VRIJE
UNIVERSITEIT
BRUSSEL

Next Generation User Interfaces

Gesture-based Interaction

Prof. Beat Signer

Department of Computer Science
Vrije Universiteit Brussel

http://www.beatsigner.com

Gesture-based Interaction

Microsoft Kinect, skeleton tracking

Minority Report, glove-based tracking

Ninteno Wii, accelerator-based tracking

American sign language

What is a Gesture?

- A *motion of the limbs or body* to *express* or help to express thought or to emphasise speech
- The act of moving the limbs or body as an *expression of thought or emphasis*
- A *succession of postures*

Formal Gesture Definition

A gesture is a form of non-verbal communication or non-vocal communication in which visible bodily actions communicate particular messages, either in place of, or in conjunction with, speech. Gestures include movement of the hands, face, or other parts of the body. Gestures differ from physical non-verbal communication that does not communicate specific messages, such as purely expressive displays, proxemics, or displays of joint attention.

A. Kendon, *Gesture: Visible Action as Utterance*, Cambridge University Press, 2004

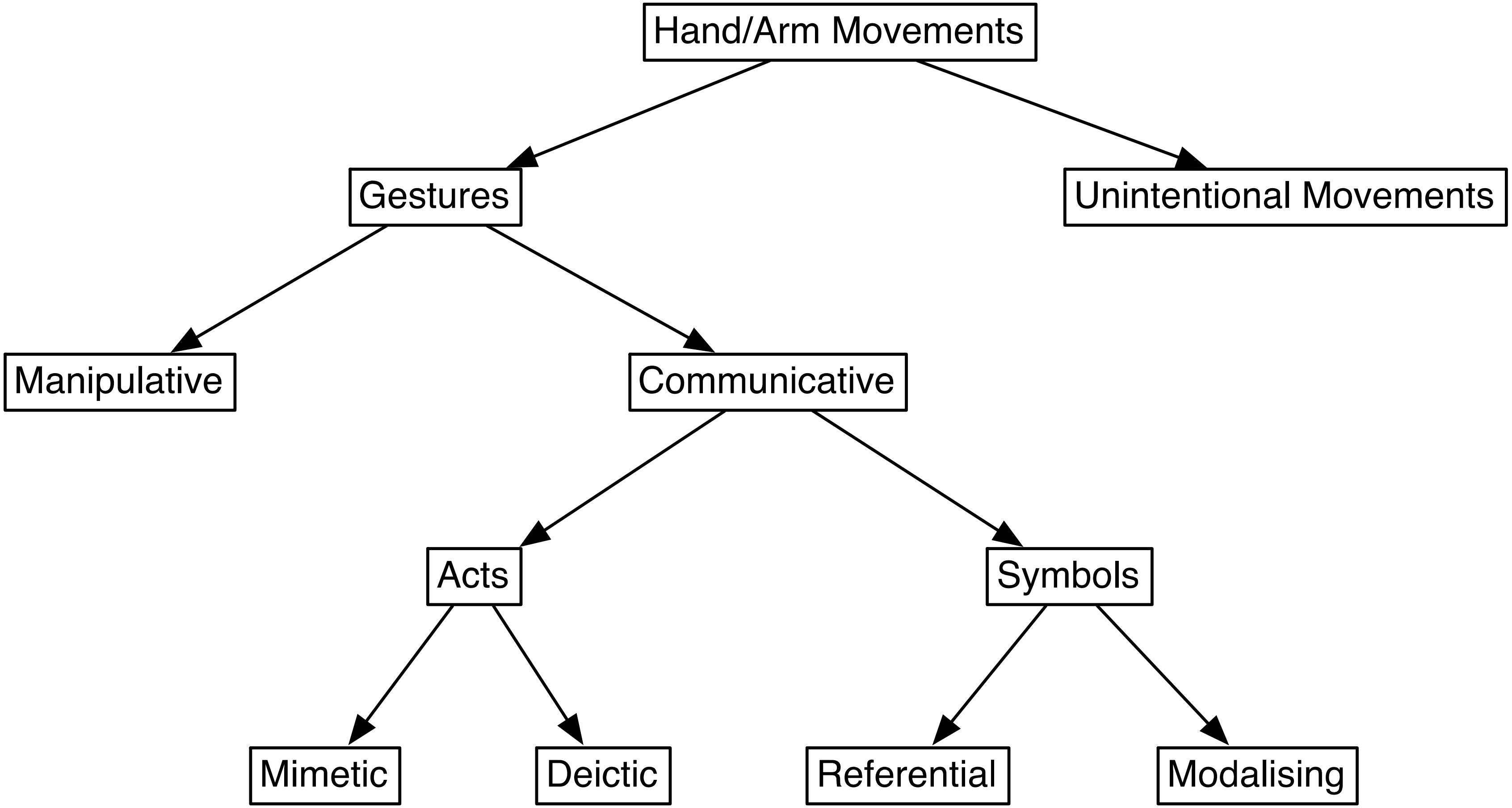
Gesture Types

- Gestures can be classified into three types of gestures according to their function (Buxton, 2011)
  - *semiotic gestures*
    - used to communicate meaningful information (e.g. thumbs up)
  - *ergotic gestures*
    - used to manipulate the physical world and create artefacts
  - *epistemic gestures*
    - used to learn from the environment through tactile or haptic exploration
- Since we are interested in human-computer interaction, we will focus on semiotic gestures

Semiotic Gestures

- Semiotic gestures can be further classified into
  - *symbolic gestures (emblems)*
    - culture-specific gestures with single meaning (e.g. *"OK"* gesture)
    - only symbolic gestures can be interpreted without contextual information
  - *deictic gestures*
    - pointing gesture (e.g. Bolt's *"put-that-there"*)
  - *iconic gestures*
    - used to convey information about the size, shape or orientation of the object of discourse (e.g. *"the plane flew like this"*)
  - *pantomimic gestures*
    - showing the use of movement of some invisible tool or object in the speaker's hand (e.g. *"I turned the steering wheel hard to the left"*)

Taxonomy of Hand/Arm Gestures



Taxonomy of Hand/Arm Gestures …

- Gestures vs. unintentional movements
  - unintentional movement does not convey meaningful information
- Communicative vs. manipulative gestures
  - manipulative (ergotic) gestures are *used to act on objects* in an environment (e.g. move or rotate an object)
  - communicative (semiotic) gestures have an *inherent communicational purpose*
- Acts are gestures that are directly related to the *interpretation of the movement itself*
  - *imitating* some actions
    - mimetic (pantomimic) gestures
  - *pointing acts* (*"put-that-there"*)
    - deictic gestures

Taxonomy of Hand/Arm Gestures …

- Symbols are gestures that have a *linguistic role*
  - symbolise *some referential action* (e.g. circular motion of the index finger to reference to a wheel)
  - used as *modalisers*, typically linked with speech *("Look at that wing!"* and a modalising gesture showing the wing vibrating)

Gesture Recognition Devices

- Wired gloves
- Accelerometers
- Camcorders and webcams
- Skeleton tracking
- Electromyography (EMG)
- Single and multi-touch surfaces
  - see lecture on Interactive Tabletops and Surfaces
- Digital pens
- …

Wired Gloves

- Wired glove (also data-glove or cyberglove) to retrieve the position of the hand and fingers
  - *magnetic sensors* or *inertial tracking sensors* to capture the movements of the glove

Power Glove for Nintendo, Mattel, 1989

- May provide *haptic feedback* which is useful for virtual reality applications
- In many application domains wired gloves are more and more replaced by camera-based gesture recognition

Accelerometers

- Accelerometers measure the *proper acceleration* of a device *in one direction*
  - use three accelerometers to measure the acceleration in all three dimensions
  - note that the *gravity g* is also measured
- Accelerometers are relatively cheap components which are present in many consumer electronic devices
  - smartphones
    - screen orientation (landscape or portrait)
  - laptops
    - active hard-drive protection in case of drops
  - cameras and camcorders
    - image stabilisation

Accelerometers …

- gaming devices (e.g. Nintendo Wii Remote)
  - note that the pointing with a Wii Remote is not recognised through the accelerometer but via an infrared camera in the head of the Wii Remote
- Accelerometers can be used to recognise *dynamic gestures* but not for the recognition of postures
  - record the 3-dimensional input data, pre-process and vectorise it
  - apply pattern recognition techniques on the vectorised data
- Typical recognition techniques
  - dynamic time warping (DTW)
  - neural networks
  - Hidden Markov Models (HMM)
- All these techniques require some *training data*

Camcorders and Webcams

- Standard camcorders and webcams can be used to record gestures which are then recognised based on *computer vision techniques*

Advantages

- relatively inexpensive hardware
- large range of use cases
  - fingers, hands, body, head
  - single user or multiple users

