# Motion Capture

**N. Alberto Borghese**
**Laboratory of Human Motion Analysis and Virtual Reality (MAVR)**
**Department of Computer Science**
**University of Milano**



# A motion capture system

# Outline

Introduction: what is Motion Capture?

History and Motion Capture technologies.

Passive Markers Motion Capture.

Specialized motion capture: face, gaze and hand.

From MoCap to Animation (post-processing)



# 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….

Specific body parts: gloves, gaze trackers….

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

# Motion Capture and Synthesis

Reproduce digitally the motion of the body.

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

Analysis

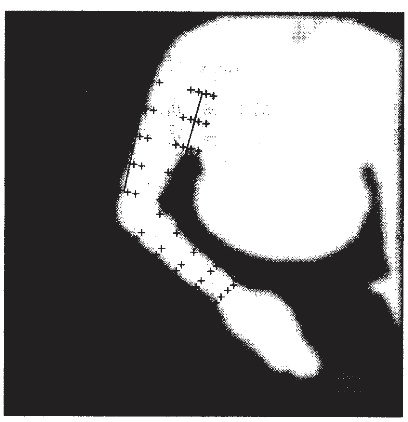
Application of the time series to a 3D digital model of the body.

Synthesis



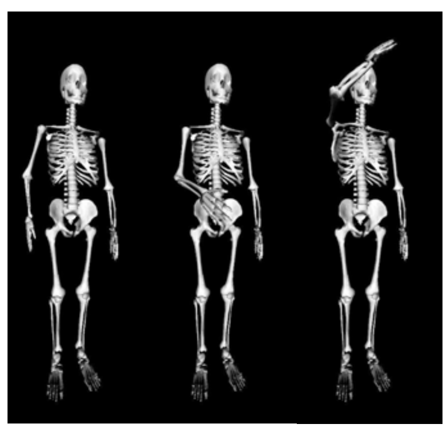
# What is captured?

Silhouette (-> Skeleton)



Computer vision techniques

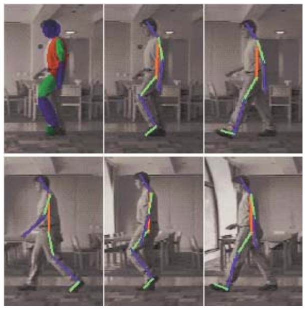
Skeleton



Bony segments or articulations (marker-based systems)

# Computer vision techniques

Silhouette (-> Skeleton)



C. Bregler, P. Fua, ….

***Set of difficult problems:***

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

Multi-view invariants.

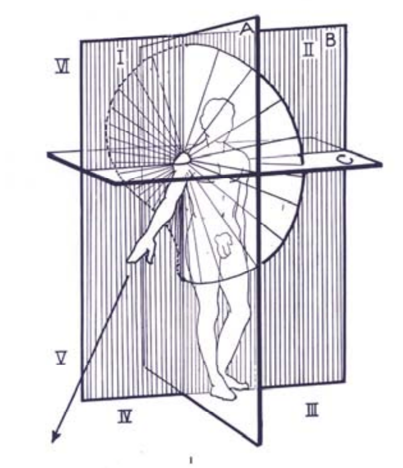
Smooth motion -> temporal filtering.

Skeleton fitting (different rigid motion for different segments).

***Pre-prototype research.***



# Description of the human skeleton



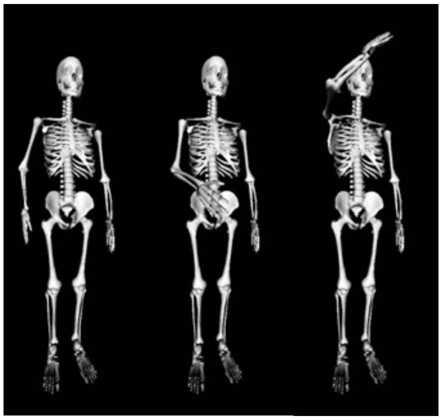

A – Frontal plane
B – Sagittal plane
C – Horizontal plane

Abduction/adduction
Flexion/extension
Axial rotation (V)

Definition of the interesting degrees of freedom.

# Marker-based techniques

Skeleton



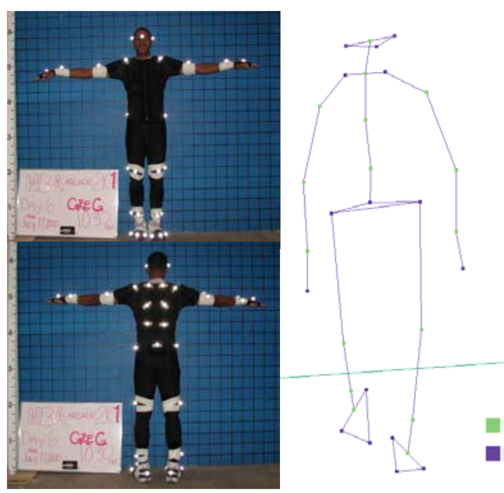
Bony segments or articulations.

Here, the problem is to find a suitable ***marker*** for the segments and a suitable HW/SW system for ***marker detection***.



# How the motion of the skeleton is captured?

Markers on the bony segments



Markers on the body joints

# Outline

Introduction: what is Motion Capture?

History and Motion Capture technologies.

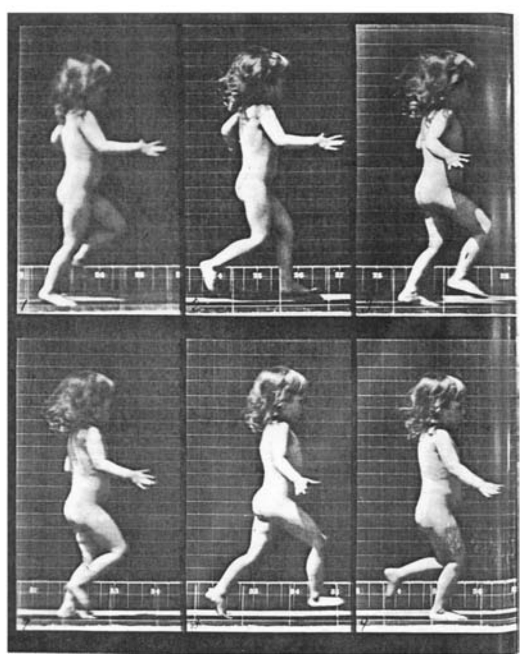
Passive Markers Motion Capture.

Specialized motion capture: face, gaze and hand.

From MoCap to Animation (post-processing)



# Edward Muybridge 1878



Laboratory of Motion Analysis & Virtual Reality, MAVR http://homes.dsi.unimi.it/~borghese/

# History

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

Optoelecontric active markers: Selspot™ 1977 (Selspot II 1993), Watsmart™ 1985, Optotrack™ 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/



# Where are we now (optoelectronic)?

Optotrack, 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.

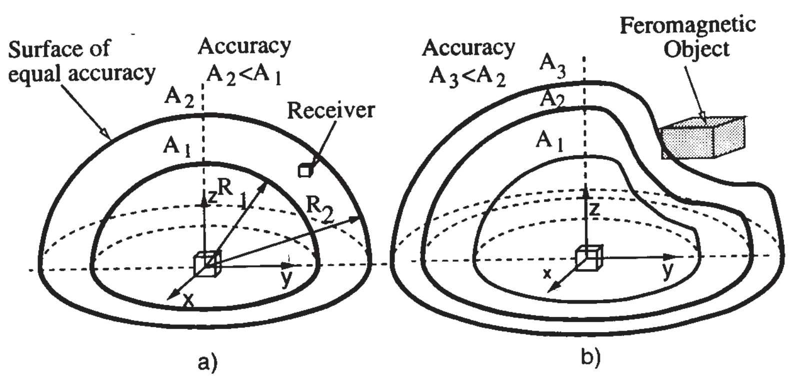
The transmitter is a triad of electromagnetic coils, enclosed in a plastic shell, that emits the magnetic fields. The transmitter is the system's reference frame for receiver measurements.

