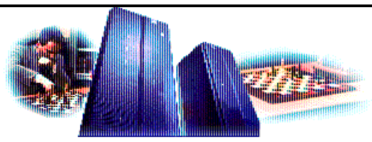
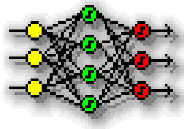




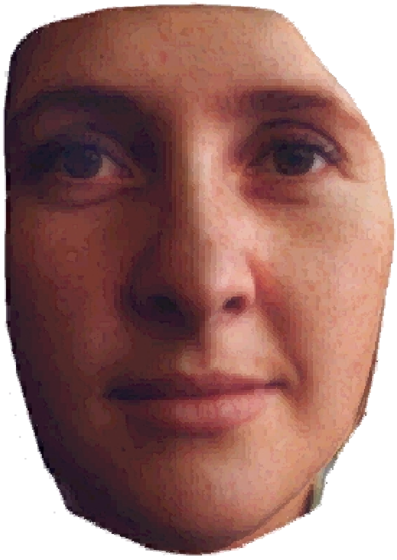
Introduzione alla Realtà Virtuale
Parte I

Prof. Alberto Borghese

Sommario

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

Which is real, which is virtual?



Historical Perspective

• *Virtual Worlds or Synthetic Environments*
• *Philosophical* and *Technologial* 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.

Historical Perspective (II)

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.

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
- 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.

Position 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).
• Gloves (specialized for hands).
• Gaze trackers.

Adopted technology

• Optoelectronics
  •Marker based
  •Computer vision.
• Magnetical
• Acoustical
• Mechanical

Input position system::measurements

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.

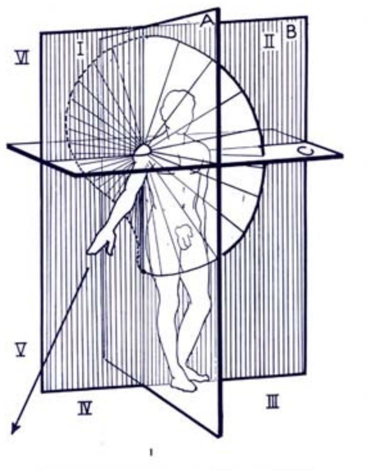
Time series of the position of the body segments
or
Time series of the motion of the articulations.

Analysis

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

Synthesis

Description of the human skeleton



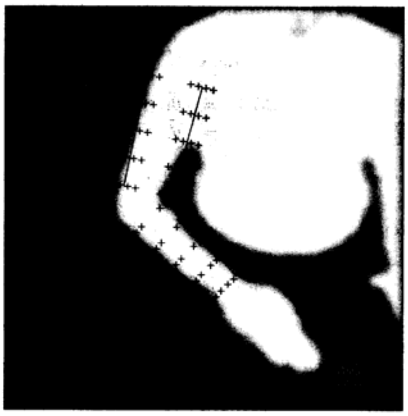
A – Frontal plane
B – Sagittal plane
C – Horizontal plane

Abduction/adduction
Flexion/extension
Axial rotation (V)

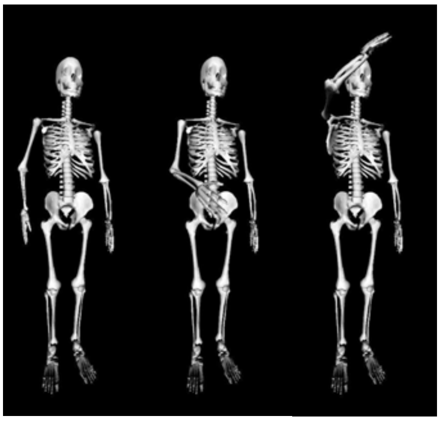
Definition of the interesting degrees of freedom.

What is captured?

Silhouette (-> Skeleton)



Skeleton

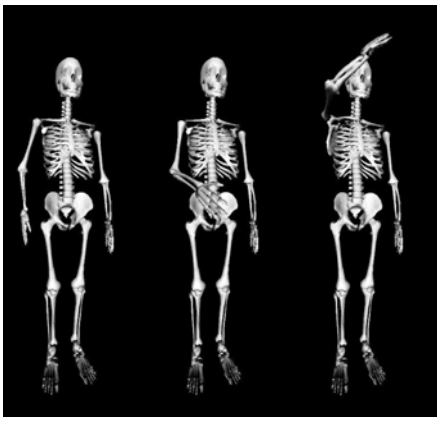


Computer vision techniques

Bony segments or articulations (marker-based systems)

Marker-based techniques

Skeleton



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

Bony segments or articulations.

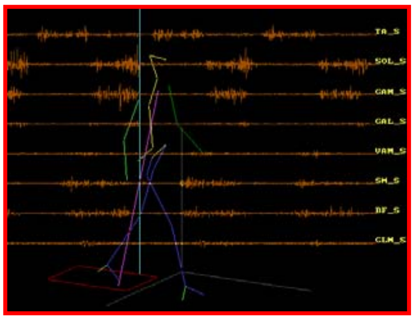
Clinical Motion Analysis

MOTION ANALYSER

FORCE TRANSDUCER

MATHEMATICAL MODELS

EMG



JOINT KINEMATICS

JOINT KINETICS

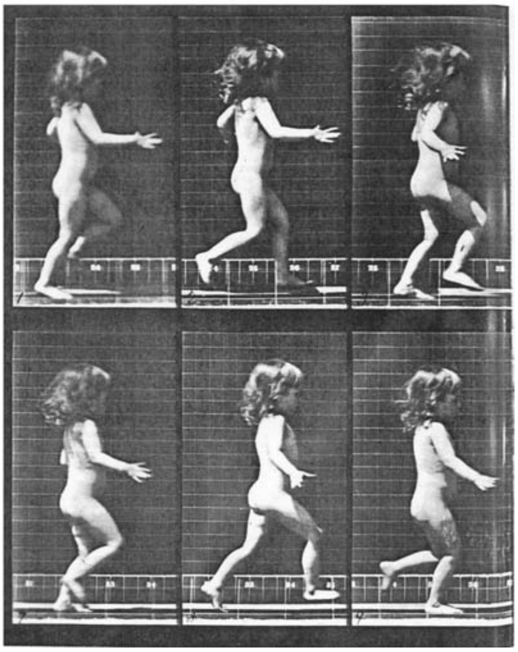
EXTERNAL FORCES

PLANTAR PRESSION

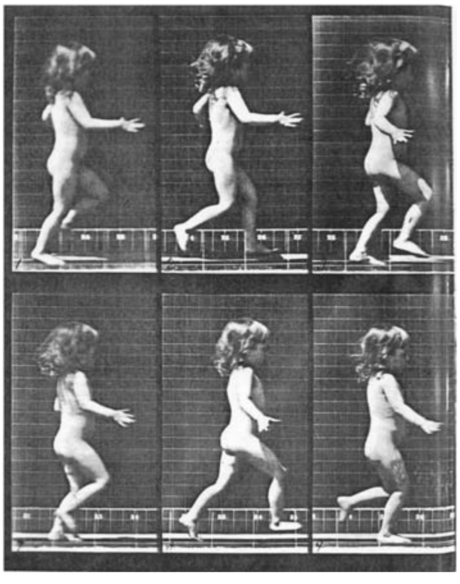
MUSCLE ACTIVATION AND FORCE

Input position system::history&technology

Edward Muybridge 1878-1901



Zoopraxinoscopio



Zoetrope, 1820 circa

E. Muybridge,
Humans figures in motion, 1901
+
zoopraxinoscope

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.

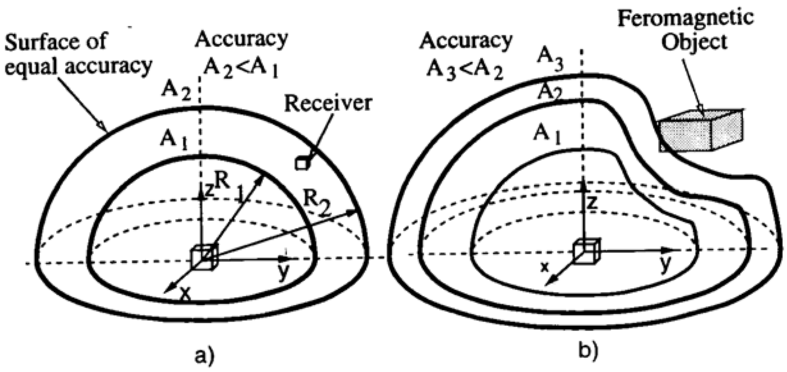
Electro-magnetic induction.

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.

