Gesture Recognition

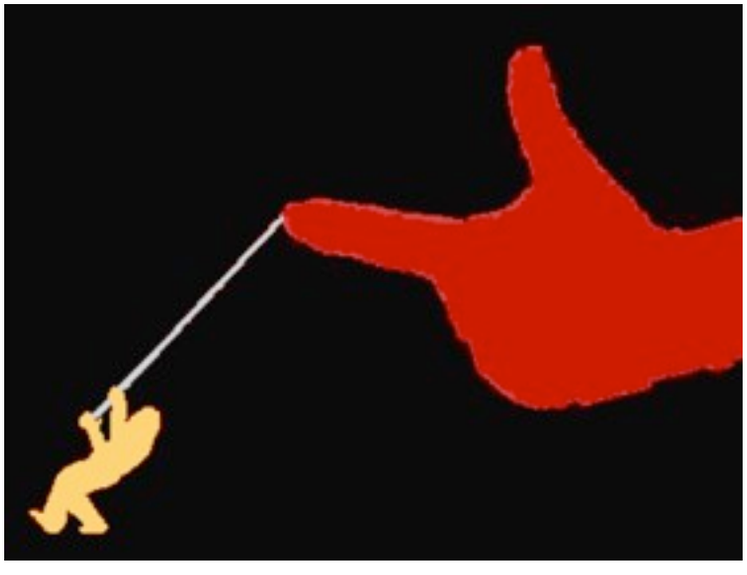
Adrian Kündig
adkuendi@student.ethz.ch

The beginning of gestures based interfaces

Gesture Recognition

- 1970 Myron W. Krueger and VideoPlace

http://sofa23.net/index.php?m=1&sm=&t=23&sp=18&spic=43&me=show%20all&s=
http://www.inventinginteractive.com/2010/03/22/myron-krueger/

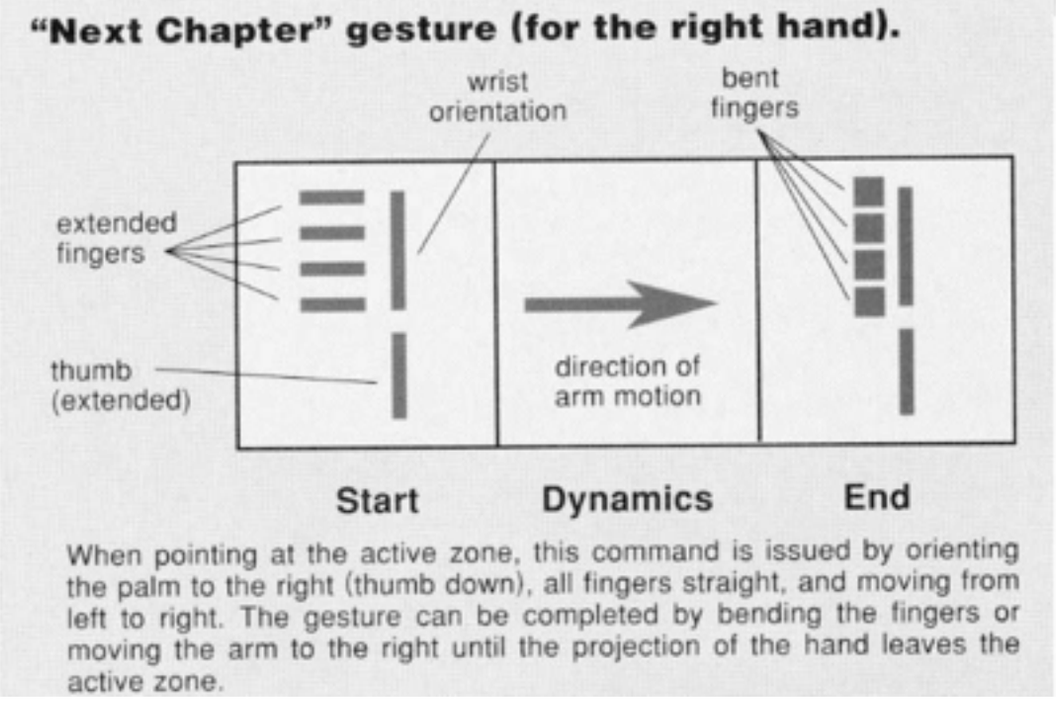


One of the first
prototyped VR
Using cameras for recognition
Simple ideas

(Baudel and Beaudouin-Lafon, 1993)

Gesture Recognition

- 1970 Myron W. Krueger and VideoPlace
- 1993 Charade

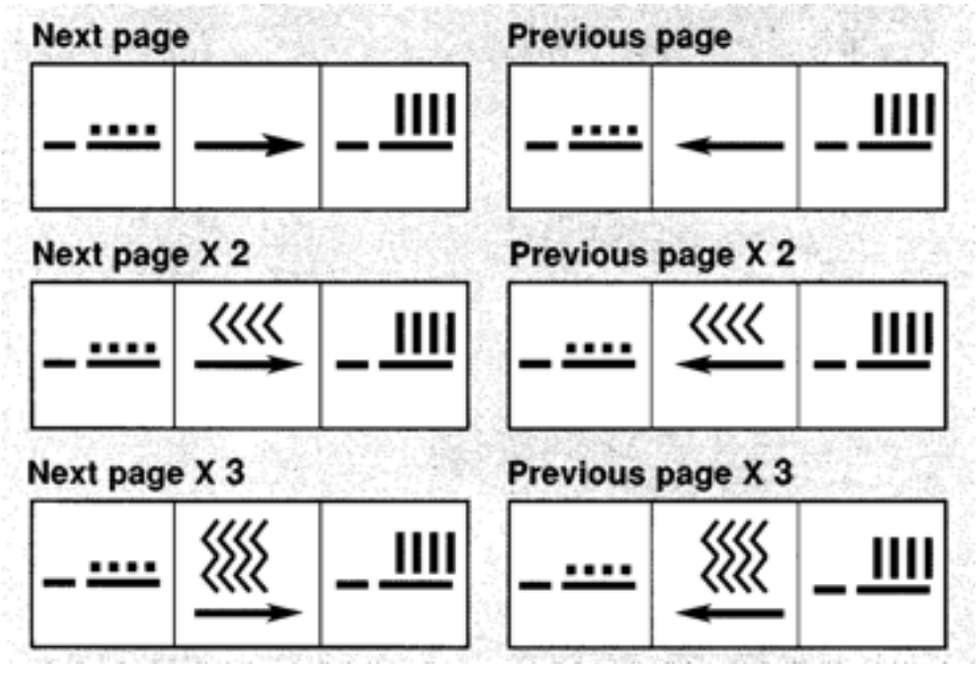


First formal definition of gestures
Control PowerPoint
dataglove
4 line = fingers, 1 line = thumb

(Baudel and Beaudouin-Lafon, 1993)

Gesture Recognition

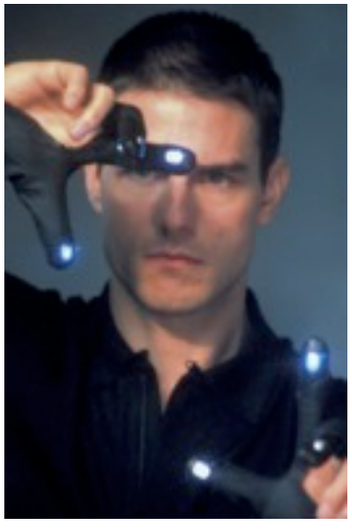
- 1970 Myron W. Krueger and VideoPlace
- 1993 Charade



Selection of gestures

Gesture Recognition

- 1970 Myron W. Krueger and VideoPlace
- 1993 Charade
- 2002 Minority Report



http://thomaspmbarnett.com/globlogization/2013/2/5/times-battleland-terrorism-minority-report-has-finally-arriv.html
http://7thperbmmrblog.blogspot.ch/2011/01/william-bermudez.html

Hollywood movie from Steven Spielberg
Rooted in Research from John Underkoffler
“like conducting an orchestra”
tom cruise

Gesture Recognition

- 1970 Myron W. Krueger and VideoPlace
- 1993 Charade
- 2002 Minority Report
- 2009 Oblong Industries

Last step in our history of gesture based interfaces
Commercial company founded by John Underkoffler
developed g-speak
Intended for big data analysis
Requires specialized applications

Oblong Industries - Demo

http://oblong.com/g-speak/

Orientation in 3D
Selection
Segmentation

Oblong Industries - Demo

http://oblong.com/g-speak/

Orientation in 3D
Selection
Segmentation

Common Factor

http://www.5dt.com/DataGloveImages.html

most shown systems have in common: data glove
Hand tracking
Hand reconstruction
Feedback

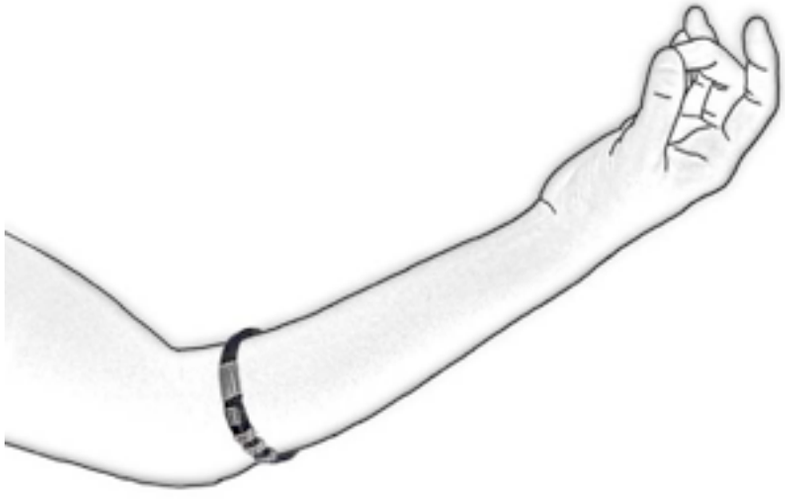
How can we get rid of the Data Glove?