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.



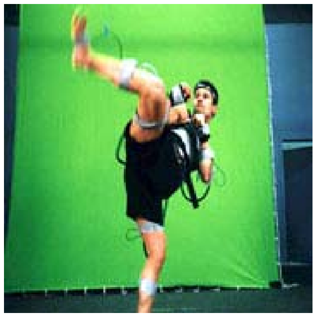
# Fast-track Motion Capture

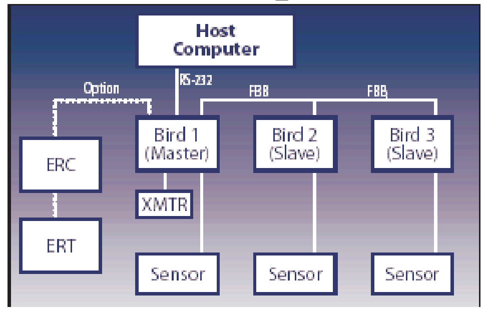
•Higher accuracy through oversampling and DSP signal processing (0,5” and 1.8mm accuracy). Range of 75cm for high accuracy.
•Sensitive to ferromagnetic (metallic) objects.




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

# 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 (exprimental):

larger field (single transmitter at a time)

higher accuracy (time-slicing)

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



# Outline

Introduction: what is Motion Capture?

History and Motion Capture technologies.

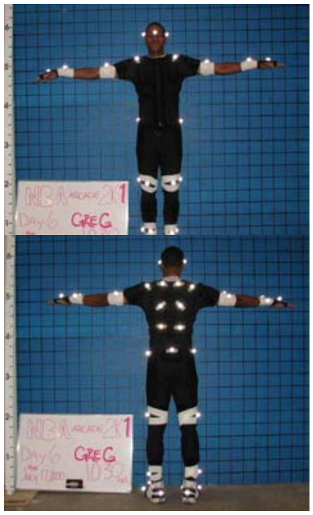
Passive Markers Motion Capture.

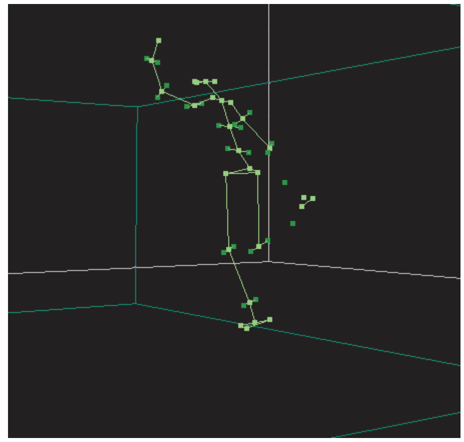
Specialized motion capture: face, gaze and hand.

From MoCap to Animation (post-processing)

# Motion Capture with passive markers

Goal: reconstruction of the 3D motion of a set of markers

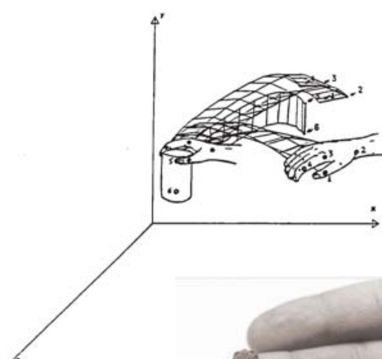


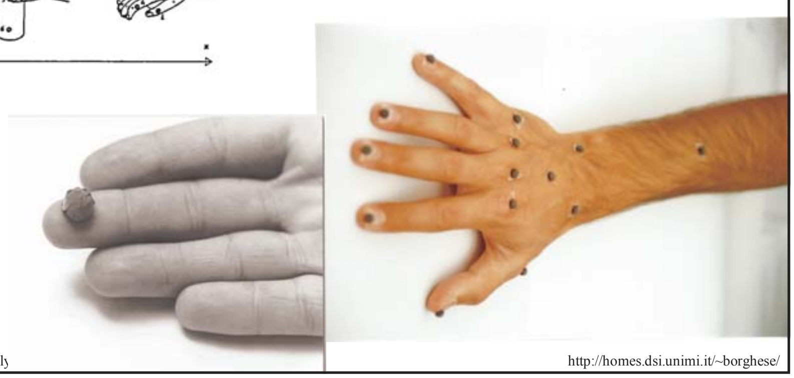


# Why passive markers?

No encoumbrance 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.



# 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.



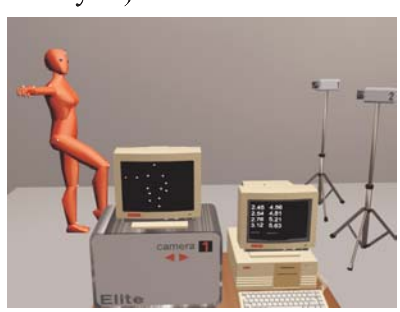
Video-cameras are equipped with a co-axial flash.

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



# Constituents of a Motion Capture system with passive markers

•Markers
•Cameras
•Flash (synchronous with frame signal)
•Connections (Fast Ethernet for Motion Analysis)
•Hub
•PC host for processing and display.




***Where is marker detection?***

PC (e-motion™)
Before the Hub (Vicon™, Eagle™, Elite™).

# 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

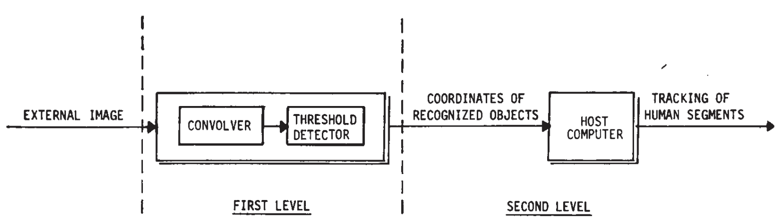
4. Matching on multiple cameras.
5. 3D Reconstruction (*accuracy*).

High-level Vision

An implicit step is CALIBRATION.



# Two-levels architecture



# 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 looses track of it.

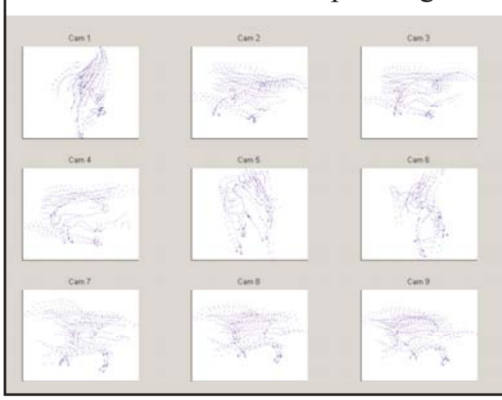
The multiple set of 2D data have to be correctly labaled and associated to the 3D markers.

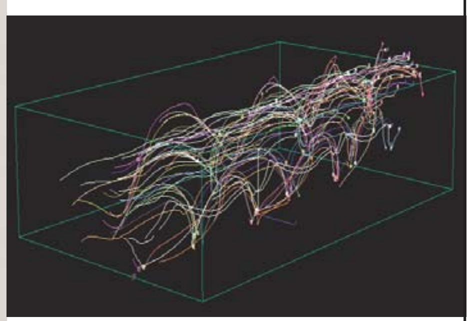


# Disadvantages of motion capture systems based on passive markers

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 looses track of it.

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





# The difficulties in data processing

1. Twists and rotations make the movement of the human body fully three-dimensional.
2. Each body part continuously moves in and out occlusion from the view of the cameras, such that each of them can see only a chunk of the whole trajectory.
3. Some body parts can be hidden to the view by other parts. Whenever it happens, the system should be able to correctly recognize the hidden markers as soon as they reappear without any intervention by the operator.
4. Chunks from the different cameras have to be correctly matched and integrated to obtain a complete motion description.
5. Each trajectory has to be associated with the corresponding body marker (labeling).
6. Reflexes, which do appear in natural environment and are erroneously detected as markers, have to be automatically identified and discarded.