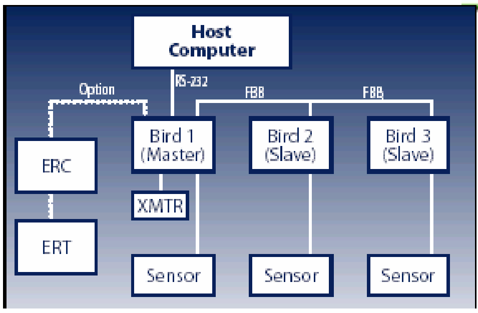
Fast-track Motion Capture

•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)

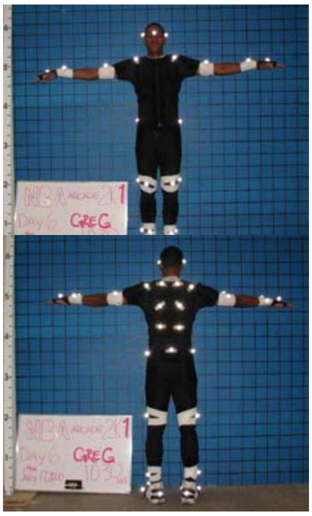
higher accuracy (time-slicing)

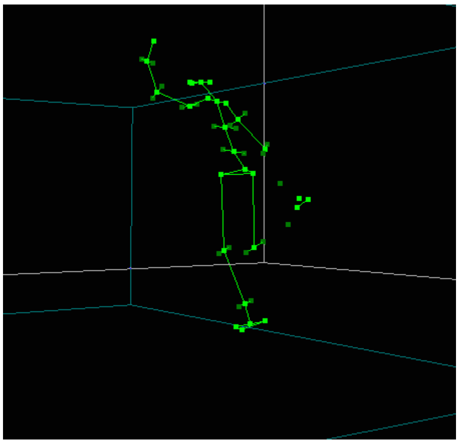
Not really un-obtrusive! Low accuracy. Real-time.

Input position system::optical motion capture

Motion Capture with passive markers

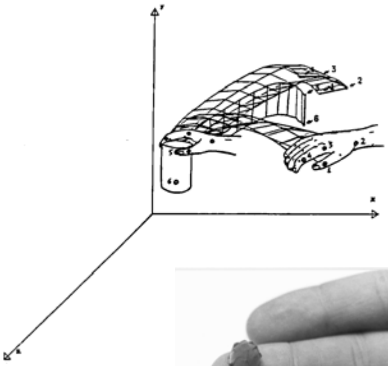
Goal: reconstruction of the 3D motion of a set of markers





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^{TM}). 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 (Smart^{TM})
Before the Hub (Vicon^{TM}, Eagle^{TM}, Elite^{TM}).

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

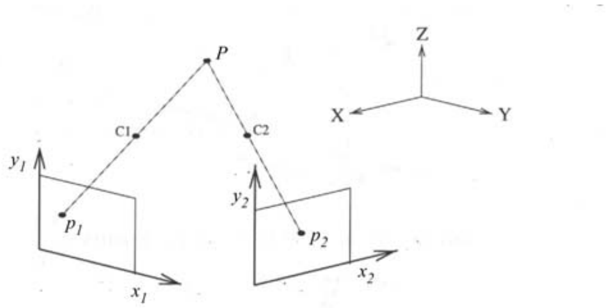
4. Matching on multiple cameras.
5. 3D Reconstruction (*accuracy*).
6. Model fitting (*labelling, classification*).

High-level Vision

An implicit step is CALIBRATION.

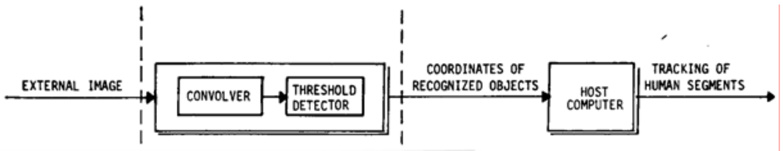
From 2D to 3D

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

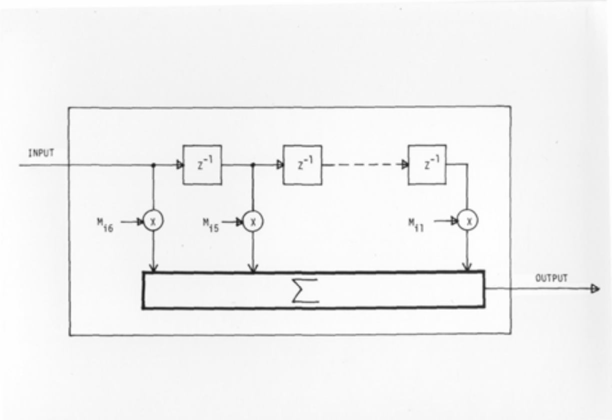


Geometrical parameters known.
Main difficulty is correct matching between multiple markers and multiple cameras.

Two-levels architecture



SECOND LEVEL

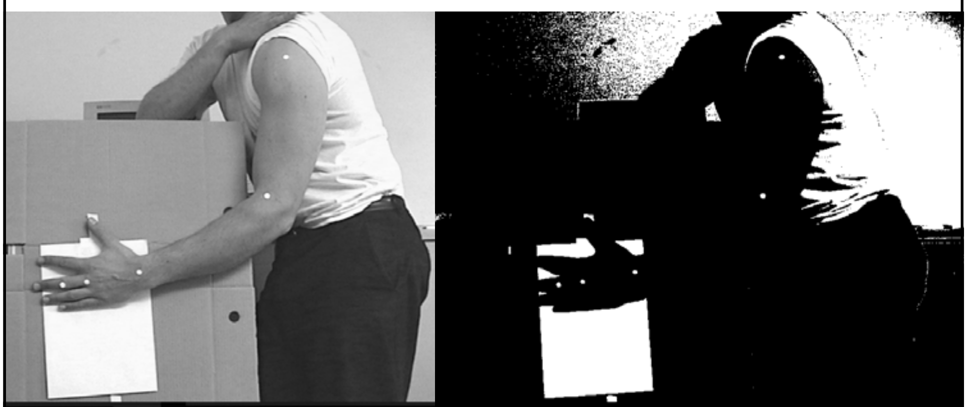


Implementable with a DSP or with a PC

➔

HUB for a set of cameras.

Markers extraction through thresholding (without flash)

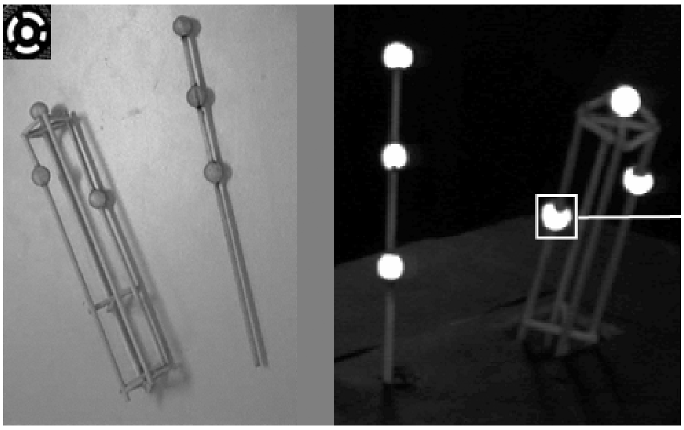


Threshold detection may be not sufficient (high contrast thanks to flashes).

- Cluster dimension.
- Shape.
- Software protection of bright target regions.

Soluzione

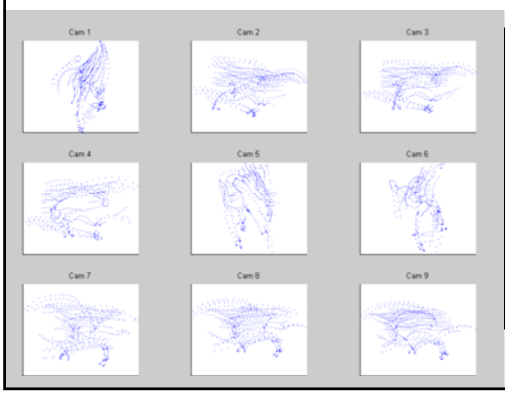
- Retro-reflective markers (or white markers).
- Short lens opening time.
- Coaxial flashes

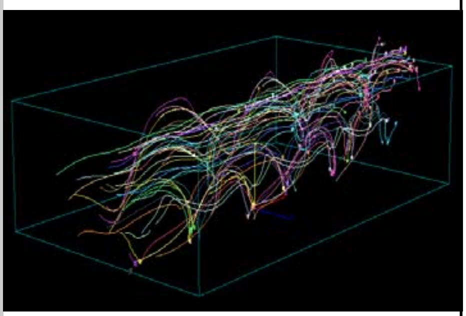


Disadvantages of motion capture systems based on passive markers

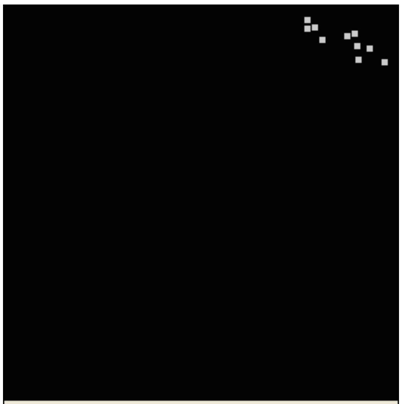
When a marker is hidden to the cameras by another body part (e.g. the arm which swings over the hip during gait), the motion capture looses track of it.