Disadvantages

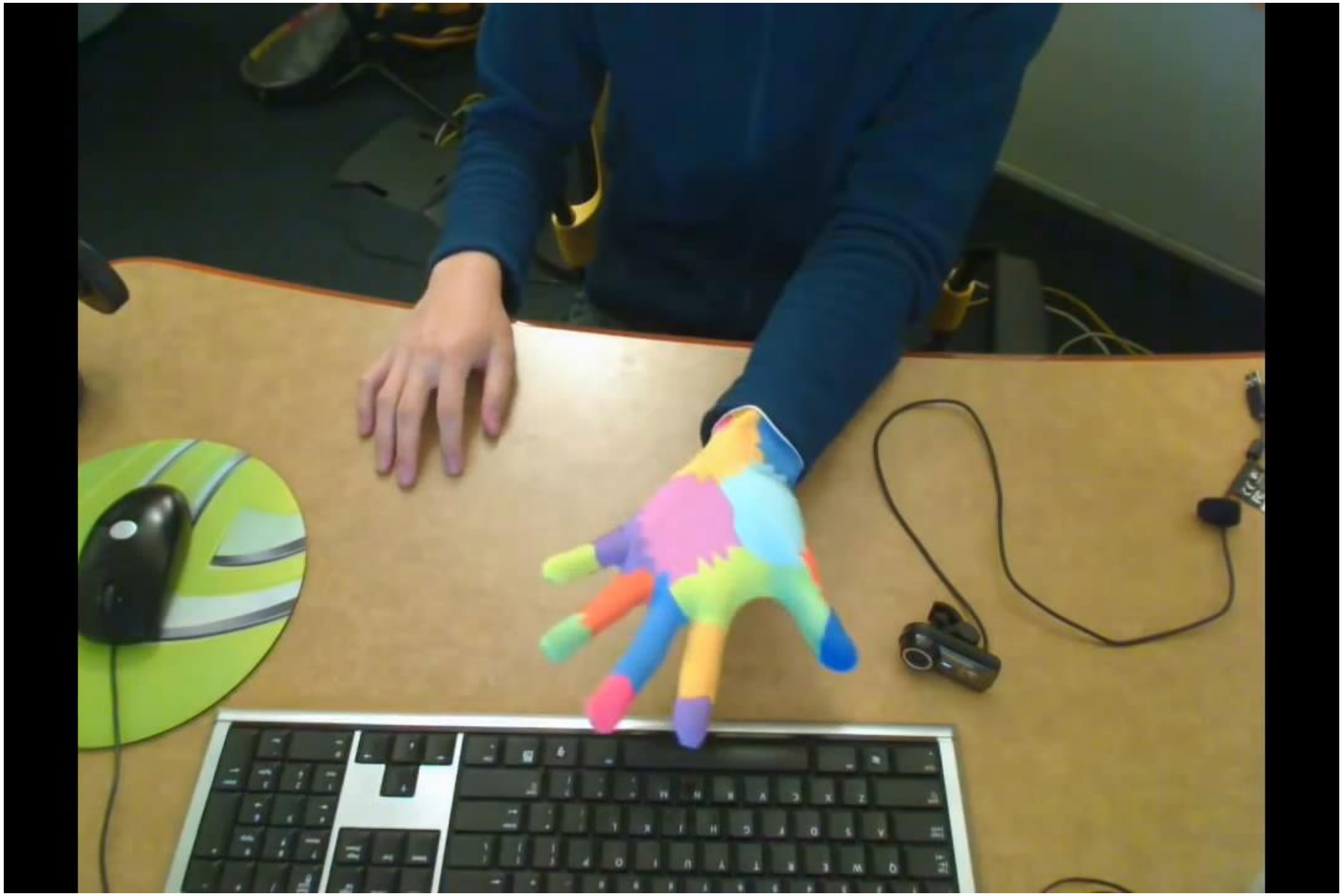
- we first have to detect the body or body part before the recognition process can start
- difficult to retrieve depth (3D) information

Vision-based Hand Gesture Example

- Hand gesture detection based on multicolour gloves
  - developed at MIT
- Colour pattern designed to simplify the pose estimation problem
- Nearest-neighbour approach to recognise the pose
  - database consisting of 100’000 gestures

Wang and Popović, 2009

Video: Colour Glove Hand Tracking

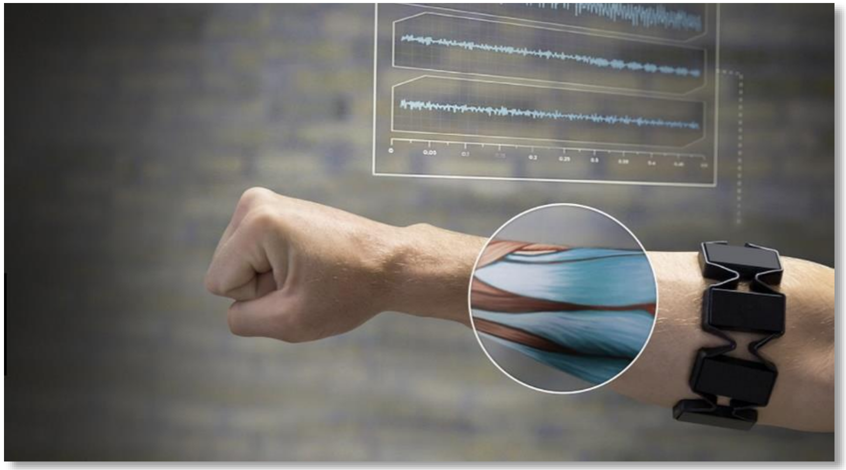


Skeleton Tracking

- So-called range cameras provide a 3D representation of the space in front of them
  - before 2010 these cameras were quite expensive
- Since 2010 the Microsoft Kinect sensor offers full-body gesture recognition for ~150€
  - infrared laser projector coupled with an infrared camera and a "classic" RGB camera
  - multi-array microphone
  - infrared camera captures the depth of the scene
  - skeleton tracking through fusion of depth data and RGB frames
- Two SDKs are available for the Kinect
  - OpenNI and the Microsoft Kinect SDK

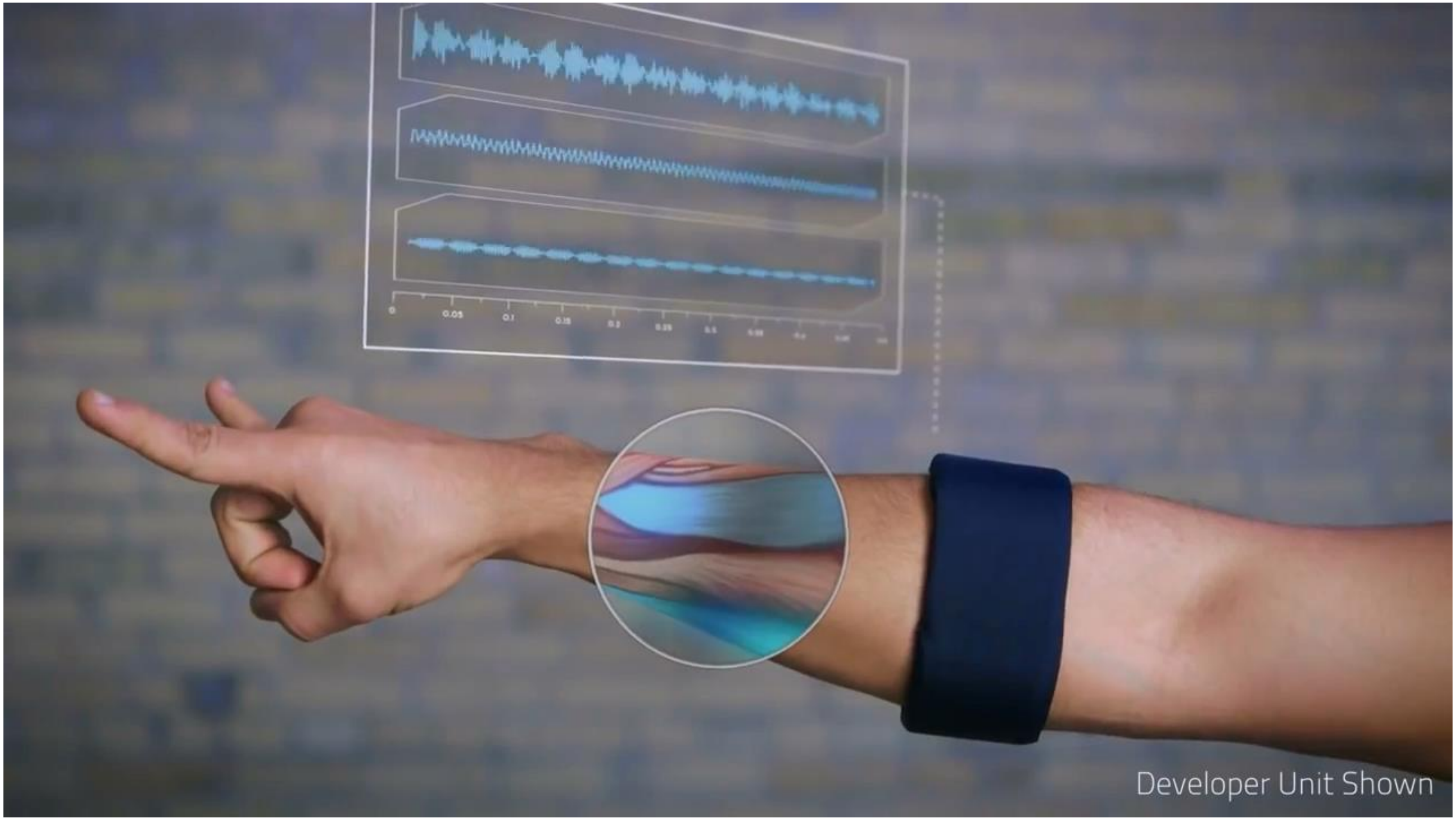
Video: Kinect Depth Sensor

Electromyography (EMG)



- MYO electromyography bracelet
  - 93 grams
  - ARM Cortex M4 processor
  - haptic feedback
  - Bluetooth 4.0
  - three-axis gyroscope
  - three-axis accelerometer
  - three-axis magnetometer
- Potential applications
  - gesture-based remote control
  - handwriting recognition or sign language translation
  - …

Video: Myo Wearable Gesture Control

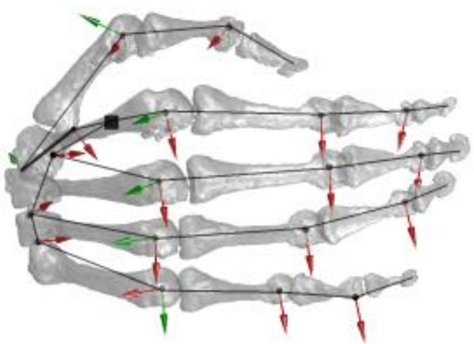


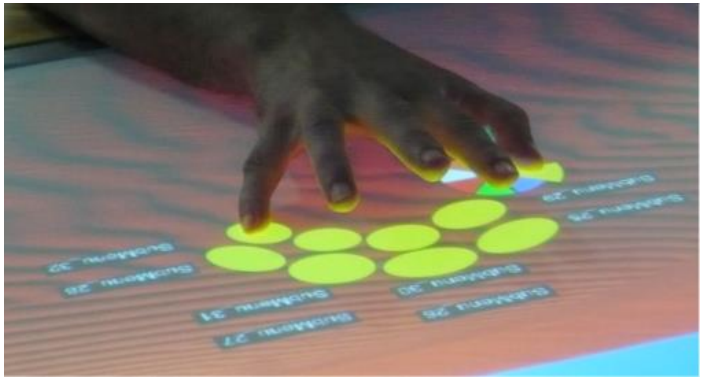
Gesture Recognition Algorithms

- Three broad families of algorithms
  - template-based algorithms
    - Rubine
    - Dynamic Time Warping (DTW)
    - $1 recogniser / $N recogniser
  - machine learning-based algorithms
    - Hidden Markov Models (HMM)
    - neural networks
  - rule-based approaches
    - LADDER
- Some approaches mix these families to keep the strengths of each
  - e.g. Mudra

Gesture Vocabularies

- Choosing a good gesture vocabulary is not an easy task!
- Common pitfalls
  - gestures might be *hard to perform*
  - gestures might be *hard to remember*
  - a user's arm might begin to feel fatigue (*"gorilla arm"*)
- The human body has *degrees of freedom* and *limitations* that have to be taken into account *and can be exploited*