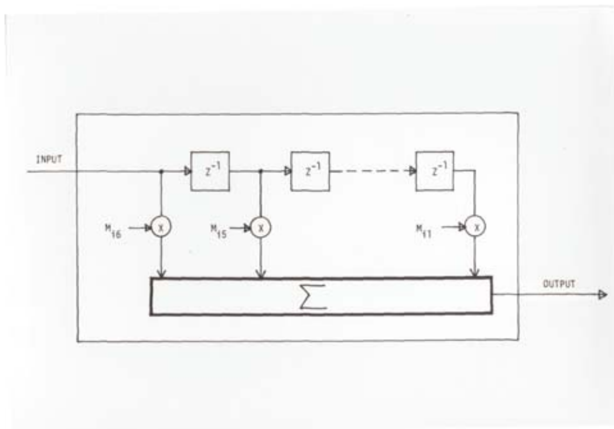
# Markers detection

Threshold detection may be not sufficient.

- Cluster dimension.
- Shape.

# Markers extraction (low-level)

Correlation implemented by convolution (template matching or feature extraction)

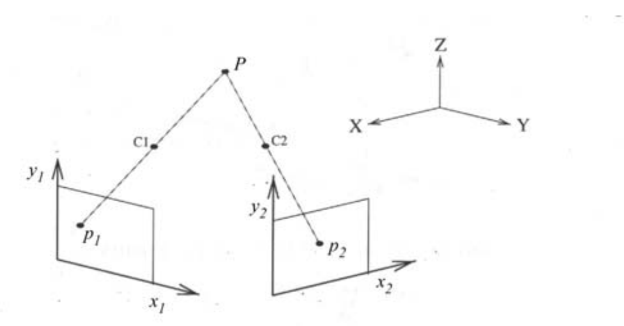



Implementable with a DSP



# From 2D to 3D

Each camera measures a geometrical transformation (projection)
Triangulation (ray intersection)




Geometrical parameters known.
Main difficulty is correct matching between multiple markers and multiple cameras.

# Tracking difficulties

<u>It is a complex problem because:</u>

• 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.

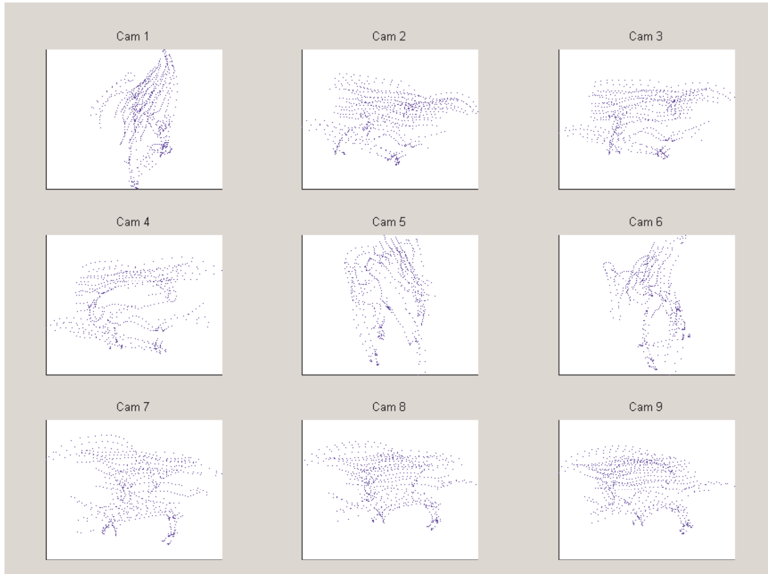
# Tracking processing

<u>ACQUISITION OF 2D POINTS</u>

<u>TRACKING:</u>
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.

<u>RECTIFY:</u>
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

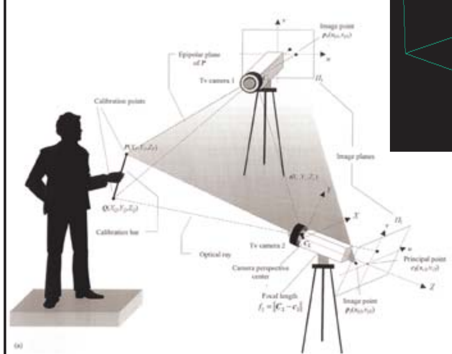


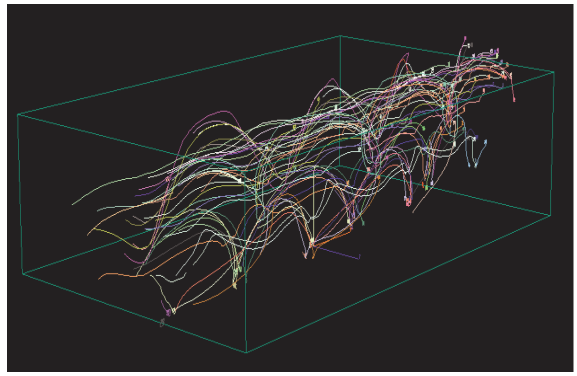



# 1) Creation of 2D strings

# 2) Matching 2D strings

Epipolarity constraint





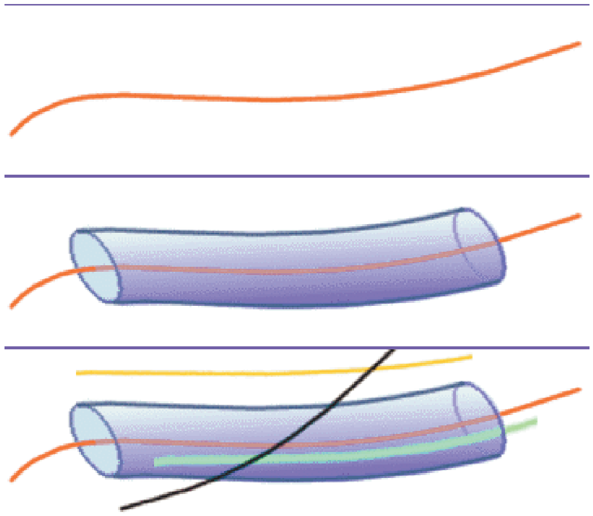
3D strings



# 2) 3D strings

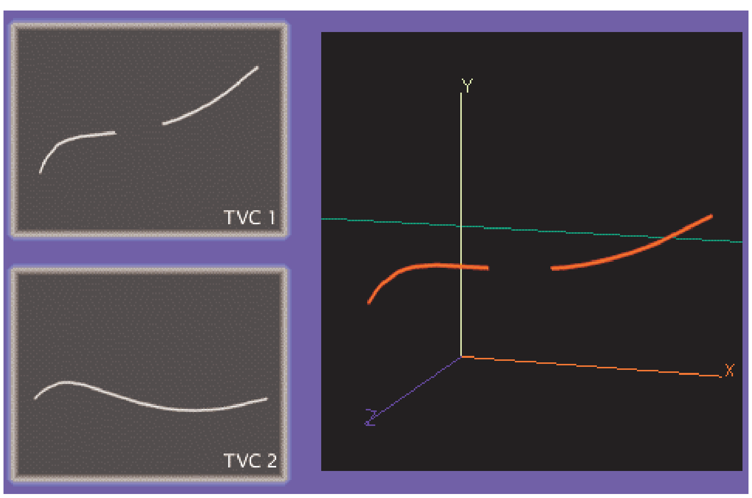
3D strings already contain motion 3D information