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





Tracking difficulties



rectify_run_cam_6.avi

<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.

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.

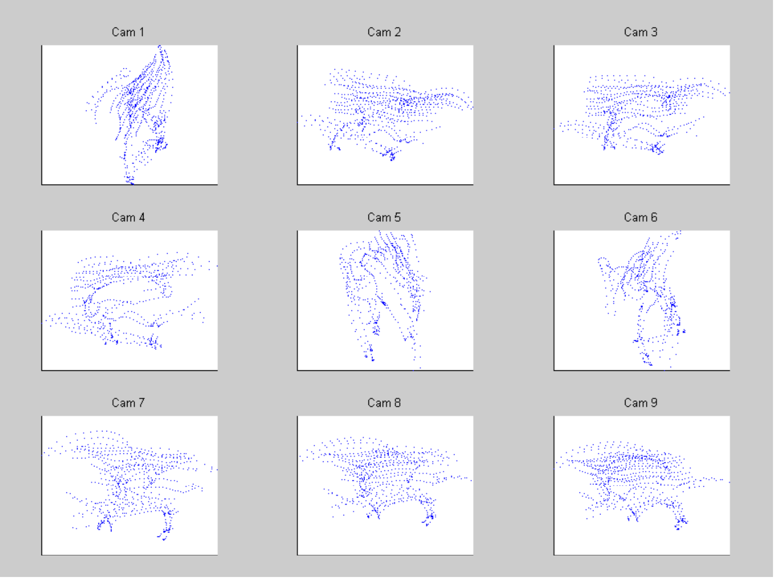
Tracking processing steps

ACQUISITION OF 2D POINTS

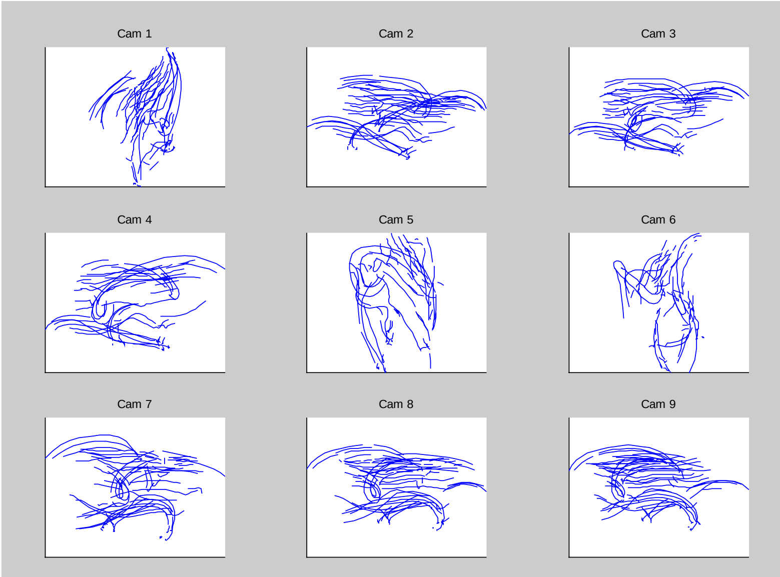
TRACKING:
1) From 2D points to 2D strings.
2) Pairing 2D strings with the epipolar constraint to create 3D strings.
3) Condensation of 3D strings.
4) Joining 3D strings.