Defining the Right Gesture Vocabulary

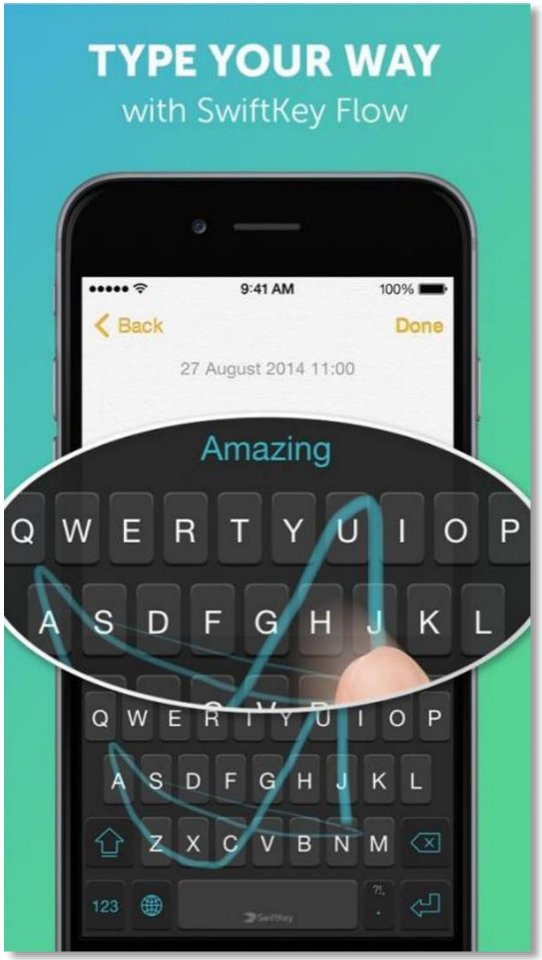
- Use the foundations of interaction design
- *Observe the users* to explore gestures that make sense
- Gestures should be
  - easy to perform and remember
  - intuitive
  - metaphorically and iconically logical towards functionality
  - ergonomic and not physically stressing when used often
- Implemented gestures can be *evaluated* against
  - semantic interpretation
  - intuitivity and usability
  - learning and memory rate
  - stress

Defining the Right Gesture Vocabulary …

- From a technical point the following things might be considered
  - different gestures should not look too similar
    - better recognition results
  - gesture set size
    - a large number of gestures is harder to recognise
- Reuse of gestures
  - same semantics for different applications
  - application-specific gestures

Shape Writing Techniques

- Input technique for virtual keyboards on touchscreens
  - e.g. mobile phones or tablets
- No longer type individual characters but perform a *single-stroke gesture over the characters of a word*
- Gestures are automatically mapped to specific words
  - e.g. SwiftKey uses a neural network which learns and adapts its prediction over time
- Single-handed text input
  - for larger screens the keyboard might float

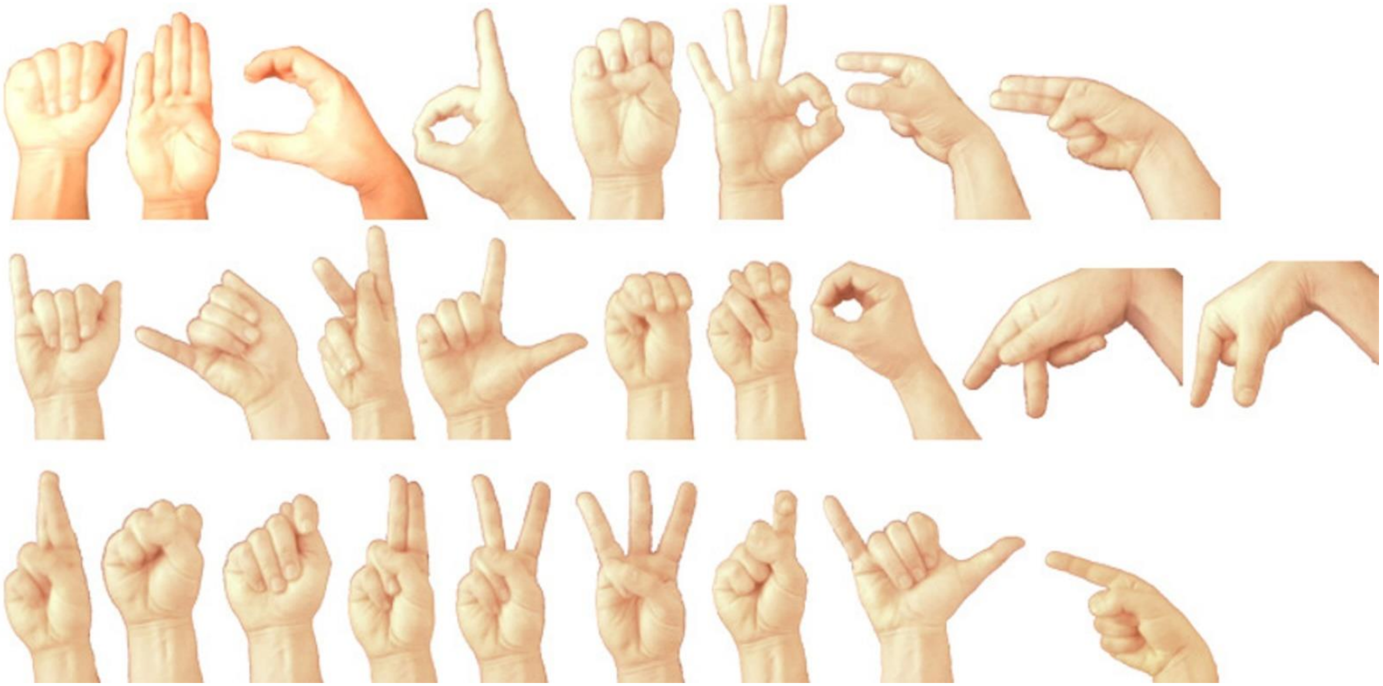


"Fat Finger" Problem

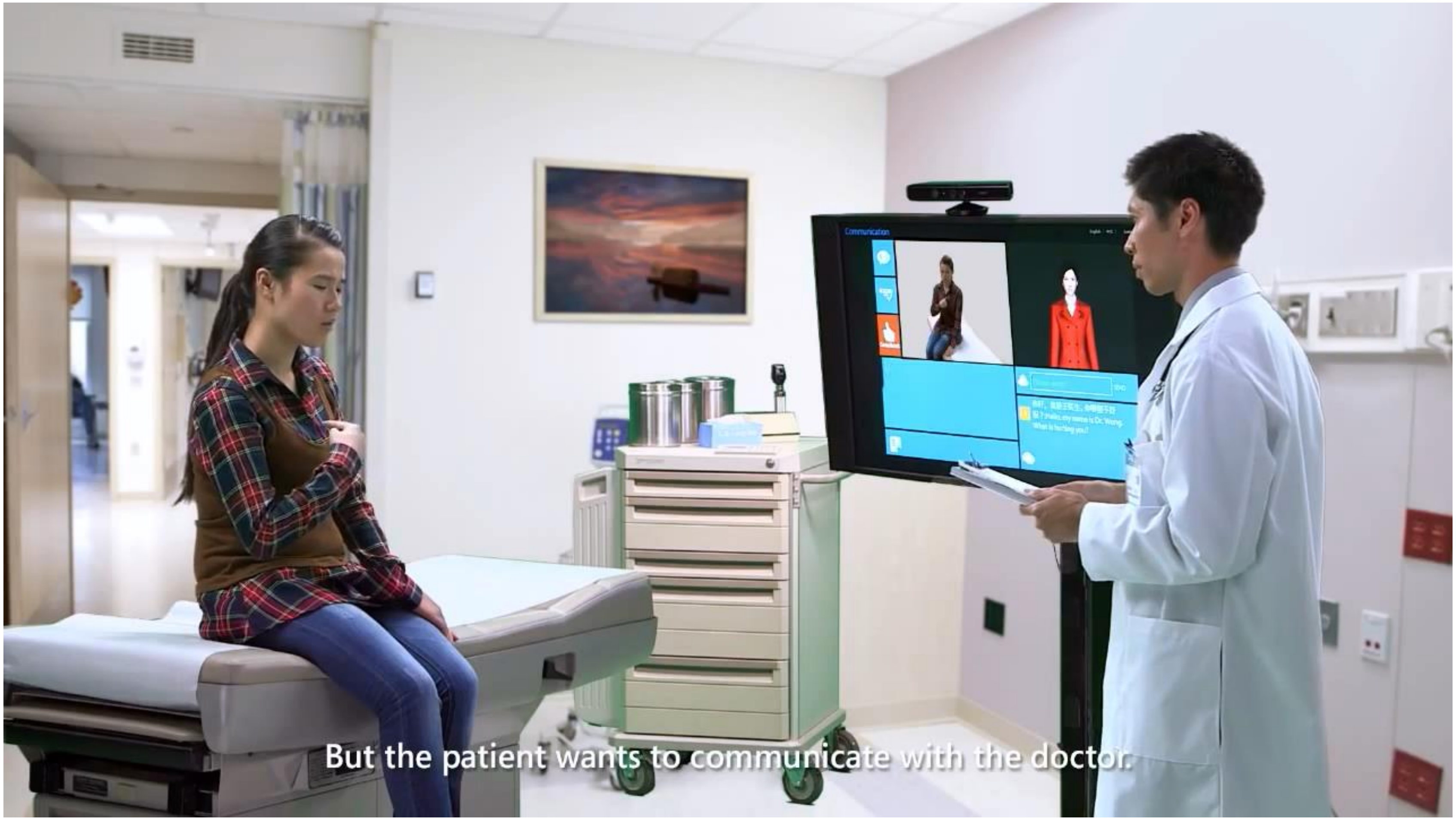
- *"Fat finger" problem* is based on two issues
  - finger makes contact with a relatively large screen area but only *single touch point* is used by the system
    - e.g. centre
  - users *cannot see* the currently computed *touch point* (occluded by finger) and might therefore miss their target
- Solutions
  - make elements larger or provide feedback during interaction
  - adjust the touch point (based on user perception)
  - use *iceberg targets* technique
  - …