# 3) Condensation of 3D strings



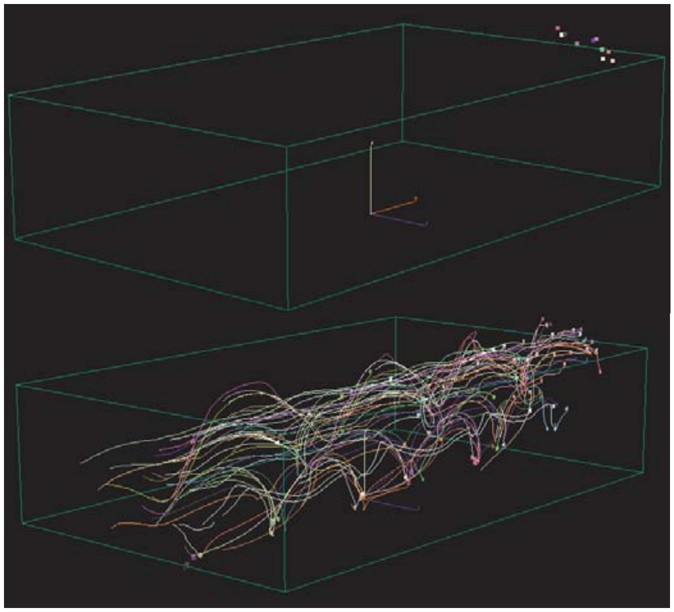


# 4) Joining 3D strings



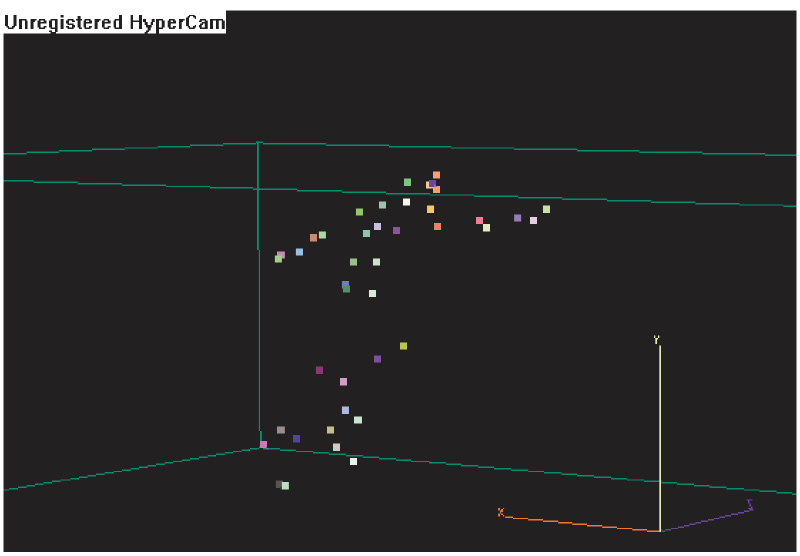

Laboratory of Motion Analysis & Virtual Reality, MAVR http://homes.dsi.unimi.it/~borghese/

# 3D strings

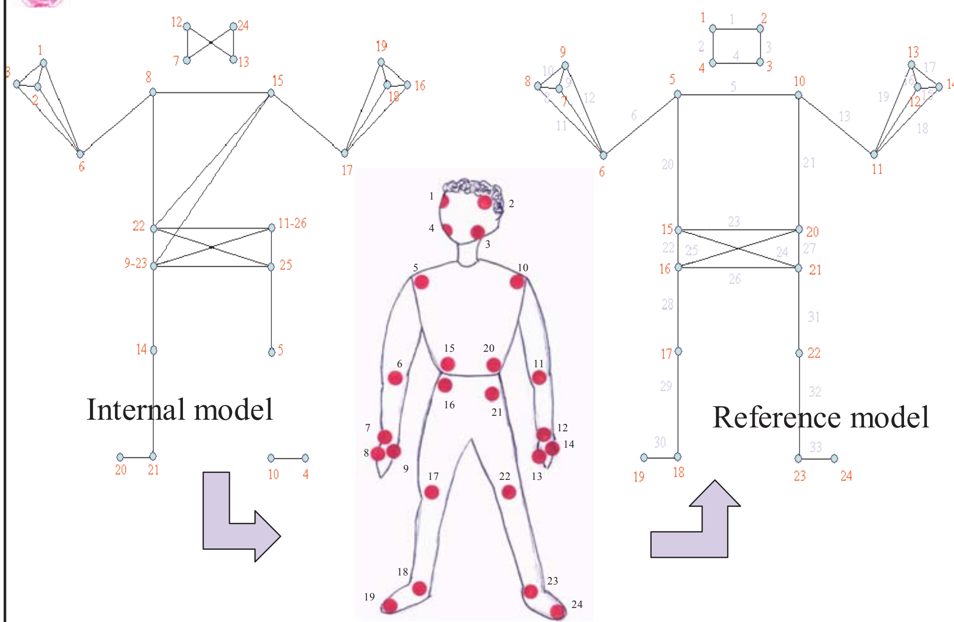




# Markers Classification

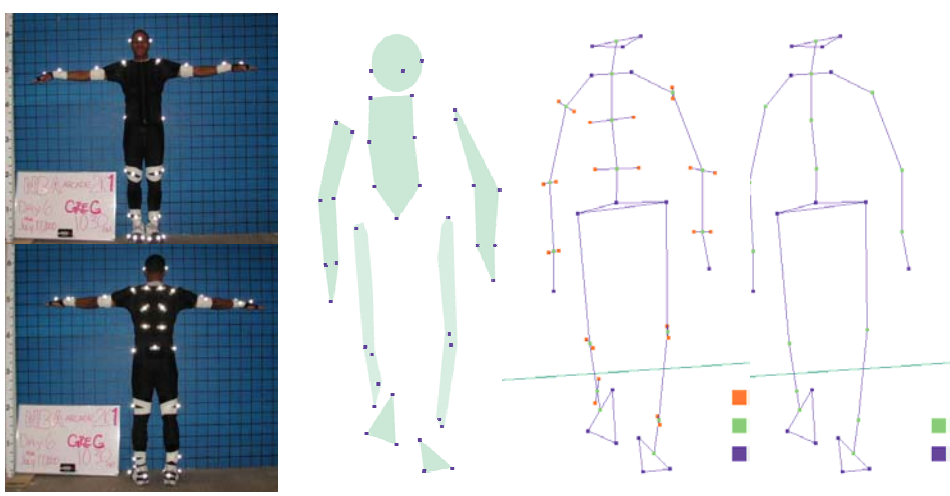


# Model fitting



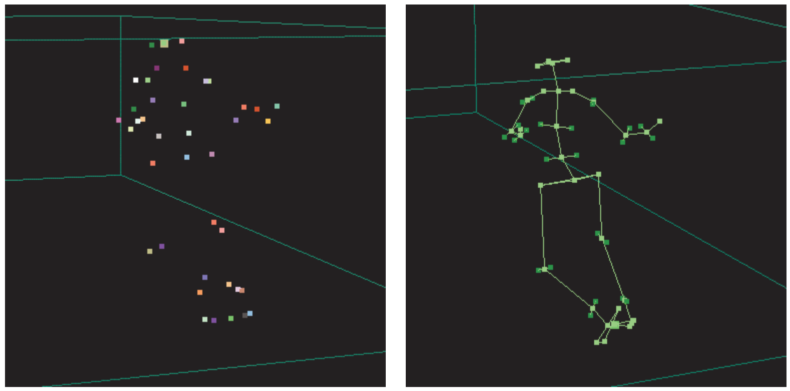



# What a model represents?



Markered subject | Modello 3D | Modello a stick | Modello hidden

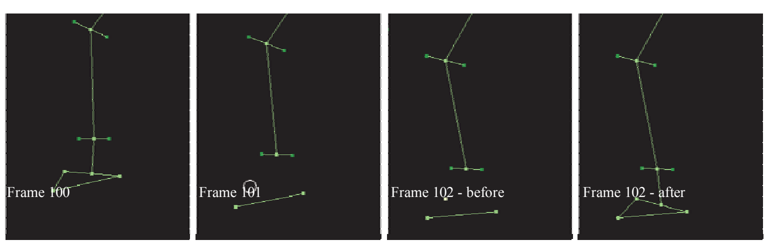
# 5) Initial classification





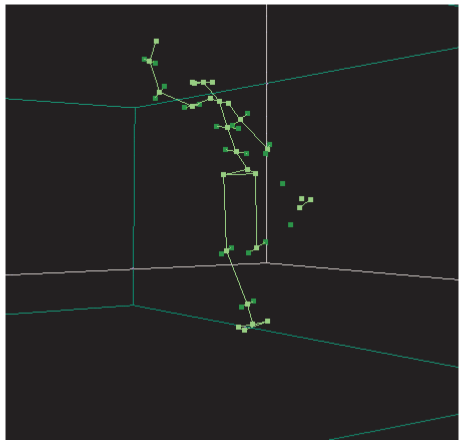
# 6) Classification extension

3D strings are automatically extended in this phase.