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

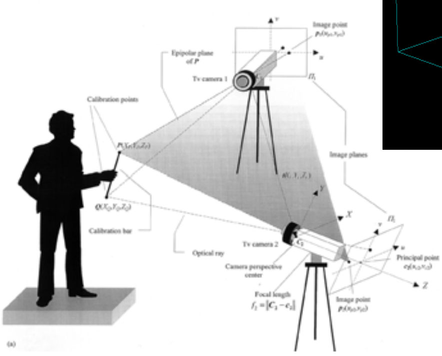


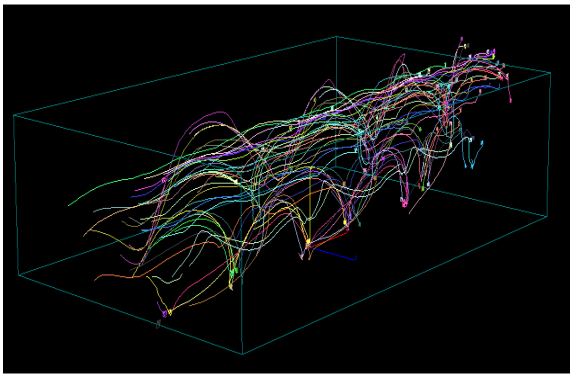
1) Creation of 2D strings



2) Matching 2D strings

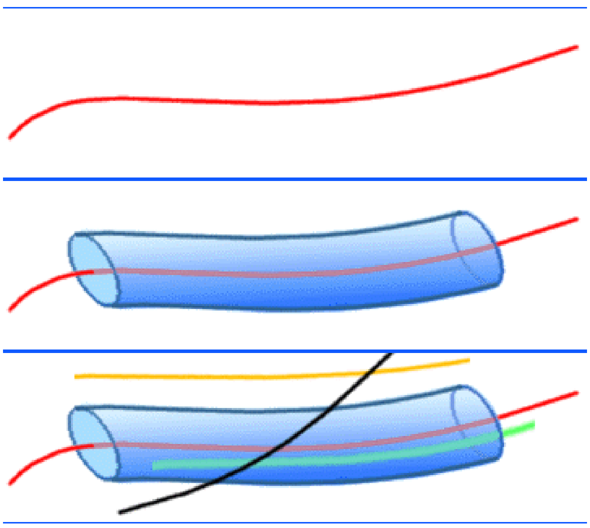
Epipolarity constraint



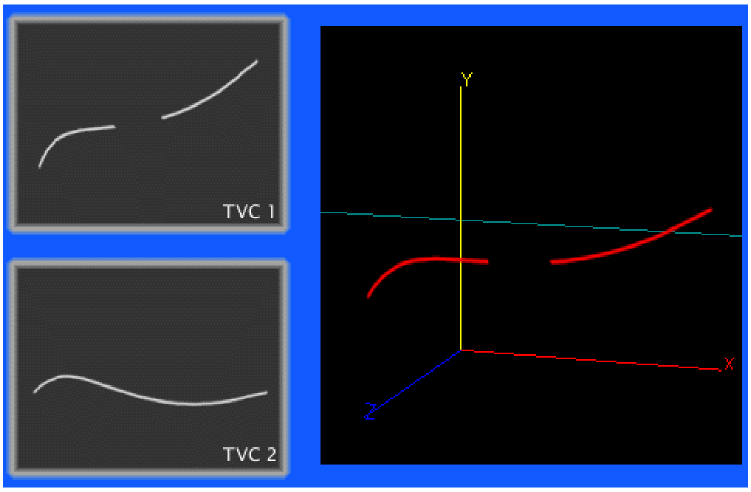


3D strings

3) Condensation of 3D strings



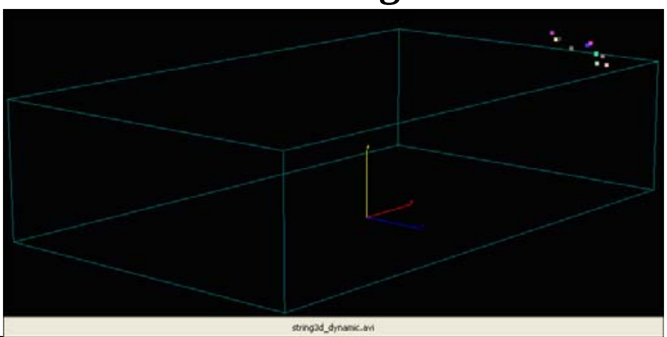
4) Joining 3D strings

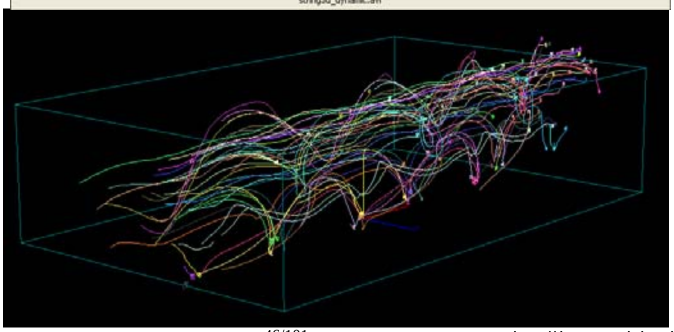


3D strings

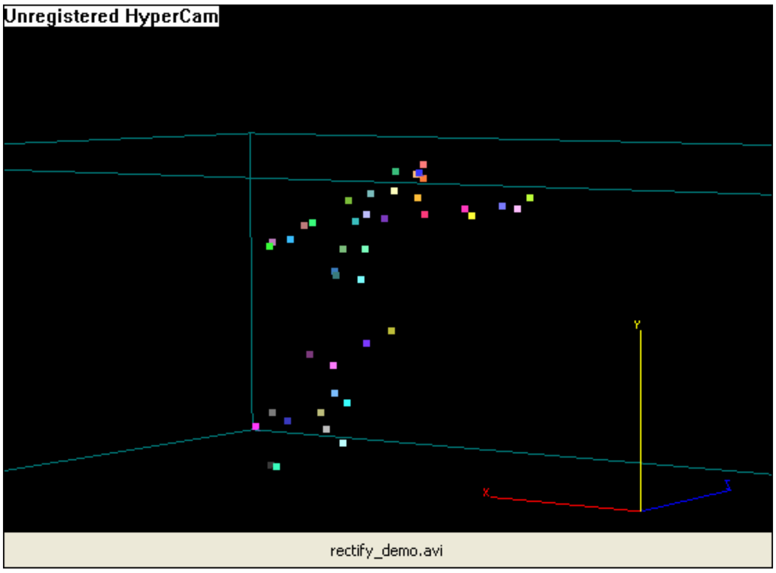
3D strings already contain motion 3D information

3D strings

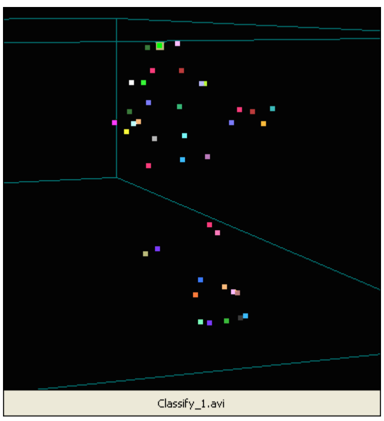


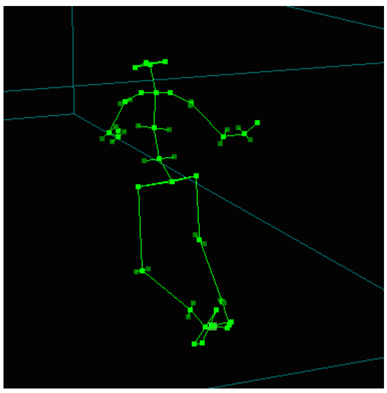


Markers Classification

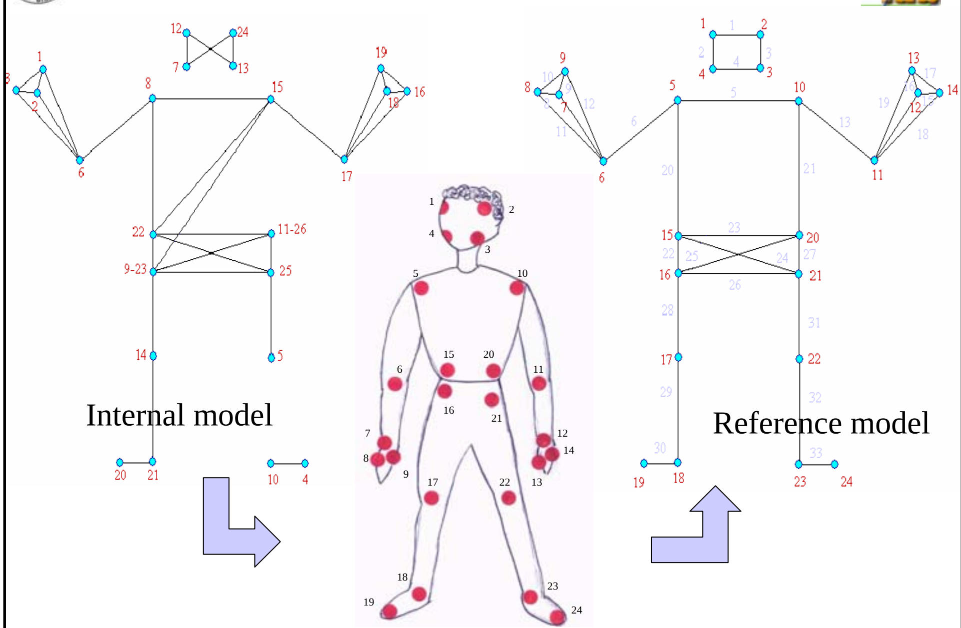


5) Initial classification

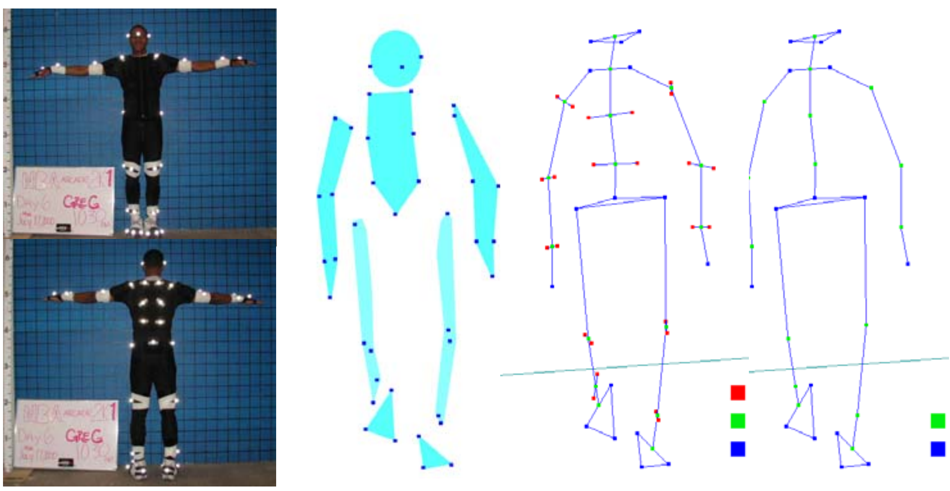




Model fitting



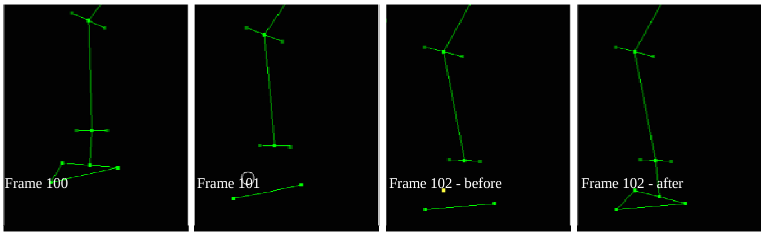
What a model represents?



Markered subject Modello 3D Modello a stick Modello hidden

6) Classification extension

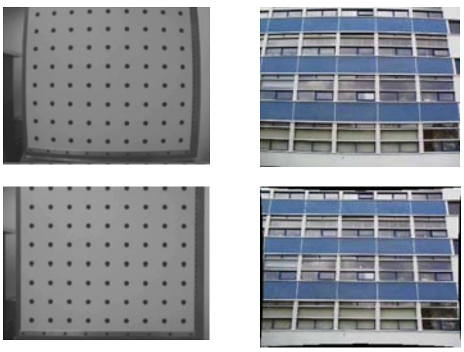
3D strings are automatically extended in this phase.



