Introduzione alla Realtà Virtuale

Alberto Borghese

Sommario

- Introduzione
- Sistemi di Input
- Generatori di mondi
- Motore di calcolo
- Sistemi di Output
- Conclusioni
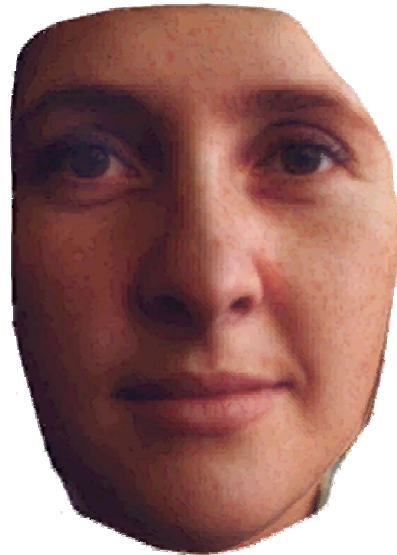
Which is real, which is virtual?

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.
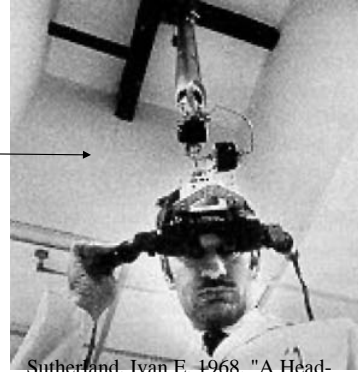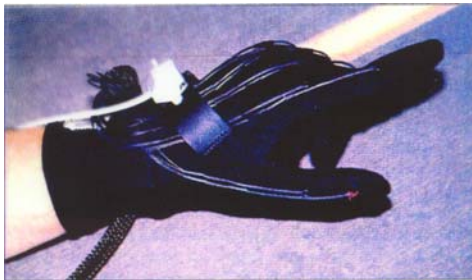• Hegel ("what is rational is real..") 1770-1831.
• New age.

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

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 …)

6

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
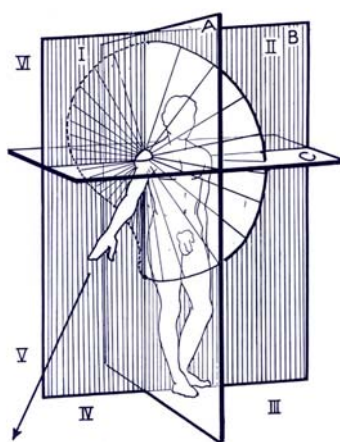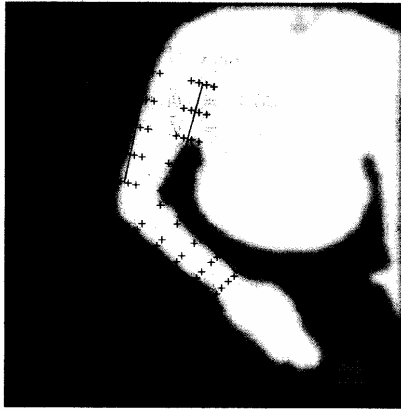

A – Frontal plane
B – Sagittal plane
C – Horizontal plane

Abduction/adduction
Flexion/extension
Axial rotation (V)
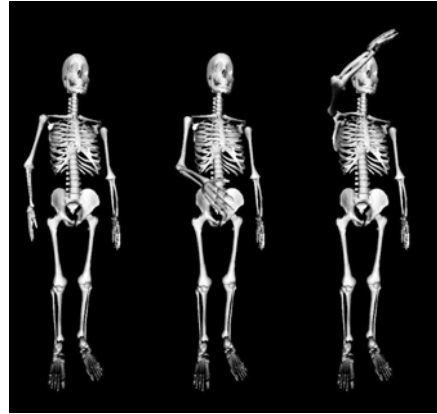
Definition of the interesting degrees of freedom.

What is captured?

