

Introduzione alla Realtà Virtuale Parte I

Prof. Alberto Borghese



1/72

<http://borghese.di.unimi.it/>



Sommario



- **Introduzione**
- Sistemi di Input
- Generatori di mondi
- Motore di calcolo
- Sistemi di Output
- Conclusioni

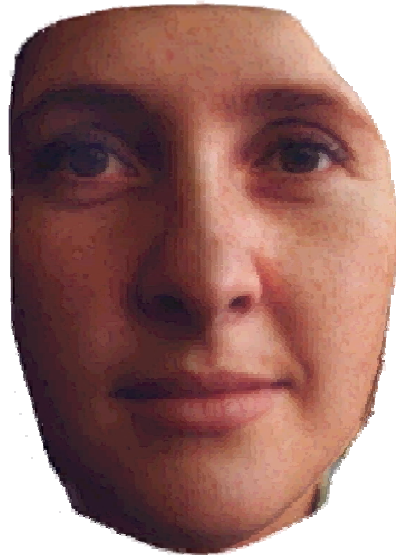
A.A. 2012-2013

2/72

<http://borghese.di.unimi.it/>



Which is real, which is virtual?



A.A. 2012-2013

3/72

<http://borghese.di.unimi.it/>



Historical Perspective (I)



- The name “Virtual Reality” has been attributed to Jaron Lanier (VPL), 1986.
- *Virtual Worlds or Synthetic Environments*
- *Philosophical and Technological* origin.

Philosophical background

Ontology and Gnoseology.

- Plato (world of the ideas) 428-348 a.C.
- Berkeley (sensorial experience is too limited) 1685-1753.
- Hegel (“what is rational is real..”) 1770-1831.
- New age.

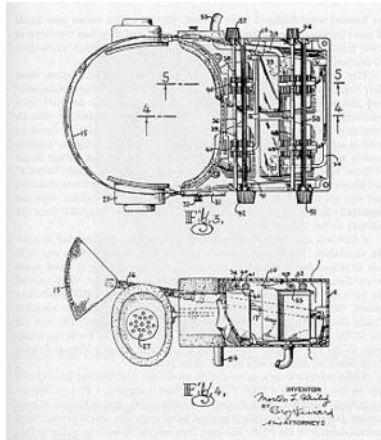
A.A. 2012-2013

4/72

<http://borghese.di.unimi.it/>



Historical Perspective (II)



Morton Heilig 1956,
patented in 1961
Non fu mai costruito

projected film,
audio, vibration,
wind, odors.

A.A. 2012-2013

5/72

<http://borghese.di.unimi.it/>



Historical Perspective (III)



Technological background

- *Philco HMD, 1961.*
- *"Ultimate display", Sutherland, 1970.*
- *Data Glove, VPL Research, 1988.*



Sutherland, Ivan E. 1968. "A Head-Mounted Three Dimensional Display," pp. 757-764 in Proceedings of the Fall Joint Computer Conference. AFIPS Press, Montvale, N.J.

A.A. 2012-2013

6/72

<http://borghese.di.unimi.it/>



Virtual Reality Systems



Key characteristics are:

- Immersivity.
- Interactivity.

VR should be able to stimulate the human sensorial systems
In a coordinated way.



A typical VR system



VR systems are constituted of:

- *Input systems* (measure the position *in* the environment and force *over* the environment).
- *World generators* (provides a realistic virtual world in which to act).
- *Computational engine* (computes the output, given the input and the virtual world).
- *Output systems* (outputs sensorial stimuli *on* the subject. Vision, sound, force ... are generated as if they were provided *by* the virtual environment).



Sommario



- Introduzione
- **Sistemi di Input (trackers)**
- Generatori di mondi
- Motore di calcolo
- Sistemi di Output
- Conclusioni



Input systems



Measure human actions on the virtual environment.

- **Position.** Measure the position of the body segments inside the virtual environment.
- **Force.** Measure the force exerted by the body segments when in contact with a virtual object.
- Estimate the motor output of the human muscle-skeleton system.



Tracking systems



- Measure the position of the body segments inside the virtual environment.
- **Motion capture** (batch, complete information on the movement).
- **Real-time trackers** (real-time position of the body).
- **Gloves** (specialized for hands).
- **Gaze trackers**.

Adopted technology

- Optoelectronics (video-camera based)
 - Marker based
 - Computer vision
 - Scanner based.
- Magnetical
- Acoustical
- Mechanical
- Inertial



What is motion capture?



Ensemble of techniques and methodologies to acquire **automatically** the motion of the objects of interest.

Characteristics: sampling rate, accuracy, 2D/3D, real-time, motion amplitude, invasivity,....

Technology: opto-electronical, magnetical, ultrasound, inertial

Specific body parts: gloves, gaze trackers....

Applications are increasing (medical applications at the origin, now interest in the entertainment, robotics, reverse engineering ...)



Motion Capture and Synthesis

Reproduce digitally the motion of the body (in real-time in case of tracker).

Time series of the position of the body segments
or
Time series of the motion of the articulations.

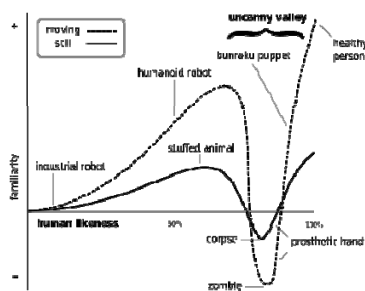
Analysis
Info extraction

Application of the time series to a
3D digital model of the body.