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

Calibration is a pre-requisite

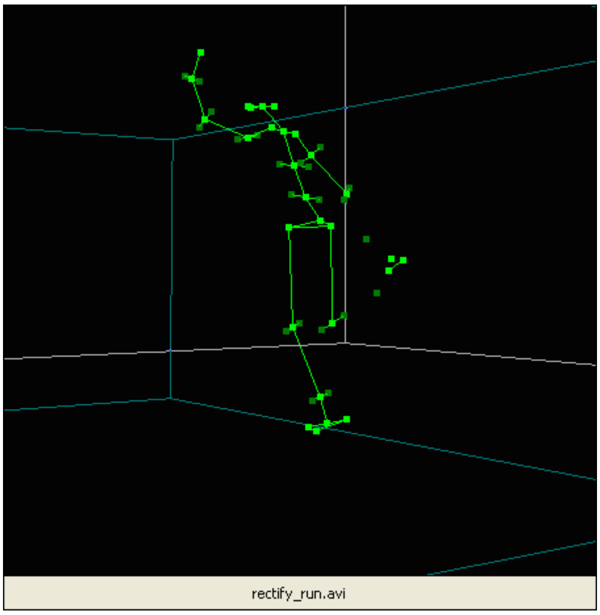
Camera calibration



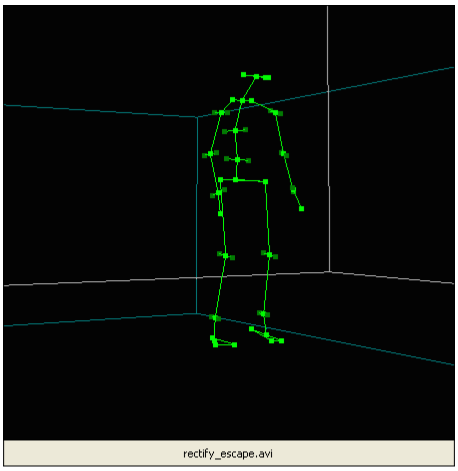
Set-up calibration

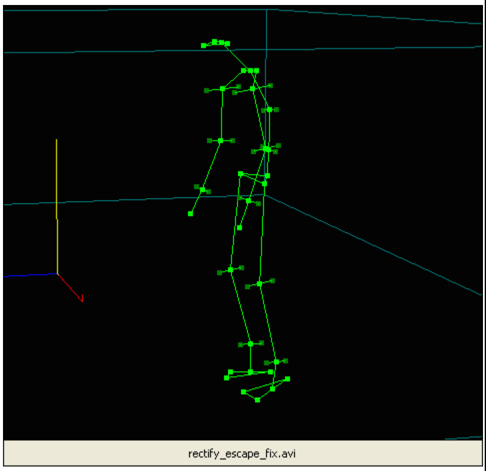
Excellent for special effects, not so good for measurements….
Cameras are not metric.

Risultati: run

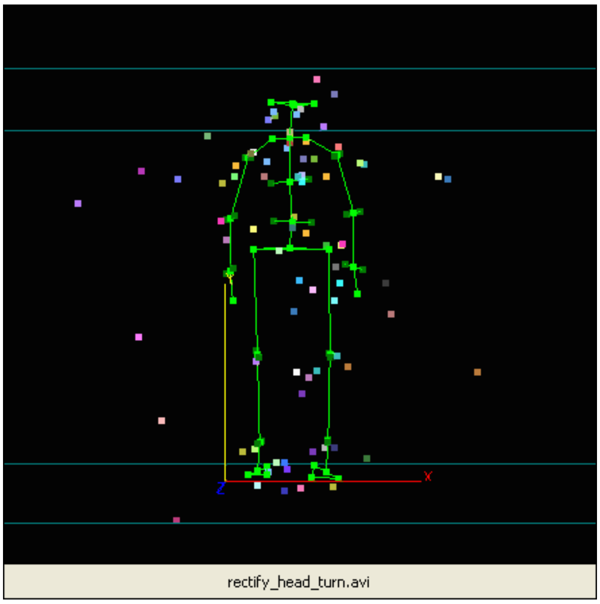


Risultati: escape

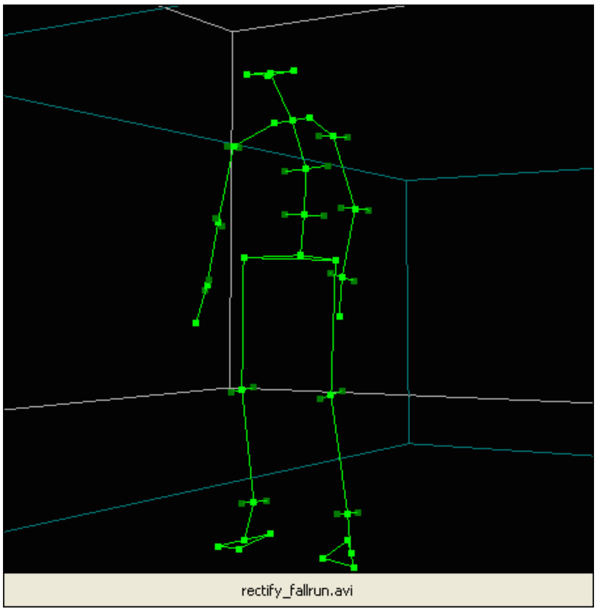




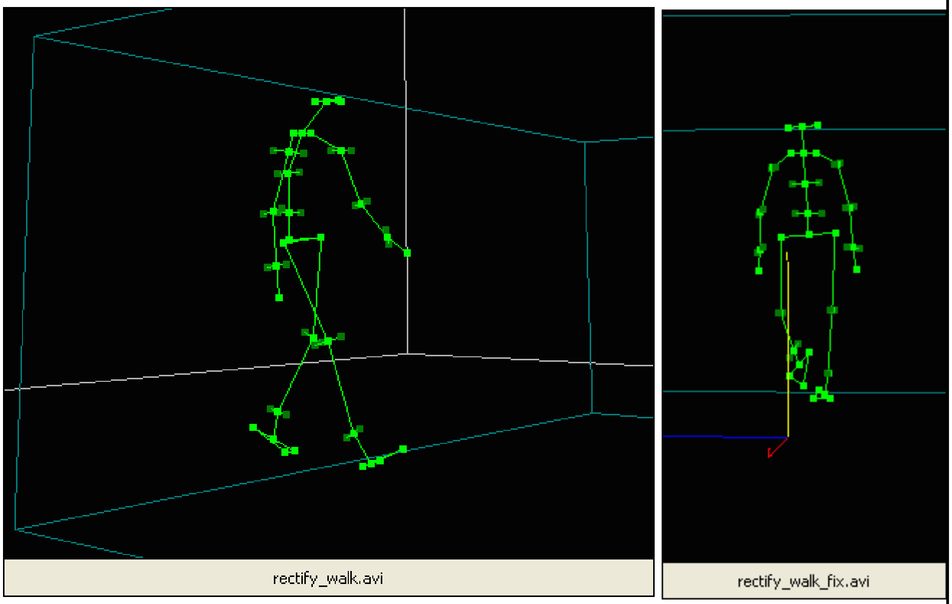
Risultati: head_turn



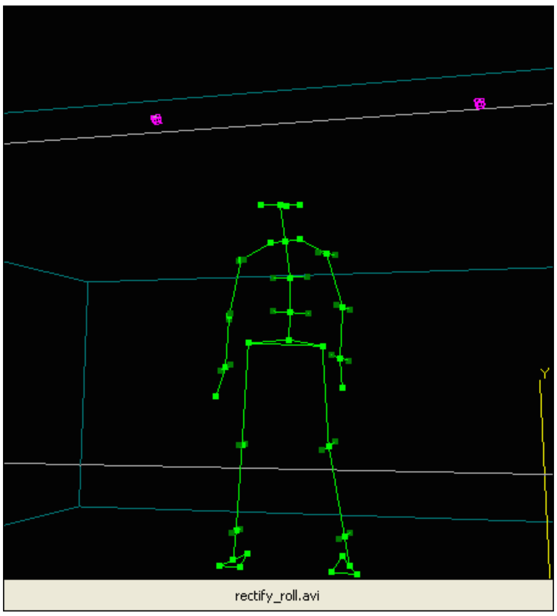
Risultati: fall_run



Risultati: walk



Risultati: roll



A motion capture system

Spectrum portfolio

Problems with motion capture

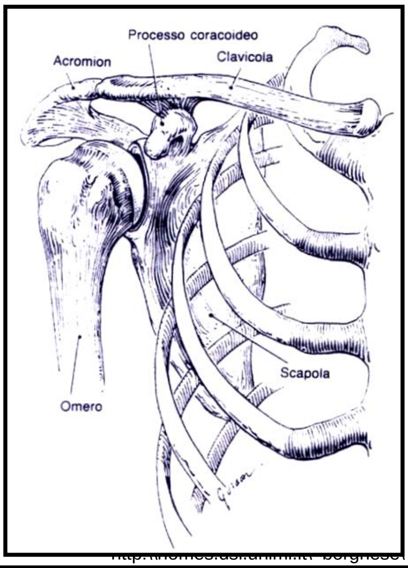
- Joints are points inside the body, markers are attached on the body surface.
- Joint are not fixed points: two adjacent bones rotate and slide.
- Joint are not spherical.
- Skin artifacts.