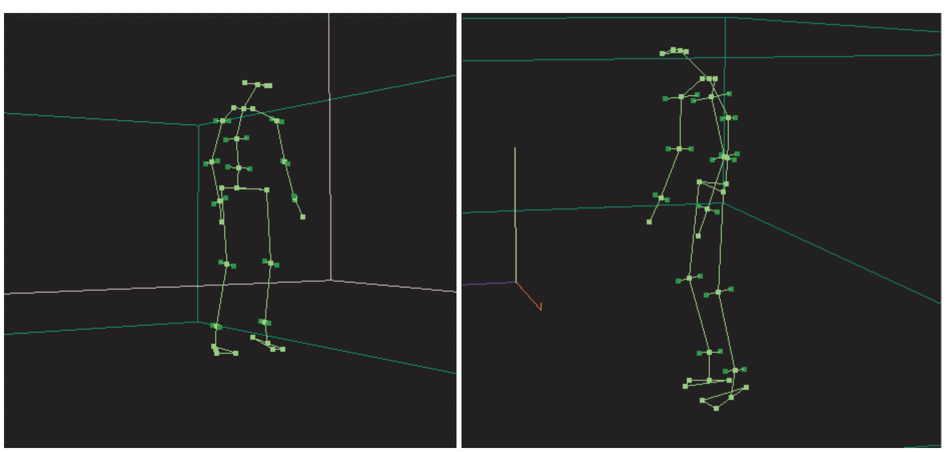

Two strings are joined on the base of:
- Smooth motion.
- Model checking (a dynamic prioirty is coded in the number of links).

# Risultati: run

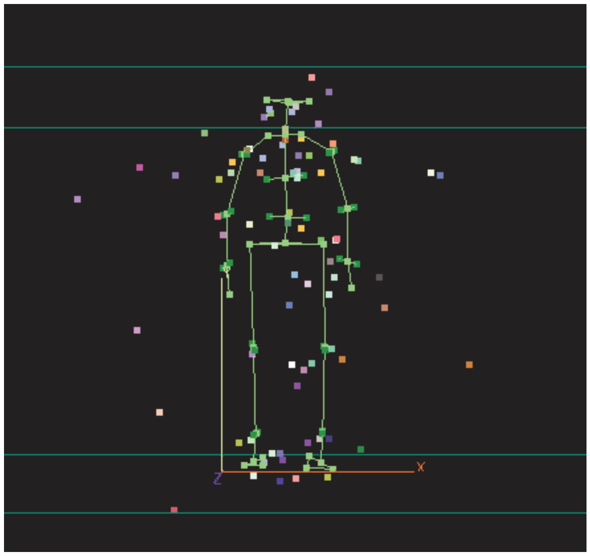




# Risultati: escape

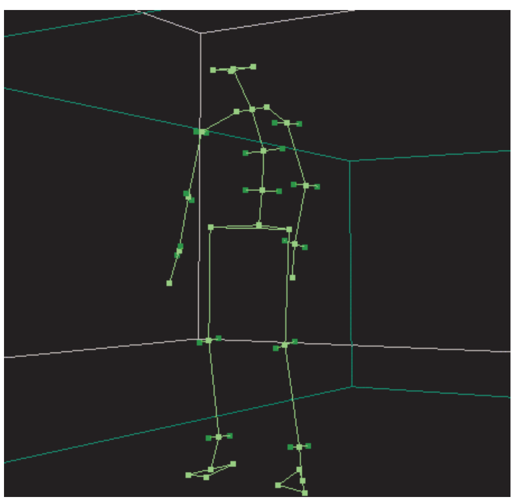


# Risultati: head_turn

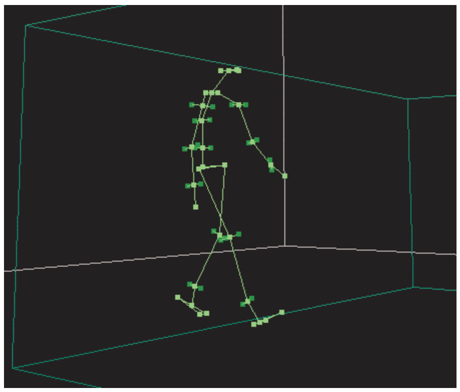


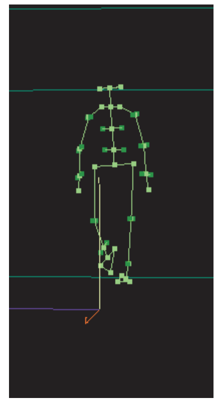



# Risultati: fall_run



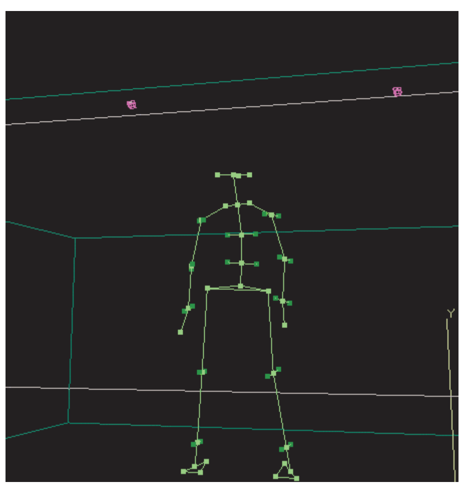
# Risultati: walk







# Risultati: roll



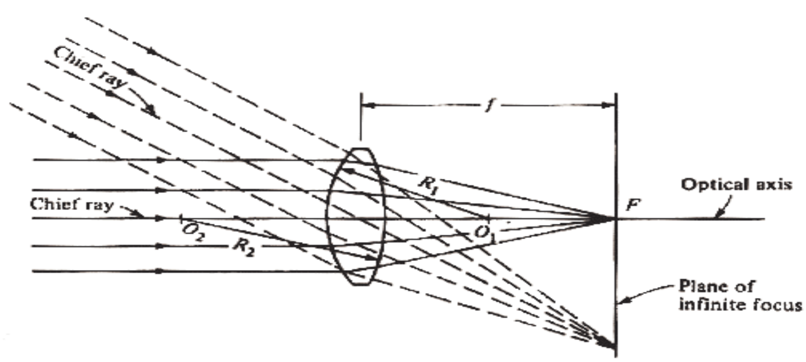
# Calibration

2D calibration (camera calibration, estimate of interior parameters).

3D calibration (estimate of the exterior parameters).

