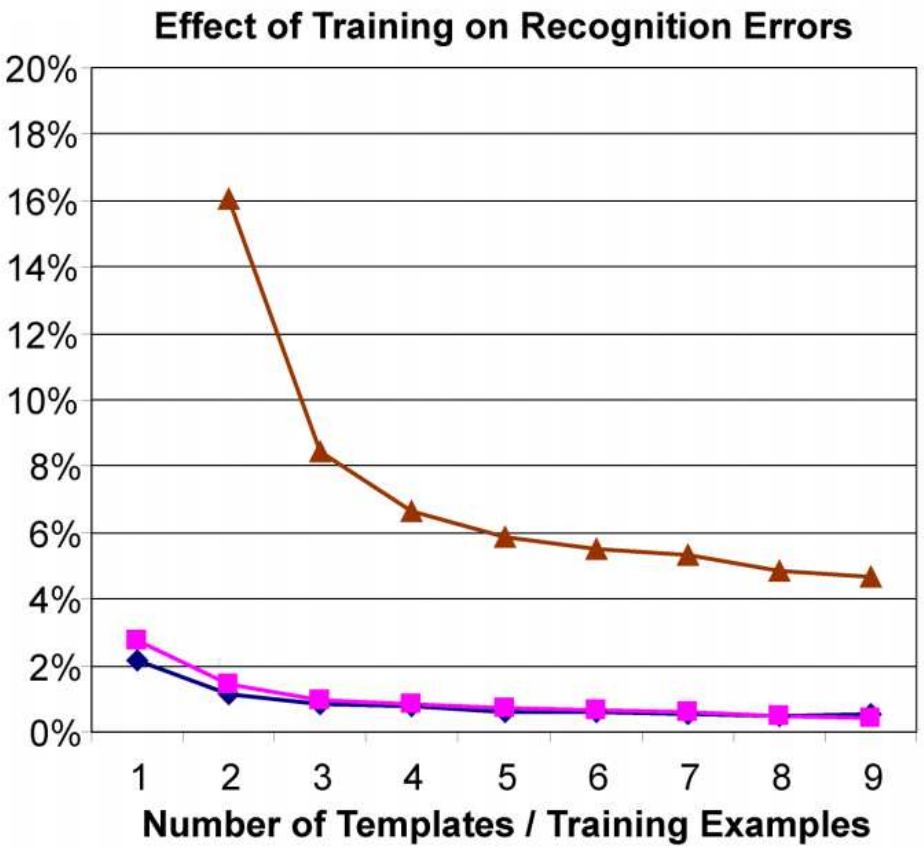
The Dollar Family: $1 recognizer

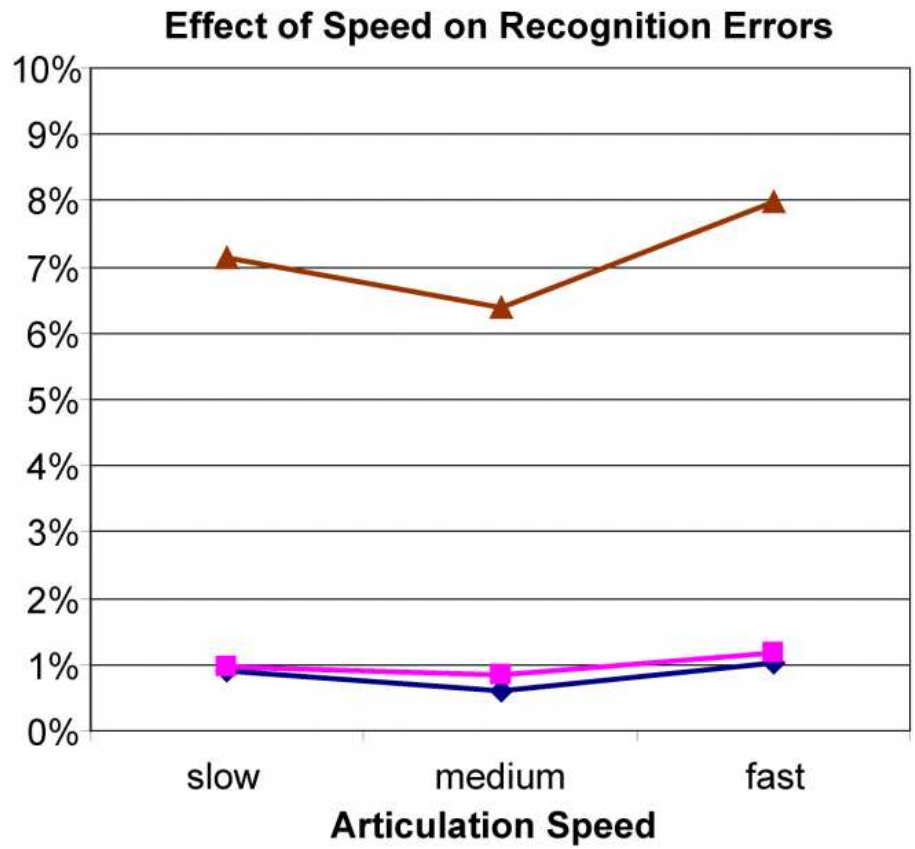
Preprocessing: resampling, rotation, scaling, translation