Free up hands
Remove instrumentation

(Saponas et al, 2009)

Muscle Computer Interface

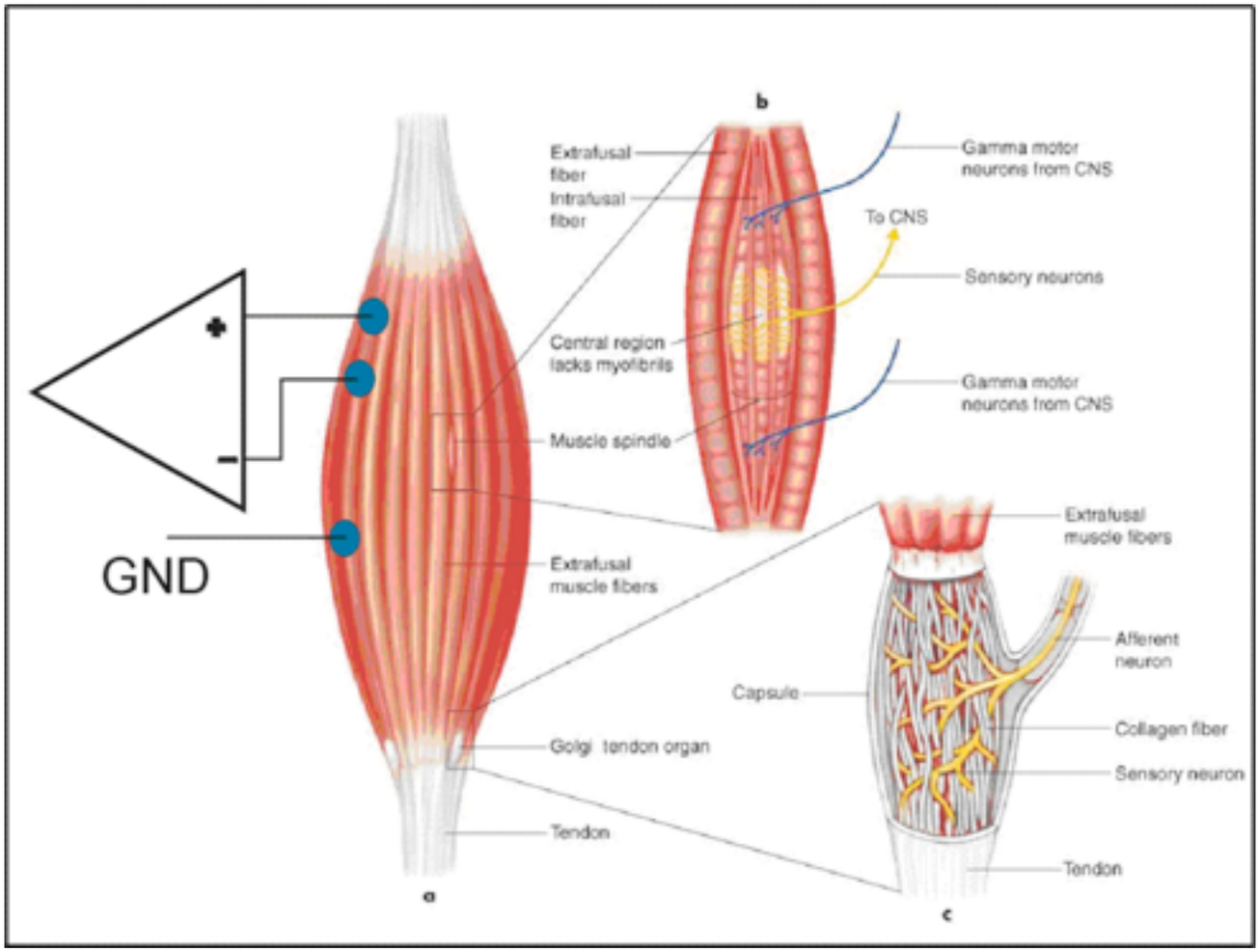
- Hands free gestures while holding an object
- Arm band like design
- Sensing muscle activity



Hands free
Muscle sending

(Saponas et al, 2009)

Muscle Computer Interface - Technology

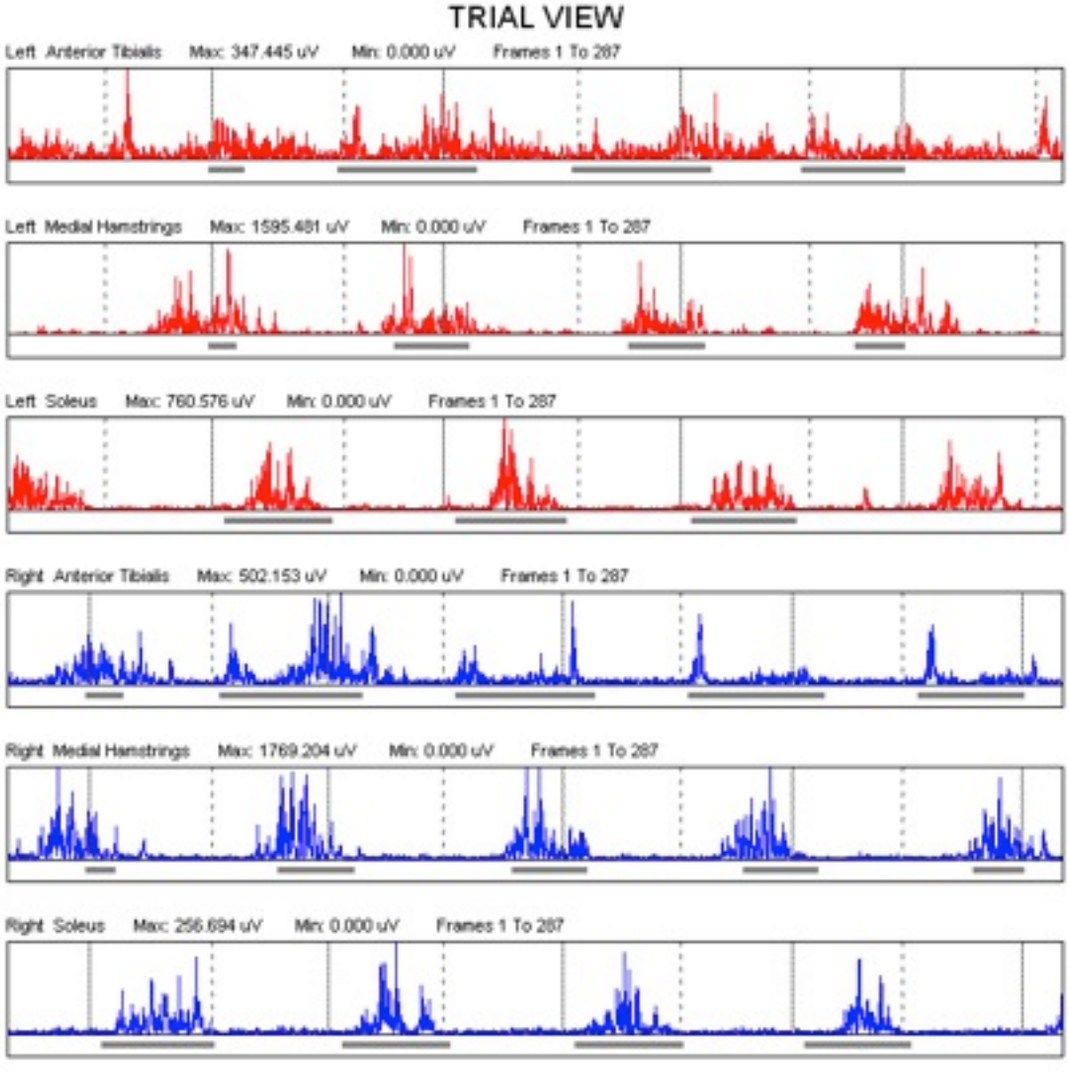


http://painmd.tv/wp-content/uploads/2011/04/emg-muscle-configuration.gif

EMG or Electromyography
primarily in Medical therapy (muscle function assessment, controlling prosthetics)
Action Potential generated by muscle when signal arrives from Motor Neuron
Invasively by inserting a needle into the muscle
Non invasively by sensing on the skin

(Saponas et al, 2009)

Muscle Computer Interface - Technology

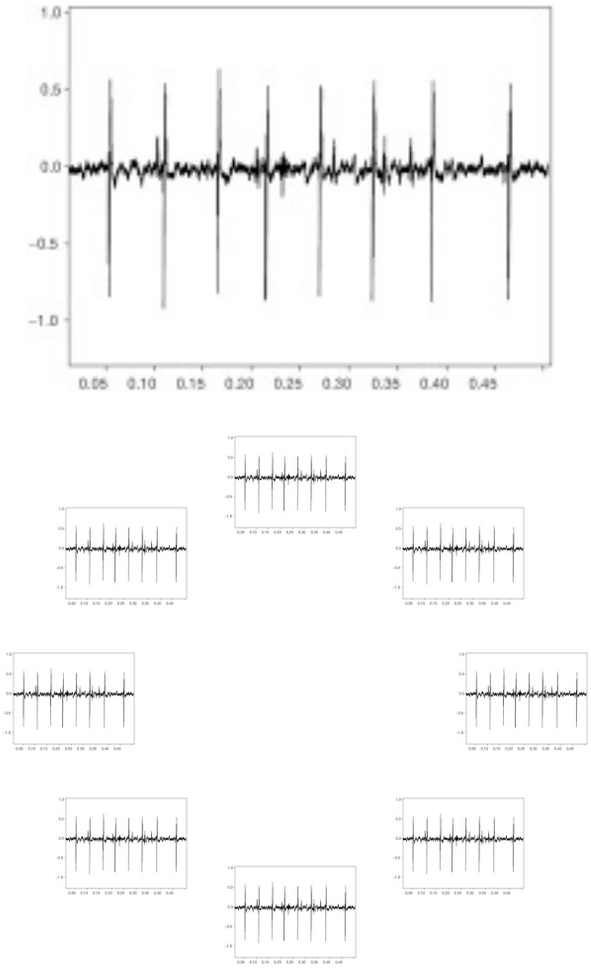


http://www.emgsrus.com/graphics/emg_trial_rect_page.png

here measured activity
6 Different muscles
Peaks of action potentials

(Saponas et al, 2009)