Sign Language

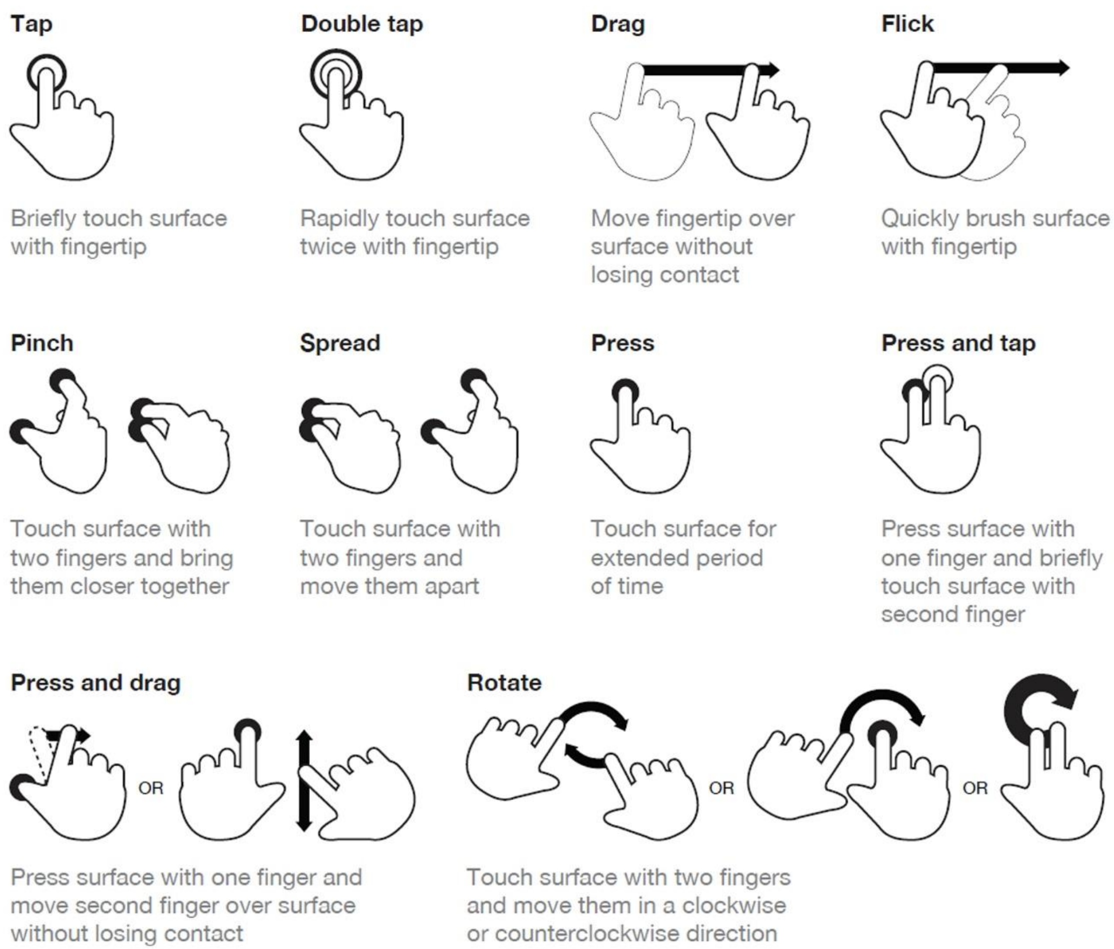
- American Sign Language (ASL) has gestures for the alphabet as well as for the representation of concepts and objects



Video: Kinect Sign Language Translator

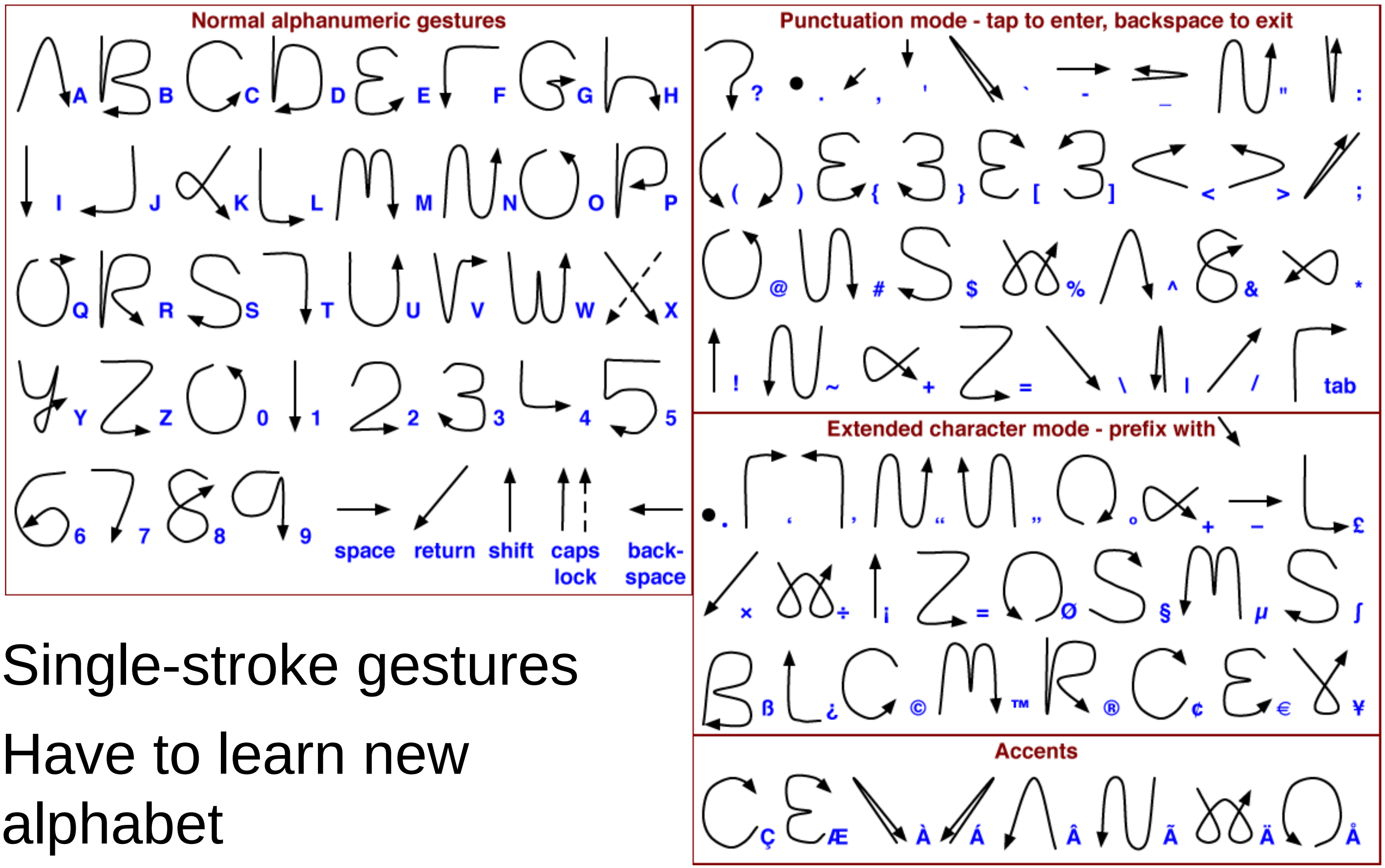


Standard Single and Multi-Touch Gestures



Touch Gesture Reference Guide

Graffiti Gestures (Palm OS)



- Single-stroke gestures
- Have to learn new alphabet

Microsoft Application Gestures

scratch-out	erase content
triangle	insert
square	action item
star	action item
check	check-off
curlicue	cut
double-curlicue	copy
circle	application-specific

double-circle	paste
left-semicircle	undo
right-semicircle	redo
caret	past/insert
inverted-caret	insert
chevron-left	application-specific
chevron-right	application-specific
arrow-up	application-specific

Customised Gestures

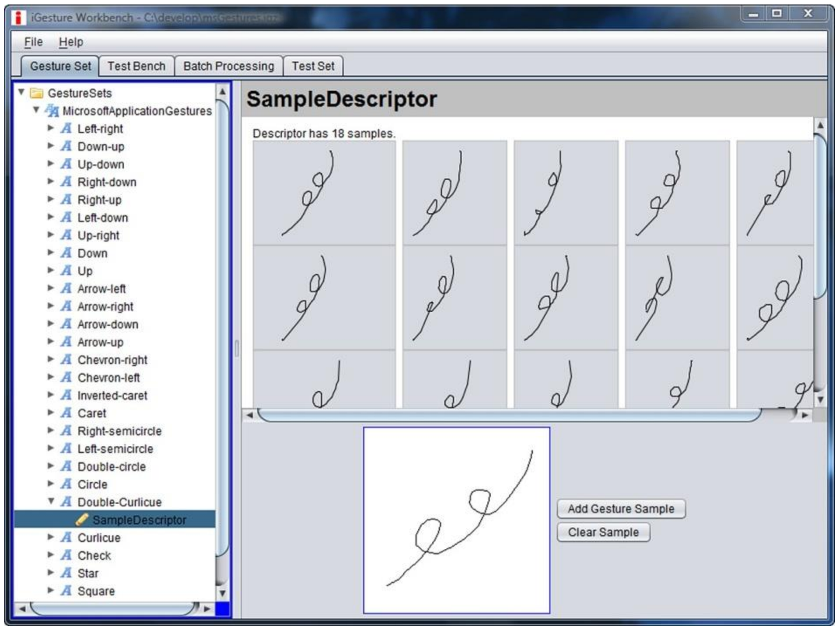
paper have ensured its retention as a key medium for reading and annotating documents. Paper has many advantages over digital media in terms of how people can work with it, both individually and in groups. It is portable, light, cheap, flexible and robust. Furthermore, various forms of paper-based collaboration and interaction are nearly impossible to support in digital environments [12].