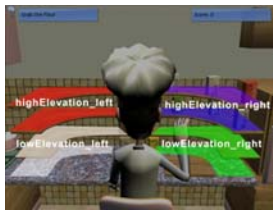
Synthesis
Avatar animation



Avatar designed avoiding the "uncanny" valley



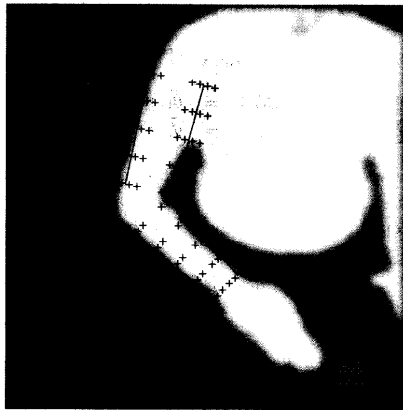
Mori, Masahiro (1970). Bukimi no tani The uncanny valley (K. F. MacDorman & T. Minato, Trans.). Energy, 7(4), 33-35. (Originally in Japanese)





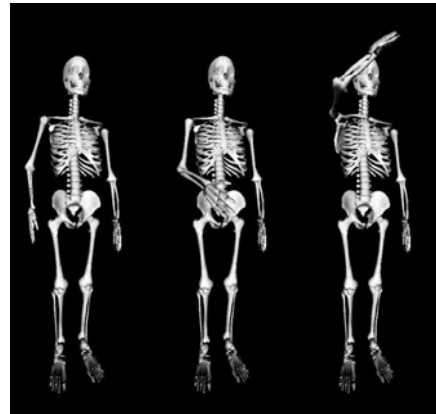
What is captured?

Silhouette (-> Skeleton)



Computer vision techniques
(silhouette)

Skeleton



Bony segments or articulations
(marker-based systems / Kinect)

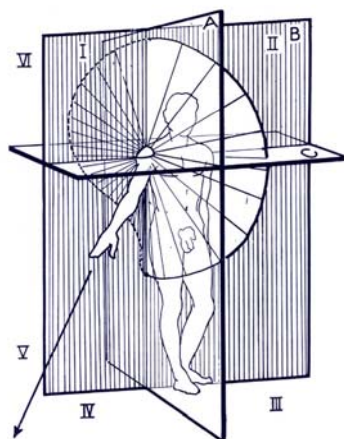
A.A. 2011-2012

15/71

<http://homes.dsi.unimi.it/~borghese/>



Description of the human skeleton



- A – Frontal plane
- B – Sagittal plane
- C – Horizontal plane

- Abduction/adduction
- Flexion/extension
- Axial rotation (V)
- Quaternions for 3D rotations

3D position of joint extremes

Definition of the interesting degrees of freedom.

A.A. 2011-2012

16/71

<http://homes.dsi.unimi.it/~borghese/>

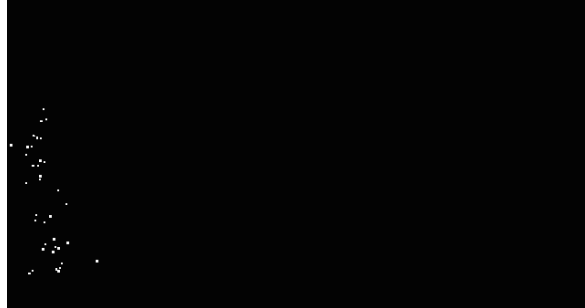


Tracking difficulties



It is a complex problem because:

- Dense set of markers. These may come very close one to the other in certain instants.
- Motion can be easily complex, as it involves rotation and twists of the different body parts (thing at a gymnastic movement).
- Multi-camera information and temporal information is required to achieve a robust tracking.



Motion Capture through markers



<http://www.vicon.com/applications/games.html>

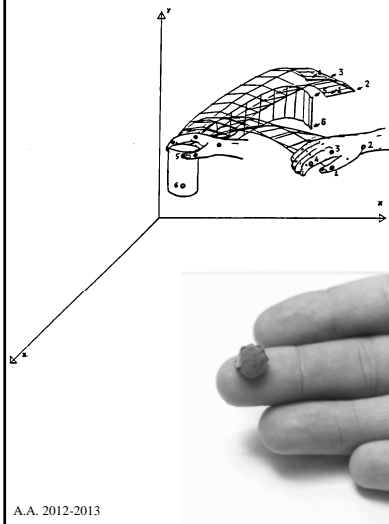
Vicon system from Oxford Metrix



Why passive markers?



Minimum encumbrance on the subject: markers do not require any powering and are hardly sensed by the subjects.



No constraint on the dimension of the working volume is prescribed.



A.A. 2012-2013

<http://borghese.di.unimi.it/>



How passive markers work?



Passive markers are constituted of a small plastic support covered with retro-reflecting material (3M™). It marks a certain repere point.



Video-cameras are equipped with a co-axial flash.

Markers appear much brighter than the background making their detection, on the video images, easier.

A.A. 2012-2013

20/72

<http://borghese.di.unimi.it/>



Sequential processing



1. Surveying the image of the moving subject on multiple cameras (*frequency & set-up*).
2. Markers extraction from the background scene (*accuracy & reliability*).
3. Computation of the “real” 2D position of the markers (*accuracy <- distortion*).

Low-level
Vision

4. **Matching on multiple cameras.**
5. 3D Reconstruction (*accuracy*).
6. **Model fitting (*labelling, classification*).**

High-level
Vision

An implicit step is CALIBRATION.

A.A. 2012-2013

21/72

<http://borghese.di.unimi.it/>

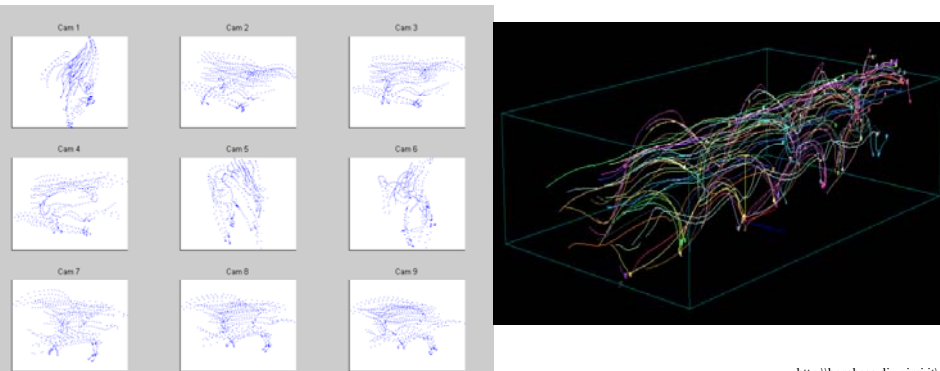


Disadvantages of motion capture systems based on passive markers



When a marker is hidden to the cameras by another body part (e.g. the arm which swings over the hip during gait), the motion capture loses track of it.