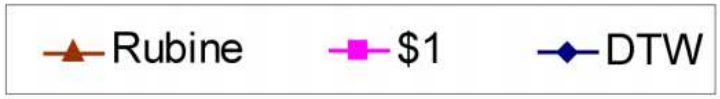


$$D(\boldsymbol{g}, \boldsymbol{t}) = \arg\min_{-\frac{\pi}{4} \leq \theta \leq \frac{\pi}{4}} \frac{1}{N} \sum_{i=1}^{N} \|g_i - t_i(\theta)\|$$

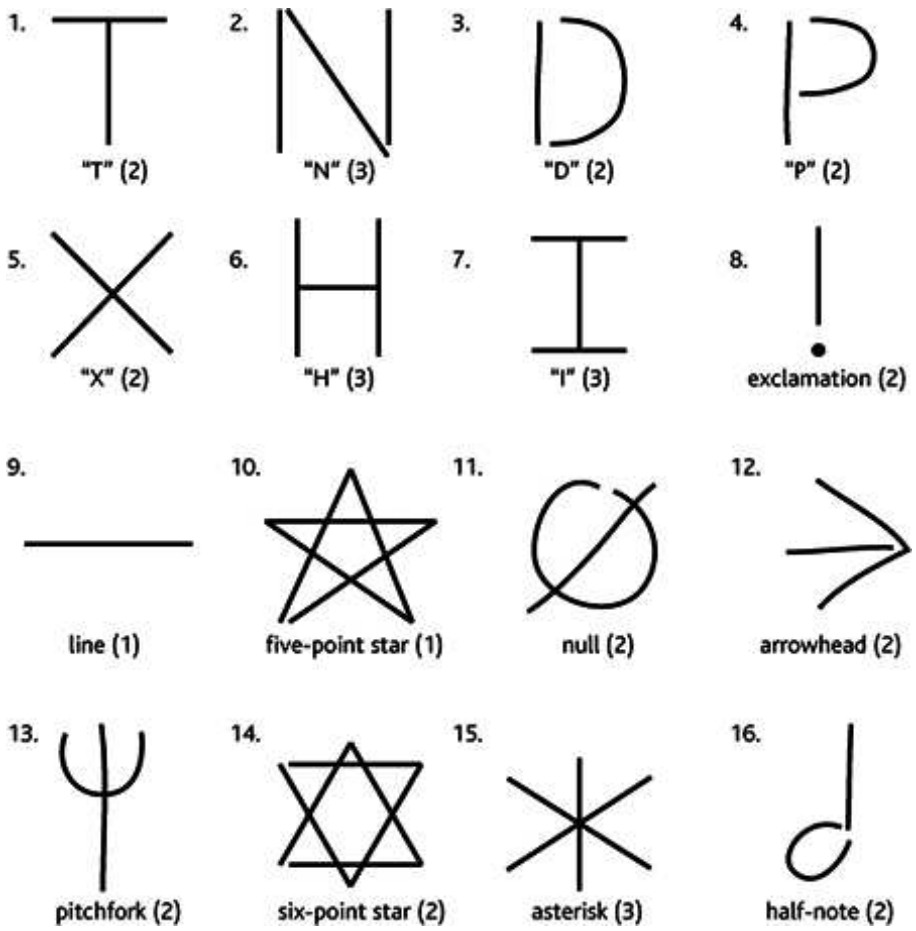
The Dollar Family: $1 recognizer







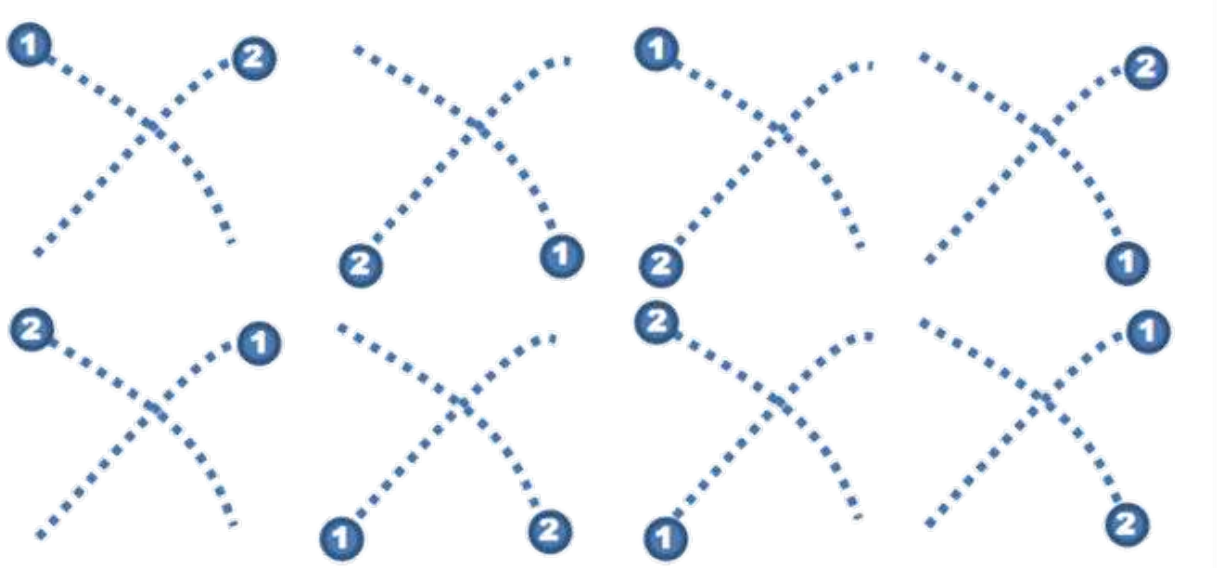
The Dollar Family: $N recognizer

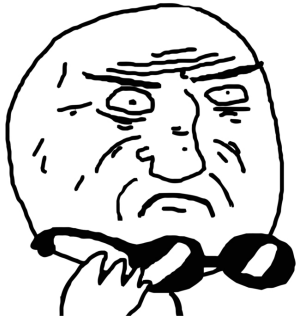


by Anthony and Wobbrock (2010)