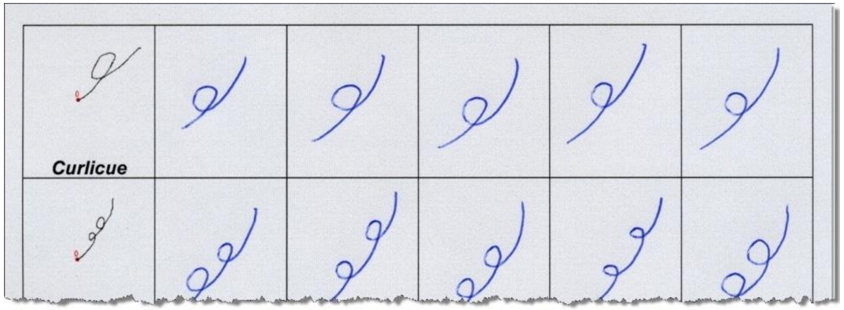
A set of reading-related affordances of paper documents are pointed out by Sellen and Harper in their book *The Myth of the Paperless Office* [18]. First, paper allows for quick and flexible navigation through a document. The size of a document acts as a rough indicator for the amount of information stored in it and provides a spatial orientation

Pen-based Gesture Recognition

- Offline recognition algorithms
  - static image
- Online recognition algorithms
  - spatio-temporal representation
- Recognition methods
  - statistical classification, neural networks, …
- Supported gesture types
  - single-stroke or multi-stroke

iGesture Framework





- iGesture Workbench
  - create/test gesture sets and algorithms
- Different modalities
  - digital pen, tablet PC, mouse, Wii remote, …
  - multimodal gestures
- Open Source
  - http://www.igesture.org

Rubine Algorithm, 1991

- Statistical classification algorithm for single stroke gestures (training / classification)
- A gesture G is represented as vector of P sample points

$$G = \{s_0, \ldots s_{P-1}\}, \text{with } s_i = \{x_i, y_i, t_i\}$$

- Feature vector f extracted from G

$$f = \{f_1, \ldots f_F\}$$

Rubine Features

- Original Rubine algorithm defines 13 features
  - f_1: cosine of the initial angle
  - f_2: sine of the initial angle
  - f_3: length of the bounding box diagonal
  - f_4: angle of the bounding box diagonal
  - f_5: distance between the first and last point
  - f_6: cosine of the angle between the first and last point
  - f_7: sine of the angle between the first and the last point
  - f_8: total gesture length
  - f_9: total angle traversed
  - f_{10}: the sum of the absolute angle at each gesture point
  - f_{11}: the sum of the squared value of these angles
  - f_{12}: maximum speed (squared)
  - f_{13}: duration of the gesture

Rubine Features ...

$$f_1 = \cos\alpha = \frac{(x_2 - x_0)}{\sqrt{(x_2 - x_0)^2 + (y_2 - y_0)^2}}$$

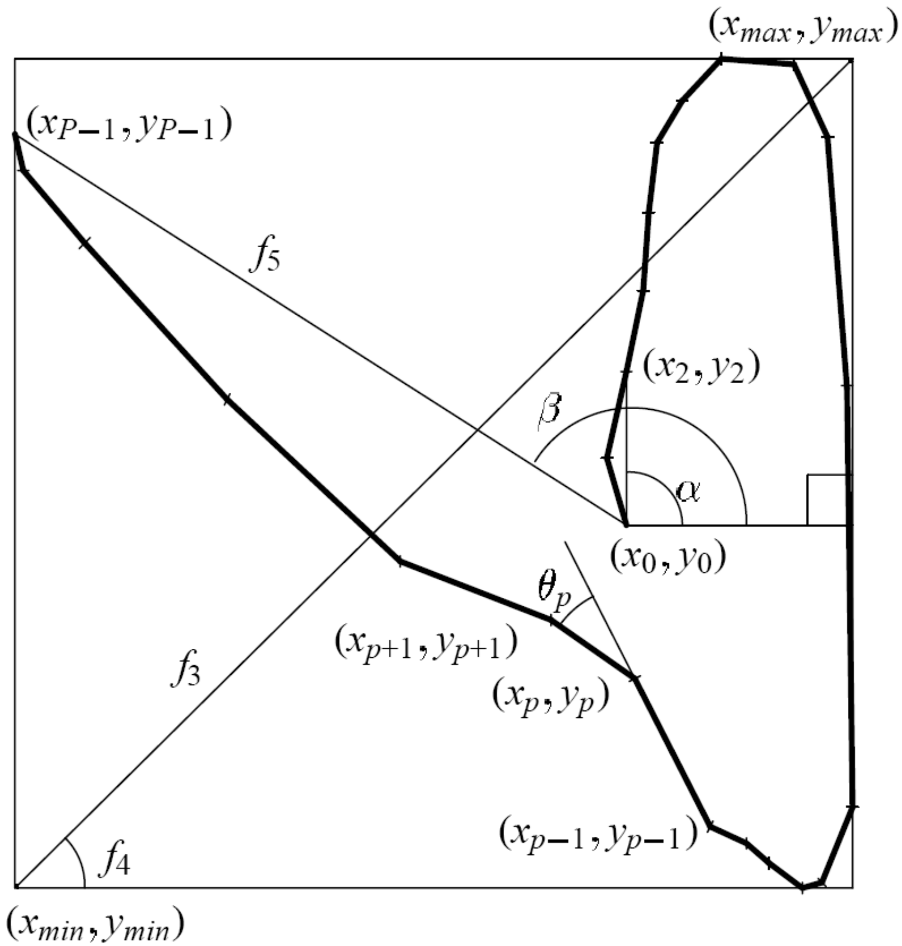
$$f_2 = \sin\alpha = \frac{(y_2 - y_0)}{\sqrt{(x_2 - x_0)^2 + (y_2 - y_0)^2}}$$

$$f_3 = \sqrt{(x_{\max} - x_{\min})^2 + (y_{\max} - y_{\min})^2}$$

$$f_4 = \arctan\frac{y_{\max} - y_{\min}}{x_{\max} - x_{\min}}$$

$$f_5 = \sqrt{\left(x_{P-1} - x_0\right)^2 + \left(y_{P-1} - y_0\right)^2}$$

$$f_6 = \cos\beta = \frac{(x_{P-1} - x_0)}{f_5}$$



Rubine Features …

$$f_7 = \sin\beta = \frac{(y_{P-1} - y_0)}{f_5}$$

$$\text{Let } \Delta x_i = x_{i+1} - x_i \quad \Delta y_i = y_{i+1} - y_i$$

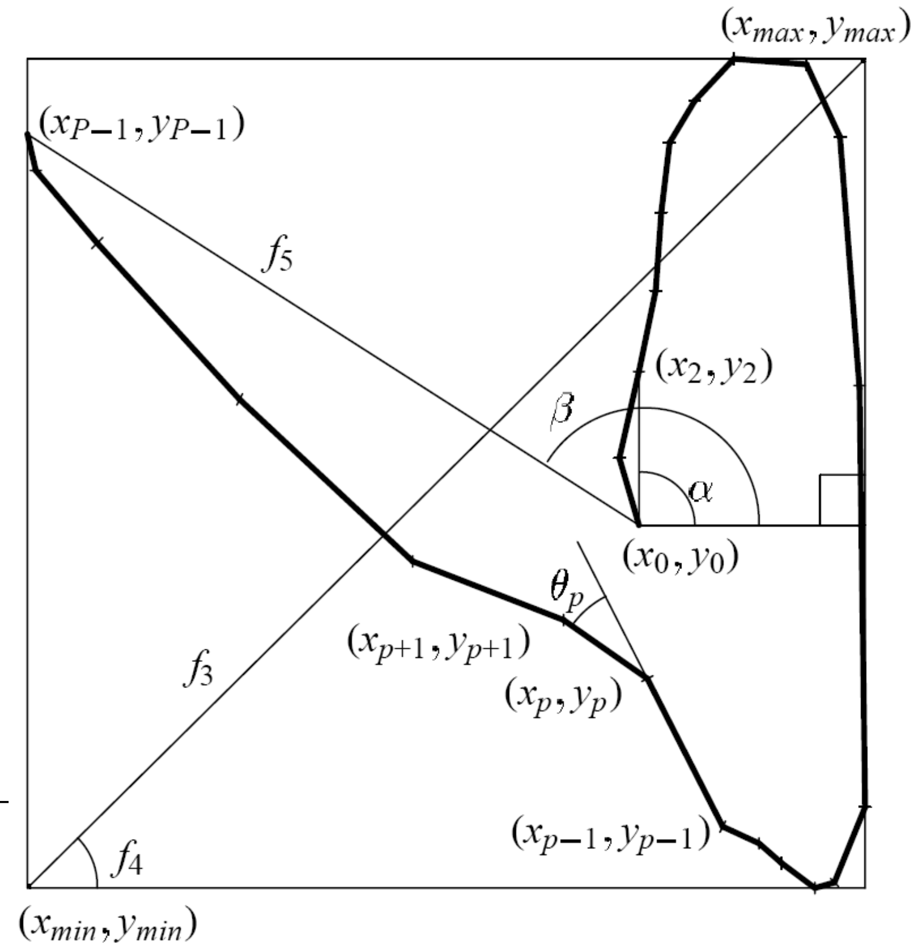
$$f_8 = \sum_{i=0}^{P-2} \sqrt{\Delta x_i^{\,2} + \Delta y_i^{\,2}}$$

$$\text{Let } \theta_i = \arctan \frac{\Delta x_i \Delta y_{i-1} - \Delta x_{i-1} \Delta y_i}{\Delta x_i \Delta x_{i-1} - \Delta x_i \Delta y_{i-1}}$$

$$f_9 = \sum_{i=1}^{P-2} \theta_i \quad f_{10} = \sum_{i=1}^{P-2} |\theta_i| \quad f_{11} = \sum_{i=i}^{P-2} \theta_i^{\,2}$$

$$\text{Let } \Delta t_i = t_{i+1} - t_i \quad f_{12} = \max_{i=0}^{P-2} \frac{\Delta x_i^{\,2} + \Delta y_i^{\,2}}{\Delta t_i^{\,2}}$$

$$f_{13} = t_{P-1} - t_0$$



Rubine Training / Classification

- Training phase

gesture samples for class c → Optimal Classifier → $w_{\hat{c}} = \{w_{\hat{c}0}, \ldots, w_{\hat{c}F}\}$

- Recognition / classification phase

$$v_{\hat{c}} = w_{\hat{c}0} + \sum_{i=1}^{F} w_{\hat{c}i} f_i$$

Evaluation Grafitti Letters (1)

	E-Rubine	Rubine	SiGrid
Correct	**334**	**280**	**273**
Error	**52**	**107**	**114**
Reject	4	3	3
Precision	0.865	0.724	0.705
Recall	0.988	0.989	0.989
F-Measure	**0.923**	**0.836**	**0.824**

Number of gesture classes: 26

Training: 15 examples for each gesture class (collected by 1 person)

Test Samples: 390 (collected by 3 persons)