Muscle Computer Interface - Technology



- Root mean square
- Frequency energy
- Phase Coherence

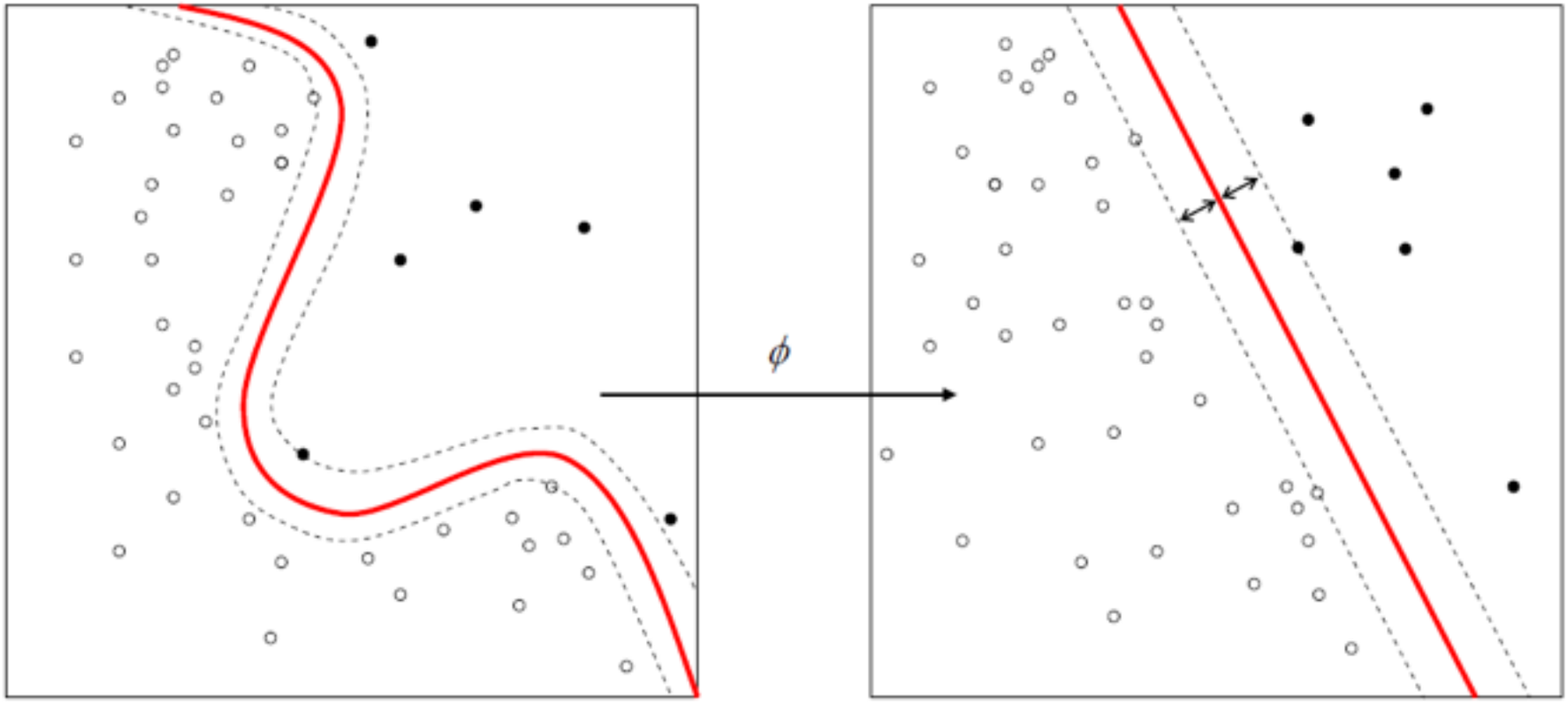
Support Vector Machine

http://www.nature.com/gimo/contents/pt1/fig_tab/gimo32_F2.html

6 Sensors and 2 ground electrodes
Features extracted from 31ms sample
- Root Mean Square of amplitude per channel and ratio of pair of channels sqrt(1/n * (x1^2 + x2 ^ 2 + ...))
- Frequency energy via FFT
- Relationship between channels
Classified from SVM into gestures

Support Vector Machines

- Binary Linear Classifier
- Extended to multiple classes



https://en.wikipedia.org/wiki/File:Kernel_Machine.png

Function phi transforms feature space, such that it is possible to lay a hyper plain between two classes
Try to lay separator such that separation is most clear
Multiple classes by (one vs rest) or pairwise (one vs one)

(Saponas et al, 2009)

Muscle Computer Interface - Demo

Guitar hero
input is sent as soon as user touches both fingers

(Saponas et al, 2009)

Muscle Computer Interface - Demo

Guitar hero
input is sent as soon as user touches both fingers

(Saponas et al, 2009)

Muscle Computer Interface

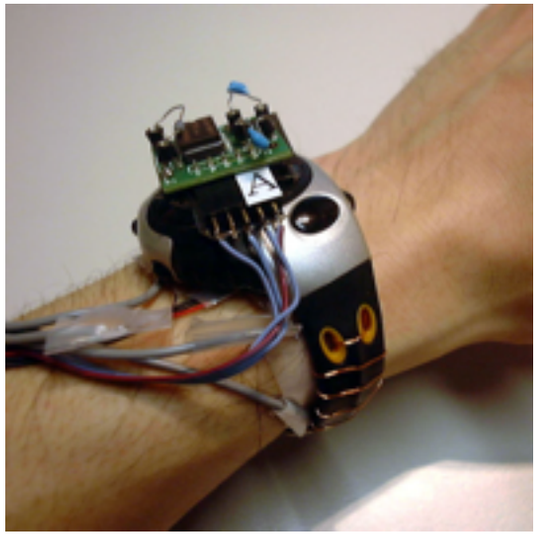
- Pro
  - No instrumentation of hand
  - Hidden near elbow
- Contra
  - Inaccurate compared to some following papers
  - Muscle activity required

79 % accuracy

(Rekimoto, 2001)

Gesture Wrist

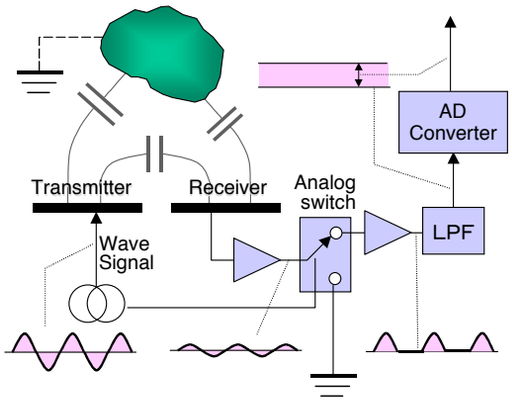
- Hands free gestures
- Embed sensing device in wrist watch
- Feedback on gesture

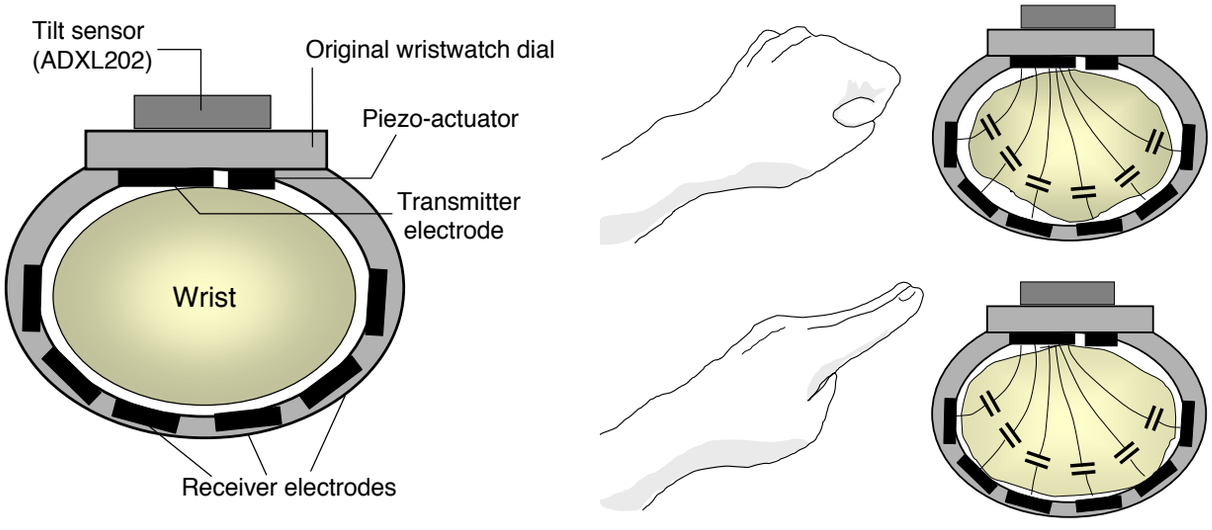


(Rekimoto, 2001)

Gesture Wrist - Technology

- Wave signal is transmitted
- The receivers are synchronized
- The received strength is proportional to the distance



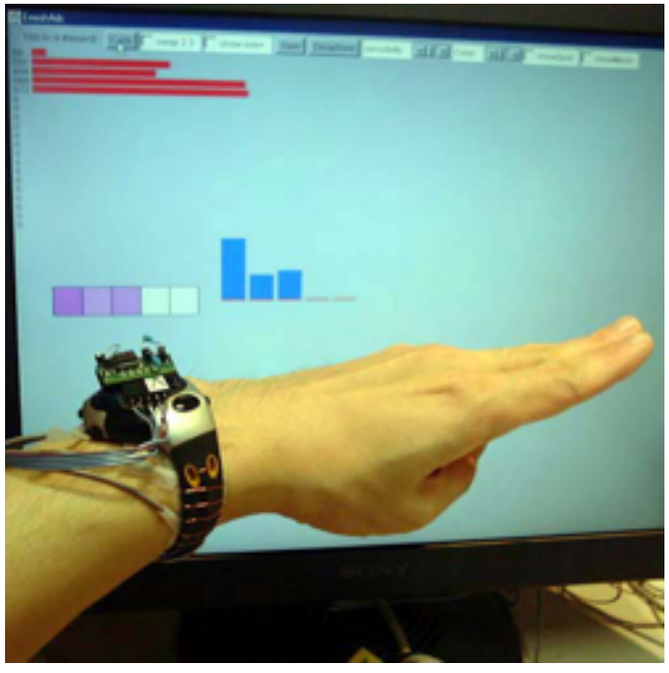


Actuator vibrates
measure the capacitance of the wrist and the receiver electrodes
measuring the distance between wristband an wrist

(Rekimoto, 2001)

Gesture Wrist - Technology

- Distinguish ‘Point’ and ‘Fist’ pose



Clear difference between point and fist
Only two gestures used to differentiate gestures

(Rekimoto, 2001)

Gesture Wrist - Examples

- Distinguish ‘Point’ and ‘Fist’ pose
- Combined with an accelerometer
- Rotation also recognizable

Only two gestures used to differentiate gestures
Use rotation to control slider or knob

(Rekimoto, 2001)

Gesture Wrist

- Pro
  - Small, watch like design
  - Sensor embedded inside accessory
  - Simple recognition method
- Contra
  - Only a small set of gestures can be recognized

(Fukui et al, 2011)

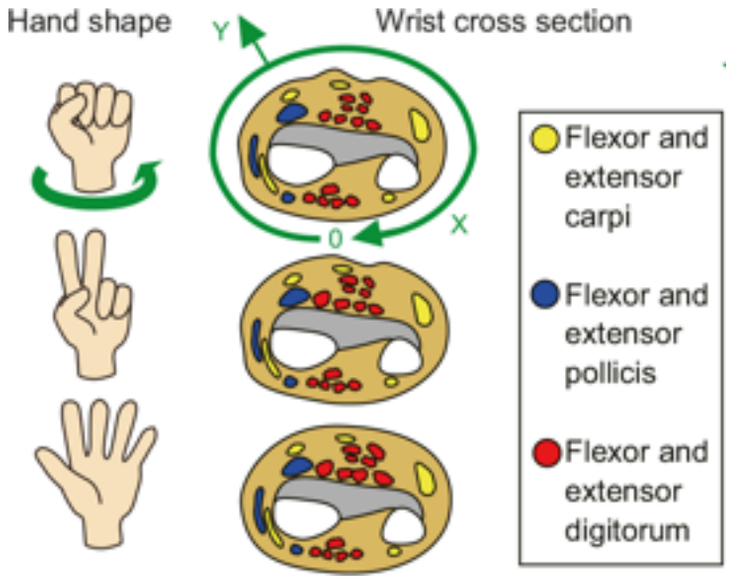
Hand Shape with Wrist Contour

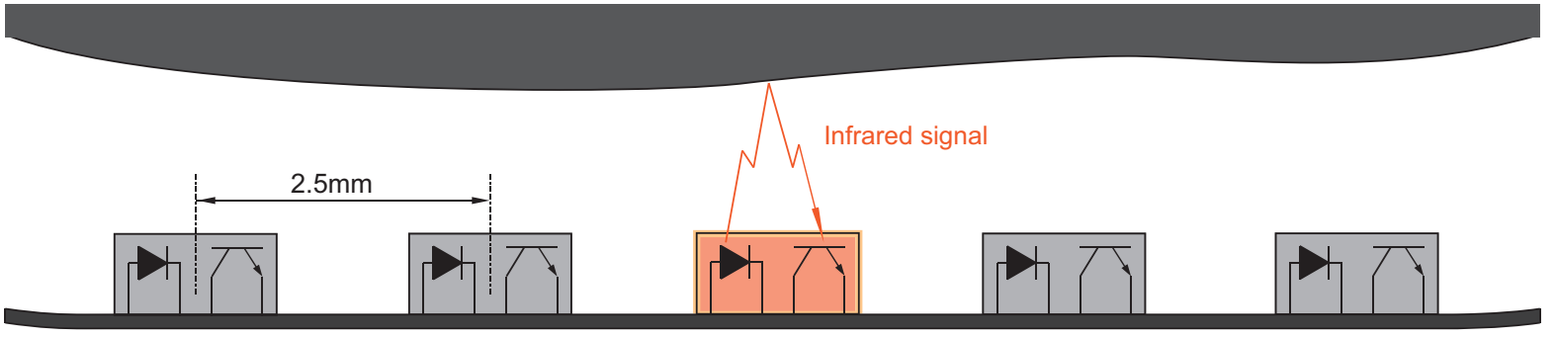
- Hands free gestures
- Wrist watch like design