Silhouette (-> Skeleton)   Skeleton

Computer vision techniques   Bony segments or articulations
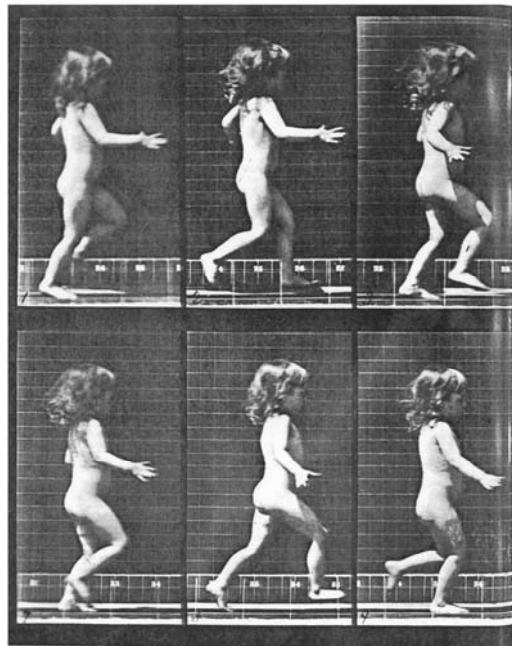   (marker-based systems)

Marker-based techniques

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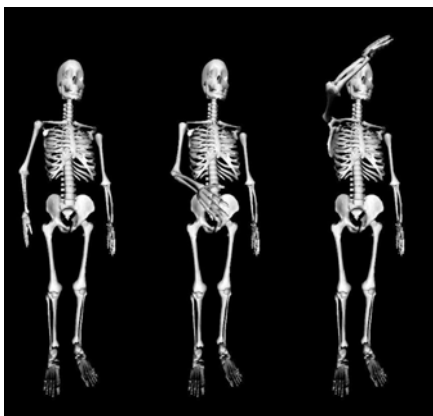
Here, the problem is to find a suitable *marker* for the segments and a suitable HW/SW system for *marker detection*.
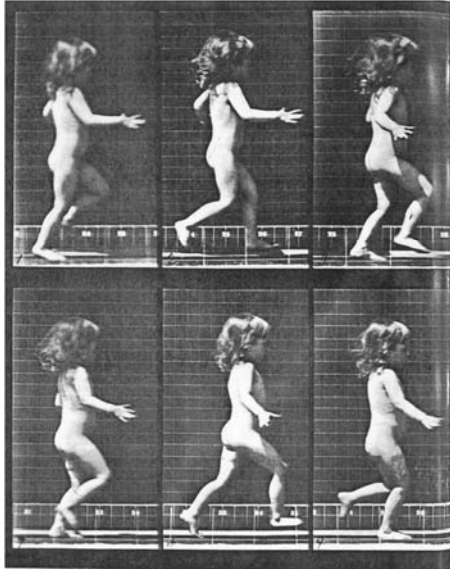
Bony segments or articulations.

Input position
system::history&technology

Edward Muybridge 1878-1901

9

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.

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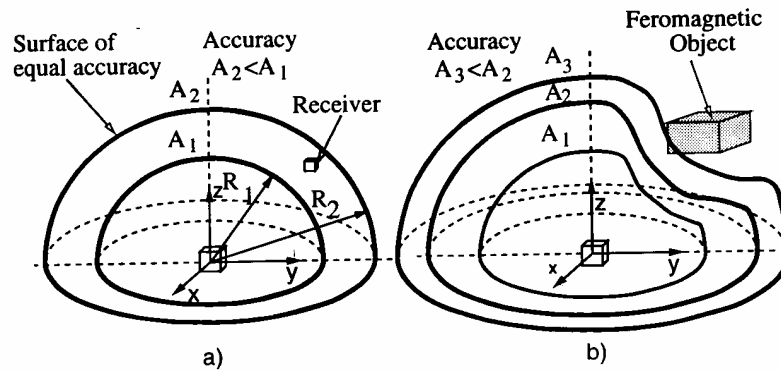
•Latency: 4msec.
•Sampling rate: 120Hz. Rate drop with multiple receivers because of multiplexing.