The Dollar Family: $N recognizer

$N is $1 with combinatory overhead: $\mathcal{O}(n\ s\ 2^s)$ per template

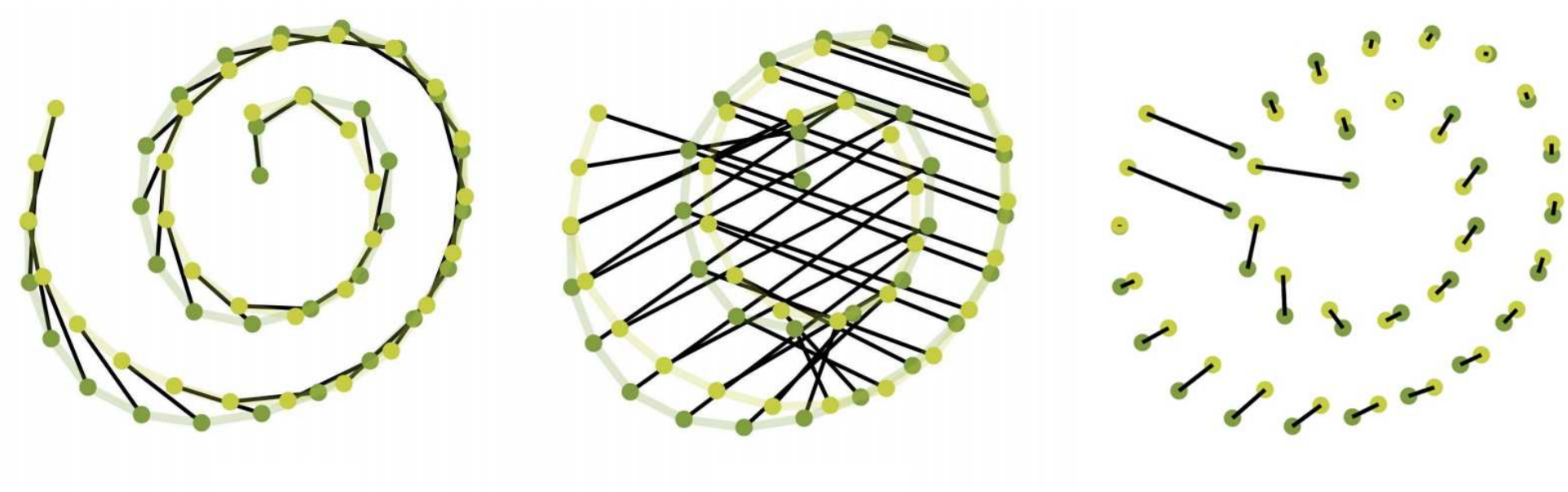




Memory drained out with 20 templates ($n = 32$ pts)

in a quad-core computer with 4 GB RAM.

The Dollar Family: $P recognizer



by Vatavu et al. (2012)

The Dollar Family: $P recognizer

Variation of the Hungarian algorithm:

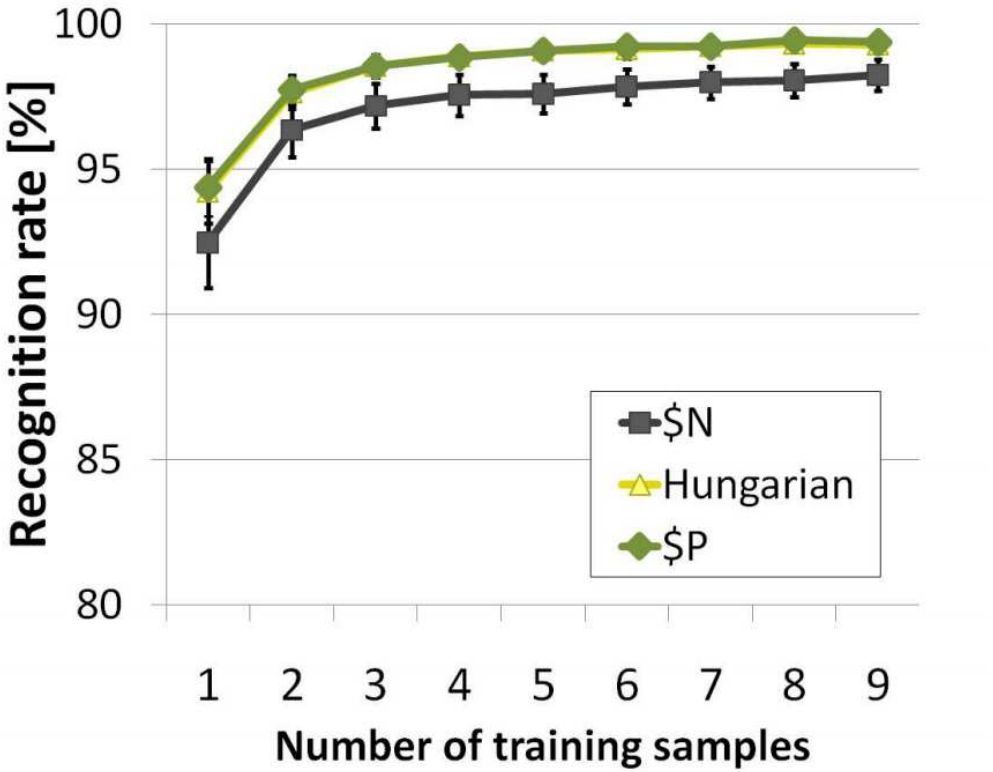
$$D(\boldsymbol{g}, \boldsymbol{t}) = \min \left\{ \|\boldsymbol{g} - \boldsymbol{t}\|, \|\boldsymbol{t} - \boldsymbol{g}\| \right\}$$