The human skeleton has complex articulations

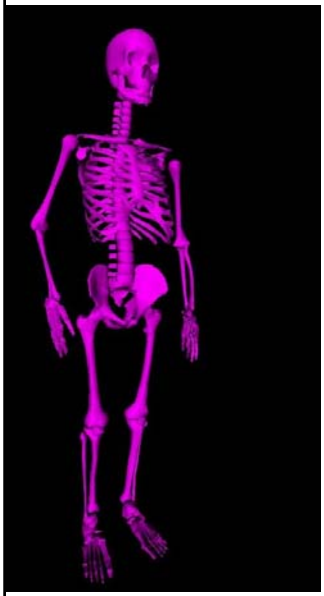
"Rigid" bones connected. Tendons keep the bones in place.

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

Animating with MoCap data

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

We measure the position and orientation in time of the segments. What does this measure represent?

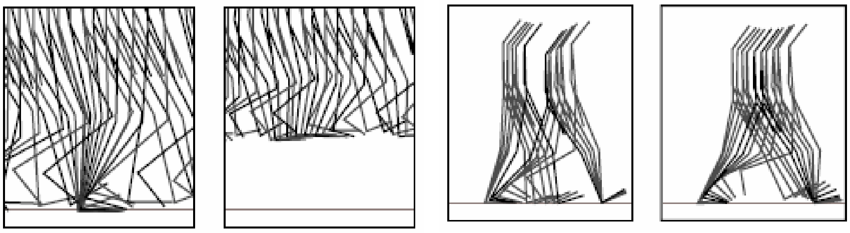
We can transform the measurements in angle time sequence but:

If we animate a skeleton constituted of segments of fixed length, which are the proper angles sequence?

If we change the segments length, how should the angles sequence change?

Motion correction & retargetting

Frame by frame techniques



Scaling
“flying effect”

Inverse kinematics:
Motion discontinuity

Space time solution

Minimal deviation of the control actions (the angle sequence).
The angles variation time course should have the minimum frequency content.
Some constraints have to be added (e.g. Foot contact with the floor contact during waking).

Non-linear optimization with soft and hard constraints.

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

Problema di controllo ottimo discreto.

More in the course of Robotics

A. Witkin and M. Kass, Spaces time constraints. Proc. Siggraph 1988.
M. Cohen. Interactive sapcetime control for animation. Proc. Siggraph 1992.
A. Witkin and Z. Popovic, Motion Warping. Proc. Siggraph 1995.
M. Gleicher, Retargetting Motion to New Characters, Proc. Siggraph 1998.

The future

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

Color-coded markers.

Mixed vision/marker techniques or pure vision techniques (marker-less).

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?

Input systems::Video motion capture

Optical systems (computer vision)

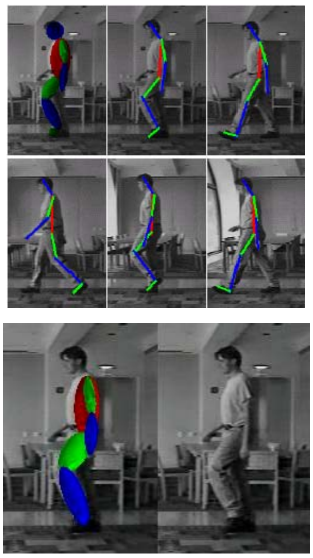
•**Advantage:** complete freedom of motion to the subjects. The scene is surveyed by standard videocameras.

•**Disadvantage:** ill-posed problems (high sensitivity to limited resolution, noise and lighting conditions).

• **Solution**: hierarchical multi-stage processing.

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).

Pre-prototype research.

http://movement.stanford.edu/

Face motion from footage

Reconstructing talking faces from footage (range points -> 3D model -> deformation) + Estimate of the camera geometry.

3D model construction through image processing techniques:
- Cross-correlation matching.
- Area matching.
to identify features or virtual markers.

Initialization through a semantic model or several manual identified points.

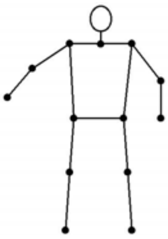
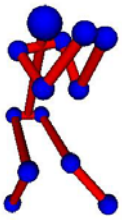
3D reconstruction through:
- Bundle Adjustment.
- Reinforcement of the matching through multi-view geometry.

Robust and Rapid Generation of Animated Faces From VideoImages: A Model-Based Modeling Approach
Zhengyou Zhang, Zicheng Liu, Dennis Adler, Michael F. Cohen, Erik Hanson, Ying Shan Technical Report: MSR-TR-2001-101

Body motion from footage

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 photogrammetric approach

 +  -> 

Anthropometry is defined as stick diagrams. Points play the role of markers.
Identification of key positions of the model (eventually by image processing)
Calibration and refined interpolation to obtain continuos motion.
Extension of the Bundle-adjustment method to incorporate motion parameters.

http://www.photogrammetry.ethz.ch

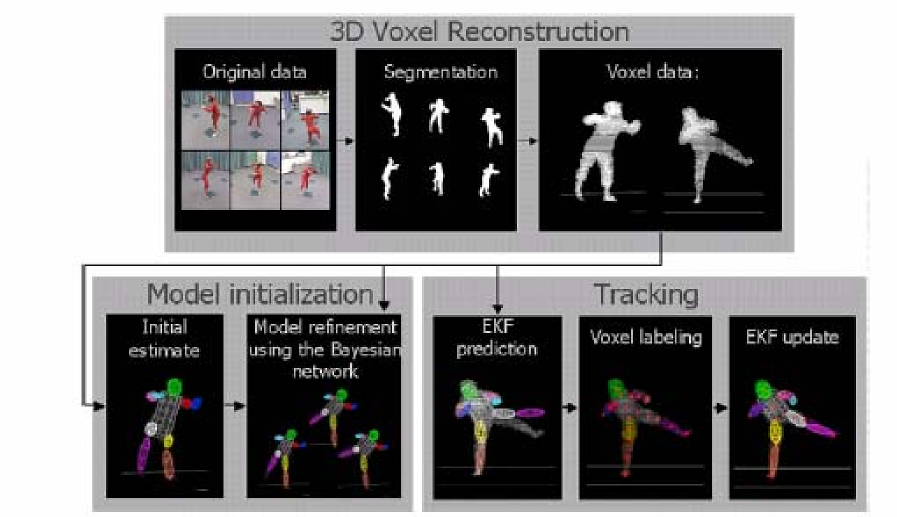
The volumetric approach: a possible solution

Mikic et al., Human Body Model Acquisition and Tracking Using Voxel Data, Int. J. Computer Vision, 53(3), 2003.

Cheung et al., A real time system for robust 3D voxel reconstruction of human motions. Proc. Ieee Conf. CVPR, 2000.

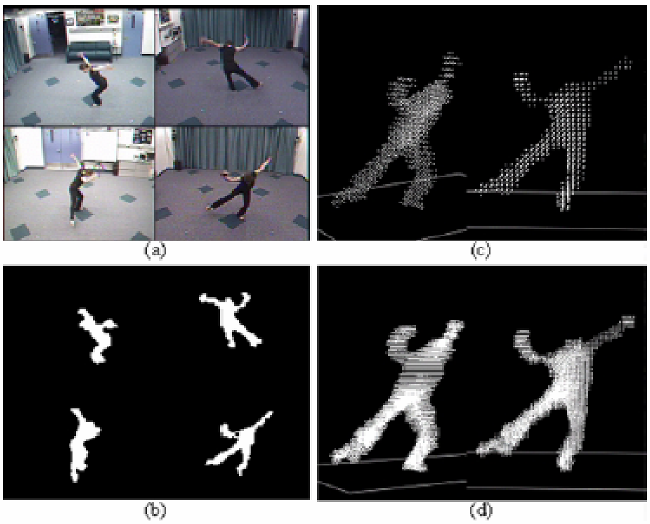
Jain et al., 3D video, Proc. VRAIS, 1994.

I passi di elaborazione (Mikic et al.)



Voxel reconstruction

- Extraction of silhouette from the background (median filtering, produces high quality silhouettes).
- Compute the 3D bounding boxes associated to multiple silhouette.
- Voxel carving through Octree processing to compute the voxel reconstruction.
- Multiple cameras allow increasing resolution.

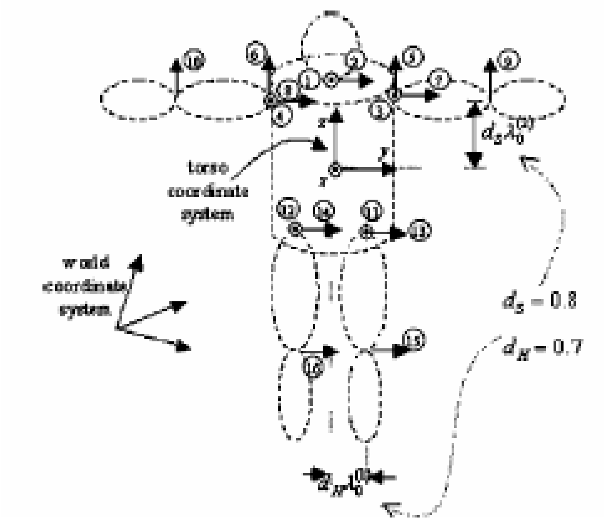


c) 50mm d) 25mm

Model adopted

Ellissoidi incernierati per gli arti, cilindro per il tronco e sfera per la testa. Movimenti relativi al tronco: catena cinematica.