Flock of birds Motion Capture
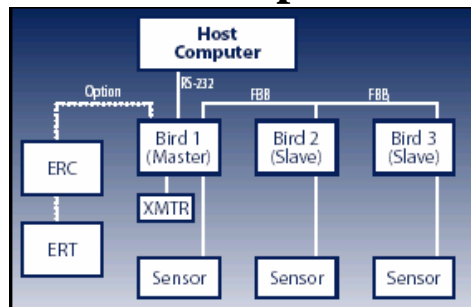
•Each receiver has its own DSP.
•All the DSP are connected with a fast internal bus.
•Latency is increased (8ms).

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

13
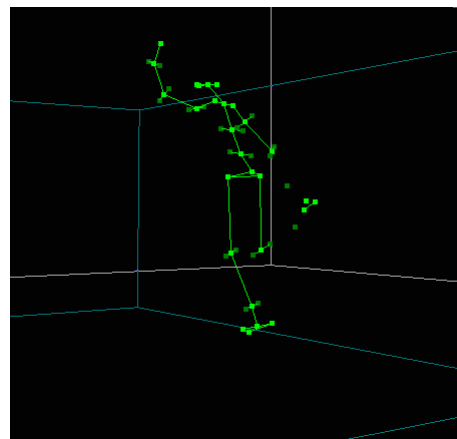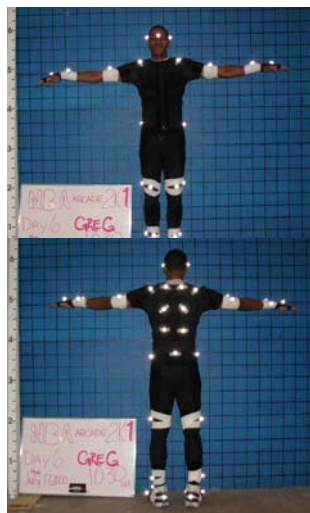
Why passive markers?

No encoumbrance on the subject: markers do not require any
powering and are hardly sensed by the subjects.
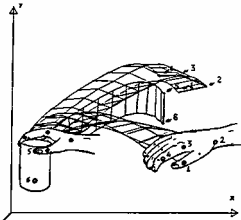
No constraint on the dimension of
the working volume is prescribed.
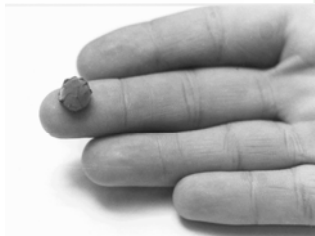
How passive markers work?

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*).

High-level Vision

An implicit step is CALIBRATION.