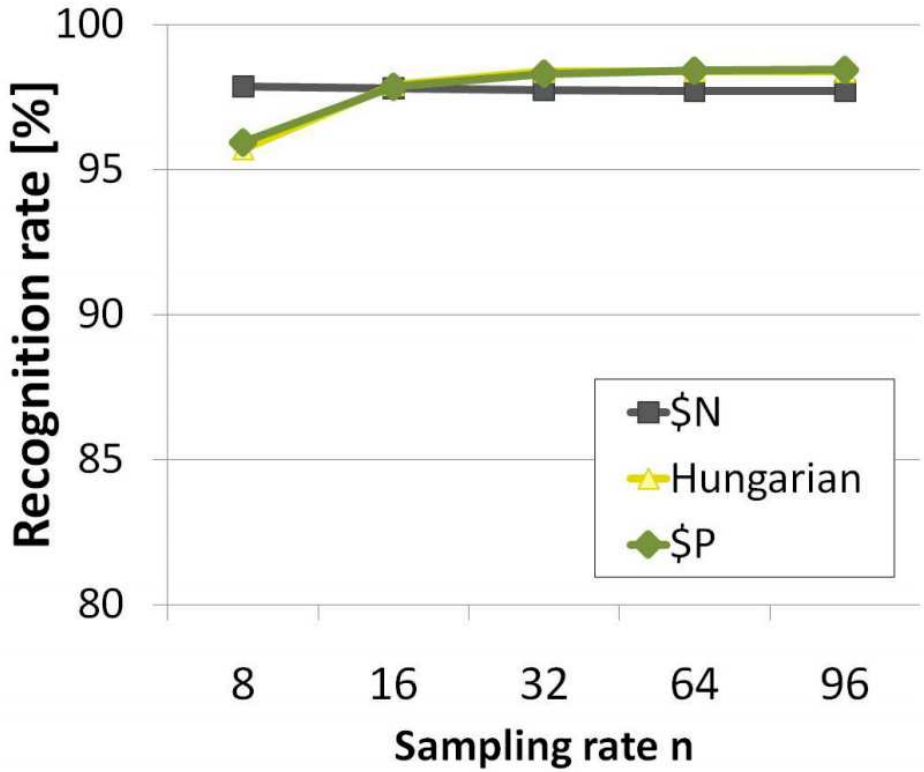
Haussdorf's alternatives:

$$D(\boldsymbol{g}, \boldsymbol{t}) = \max_i \min_j \|g_i - t_j\|$$

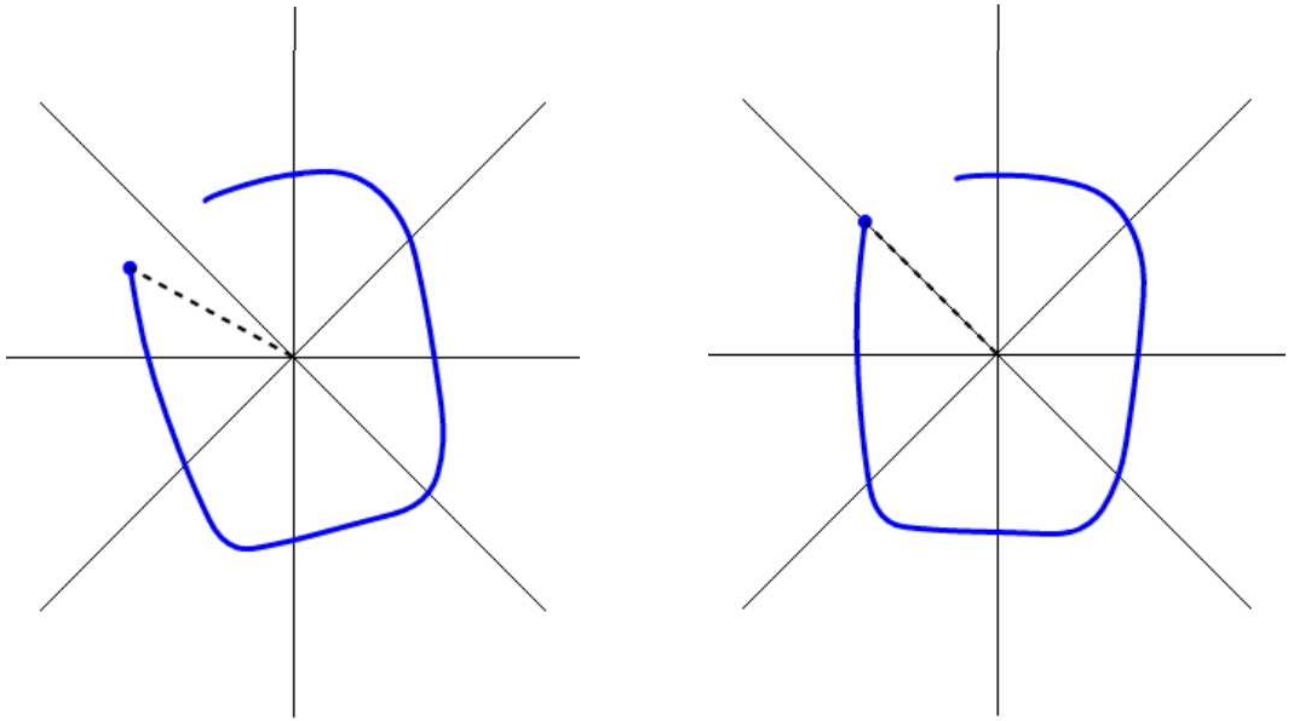
$$D(\boldsymbol{g}, \boldsymbol{t}) = \frac{1}{N} \sum_{i=1}^{N} \min_j \|g_i - t_j\|$$