Evaluation MS Application Gestures

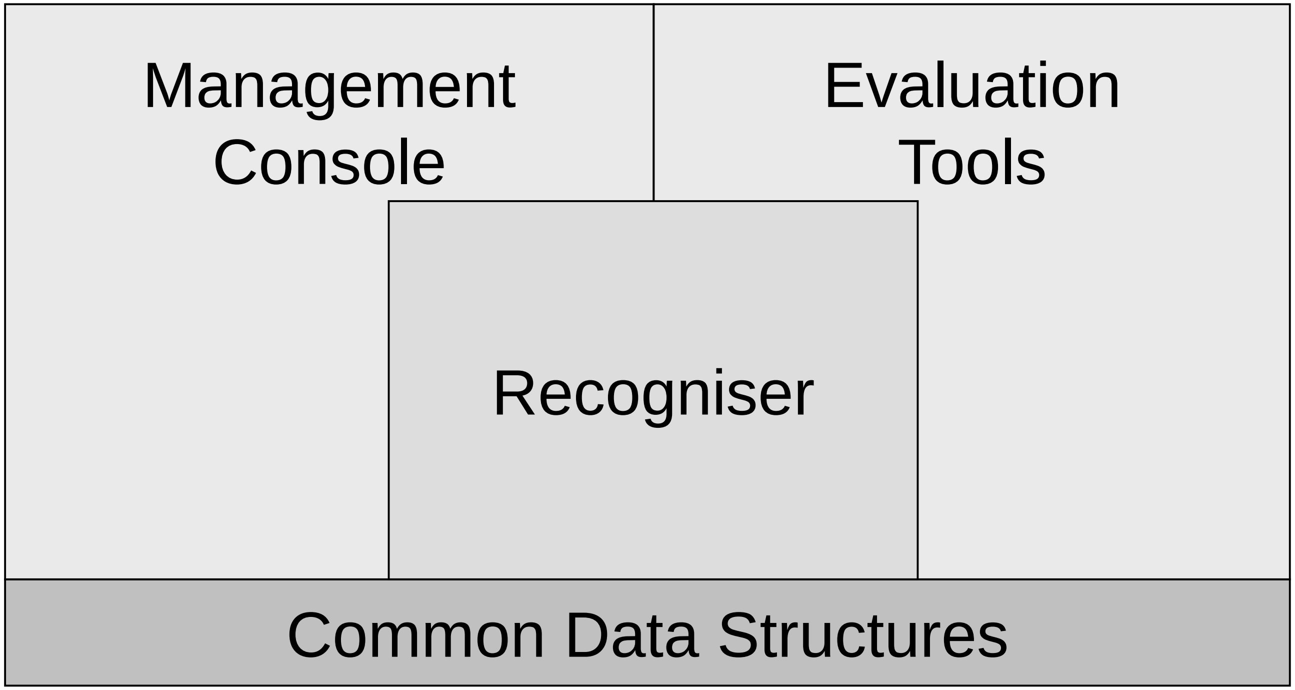
	E-Rubine	Rubine	SiGrid
Correct	**196**	**178**	**145**
Error	**4**	**19**	**32**
Reject	0	3	23
Precision	0.980	0.904	0.819
Recall	1.000	0.983	0.863
F-Measure	**0.990**	**0.942**	**0.840**

Number of gesture classes: 40

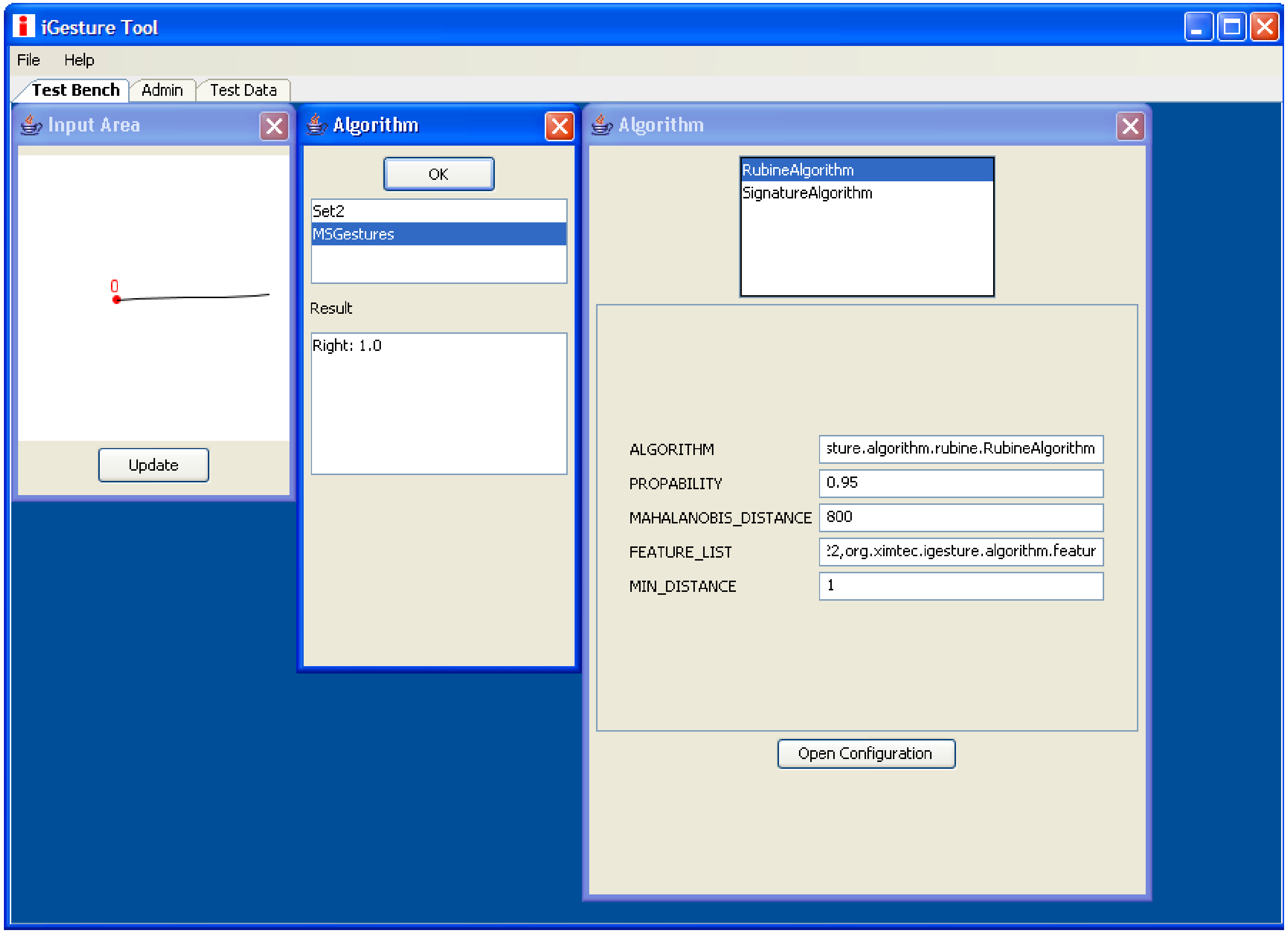
Training: 15 examples for each gesture class (collected by 1 person)

Test Samples: 200 (collected by 1 person)