Inizializzazione: corpo a gambe e braccia tese con le braccia rivolte verso l'esterno.



Model acquisition

Model "anthropometric" parameters have to be adjusted.
The different anatomical segments have to be identified.

Two steps-process:

Initialization

Refinement (given all the measured segment lengths, which would be the most probable length of each segment?).

Model initialization

1) **Head:**

Spherical template with minimum and maximum radius.

The center of the sphere is computed as the point, which maximizes the inner voxels.

The neck is computed as the average position of the head voxels, which have an adjacent non-head voxel.

2) **Torso:**

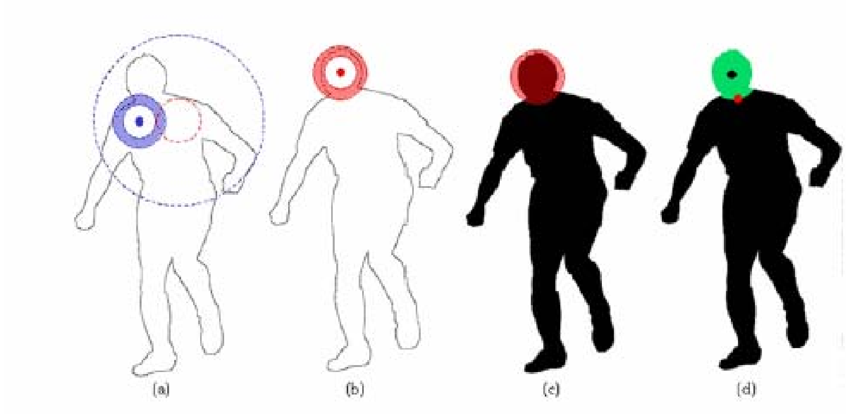
An average cylinder is attached to the neck and oriented as the centroid of the remaining voxels.

It is then shrunk and grown again until it incorporates empty voxels.

Every k-steps, its orientation is recalculated.

3) **Limbs:**

Head initial identification

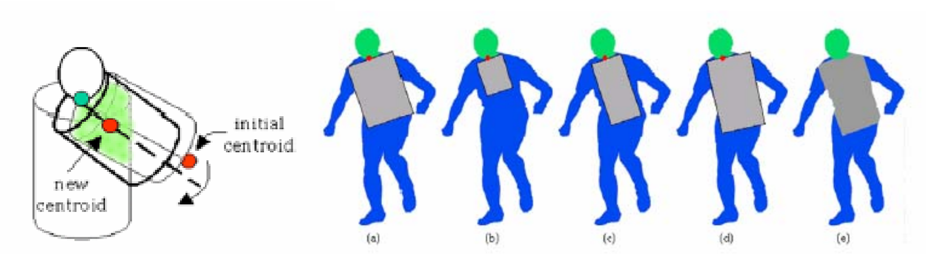


Spherical template with minimum and maximum radius.

The center of the sphere is computed as the point, which maximizes the inner voxels.

The neck is computed as the average position of the head voxels, which have an adjacent non-head voxel.

Identification of the trunk



An average cylinder is attached to the neck and oriented as the centroid of the remaining voxels.

The inner voxels are determined and the new centroid computed. Cylinder is reoriented.

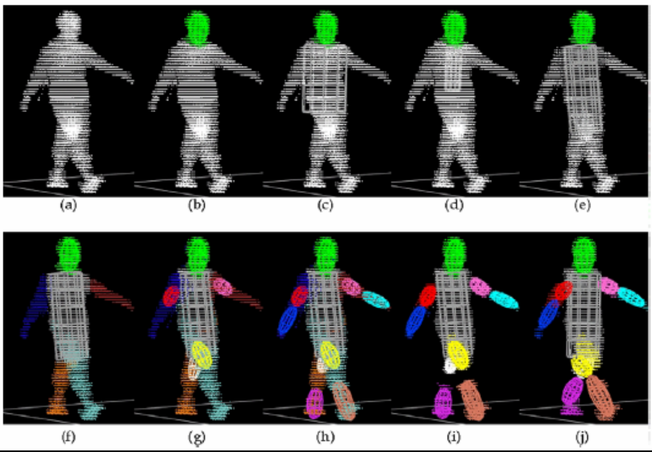
It is then shrunk and grown again until it incorporates empty voxels.

Limb segments acquisition

Are identified as the four connected regions of remaining voxels.

Hip and shoulder are the average of the position of the torso voxel adjacent to legs and arms.

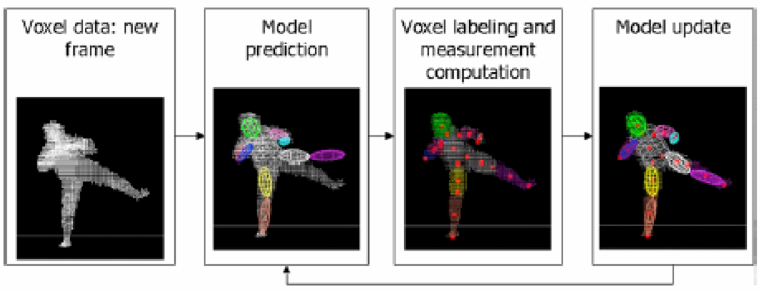
The limb segments (arm and forearm and thigh and calf) are identified with the same shrinking and growing procedure used for the trunk.



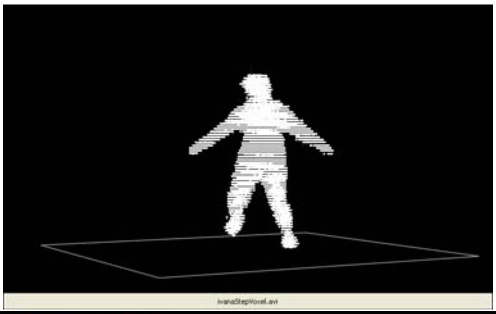
Tracking

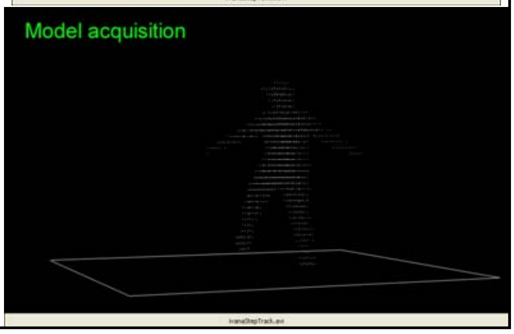
Kalman filter.

- Silhouettes → Voxel reconstruction.
- Model prediction through Kalman filter → Voxel labeling.
- Measurement on labelled voxels of body positions (end-points and centroid of the segments).
- From the difference from the measurement and the prediction, the new position of the model is determined.

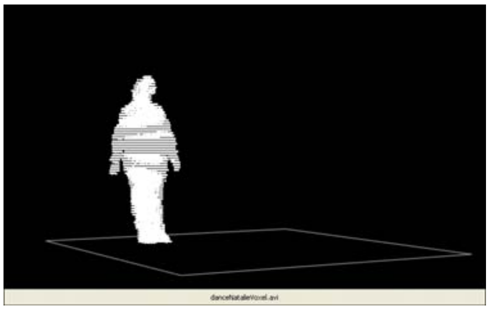


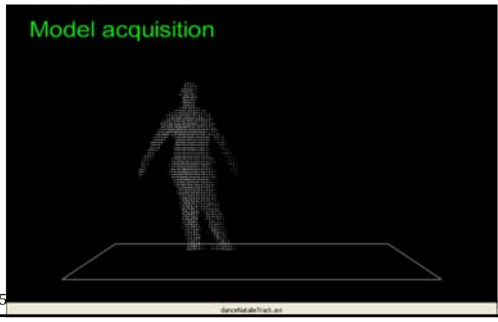
Results: stepping (640 x 480, 10Hz)