The multiple set of 2D data have to be correctly labeled and associated to their corresponding 3D markers.



<http://borghese.di.unimi.it/>

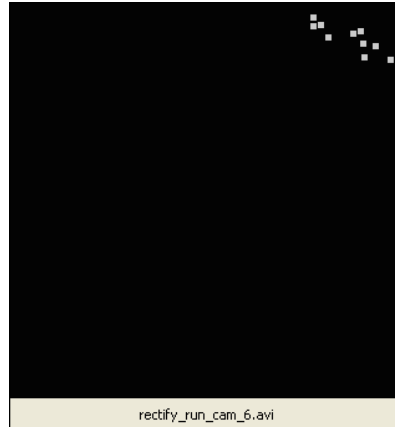


Tracking difficulties



It is a complex problem because:

- Dense set of markers. These may come very close one to the other in certain instants.



- Motion can be easily complex, as it involves rotation and twists of the different body parts (thing at a gymnastic movement).
- Multi-camera information and temporal information is required to achieve a robust tracking.



Tracking processing steps





ACQUISITION OF 2D POINTS

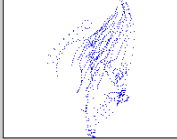

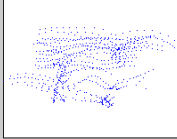
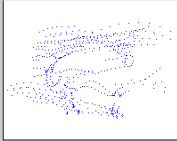
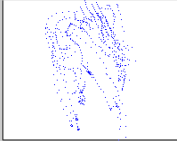



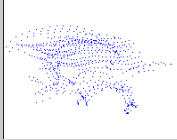
TRACKING:

- 1) From 2D points to 2D strings.
- 2) Pairing 2D strings with the epipolar constraint to create 3D strings.
- 3) Condensation of 3D strings.
- 4) Joining 3D strings.



RECTIFY:





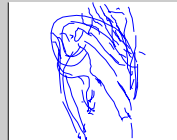




- 5) Classification of 3D strings according to the markers arrangement.
- 6) Estimate of the 3D model of the subject from the strings data.
- 7) Estensione automatica della classificazione alle altre stringhe.

 **2D tracking** 

Cam 1 	Cam 2 	Cam 3 
Cam 4 	Cam 5 	Cam 6 
Cam 7 	Cam 8 	Cam 9 

A.A. 2012-2013 25/72 <http://borghese.di.unimi.it/>

 **1) Creation of 2D strings** 

Cam 1 	Cam 2 	Cam 3 
Cam 4 	Cam 5 	Cam 6 
Cam 7 	Cam 8 	Cam 9 

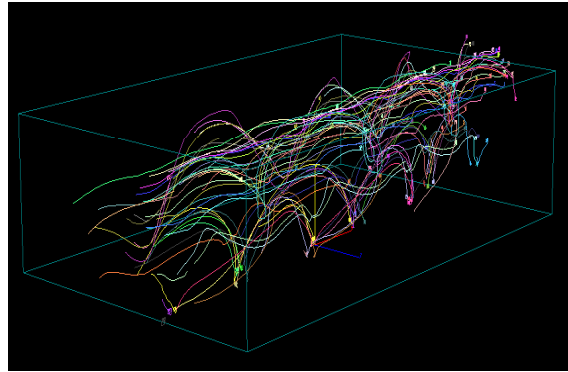
A.A. 2012-2013 26/72 <http://borghese.di.unimi.it/>