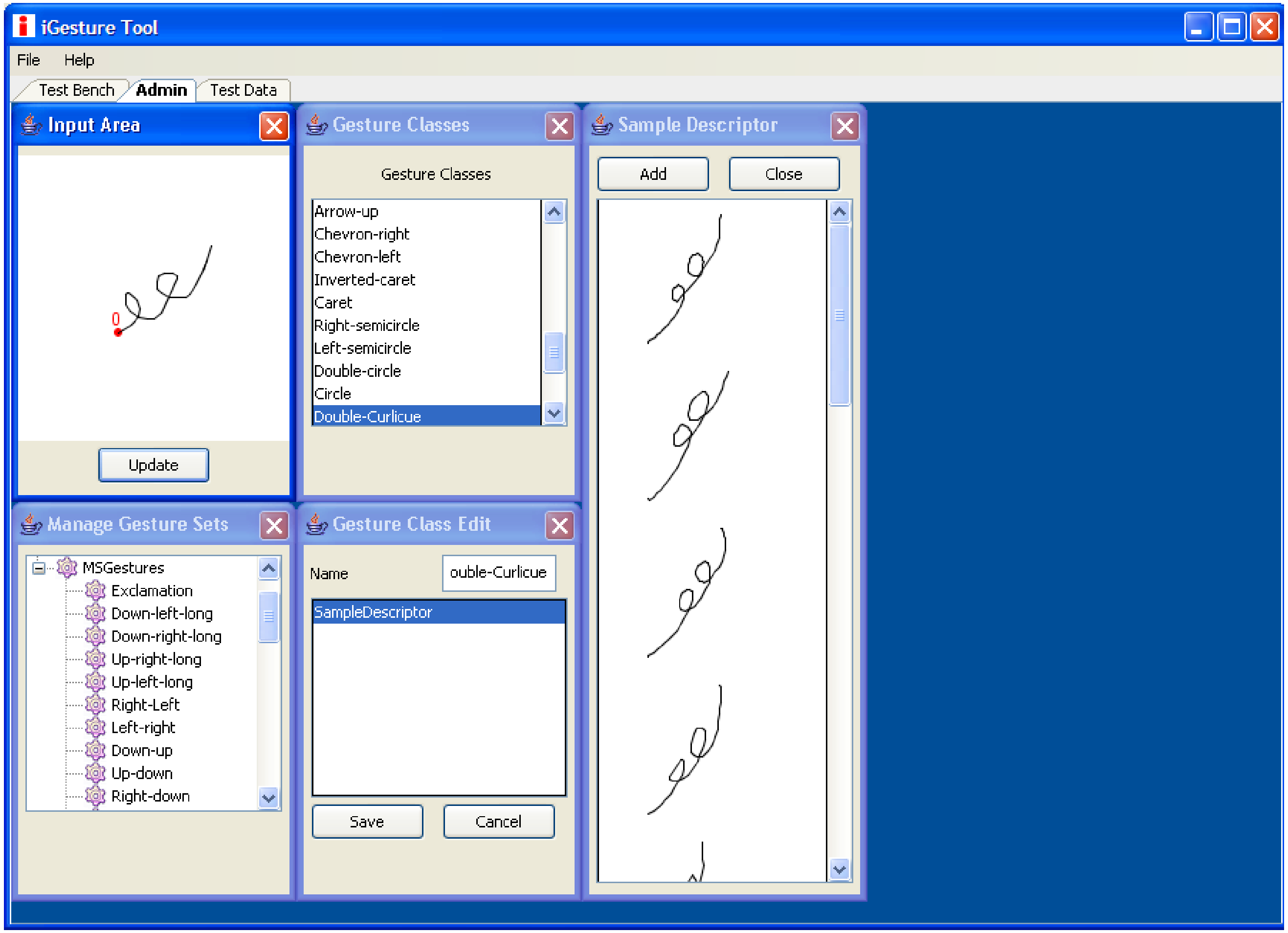
iGesture Architecture Overview



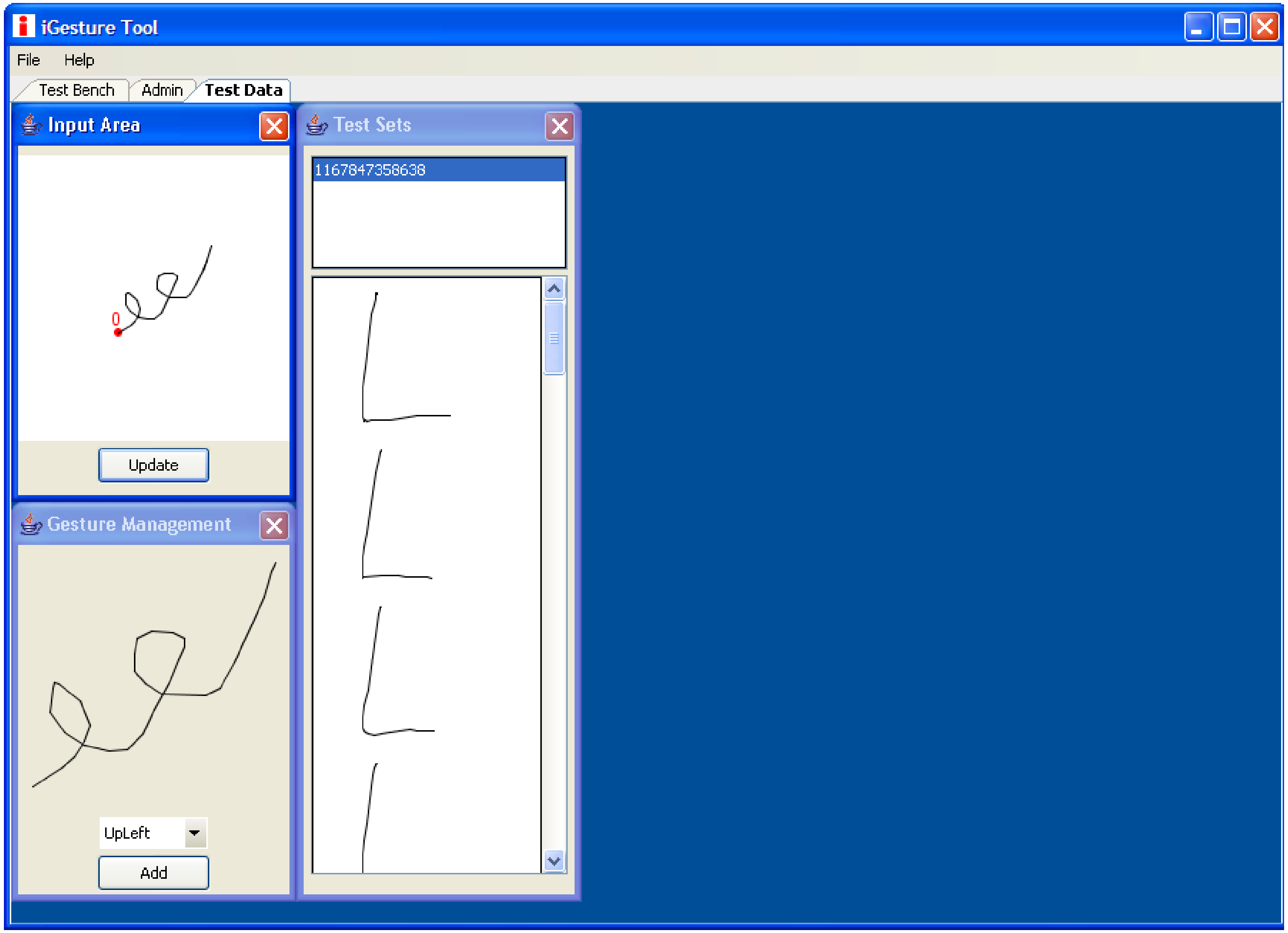
iGesture Test Bench Tab



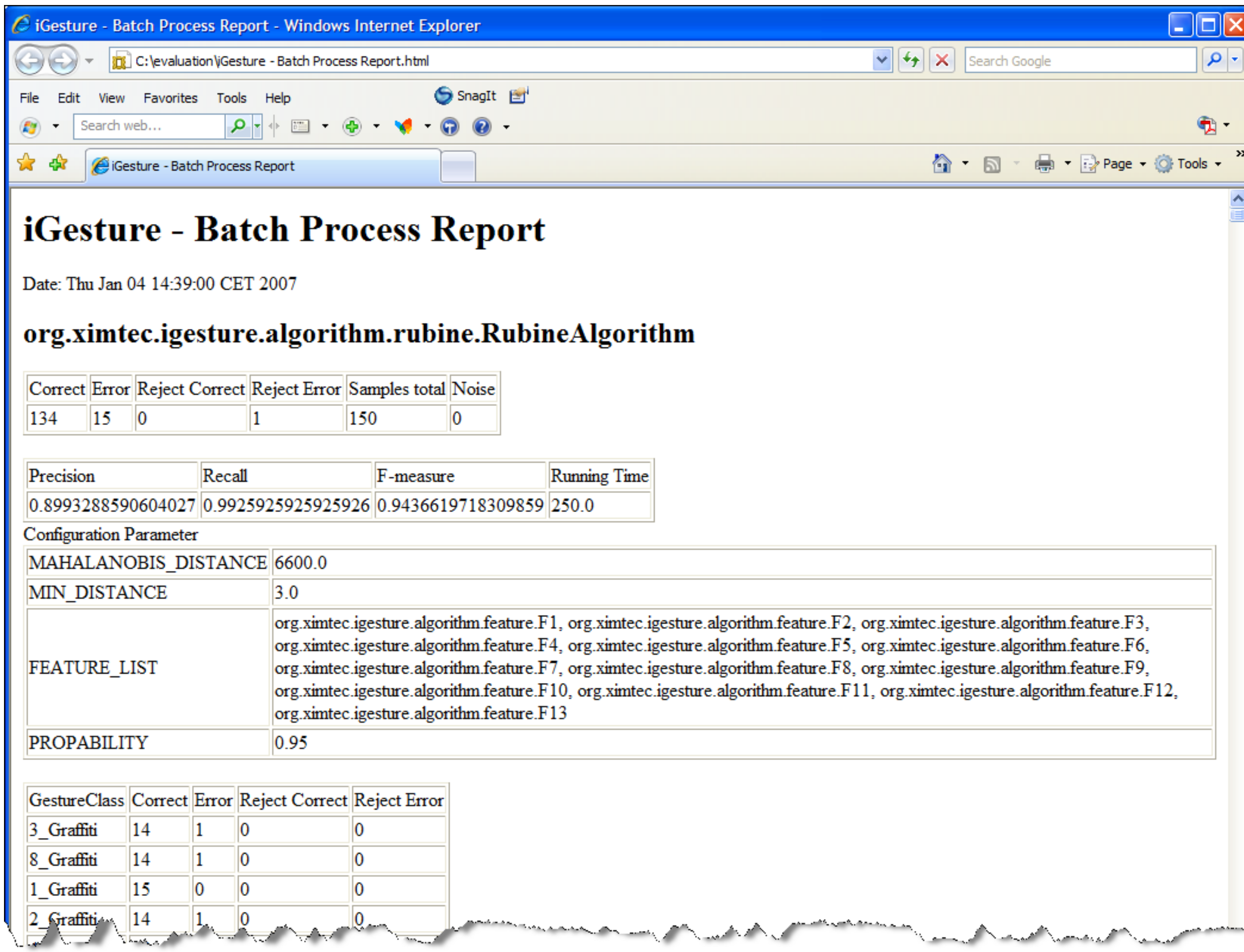
iGesture Admin Tab



iGesture Test Data Tab



Evaluation Tools



iGesture - Batch Process Report

Date: Thu Jan 04 14:39:00 CET 2007

org.ximtec.igesture.algorithm.rubine.RubineAlgorithm

Correct	Error	Reject Correct	Reject Error	Samples total	Noise
134	15	0	1	150	0

Precision	Recall	F-measure	Running Time
0.8993288590604027	0.9925925925925926	0.9436619718309859	250.0

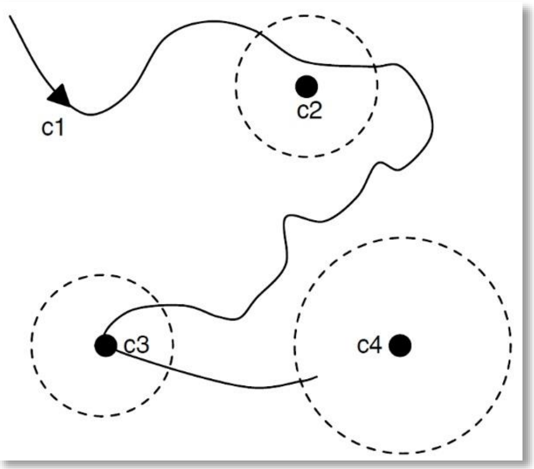
Configuration Parameter

MAHALANOBIS_DISTANCE	6600.0
MIN_DISTANCE	3.0
FEATURE_LIST	org.ximtec.igesture.algorithm.feature.F1, org.ximtec.igesture.algorithm.feature.F2, org.ximtec.igesture.algorithm.feature.F3, org.ximtec.igesture.algorithm.feature.F4, org.ximtec.igesture.algorithm.feature.F5, org.ximtec.igesture.algorithm.feature.F6, org.ximtec.igesture.algorithm.feature.F7, org.ximtec.igesture.algorithm.feature.F8, org.ximtec.igesture.algorithm.feature.F9, org.ximtec.igesture.algorithm.feature.F10, org.ximtec.igesture.algorithm.feature.F11, org.ximtec.igesture.algorithm.feature.F12, org.ximtec.igesture.algorithm.feature.F13
PROPABILITY	0.95

GestureClass	Correct	Error	Reject Correct	Reject Error
3_Graffiti	14	1	0	0
8_Graffiti	14	1	0	0
1_Graffiti	15	0	0	0
2_Graffiti	14	1	0	0