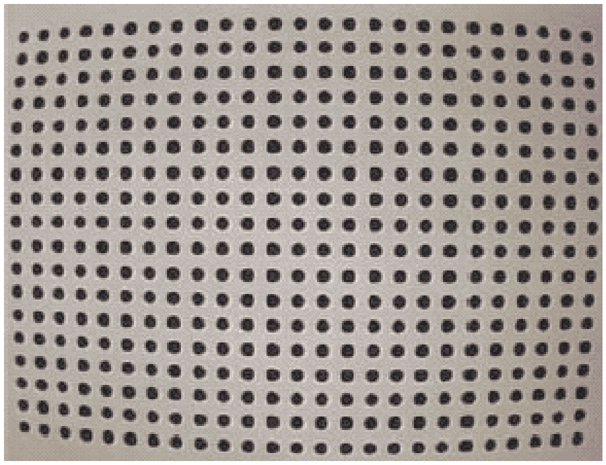


# Interior parameters (geometrical)




Focal length + principal point

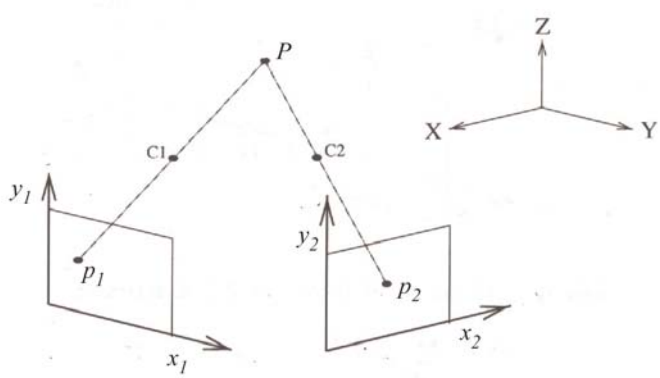
# Interior parameters (distortions)



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



# Exterior parameters




For each camera, the position and orientation with respect to a reference frame has to be determined (location + orientation).

# Set-up

Passive motion capture does not constraint cameras position.

These have to be positioned to get the best volume coverage (every marker should be surveyed by at least two cameras).

***Set-up requires that:***
•Cameras position
•Focusing (and possibly choice of a proper lens)
•Lens opening

should be set-up before calibration.

**Optimal set-up may require some time**
**and/or**
**Multiple cameras are used.**



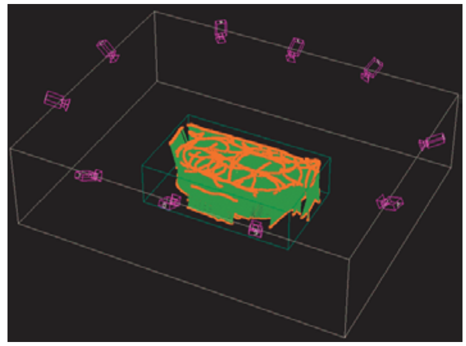
# Motion Capture for animation


• Passive markers + flash coaxial with the cameras.

• Large working volume (10m x 8m x 4m).

• Reduntant set of cameras (>10).

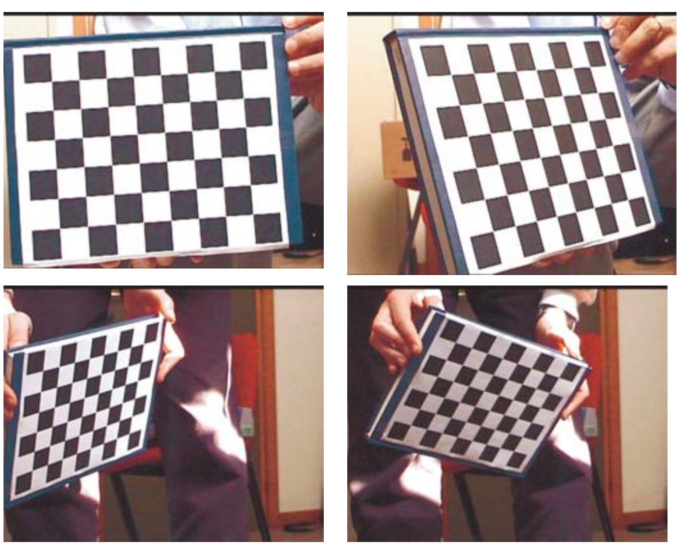
# Motion Capture, calibration procedure

Before data acquisition the subject has to be markered.



# Calibration through a chessboard



Estimated parameters:

Distortions
Internal
External

http://www.intel.com/research/mrl/research/opencv/

# Outline

Introduction: what is Motion Capture?

History and Motion Capture technologies.

Passive Markers Motion Capture.

Specialized motion capture: hand, gaze and face.

From Motion Capture to Animation (post-processing)



# Gloves

Monitor fingers position and force.

Problems with the motion of the fingers:
- overlap.
- fine movements.
- fast movements.
- rich repertoire.

# Sayre glove (1976)



# MIT glove (1977)

# Digital Data Entry Glove (1983)

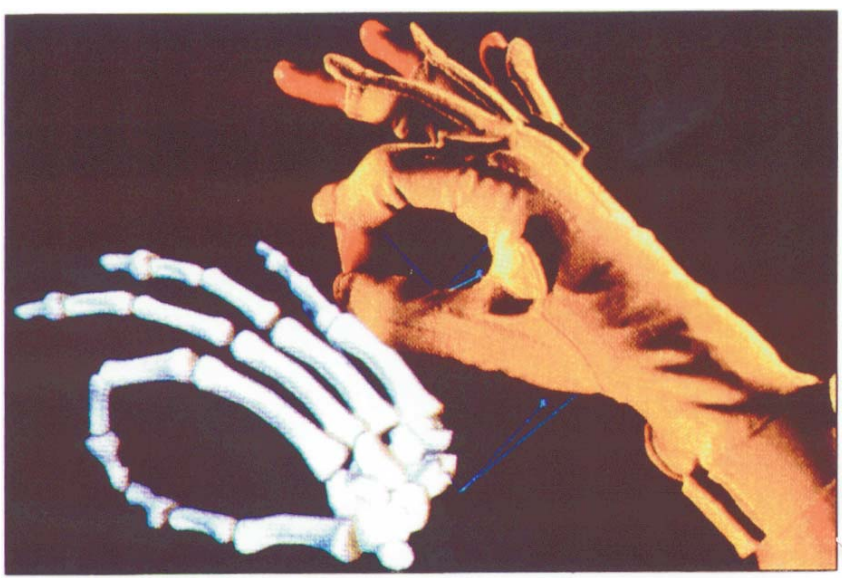


# Data Glove (1987)

## Power Glove (1990)



## Cyber Glove (1995)



## 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).



## Aptic displays

Convey to the subject the sensorial information generated in the interaction with the virtual objects: force, material texture…

Measure the force exerted by the subject on the virtual environment.

Aptic displays provide a mechanical interface for Virtual Reality applications.

Most important developments have been made in the robotics field.

## Requirements of aptic displays

• Large bandwidth.
• Low intertial and viscosity.

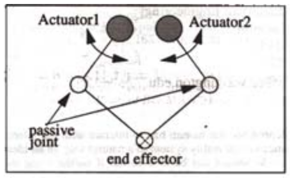
**Technological solutions:**

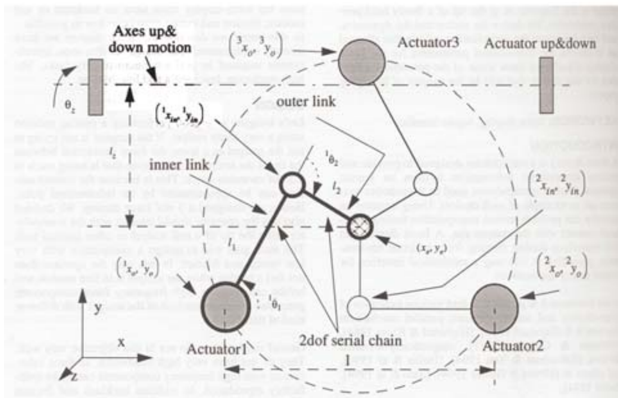
• Direct drive manipulandum (Yoshikawa, 1990), Phantom (2000).
• Parallel manipulandum (Millman and Colgate, 1991; Buttolo and Hannaford, 1995).
• Magnetic levitation devices (Salcudean and Yan, 1994; Gomi and Kawato, 1996).
• Gloves (Bergamasco, 1993).