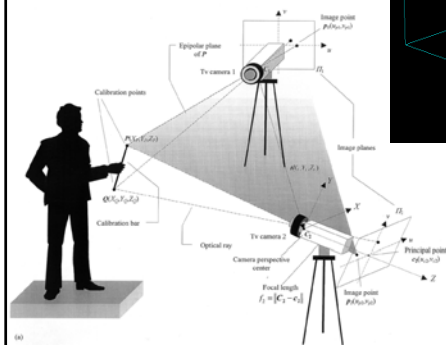
2) Matching 2D strings



Epipolarity constraint



3D strings



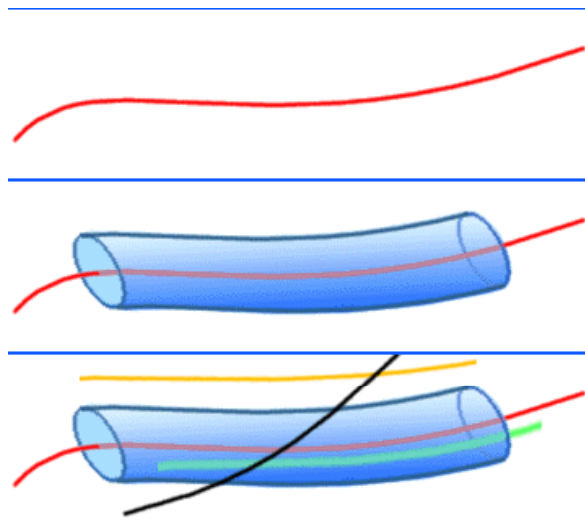
A.A. 2012-2013

27/72

<http://borghese.di.unimi.it/>



3) Condensation of 3D strings



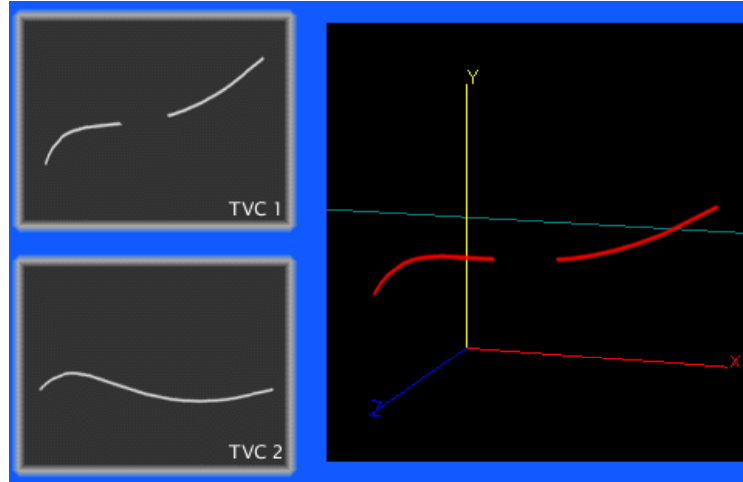
A.A. 2012-2013

28/72

<http://borghese.di.unimi.it/>



4) Joining 3D strings



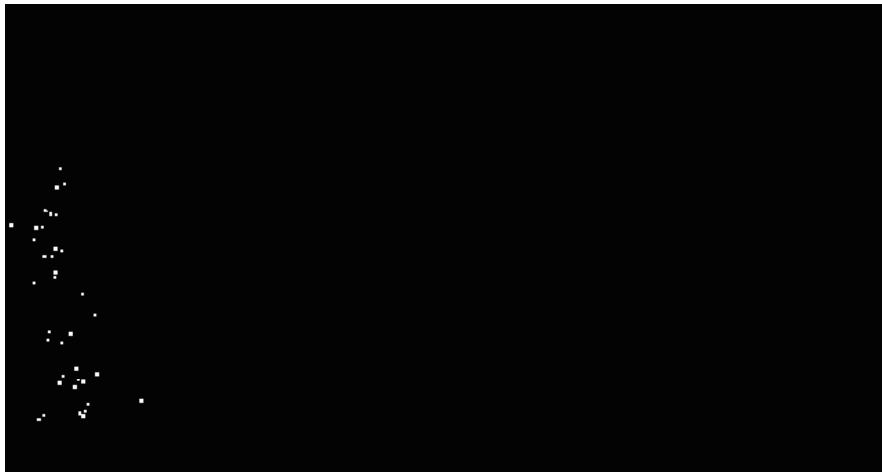
A.A. 2012-2013

29/72

<http://borghese.di.unimi.it/>



3D strings



3D strings already contain motion 3D information

A.A. 2012-2013

30/72

<http://borghese.di.unimi.it/>

3D strings

string3d_dynamic.avi

A.A. 2012-2013 31/72 http://borghese.di.unimi.it/

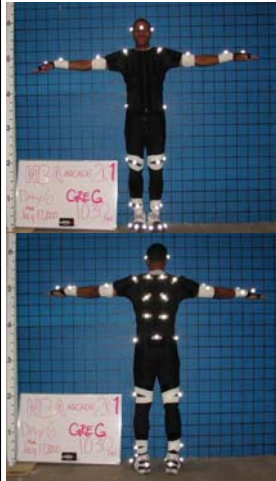
Model fitting

Internal model Reference model

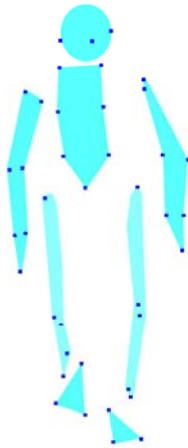
A.A. 2012-2013 32/72 http://borghese.di.unimi.it/



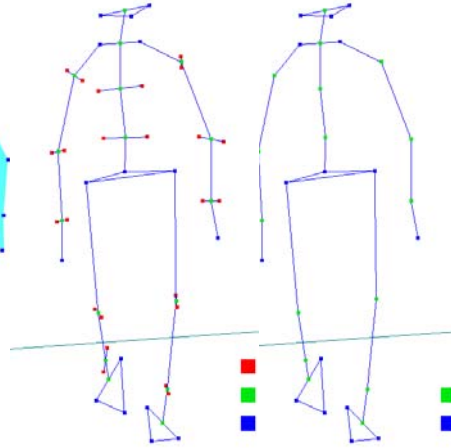
What a model represents?



Markered subject



Modello 3D



Modello a stick

Modello hidden

A.A. 2012-2013

33/72

<http://borghese.di.unimi.it/>



Problems intrinsic in body tracking



- Joints are points inside the body, markers are attached on the body surface.
- Joint are not fixed points: two adjacent bones rotate and slide.
- Joint are not spherical.
- Joints can be complex (e.g. Shoulder, spine)
- Skin artifacts.

A.A. 2012-2013

34/72

<http://borghese.di.unimi.it/>



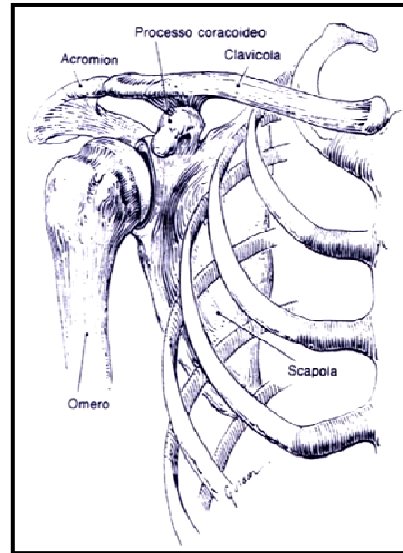
The human skeleton has complex articulations



“Rigid” bones connected. Tendons keep the bones in place.

Motion allowed can be very complex (e.g. shoulder, spine).

The reconstruction of the finest details of the motion are beyond reach, simplifying assumptions are made => ***Level of detail*** in motion analysis