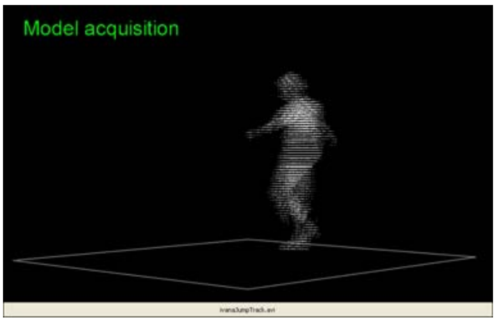


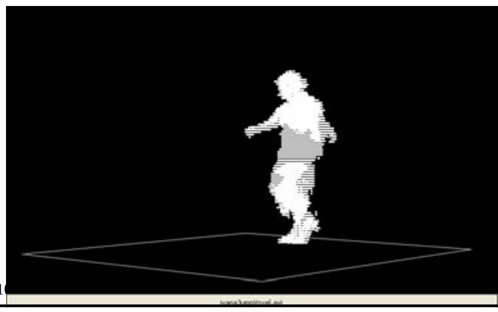
Results: dancing



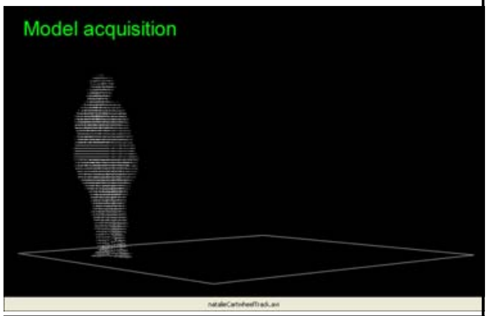


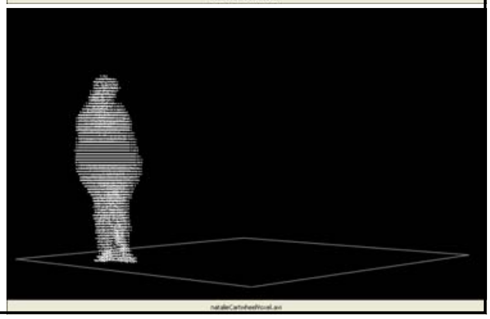
Results: jumping





Results: cartwheel





Motion capture for animation

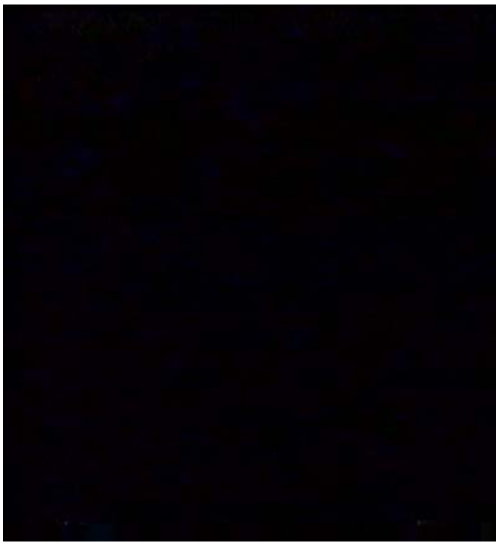
•Motion capture

•Definition of a 3D model.

•Mapping of the motion onto the 3D model.

•Animation.

Video by Superfluo



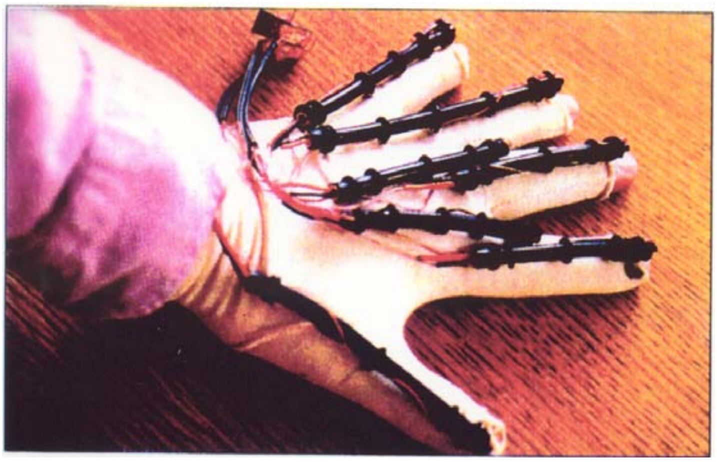
Input systems::Gloves&Gaze trackers

Gloves

Monitor fingers position and force.

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

Sayre glove (1976)



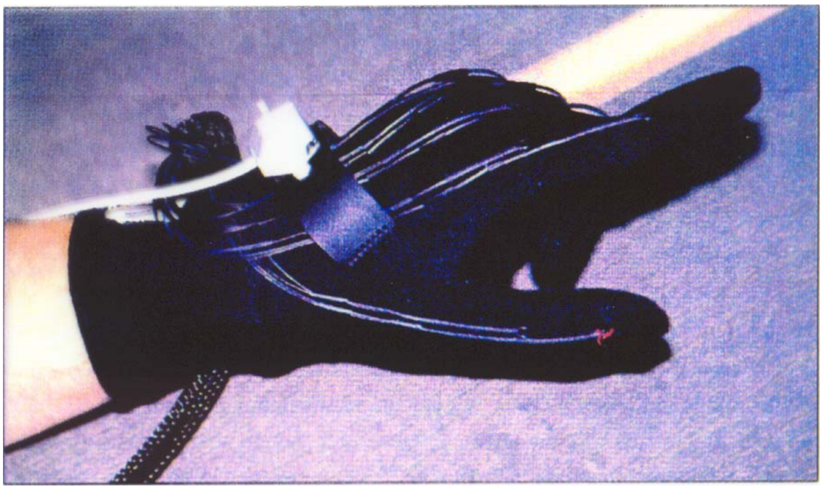
MIT glove (1977)

A.A. 2006-2007 93/101 http:\\homes.dsi.unimi.it\~borghese\

Digital Data Entry Glove (1983)

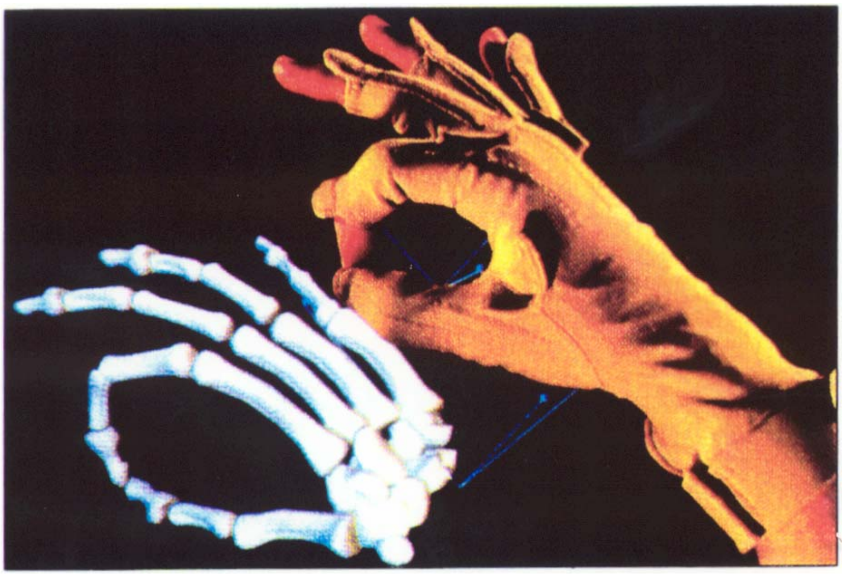
A.A. 2006-2007 94/101 http:\\homes.dsi.unimi.it\~borghese\

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).

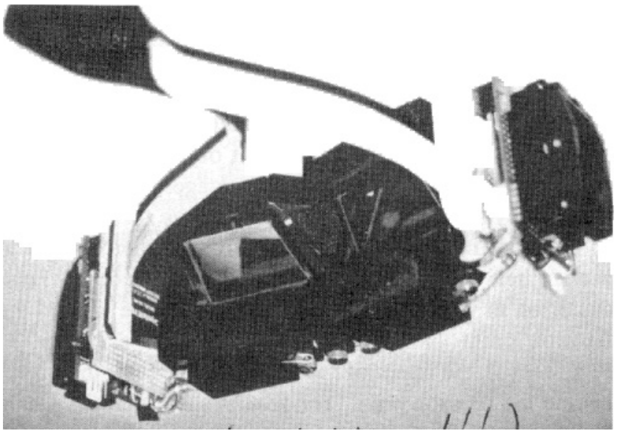
Estimate of the parameters, which describe distortions introduced by the measurement system.

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

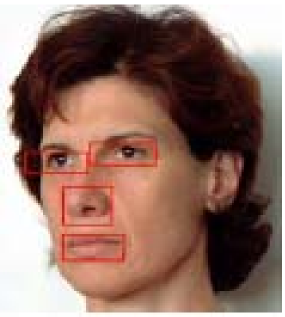
Algorithms adopted: polynomial, local correction (neural networks, fuzzy).

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).



Vision based eye trackers



Color information
Geometry information (circles, relative position...)
Histogram analysis on gray level.
...

Sommario

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