The Dollar Family: $P recognizer





The Dollar Family: Protractor



by Li (2010)

The Dollar Family: Protractor

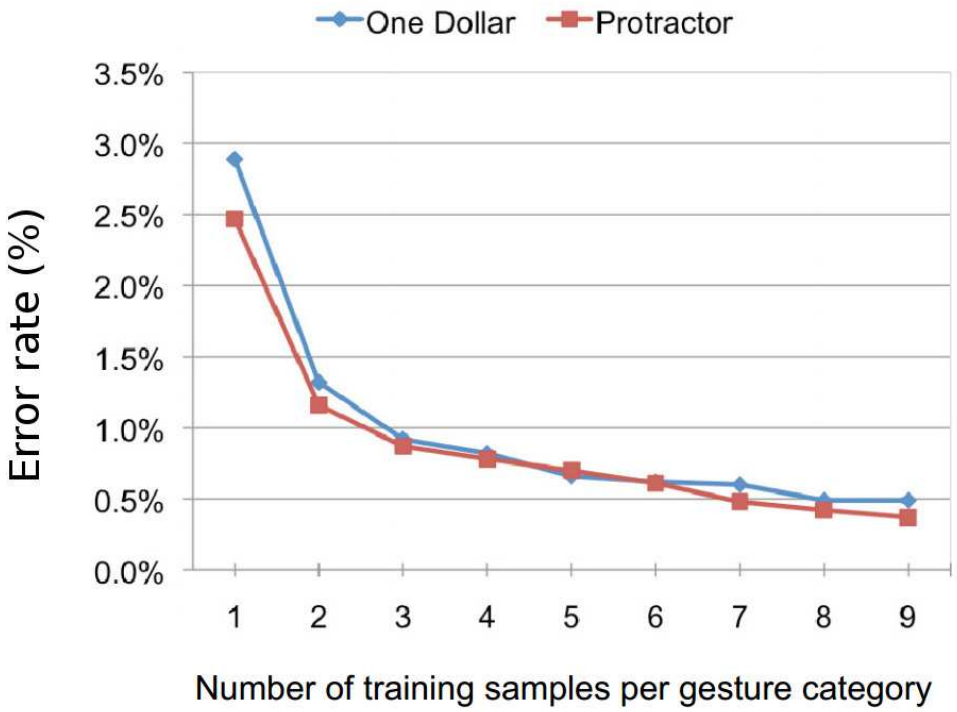
Closed-form solution, minimum angular distance

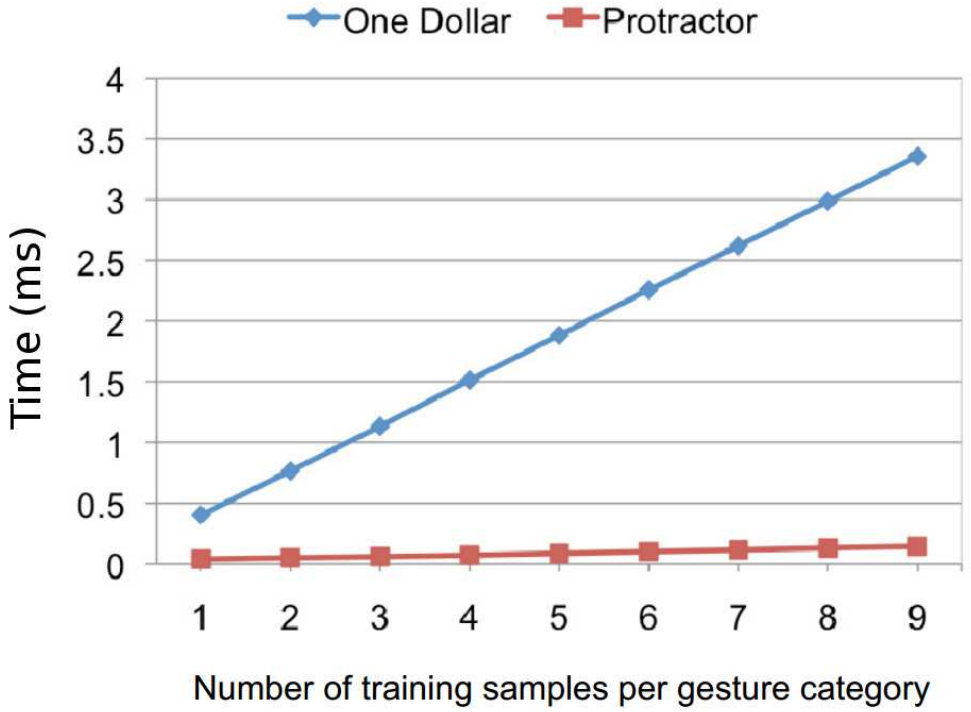
$$D(\boldsymbol{g}, \boldsymbol{t}) = \frac{1}{\arccos(a \cos\hat{\theta} + b \sin\hat{\theta})}$$