A.A. 2012-2013

35/72



Video by Superfluo



superfluo3.wmv

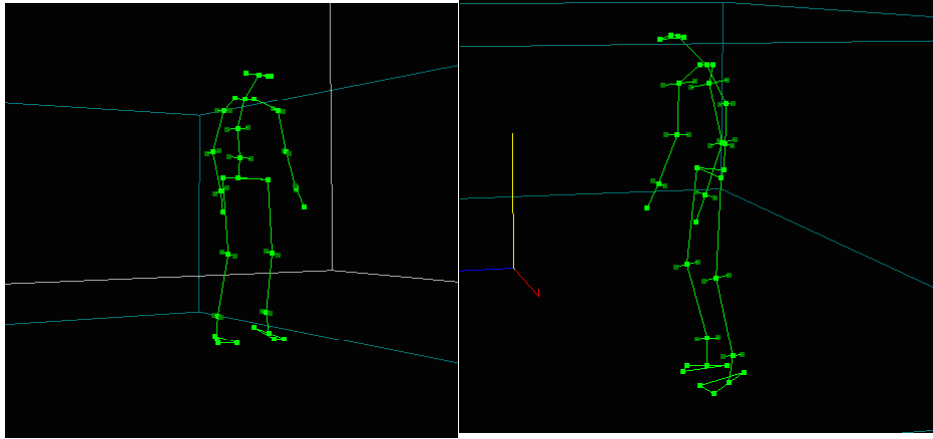
A.A. 2012-2013

36/72

<http://borghese.di.unimi.it/>



Risultati: escape



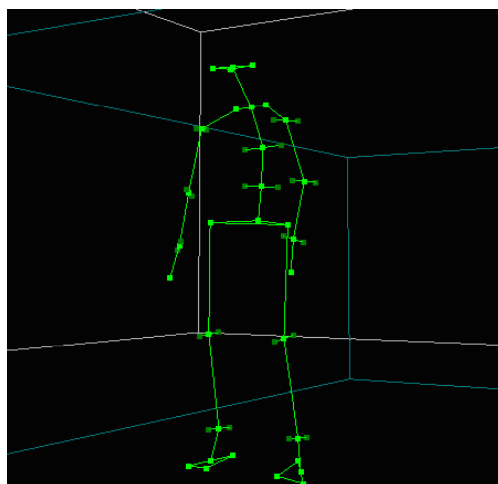
A.A. 2012-2013

37/72

<http://borghese.di.unimi.it/>



Risultati: fall_run



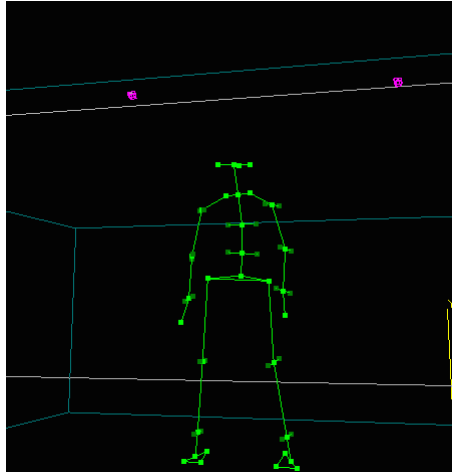
A.A. 2012-2013

38/72

<http://borghese.di.unimi.it/>



Risultati: roll



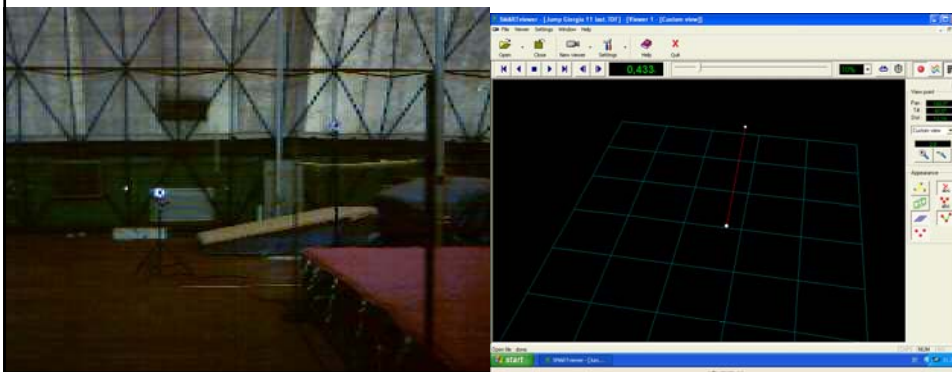
A.A. 2012-2013

39/72

<http://borghese.di.unimi.it/>



High jump – top athletes



A.A. 2012-2013

40/72

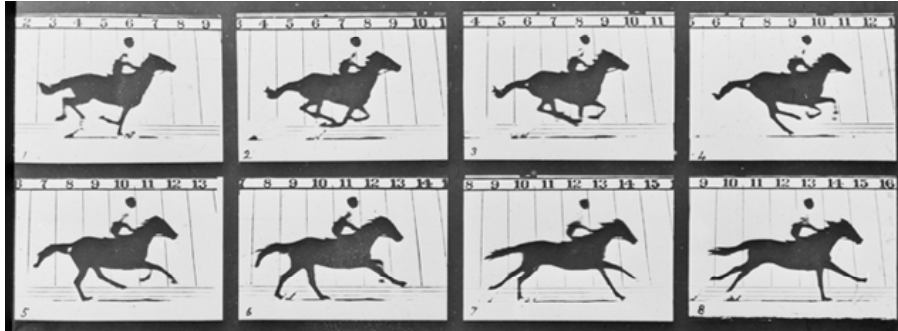
<http://borghese.di.unimi.it/>



Can we work without markers?



Eadweard Muybridge 1878-1901



<http://www.eadweardmuybridge.co.uk/>

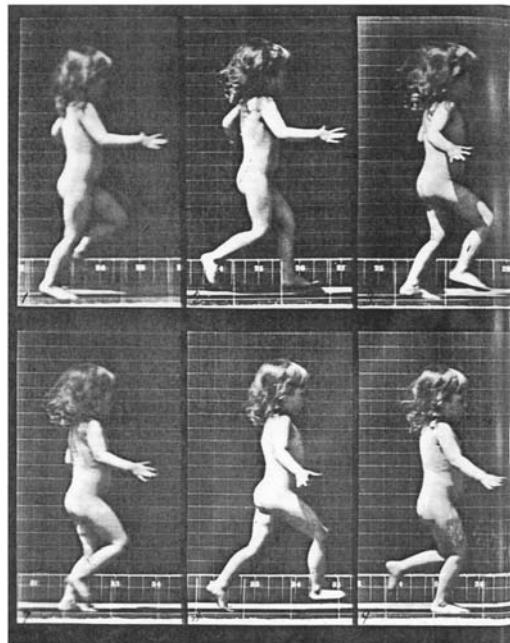
A.A. 2012-2013

41/72

<http://borghese.di.unimi.it/>



Edward Muybridge 1878-1901



A.A. 2012-2013

42/72

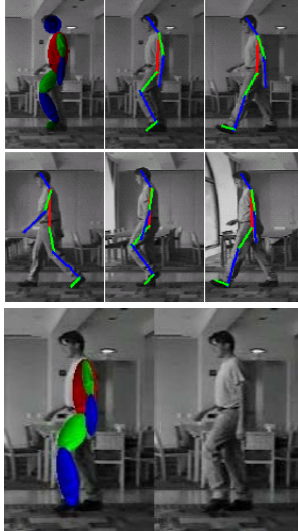
<http://borghese.di.unimi.it/>



Computer vision techniques



Silhouette (-> Skeleton)



Set of difficult problems:

2D Image processing (silhouette identification, optical flow detectors...)