## Direct drive manipulandum (phantom)

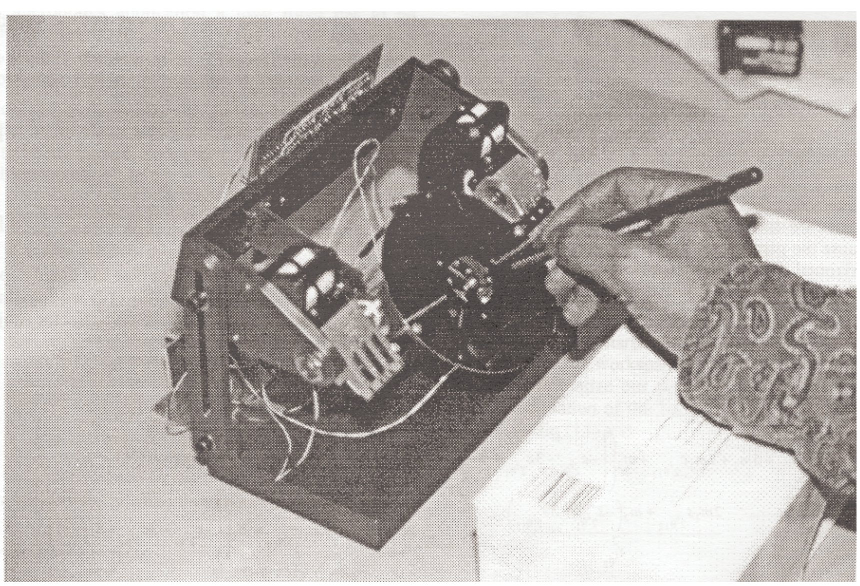
# Parallel manipulandum (schema)



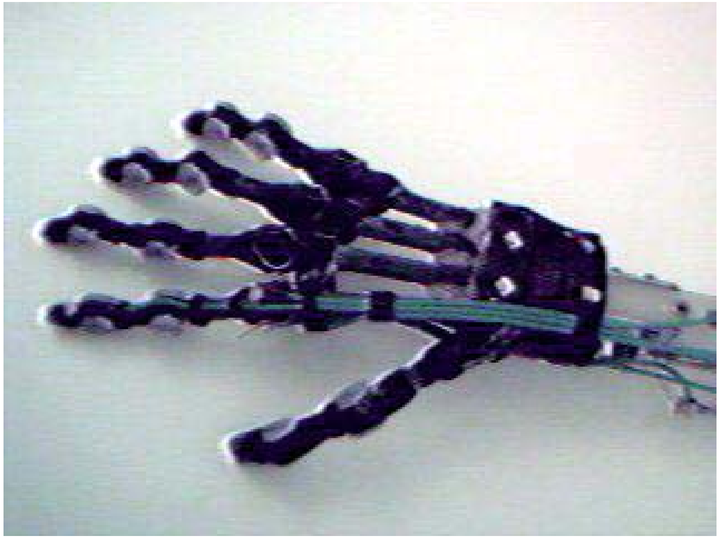






# Pen aptic display

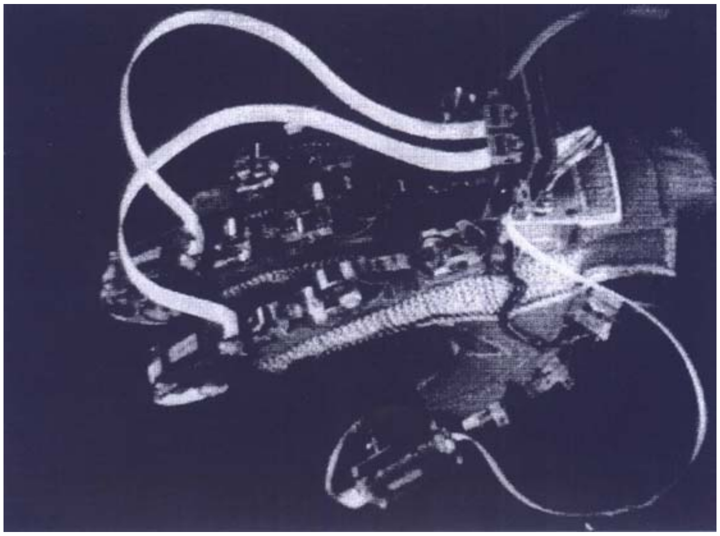


## Gloves (blackfinger, 2000)





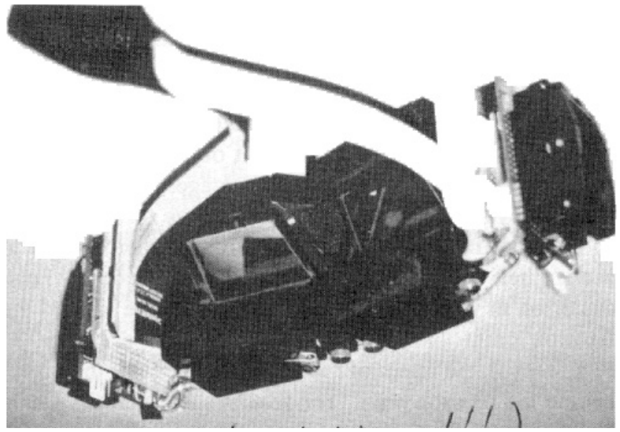
## Percro gloves (Begamasco, 1993)



Laboratory of Motion Analysis & Virtual Reality, MAVR http://homes.dsi.unimi.it/~borghese/

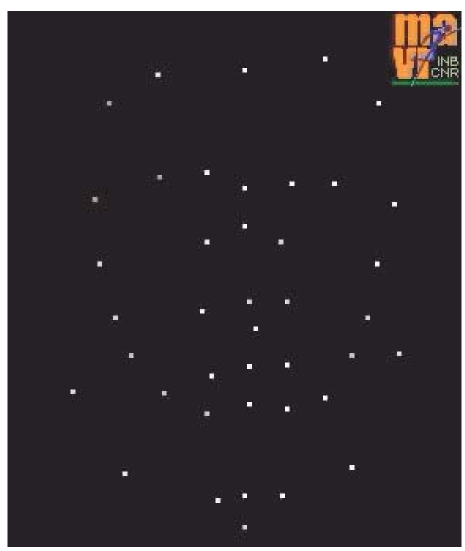
# Gaze input

- 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).





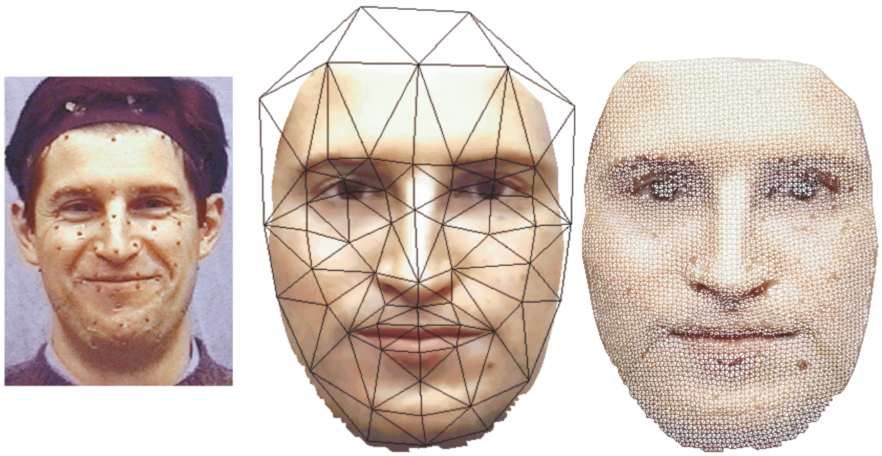
# Face Animation





***Field morphing.***

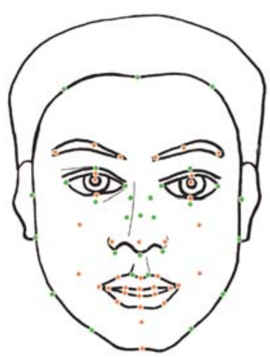
# Two-layers technique



•Deformation of a topological mesh induced by a control mesh.
•The control mesh connects the marker points.