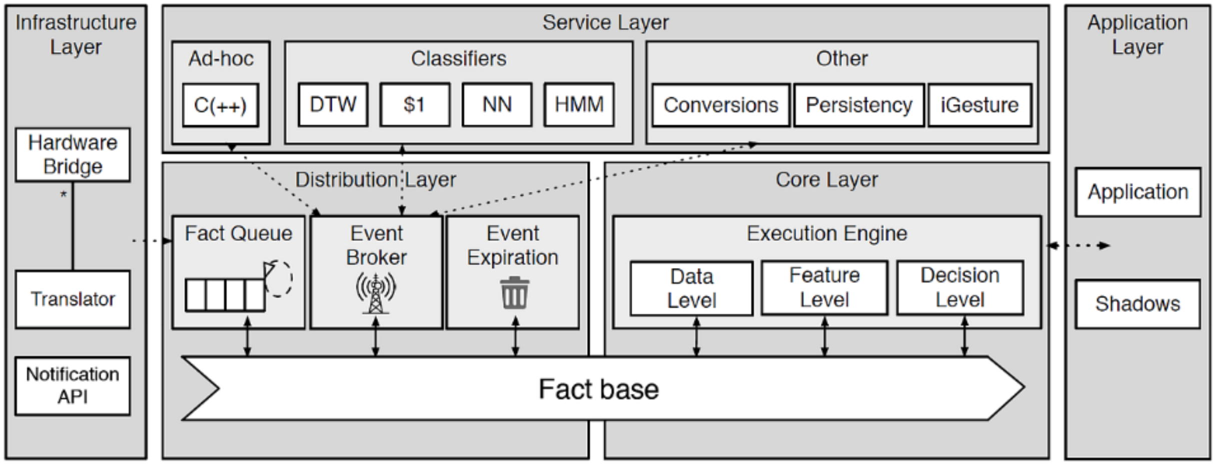
Gesture Spotting / Segmentation

- Always-on mid-air interfaces like the Microsoft Kinect do not offer an explicit start and end point of a gesture
- How do we know when a gesture starts?
  - use another modality (e.g. pressing a button or voice command)
    - not a very natural interaction
  - try to continuously spot potential gestures
- We introduced a *new gesture spotting approach* based on a human-readable representation of automatically inferred spatio-temporal constraints
  - potential gestures handed over to a gesture recogniser



Hoste et al., 2013

Mudra

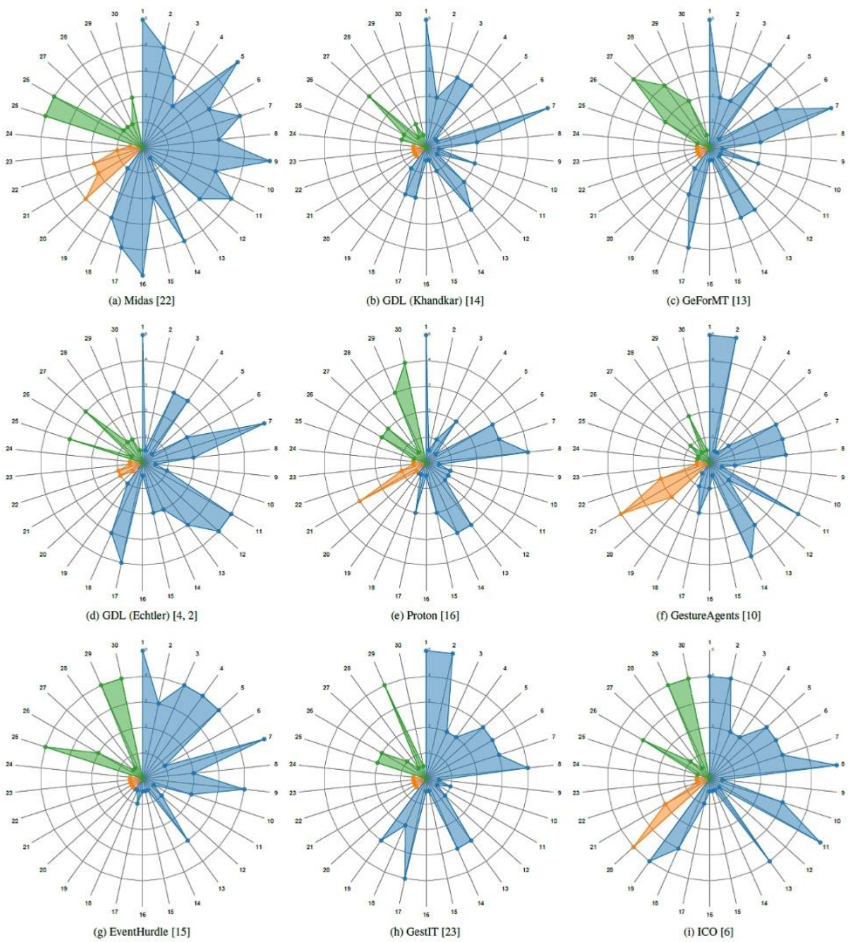


Hoste et al., 2011

- Fusion across different levels of abstraction
  - via fact base
- Interactions defined via declarative rule-based language
- Rapid prototyping
  - simple integration of new input devices
  - integration of external gesture recognisers

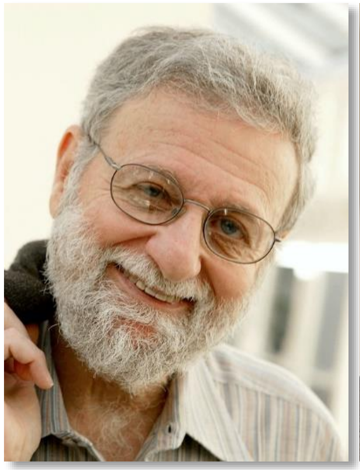
Challenges and Opportunities

- Various (declarative) domain-specific languages have been proposed over the last few years
- Challenges
  - gesture segmentation
  - scalability in terms of complexity
  - how to deal with uncertainty
  - …



A Step Backward In Usability

- Usability tests of existing gestural interfaces revealed a number of problems
  - lack of established guidelines for gestural control
  - misguided insistence of companies to ignore established conventions
  - developers' ignorance of the long history and many findings of HCI research
    - unleashing untested and unproven creative efforts upon the unwitting public
- Several fundamental principles of interaction design are disappearing from designers' toolkits
  - weird design guidelines by Apple, Google and Microsoft



Don Norman

Jacob Nielsen

A Step Backward In Usability …

- Visibility
  - *non-existent signifiers*
    - swipe right across and unopened email (iPhone) or press and hold on an unopened email (Android) to open a dialogue
  - *misleading signifiers*
    - some permanent standard buttons (e.g. menu) which do not work for all applications (Android)
- Feedback
  - back button does not only work within an application but moves to the "activity stack" and might lead to "leaving" the application without any warning
    - forced application exit is not good in terms of usability

A Step Backward In Usability …

- Consistency and Standards
  - operating system developers have their own interface guidelines
  - proprietary standards make life more difficult for users
    - touching an image might enlarge it, unlock it so that it can be moved, hyperlink from it, etc.
    - flipping screens up, down, left or right with different meanings
  - consistency of gestures between applications on the same operating system is often also not guaranteed
- Discoverability
  - while possible actions could be explored via the GUI, this is no longer the case for gestural commands

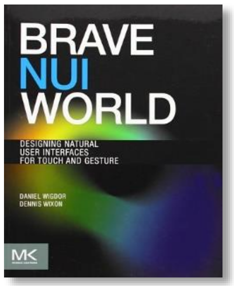
A Step Backward In Usability …

- Scalability
  - gestures that work well for small screens might fail on large ones and vice versa
- Reliability
  - gestures are invisible and users might not know that there was an accidental activation
  - users might lose their sense of controlling the system and the user experience might feel random
- Lack of undo
  - often difficult to recover from accidental selections

Homework

- Read the following paper that is available on PointCarré (papers/Norman 2010)
  - D.A. Norman and J. Nielsen, *Gestural Interfaces: A Step Backward In Usability*, interactions, 17(5), September 2010

References



Brave NUI World: Designing Natural User Interfaces for Touch and Gesture, Daniel Wigdor and Dennis Wixon, Morgan Kaufmann (1st edition), April 27, 2011, ISBN-13: 978-0123822314

- D.A. Norman and J. Nielsen, *Gestural Interfaces: A Step Backward In Usability*, interactions, 17(5), September 2010
  - http://dx.doi.org/10.1145/1836216.1836228
- A. Kendon, *Gesture: Visible Action as Utterance*, Cambridge University Press, 2004

References …

- Power Glove Video
  - https://www.youtube.com/watch?v=3g8JiGjRQNE
- R.Y. Wang and J. Popović, *Real-Time Hand-Tracking With a Color Glove*, Proceedings of SIGGRAPH 2009, 36th International Conference and Exhibition of Computer Graphics and Interactive Techniques, New Orleans, USA, August 2009
  - http://dx.doi.org/10.1145/1576246.1531369
- Real-Time Hand-Tracking With a Color Glove Video
  - https://www.youtube.com/watch?v=kK0BQjItqgw
- How the Kinect Depth Sensor Works Video
  - https://www.youtube.com/watch?v=uq9SEJxZiUg

References …

- Myo Wearable Gesture Control Video
  - https://www.youtube.com/watch?v=oWu9TFJjHaM
- Kinect Sign Language Translator Video
  - https://www.youtube.com/watch?v=HnkQyUo3134
- iGesture Gesture Recognition Framework
  - http://www.igesture.org
- B. Signer, U. Kurmann and M.C. Norrie, *iGesture: A General Gesture Recognition Framework*, Proceedings of ICDAR 2007, 9th International Conference on Document Analysis and Recognition, Curitiba, Brazil, September 2007
  - http://beatsigner.com/publications/signer_ICDAR2007.pdf

References …

- D. Rubine, *Specifying Gestures by Example*, Proceedings of SIGGRAPH 1991, International Conference on Computer Graphics and Interactive Techniques, Las Vegas, USA, July 1991
  - https://doi.org/10.1145/122718.122753
- L. Hoste, B. De Rooms and B. Signer, *Declarative Gesture Spotting Using Inferred and Refined Control Points*, Proceedings of ICPRAM 2013, International Conference on Pattern Recognition, Barcelona, Spain, February 2013
  - http://beatsigner.com/publications/hoste_ICPRAM2013.pdf

References …

- L. Hoste, B. Dumas and B. Signer, *Mudra: A Unified Multimodal Interaction Framework*, Proceedings of ICMI 2011, 13th International Conference on Multimodal Interaction, Alicante, Spain, November 2011
  - http://beatsigner.com/publications/hoste_ICMI2011.pdf
- L. Hoste and B. Signer, *Criteria, Challenges and Opportunities for Gesture Programming Languages*, Proceedings of EGMI 2014, 1st International Workshop on Engineering Gestures for Multimodal Interfaces, Rome, Italy, June, 2014
  - http://beatsigner.com/publications/hoste_EGMI2014.pdf

References …

- B. Buxton, *Gesture Based Interaction*, Draft, August 2011
  - http://www.billbuxton.com/input14.Gesture.pdf
- Touch Gesture Reference Guide
  - http://static.lukew.com/TouchGestureGuide.pdf

Next Lecture

Tangible, Embedded and Embodied Interaction