Multi-view invariants.

Smooth motion -> temporal filtering.

Skeleton fitting (different rigid motion for different segments).

3D cameras help a lot

<http://movement.stanford.edu/>

A.A. 2012-2013

43/72

<http://borghese.di.unimi.it/>



2D tracking for rehabilitation



Duck-neglect project <http://borghese.dsi.unimi.it/Research/LinesResearch/Virtual/Virtual.html>

"Magic mirror" paradigm in which video of the player is overlaid with graphics generated by the computer.



Background measurement. Thresholding.

Alternative is the difference between consecutive images (glaring and blurring require some filtering).

duckNeglet_video_v0.mov

A.A. 2012-2013

44/72

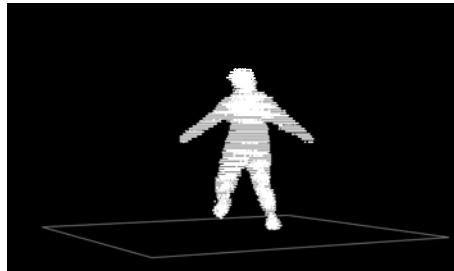
<http://borghese.di.unimi.it/>



Results: stepping (640 x 480, 10Hz)



Mikic, Trivedi, Hunter



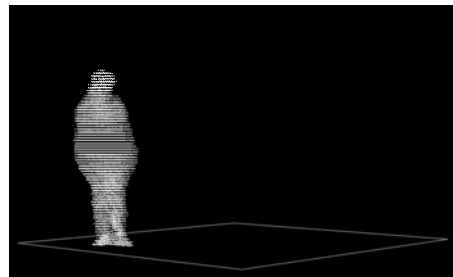
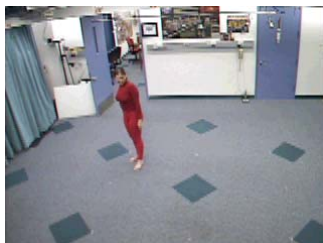
A.A. 2012-2013

45/72

<http://borghese.di.unimi.it/>



Results: cartwheel



A.A. 2012-2013

46/72

<http://borghese.di.unimi.it/>



Body motion from footage (Structure from Motion)

2 approcci:

- Probabilistico. Stima di un modello parametrizzato e dei parametri di movimento.
- Deterministico. Definisco un modello a-priori e stimo i parametri della camera e del movimento.



A.A. 2012-2013

47/72

<http://borghese.di.unimi.it/>



2D collision detection

- Collision detection with target can be checked by analyzing the overlapping between part of the motion mask only in particular regions.
- Identification of the motion mask as the outermost part of the body. Approximated collision detection defining general shapes.

Correct Hand collision area
(most left pixel in the area around first top most high pixel)

- Collision with targets gives hit, collision with distractors gives a miss.
- Same principles implemented with Sony EyeToy Webcam (2003).



A.A. 2012-2013

48/72

di.unimi.it/



2D color coded tracking

- Players could interact with a 3D scene by moving known brightly saturated colored objects that were visually tracked in PlayStation 2 (EyeToy Webcam). Threshold on color representation.
- Pose recovery can be accomplished robustly for certain shapes of known physical dimensions by measuring the statistical properties of the shape's 2D projection. In this manner, for a sphere the 3D position can be recovered (but no orientation), and for a cylinder, the 3D position and a portion of the orientation can be recovered.



- Multiple objects can also be combined for complete 3D pose recovery, though occlusion issues arise.
- Perfect recognition in all lighting conditions is difficult.

A.A. 2012-2013

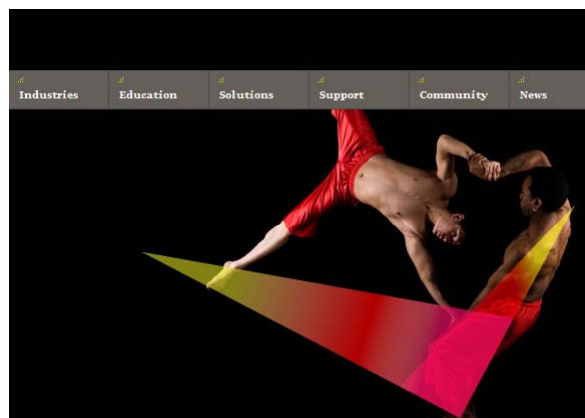
49/72

<http://borghese.di.unimi.it/>



Markerless optical motion capture

- Clustering and labeling of each image with a probabilistic framework.
- Addition of temporal and spatial constraints.
- Special stage place is required.
- Cost and complexity



www.organicmotion.com

A.A. 2012-2013

50/72

<http://borghese.di.unimi.it/>

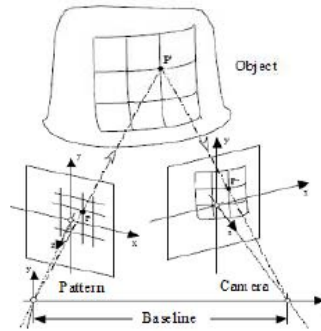


2.5D cameras (Kinect)

- 3D scanner with active pattern (IR)
- RGB camera
- Robust background separation
- Robust skeletal tracking

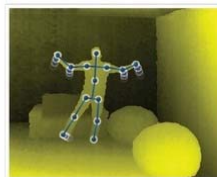
Used as a Web-cam with advanced silhouette subtraction for rehabilitation.

Come to our lab and see...