# Markers disposition



Position of the feature points according to MPEG-4 standard:
- ◆ principali
- ● secondari

*Problems with:*
Eyes and tongue.
Nose basis (visibility).

# Construction of the Control Mesh

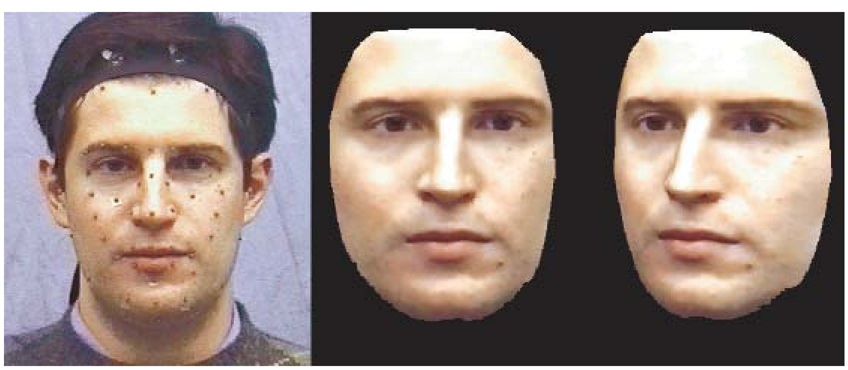

**47 markers on the skin:**
- *Problems with:*
  Eyes and tongue.
  Nose basis (visibility).

**4 markers on an elastic band:**
*To identify a local Reference Frame (LRF).*

- 51 Markers acquired (cf. MPEG-4 specifications).
- 7 virtual markers defined through the LRF (green).
- 2 Virtual markers defined through Real Markers (blue).
- 56 control points for the mesh + 4 for LRF.



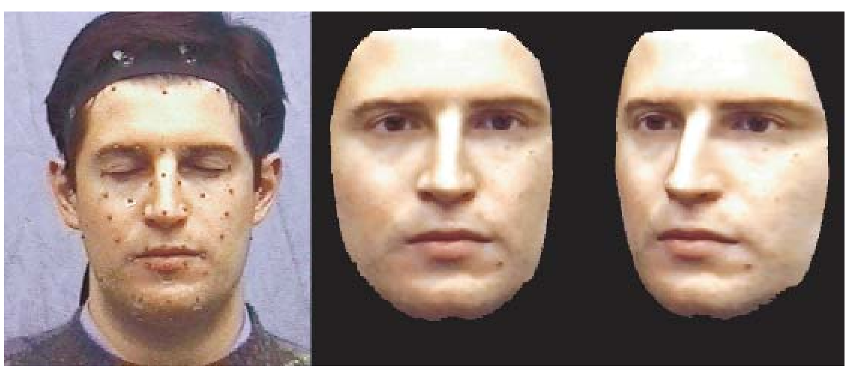
# Results: anger



# Results: surprise



# Results: disgust



# Results: happiness



# Avenues of research

Detailed biomechanical models (FEM). Not compatible with real-time for non-linear elements.

Streaming of images over the 3D mesh.

Blending 3D models of "critical" parts (tongue, teeth..) and pre-defined texture for grooves (bump mapping) with the 3D mesh.

Intersting problems:

- Impossible interviews.
- Virtual speakers for low-band transmission.
- Rehabilitation.
- …..

# Outline

Introduction: what is Motion Capture?

History and Motion Capture technologies.

Passive Markers Motion Capture.

Specialized motion capture: hand, gaze and face.

From Motion Capture to Animation (post-processing).

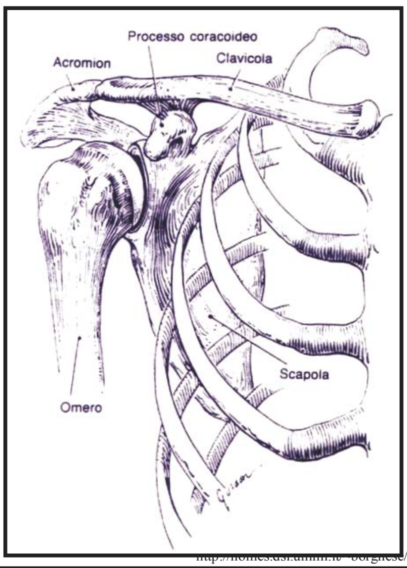


# The human skeleton has complex articulations

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

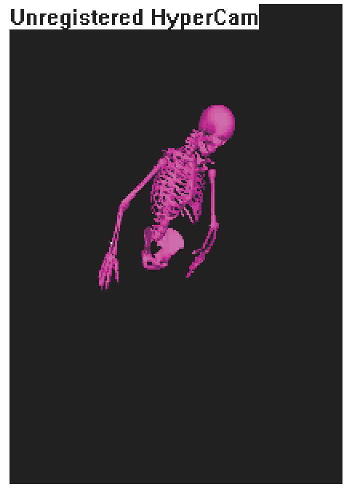
Motion allowed can be very complex (e.g. shoulder, spine).

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



# Retargetting

From Motion Capture to Virtual Motion:
3D positions → Angles
Model fitting
Motion correction






# Motion correction & retargetting

What happens if the arm of the digital character enter inside the shoulder of his girl-friend?

The problem is reframes as an optimal control problem.

Zero error in the final frame.

Minimal deviation of the control actions (the angle sequence).

$$a\underset{\{\mathbf{u}(t_k)\}}{\sum_k}\left(\mathbf{u}_k(t)\right)^2 + b(\mathbf{x}_d(t_N) - \mathbf{x}(\mathbf{t}_N))^2$$

# Motion retargetting: an example

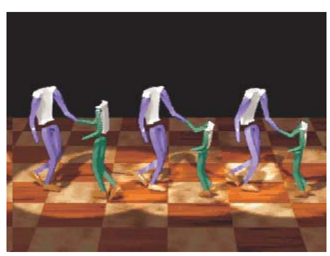


Data captured have to be adapted to a smaller female.



# Motion retargetting: an example



Laboratory of Motion Analysis & Virtual Reality, MAVR http://homes.dsi.unimi.it/~borghese/

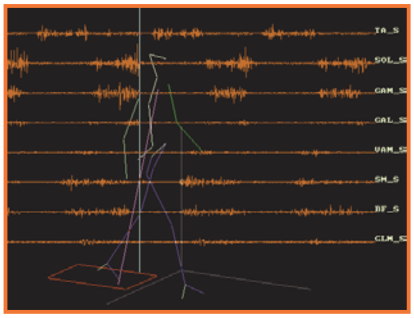
# Clinical Motion Analysis

MOTION ANALYSER

FORCE TRANSDUCER

MATHEMATICAL MODELS

EMG



JOINT KINEMATICS

JOINT KINETICS

EXTERNAL FORCES

PLANTAR PRESSION

MUSCLE ACTIVATION AND FORCE



# The future

Digital and Reality in real-time (virtual theater).

Color-coded markers.

Mixed vision/marker techniques.

Integration of gloves, gaze trackers and marker trackers.

Detailed biomechanical models.

More biology into digital characters (motion retargetting, with “biological rules”).

***Is there any future for motion capture?***

# Motion capture