(Fukui et al, 2011)

Hand Shape with Wrist Contour - Technology

- Static wrist band
- Photo reflectors
- Senses distance between band and skin





150 sensors

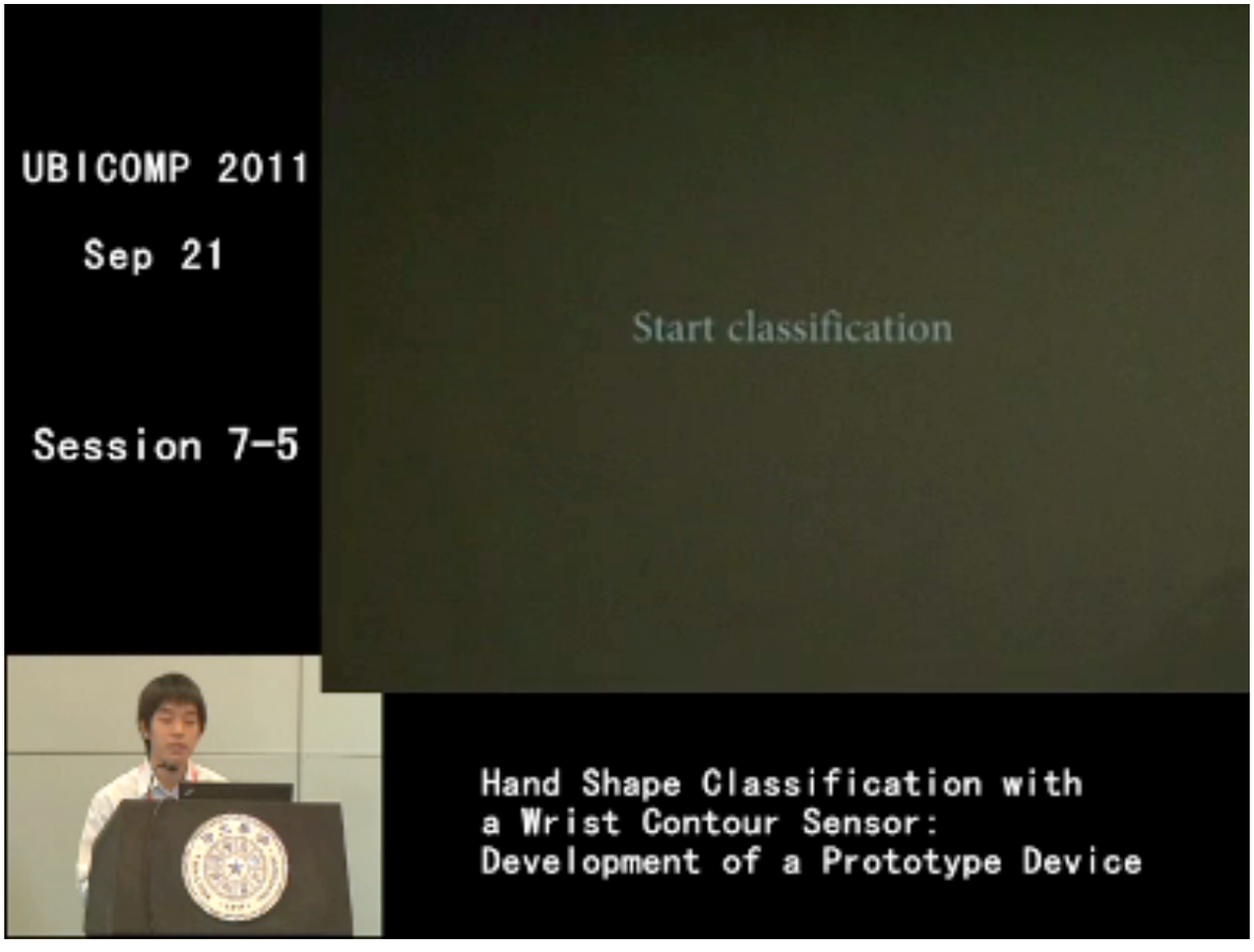
(Fukui et al, 2011)

Hand Shape with Wrist Contour - Demo

static image representing gesture

(Fukui et al, 2011)

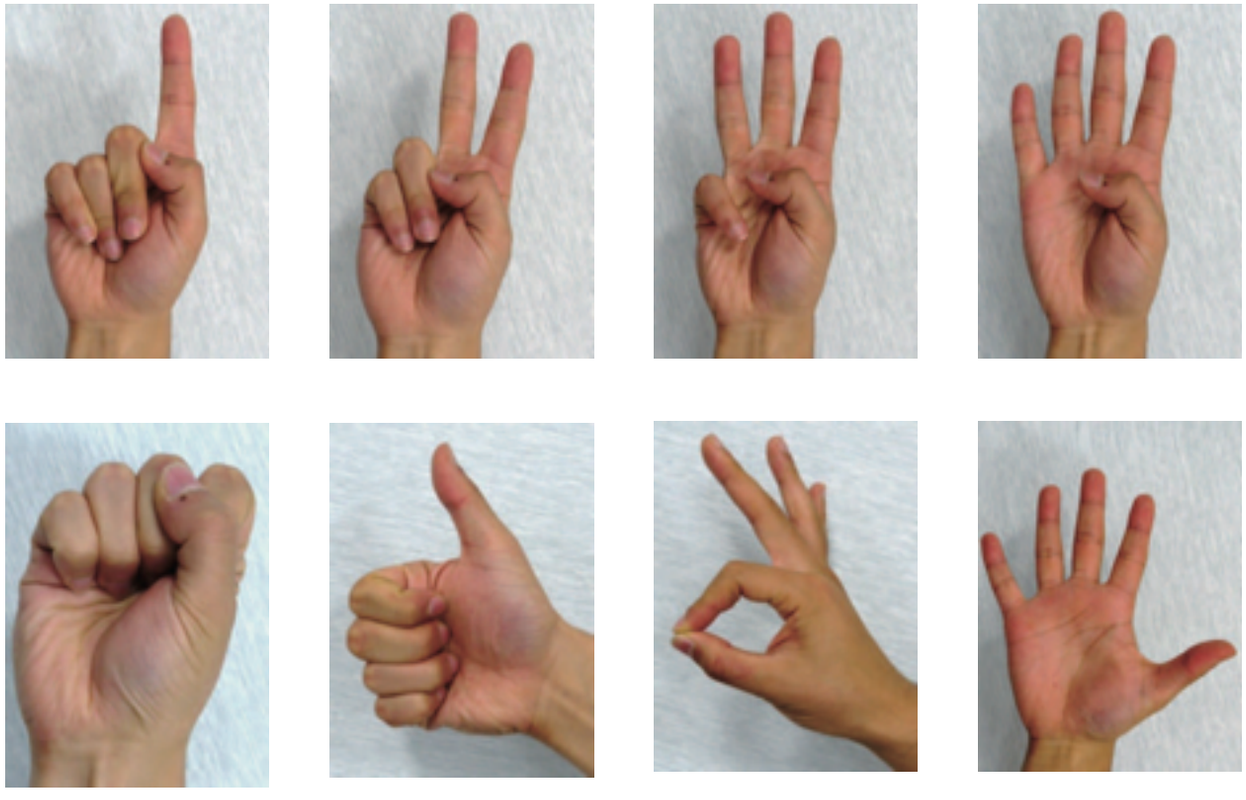
Hand Shape with Wrist Contour - Demo



static image representing gesture

(Fukui et al, 2011)

Hand Shape with Wrist Contour - Examples



The recognized gesture set
some gestures quiet similar

(Fukui et al, 2011)

Hand Shape with Wrist Contour - Accuracy

•Boosting method

Input \ Output	Fl	TU	1F	2F	3F	4F	OK	OH	Rate
Fist	291	8	0	0	0	1	0	0	97
Thumbs up	64	117	80	29	6	0	4	0	39
One finger	9	127	97	15	13	26	13	0	32.3
Two fingers	24	10	20	125	34	27	53	7	41.7
Three fingers	0	0	4	36	61	112	52	35	20.3
Four fingers	0	0	0	5	124	102	33	36	34
A-OK	0	20	43	104	17	15	55	46	18.3
Open hand	0	0	0	0	0	0	0	300	100

Average rate: 47.8%

•k-NN method

Input \ Output	Fl	TU	1F	2F	3F	4F	OK	OH	Rate
Fist	298	1	0	0	0	1	0	0	99.3
Thumbs up	30	207	15	19	0	0	29	0	69
One finger	46	59	147	29	0	1	18	0	49
Two fingers	0	63	0	74	63	7	65	28	24.7
Three fingers	0	28	0	31	79	58	72	32	26.3
Four fingers	0	21	2	9	63	51	92	62	17
A-OK	0	14	46	60	31	76	29	44	9.67
Open hand	0	0	0	0	0	66	24	210	70

Average rate: 45.6%

Confusion matrix
wide spread
boosting method and k-NN method rather simple
diagonal is correctly recognized

(Fukui et al, 2011)