$$\hat{\theta} = \arctan\frac{b}{a}$$

$$a = \sum_{i=1}^{N} (x_{gi}\, x_{ti} + y_{gi}\, y_{ti}) \qquad b = \sum_{i=1}^{N} (x_{gi}\, y_{ti} - y_{gi}\, x_{ti})$$

The Dollar Family: Protractor





MinGestures for MIUIs

Disambiguate gestures from handwritten text

LABEL	ACTION	RESULT
Substitute	Lorem Ipsum 	Lorem Ipsan
Reject	Lorem Ipsum	Lorem ...
Merge	Lorem Ipsum 	LoremIpsum
Delete	Lorem Ipsum 	Lorem
Insert	Lorem Ipsum 	Lorem et Ipsum

LABEL	ACTION	RESULT
Split	Lorem 	Lor em
Validate	Lorem Ipsum 	Lorem Ipsum
Undo	Lorem 	Lorem Ipsum
Redo	Lorem Ipsum 	Lorem
Help	Lorem Ipsum 	<help event>

by Leiva et al. (2014)

MinGestures for MIUIs

Three disambiguating features:

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} x_i y_i}$$

$$\Delta_x^- = \sum_{i=2}^{N} \max(x_{i-1} - x_i, 0) \qquad \varphi = \frac{\max(\boldsymbol{x}) - \min(\boldsymbol{x})}{\max(\boldsymbol{y}) - \min(\boldsymbol{y})}$$

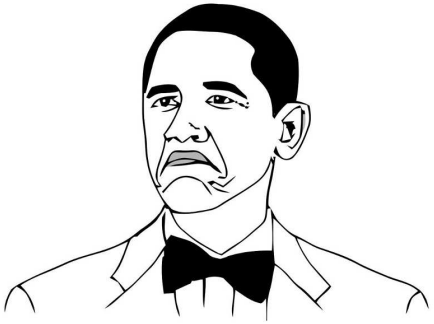
Classification rule:

$$\theta = \frac{\hat{\boldsymbol{y}} - b}{\boldsymbol{x}} \pm \epsilon$$

MinGestures for MIUIs

	E-pen		Mouse	
System	**training**	**test**	**training**	**test**
$1 recognizer	41.2	40.5	45.0	46.6
Marking Menus	18.5	19.5	12.8	13.1
Modified $1	16.0	15.6	7.48	7.56
Rubine	14.6	14.1	15.4	15.7
MinGestures	**0.77**	**1.32**	**0.26**	**0.43**

Error rates in %



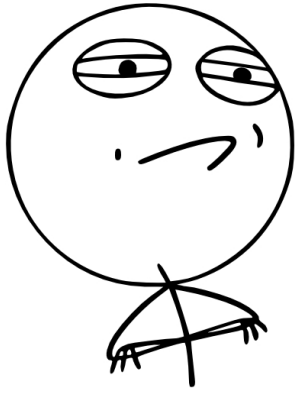
Recap

Takeaways

- Gestures shortcut tedious commands
- Many recognition techniques for many input devices
- Trade-offs: design, setup, performance, accuracy
- Gestures should be **fast** and **simple**:
  1. For humans to *perform* and *recall*
  2. For computers to *recognize*

Open Problems

1. Integration: free-form gestures in context
2. Error analysis and recovery:
   - When should the recognizer ask the user? How much to ask?
3. Segmentation: automatic gesture parts identification
4. Generation: grammar-based, kinematic theory, VAEs, etc.