2.5D First SDK for Kinect

Primesense drivers, with skeleton tracking: <http://www.primesense.com/se.com>



Color (RGB) Image



Depth Image



Audio Stream



Open Source drivers:

<http://openni.org>

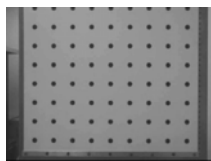
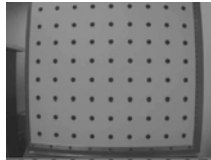
http://openkinect.org/wiki/Main_Page

Microsoft's SDK is available



Calibration is a pre-requisite

Camera calibration



Set-up calibration

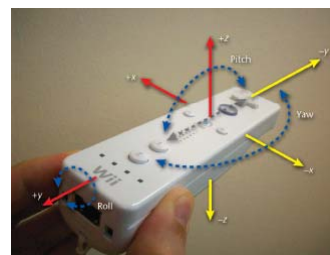
Excellent for special effects, not so good for measurements....
Cameras are not metric.



Inertial tracking::Wii

$$pitch = \arctan\left(\frac{a_z}{a_y}\right)$$

$$roll = \arctan\left(\frac{a_z}{a_x}\right)$$



Positional data are obtained through integration.

⇒Instability. A flip of the LSB for one frame generates a rotation at constant speed!!

Other devices are required to stabilize the measurements: Nunchuk (gyroscope), sensor IR-bar



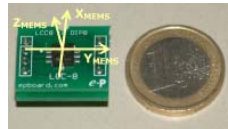


Inertial tracking::Xsens



- Xsens by Moven is a full-body, camera-less inertial motion capture (MoCap) solution. It is flexible motion capture system that can be used indoors or outdoors (on-set). With the short turnaround times MVN is a cost effective system with clean and smooth data.

- Costly



- We have used such system inside the FITREHAB project:

<http://www.innovation4welfare.eu/287/subprojects/fitrehab.html>



Where are we now (optoelectronic)?



Optotrak, 1991.

LED + cameras



- Measure the position of the joints.
- Time multiplexing for the markers (3 at 450Hz or 750Hz with additional hardware). No-tracking, real-time.
- Power for the LEDs has to be delivered on the subject's body (markers get hot on the skin!!).
- Accuracy 0.1mm (X,Y), 0.15mm (Z, depth).



Where are we now (magnetic)?



Magnetic technology: Fastrack & older Polhemus sensors.

They measure: pitch, yaw and roll; X, Y, Z of the segments.

Electro-magnetic induction.



The transmitter is a triad of electromagnetic coils, enclosed in a plastic shell, that emits the magnetic fields. The transmitter is the system's reference frame for receiver measurements.

The receiver is a small triad of electromagnetic coils, enclosed in a plastic shell, that detects the magnetic fields emitted by the transmitter. The receiver is a lightweight cube whose position and orientation are precisely measured as it is moved.

A.A. 2012-2013

57/72

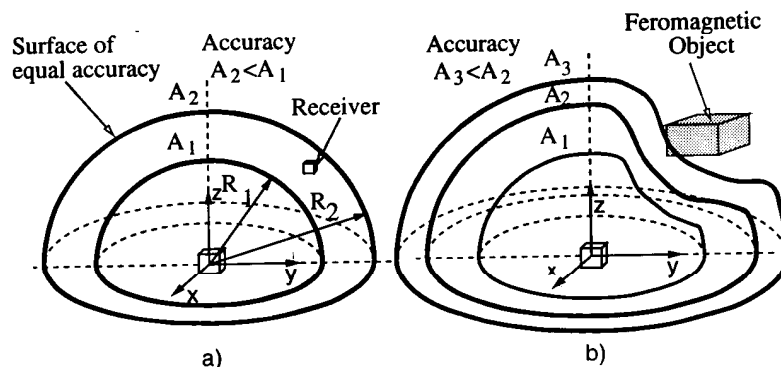
<http://borghese.di.unimi.it/>



Fast-track Motion Capture



- Higher accuracy through oversampling and DSP signal processing (0,5" and 1.8mm accuracy). Range of 75cm for high accuracy.
- Sensitive to ferromagnetic (metallic) objects.



- Latency: 4msec.
- Sampling rate: 120Hz. Rate drop with multiple receivers because of multiplexing.

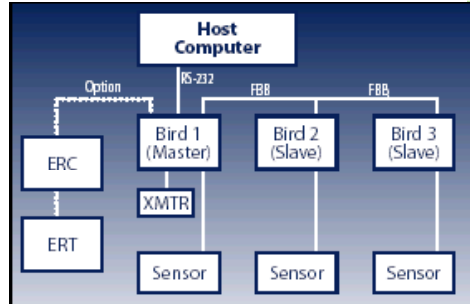
A.A. 2012-2013

58/72

<http://borghese.di.unimi.it/>



Flock of birds Motion Capture



- Each receiver has its own DSP.
- All the DSP are connected with a fast internal bus.
- Latency is increased (8ms).

When more than one transmitter is adopted (experimental):
larger field (single transmitter at a time)
higher accuracy (time-slicing)

Not really un-obtrusive! Low accuracy. Real-time.

A.A. 2012-2013

59/72

<http://borghese.di.unimi.it/>



Gloves



Monitor fingers position and force.

Problems with the motion of the fingers:

- overlap.
- fine movements.
- fast movements.
- rich repertoire.