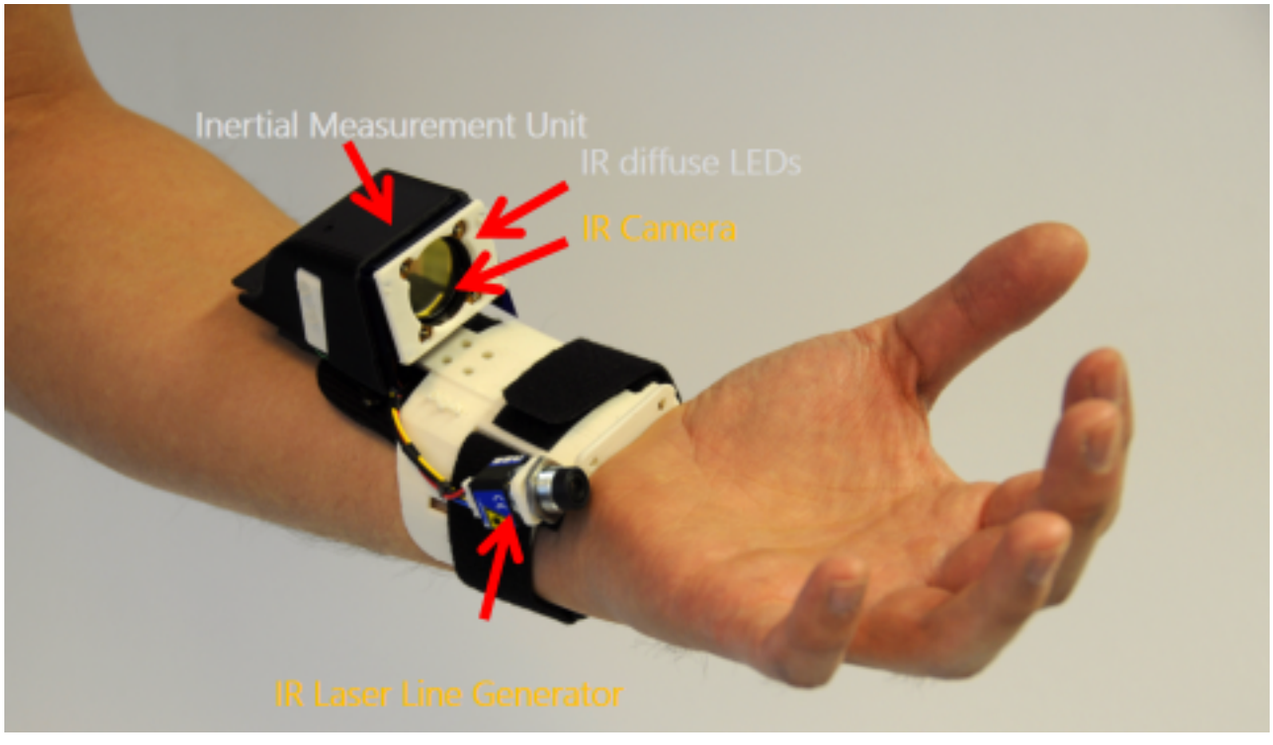
Hand Shape with Wrist Contour

- Pro
  - Small, watch like design
  - Can be hidden inside accessory
  - New approach to gesture recognition
- Contra
  - Bad recognition rate
  - Limited set of gestures

(Kim et al, 2012)

Digits

- Recover full 3D hand model
- Cheap hardware
- Low power



Already partly presented by Professor Hilliges in the introduction of the seminar
more sophisticated
imitates data glove

(Kim et al, 2012)

Digits - Technology

Background Subtraction → CCL & Tracking → 3D Laser Triangulation → Hand Pose Recovery

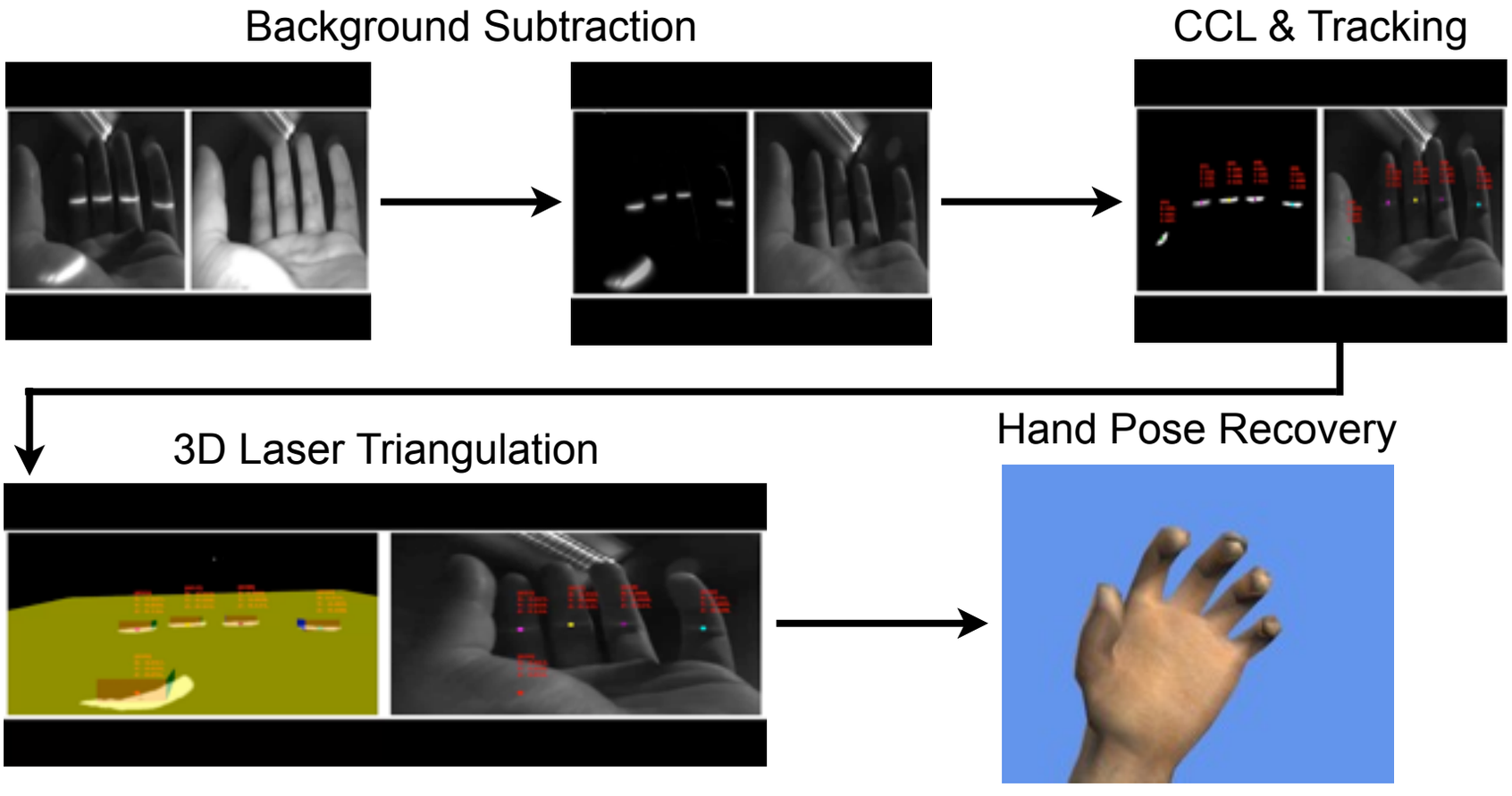
We use a number of image processing techniques to segment and track five discrete points on the fingers

Knowing the camera and laser position we can triangulate 3D positions from this information

And finally use a kinematics model to recover the full hand configuration

(Kim et al, 2012)

Digits - Technology



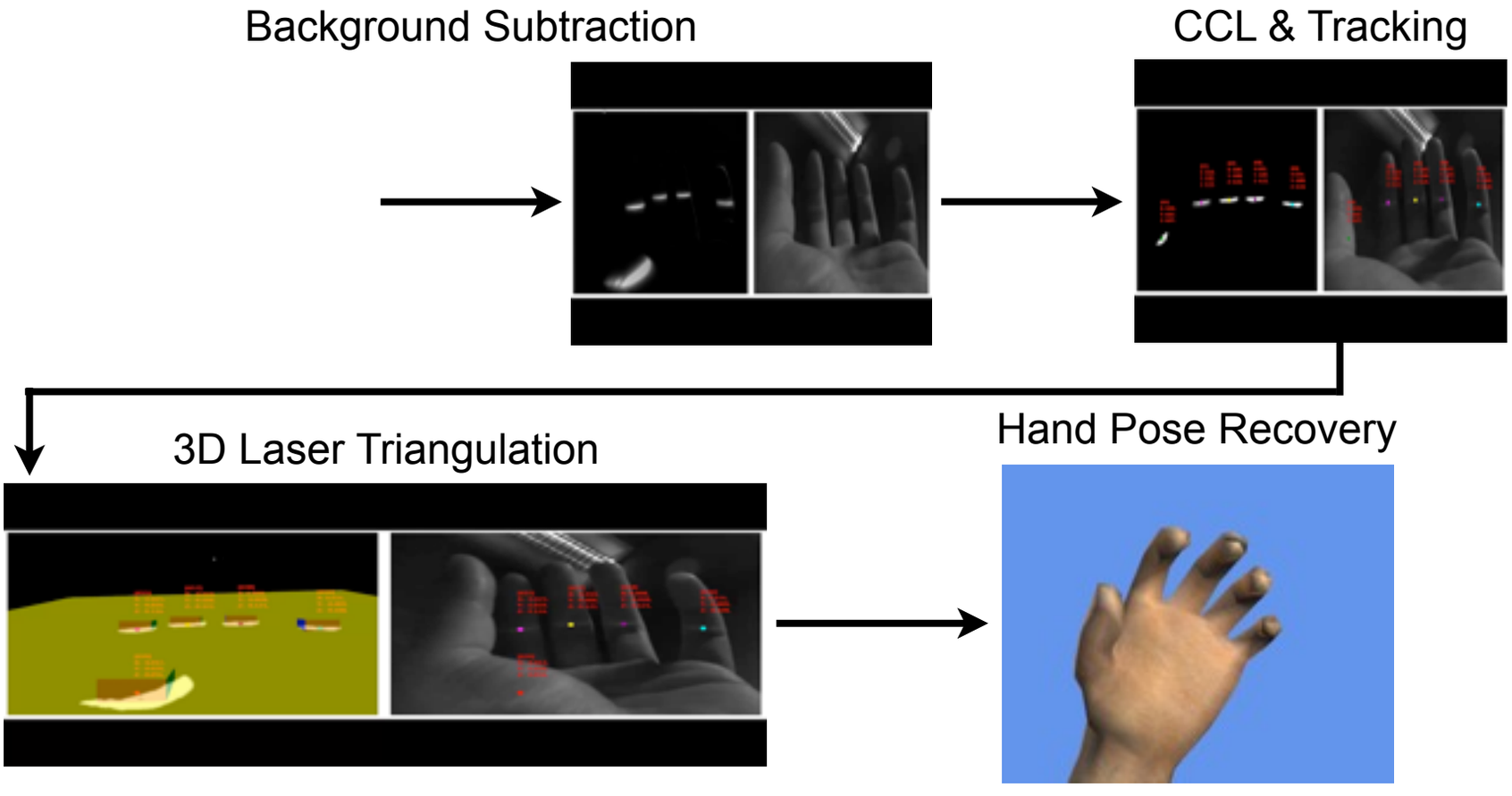
We use a number of image processing techniques to segment and track five discrete points on the fingers

Knowing the camera and laser position we can triangulate 3D positions from this information