Bibliography

I. E. Sutherland. Sketchpad: A man-machine graphical communication system. Tech. Report 296, Lincoln Laboratory, MIT, 1963.

M. Davis and T. Ellis. The RAND tablet: A man-machine graphical communication device. In *Proc. AFIPS*, 1964.

D. H. Rubine. *The Automatic Recognition of Gestures*. PhD thesis, Carnegie Mellon University, 1991.

S. Zhai, P. O. Kristensson, C. Appert, T. H. Andersen, and X. Cao. Foundations and trends in Human–Computer Interaction. *Foundational Issues in Touch-Surface Stroke Gesture Design – An Integrative Review*, 5(2), 2012.

C. C. Tappert, C. Y. Suen, and T. Wakahara. The state of the art in on-line handwriting recognition. *IEEE T. PAMI*, 12(8), 1990.

R. Plamondon and S. N. Srihari. On-line and off-line handwriting recognition: a comprehensive survey. *IEEE T. PAMI*, 22(1), 2000.

J. S. Lipscomb. A trainable gesture recognizer. *Patt. Recogn.*, 24(9), 1991.

G. P. Kurtenbach. *The design and evaluation of marking menus*. PhD thesis, University of Toronto, 1991.

O. Bau and W. E. Mackay. OctoPocus: A dynamic guide for learning gesture-based command sets. In *Proc. UIST*, 2008.

S. J. Castellucci and I. S. MacKenzie. Graffiti vs. Unistrokes: An empirical comparison. In *Proc. CHI*, 2008.

P. O. Kristensson and S. Zhai. SHARK^2: A large vocabulary shorthand writing system for pen-based computers. In *Proc. UIST*, 2004.

L. A. Leiva, V. Alabau, V. Romero, A. H. Toselli, and E. Vidal. Context-aware gestures for mixed-initiative text editing UIs. *Interact. Comput.*, 27(1), 2014.

J. O. Wobbrock, A. D. Wilson, and Y. Li. Gestures without libraries, toolkits or training: A $1 recognizer for user interface prototypes. In *Proc. UIST*, 2007.

Y. Li. Protractor: a fast and accurate gesture recognizer. In *Proc. CHI*, 2010.

L. Anthony and J. O. Wobbrock. A lightweight multistroke recognizer for user interface prototypes. In *Proc. GI*, 2010.

R.-D. Vatavu, L. Anthony, and J. O. Wobbrock. Gestures as point clouds: a $P recognizer for user interface prototypes. In *Proc. ICMI*, 2012.

F. Tian, F. Lu, Y. Jiang, X. Zhang, X. Cao, G. Dai, and H. Wang. An exploration of pen tail gestures for interactions. *Int. J. Human-Comput. Stud.*, 71, 2012.

Videography

Sketchpad. https://youtube.com/watch?v=YB3saviItTI.

RAND tablet. http://youtube.com/watch?v=LLRy4Ao62ls.

Minority Report. http://youtube.com/watch?v=NIDuYY2twkM.

Nintendo Wii. http://youtube.com/watch?v=DW3XSd93eTM.

T(ether). http://youtube.com/watch?v=P6h1boE-fsI.

MS Kinnect. http://youtube.com/watch?v=frqDo0xsAic.

Humantenna. http://youtube.com/watch?v=7lRnm2oFGdc.

Skinput. http://youtube.com/watch?v=g3XPUdW9Ryg.

Myo. http://youtube.com/watch?v=oWu9TFJjHaM.

Leap Motion. http://youtube.com/watch?v=_d6KuiuteIA.

TapTap. http://youtube.com/watch?v=GVEgaubNgN0.

Intugine. http://youtube.com/watch?v=Q4jpiSWEYPk.

Wacom gestures. http://youtube.com/watch?v=hKQFqEVK8lM.

OctoPocus. http://vimeo.com/2116172.

Marking Menus. http://youtube.com/watch?v=dtH9GdFSQaw.

Shapewriting. http://youtube.com/watch?v=WtlyuuYmFN0.

$1 recognizer. http://youtube.com/watch?v=ept2Z1UVxfw.

$N recognizer. http://youtube.com/watch?v=74n4JwPqlNA.

$P recognizer. http://youtube.com/watch?v=VI7J4CftGNg.