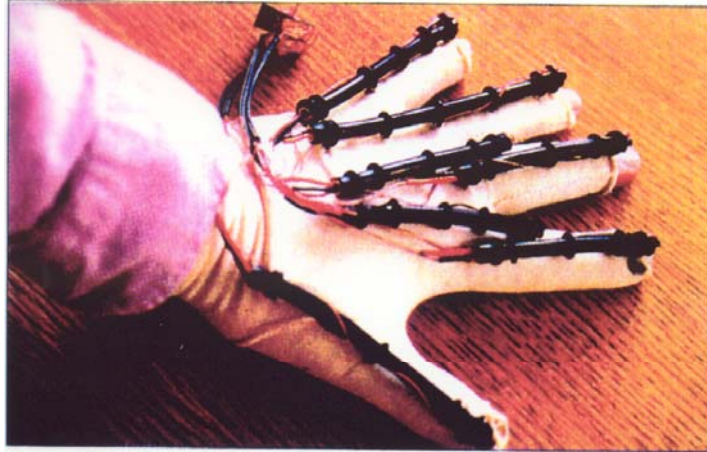
A.A. 2012-2013

60/72

<http://borghese.di.unimi.it/>



Sayre glove (1976)



A.A. 2012-2013

61/72

<http://borghese.di.unimi.it/>



MIT glove (1977)



A.A. 2012-2013

62/72

<http://borghese.di.unimi.it/>

Digital Data Entry Glove (1983)

© AT&T Bell Labs

A.A. 2012-2013 63/72 <http://borghese.di.unimi.it/>

Data Glove (1987)

A.A. 2012-2013 64/72 <http://borghese.di.unimi.it/>



Power Glove (1990)



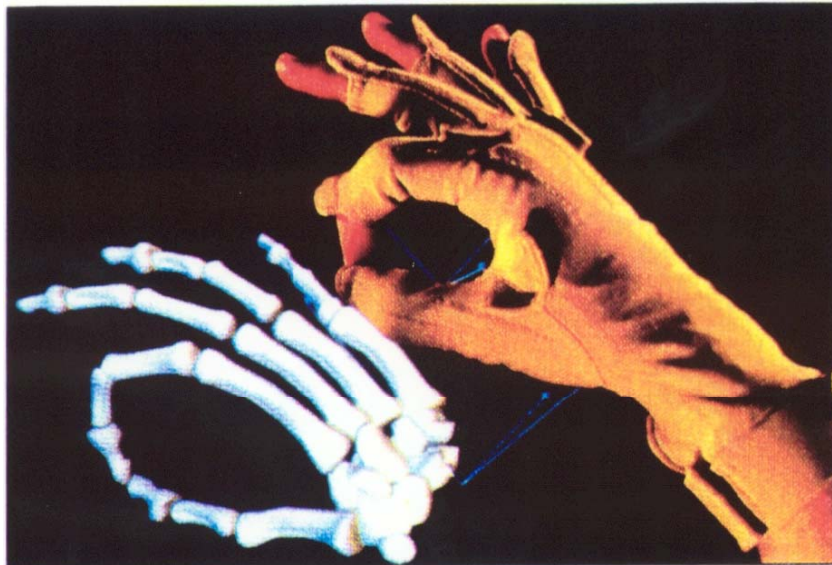
A.A. 2012-2013

65/72

<http://borghese.di.unimi.it/>



Cyber Glove (1995)



A.A. 2012-2013

66/72

<http://borghese.di.unimi.it/>



AcceleGlove / iGlove (2009)



http://www.anthronix.com/index.php?option=com_content&view=article&id=87&Itemid=138



Calibration



Estimate of the geometrical parameters in the transformation operated by the sensors (e.g. the perspective transformation operated by a video-camera).

Estimate of the parameters, which describe distortions introduced by the measurement system.

Measurement of a known pattern. From its distortion, the parameters can be computed.

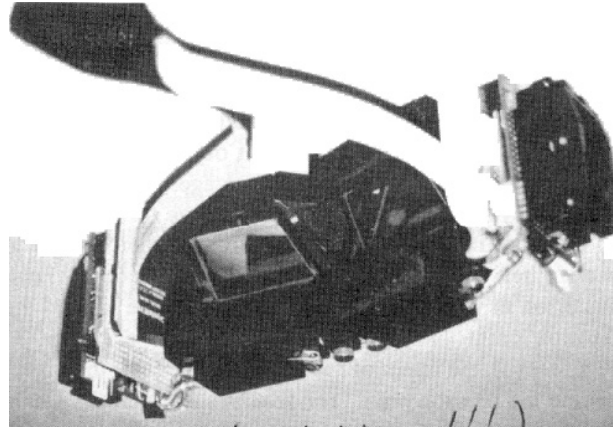
Algorithms adopted: polynomial, local correction (neural networks, fuzzy).



Gaze input



- Contact lenses carrying magnetic coils.
- Tvcameras aligned with an IR LED source.
- Stereoscopic eye-wear.
 - The direction of gaze is decided by measuring the shape of the spot reflected by the frontal portion of the cornea (Ohshima et al., 1996).



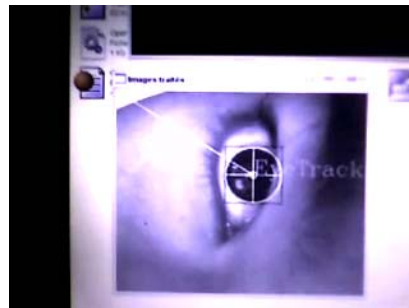
A.A. 2012-2013

69/72

<http://borghese.di.unimi.it/>



Vision based eye trackers



Logitech Quickcam 4000

- Color information
- Geometry information (circles, relative position...)
- Histogram analysis on gray level.
- Custom tool for many WEBCams
- ...

A.A. 2012-2013

70/72

<http://borghese.di.unimi.it/>



History



Video technology (semi-automatic marker detection, slow-motion, 1975)

Optoelectronic active markers: Selspot™ 1977 (Selspot II 1993), Watsmart™ 1985, Optotrak™ 1992, Polaris™ 1998. <http://www.ndigital.com/home.html>

Automatic video marker detection:

Vicon™ 1981. <http://www.oxfordmetrics.com/>

Elite™ 1988. <http://www.bts.it/>

MotionAnalysis™ 1992, Eagle™ 2001. <http://www.motionanalysis.com/>

Smart™ 2000. <http://www.motion-engineering.com/>

Magnetic systems:

Sensors: Polhemus 1987, Fastrack 1993. <http://www.polhemus.com/>

Systems: Flock of birds 1994. <http://www.ascension-tech.com/>

Intertial systems: Xmoven Xsense 2000, Wii 2008.

Video processing: organicmotion 2010.

3D video systems: Kinect 2010, <http://www.microsoft.com/en-us/kinectforwindows/>



Sommario



- Introduzione
- Sistemi di Input
- Generatori di mondi
- Motore di calcolo
- Sistemi di Output
- Conclusioni