And finally use a kinematics model to recover the full hand configuration

(Kim et al, 2012)

Digits - Technology



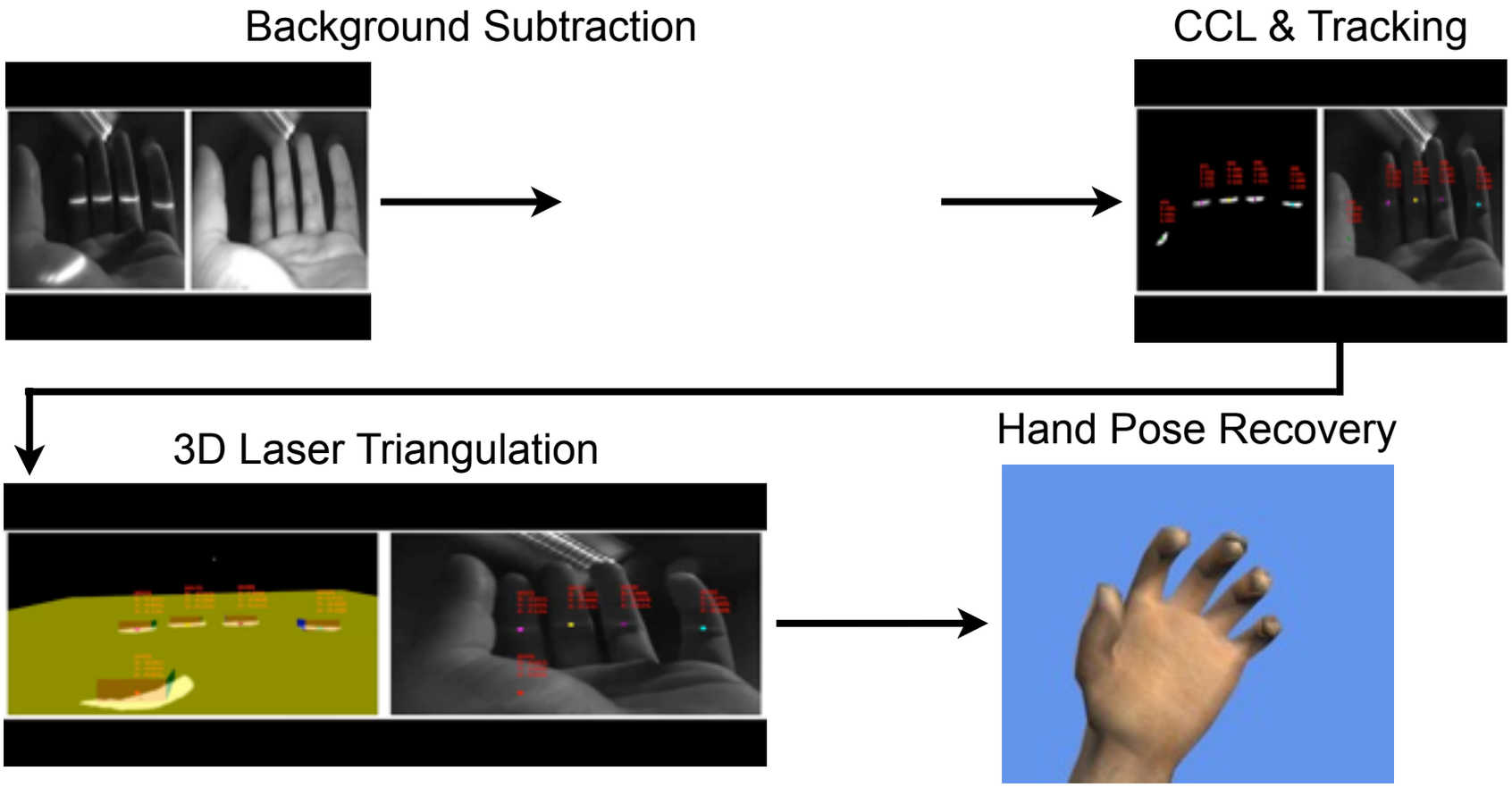
We use a number of image processing techniques to segment and track five discrete points on the fingers

Knowing the camera and laser position we can triangulate 3D positions from this information

And finally use a kinematics model to recover the full hand configuration

(Kim et al, 2012)

Digits - Technology



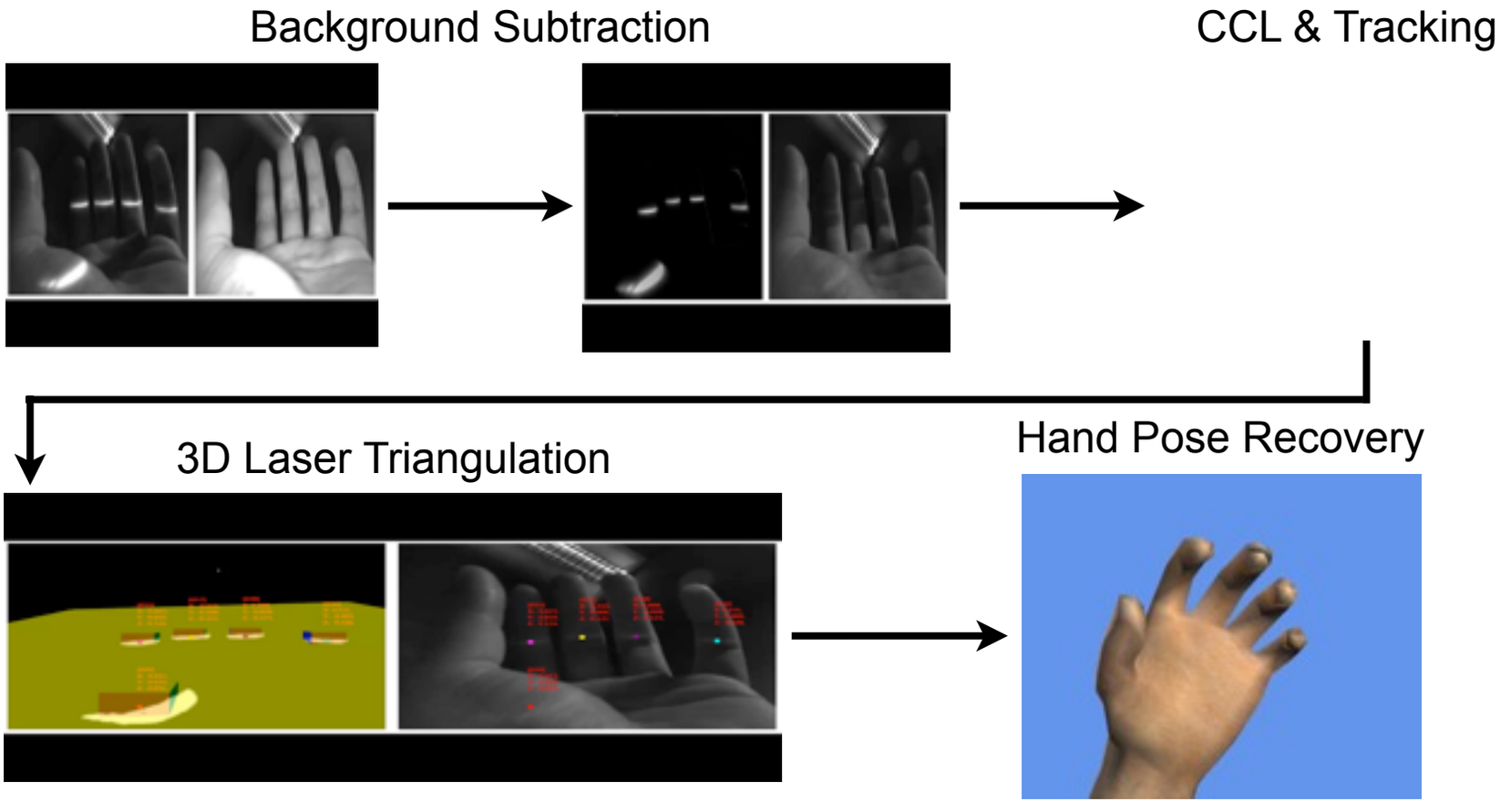
We use a number of image processing techniques to segment and track five discrete points on the fingers

Knowing the camera and laser position we can triangulate 3D positions from this information

And finally use a kinematics model to recover the full hand configuration

(Kim et al, 2012)

Digits - Technology



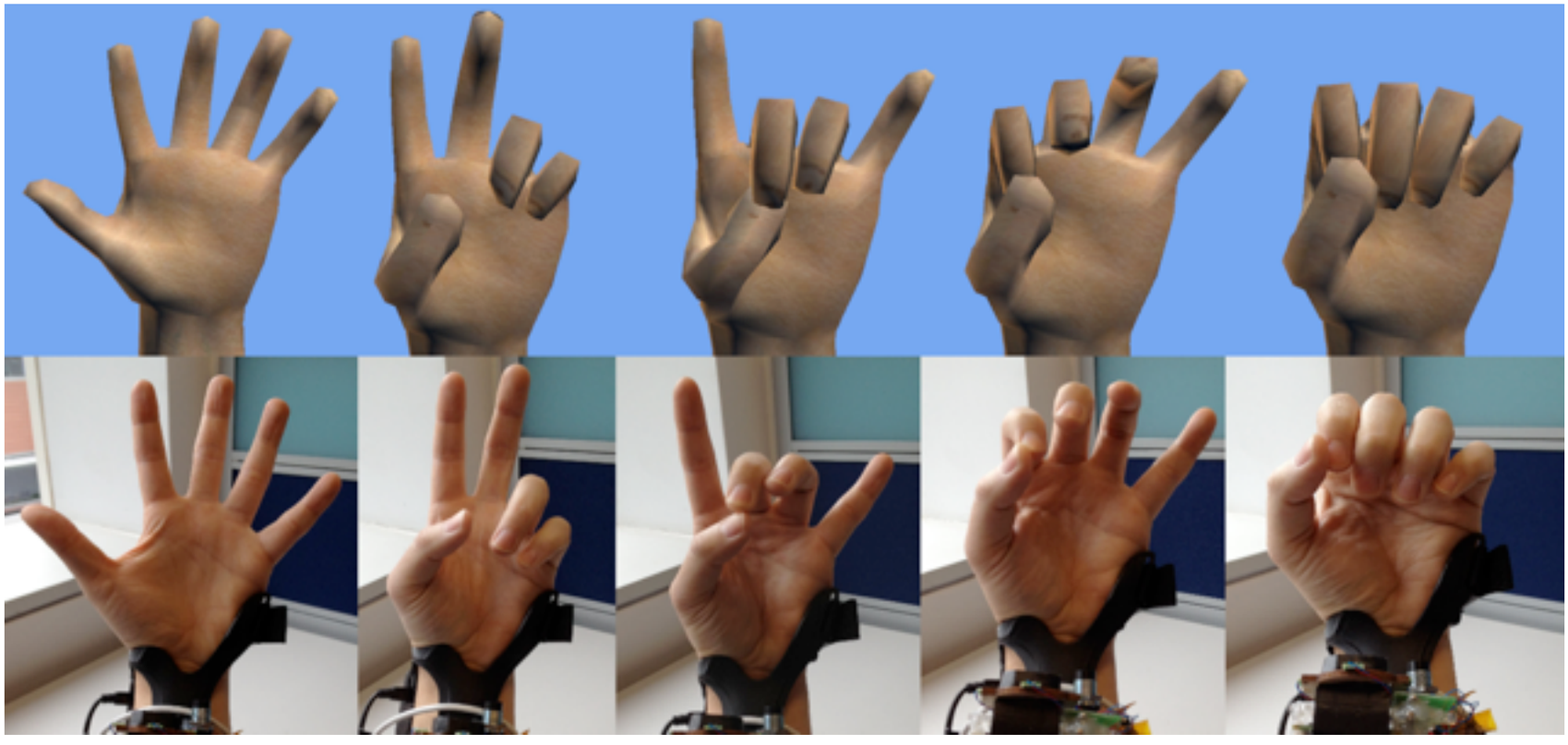
We use a number of image processing techniques to segment and track five discrete points on the fingers

Knowing the camera and laser position we can triangulate 3D positions from this information

And finally use a kinematics model to recover the full hand configuration

(Kim et al, 2012)

Digits - Examples



accurate

(Kim et al, 2012)

Digits - Demo

shooting
grabbing
pulling

(Kim et al, 2012)

Digits - Demo

shooting
grabbing
pulling

(Kim et al, 2012)

Digits

- Pro
  - Portable
  - Intern processing
  - Accurate replacement for data glove
- Contra
  - As obtrusive as a data glove
  - Occlusion is major problem

Towards bimanual gestures

previous papers all tried to reconstruct a model of the hand in a more or less accurate fashion
In the next paper we will see a move away from reconstruction towards using the second hand for input and the first hand as a trigger

(Kim et al, 2007)

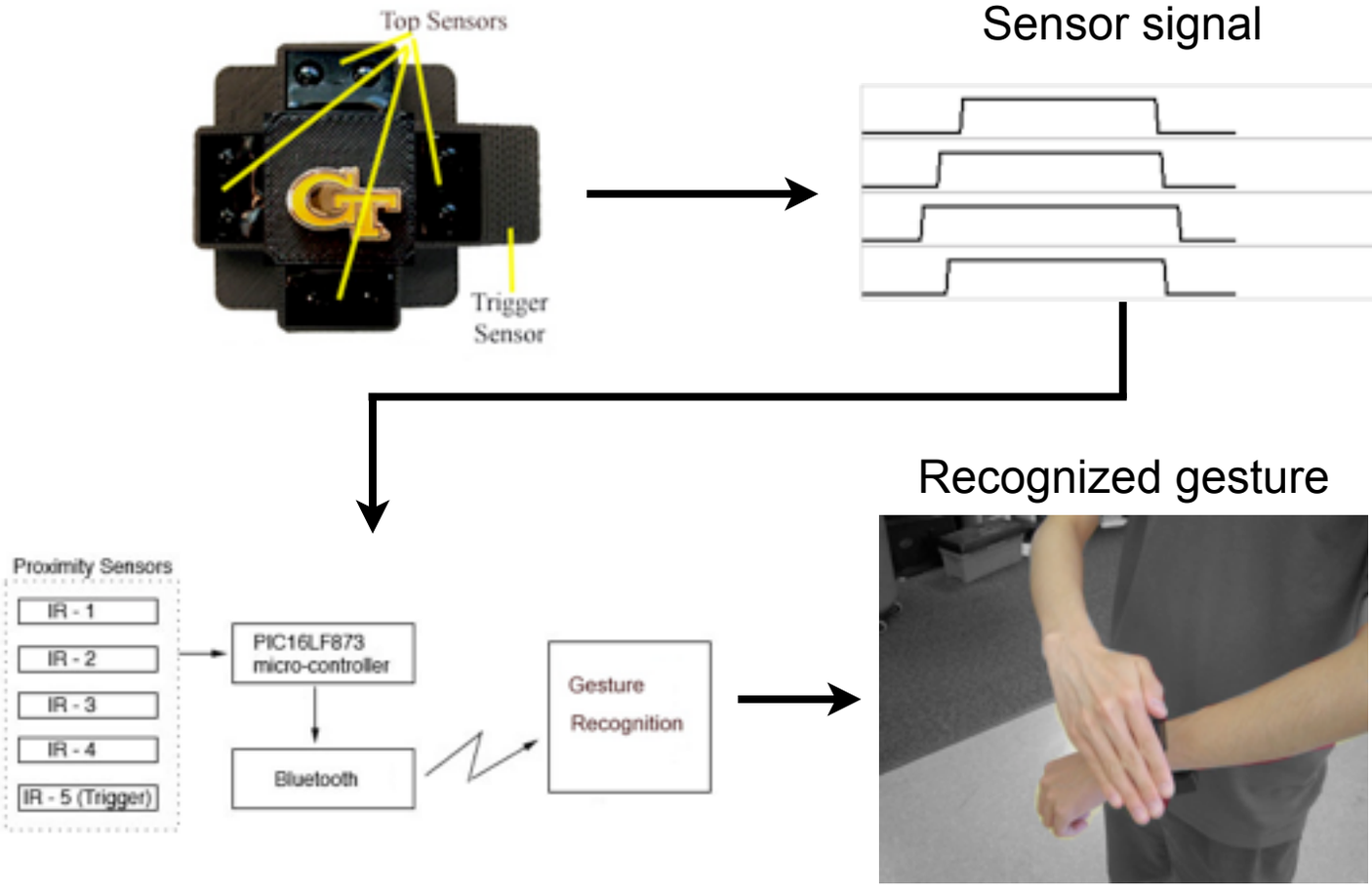
Gesture Watch

- Contact free interface
- Unobtrusive

device recognizing other hand
wearing arm used to initiate gesture

(Kim et al, 2007)

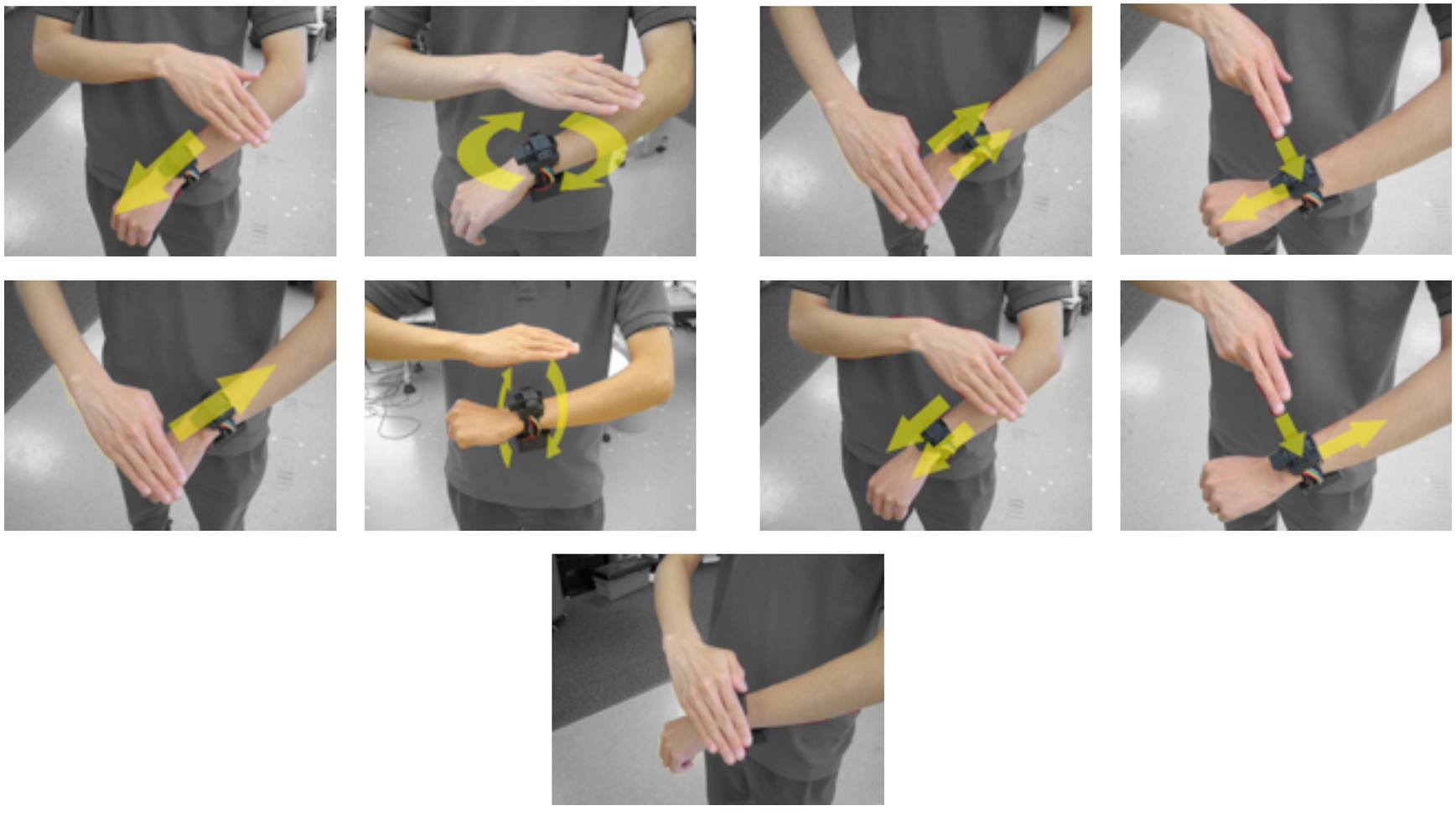
Gesture Watch - Technology



4 proximity Sensors arranged in a cross
+ 1 for initiating towards the hand
binary 0/1 sensors

(Kim et al, 2007)

Gesture Watch - Examples



proposed gestures

(Kim et al, 2007)

Gesture Watch

- Pro
  - Unobtrusive desgin
  - Sensors embedded
  - Contact free
  - Private
- Contra
  - Requires action from second hand to start gesture

private by hiding the gesture from other people

What if we could eliminate all instrumentation?

But still, instrumentation of the user is required
To get hands free
To be cheaper

(Gupta et al, 2012)

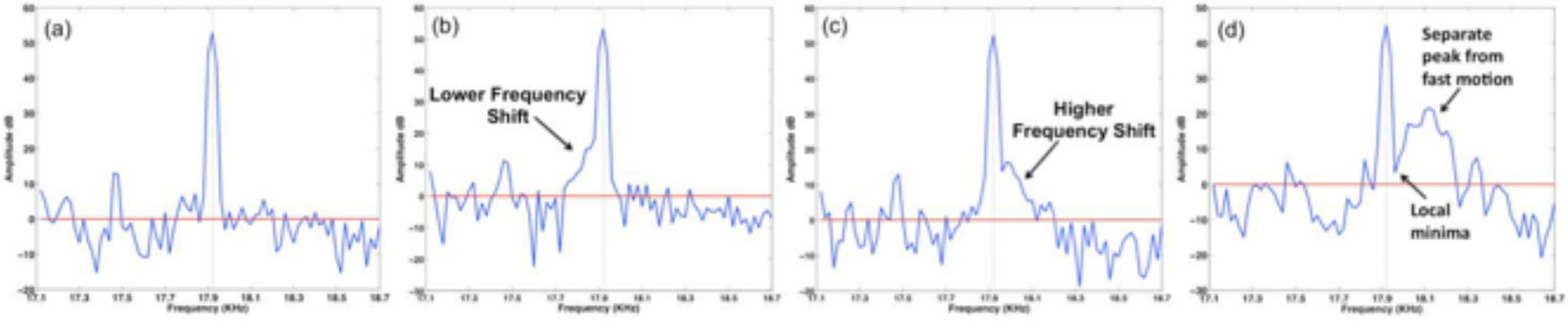
Sound Wave

- No instrumentation of user
- Reusing existing hardware

Reuses speakers and microphone from an existing laptop

(Gupta et al, 2012)

Sound Wave - Technology



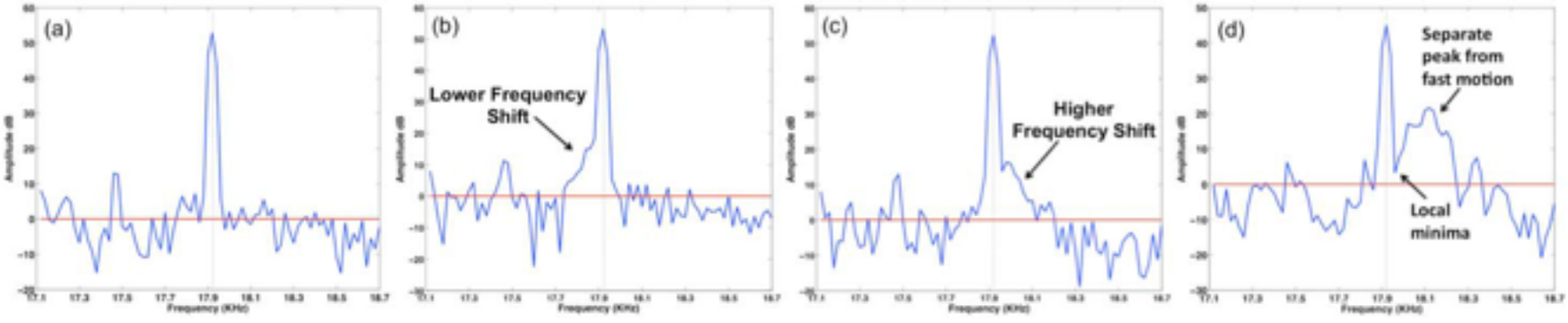
$$f_r = f_t \cdot \left(\frac{c + v}{c - v}\right)$$

$where, f_r = perceived\ frequency\ at\ microphone;$
$f_t = original\ frequency\ from\ speaker;$
$c = speed\ of\ sound\ in\ air;$
$v = velocity\ of\ target/hand$

Doppler effect
Emitted sound 18 – 22 kHz
Input sampled -> FFT
22.05kHz spectrum divided into 33 bins
scanned until amplitude drops below 10%
second scan until 30% away from pilot tone

(Gupta et al, 2012)

Sound Wave - Technology



$$f_r = f_t \cdot \left(\frac{c + v}{c - v}\right)$$

$where, f_r = perceived\ frequency\ at\ microphone;$
$f_t = original\ frequency\ from\ speaker;$
$c = speed\ of\ sound\ in\ air;$
$v = velocity\ of\ target/hand$

Doppler effect
Emitted sound 18 – 22 kHz
Input sampled -> FFT
22.05kHz spectrum divided into 33 bins
scanned until amplitude drops below 10%
second scan until 30% away from pilot tone

(Gupta et al, 2012)

Sound Wave - Demo

Wake up and sleep automatically
control media player

(Gupta et al, 2012)

Sound Wave - Demo

Wake up and sleep automatically
control media player

(Gupta et al, 2012)

Sound Wave

- Pro
  - No instrumentation of user
  - Accurate results
  - Even in noisy environments
- Contra
  - Base tone may be hearable

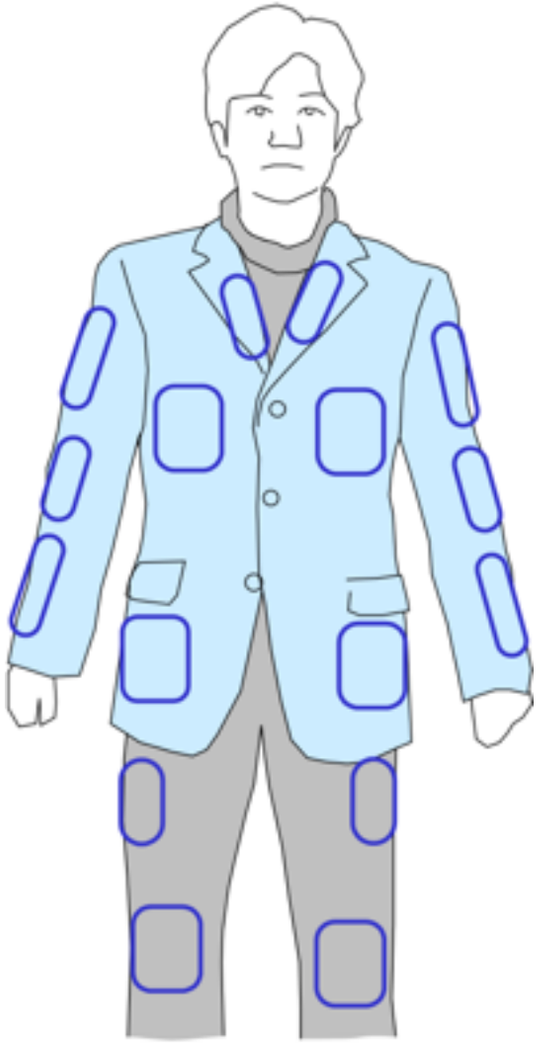
All sensors need a network

To conclude we have a look at a completely different paper that discusses how the body itself can be used as a network for communication

(Rekimoto, 2001)

Gesture Pad

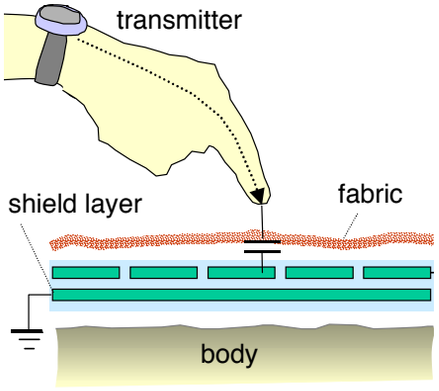
- The body as touch interface
- The body as network
- The body as transceiver

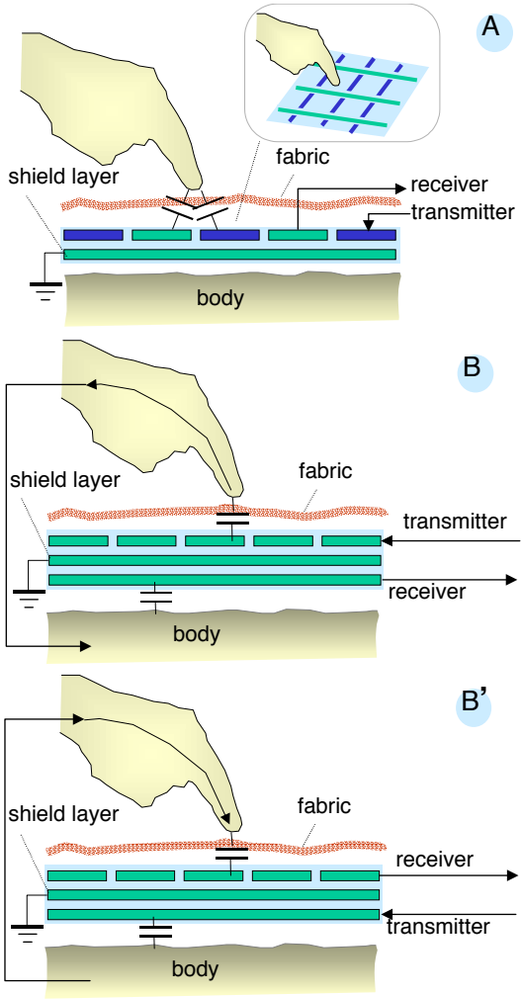


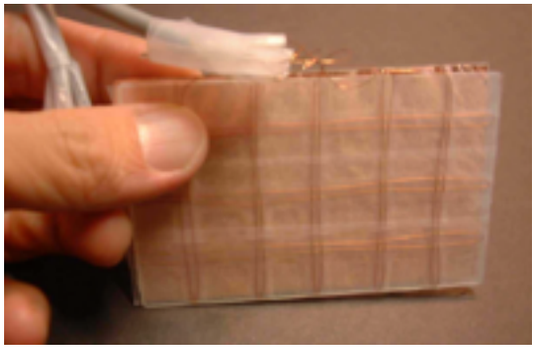
Taken from the paper of Gesture Wrist, the capacitance sensing wrist sensor
Communicate between themselfes
Send data to (touched) outside world
Humantenna inverted

(Rekimoto, 2001)

Gesture Pad







A: Transmitter/receiver multiplexed
B: Shield layer separates transmitter from receiver

(Rekimoto, 2001)

Gesture Pad

- Further Ideas
  - Use NFC transceivers inside pads
  - Identify person touching by there signal

Comparison

	Mobility	Accuracy	Instrumentation	Main Application
Muscle Computer Interface	Designed for mobile use, data sent via wifi/BT	65% busy hand, no feedback, 4 fingers 91% busy hand, feedback, 3 fingers	An arm band at the upper forearm	Gesture recognition with busy hands
Gesture Wrist (Capacity sensing)	Designed for mobile use, data sent via body network	N/A	Wrist watch like utility	Hand shape recognition, authentication
Wrist Shape (Photosensors)	Designed for mobile use, offline processing atm.	45-48%	Wrist watch like utility	Hand shape recognition
Digits (3D reconstruction)	Designed for mobile use, data sent via wifi/BT	91%, varying from finger to finger	Small camera worn at a wrist band	Reconstructing 3D model of hand
Gesture Watch (in air over hand)	Designed for mobile use, data sent via wifi/BT	95 %	Wrist watch like utility	Simple gesture recognition using one hand
Sound Wave (in air over laptop)	Bound to Laptop	90-95%	None, using existing hardware	Add simple gesture recognition to laptop

Different aspect that would maybe required from a gesture based interface

Summary and Future Technology

- Today
  - Gesture recognition is feasible
  - Ranging accuracy
  - Integration is still complicated

- In the future we ...
  - need to control unobtrusively
  - can authenticate with an accessory
  - wear touchable cloth
  - use the body as a network

Vaporware!?
commercial from myo
foresight of how gesture interaction could look like

Vaporware!?
commercial from myo
foresight of how gesture interaction could look like

“Any sufficiently advanced technology is indistinguishable from magic.”

Arthur C. Clarke