Two-levels architecture

Implementable with a DSP
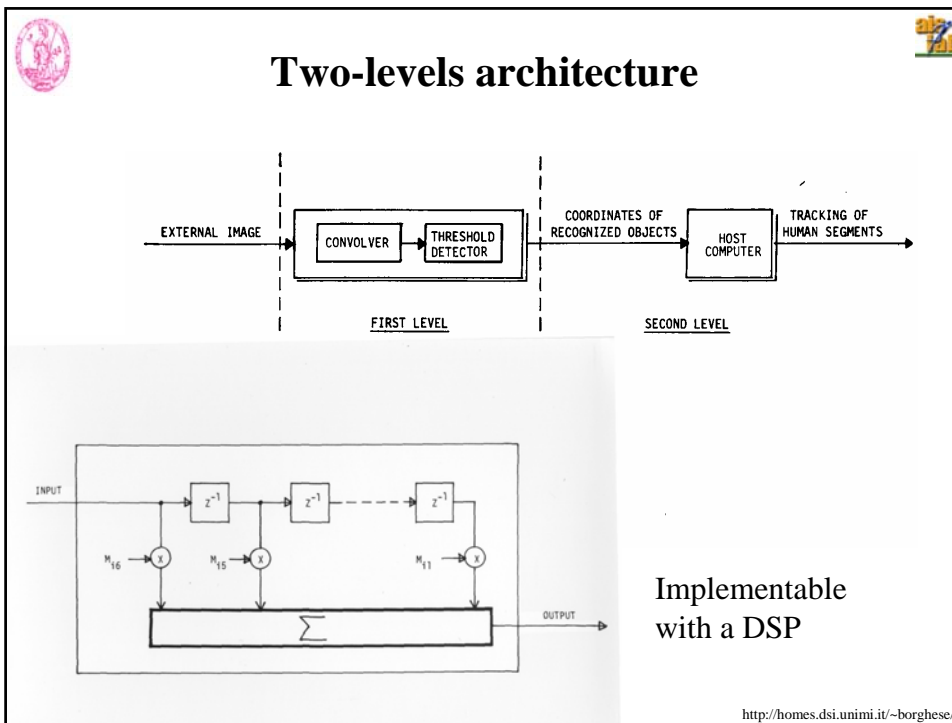
Markers extraction through thresholding (without flash)

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

Cluster dimension.

Shape.

Software protection of bright target regions.

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.

http://homes.dsi.unimi.it/~borghese/
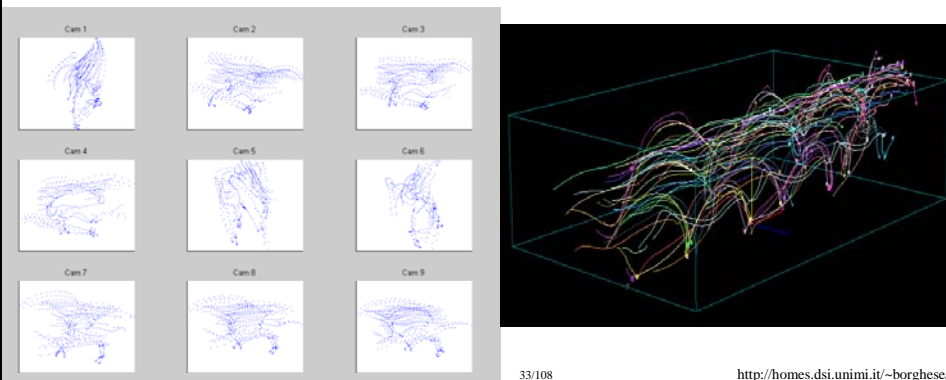
The difficulties in data processing

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.

17

From 2D to 3D

Each camera measures a geometrical transformation (projection)
Triangulation (ray intersection)
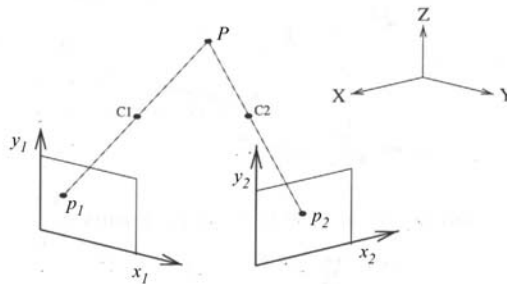
Geometrical parameters known.
Main difficulty is correct matching between multiple markers and
multiple cameras.

Tracking difficulties

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.
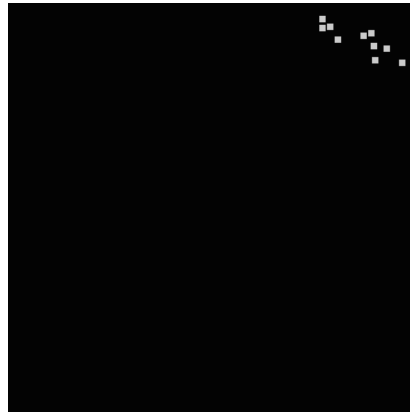
Tracking processing steps

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
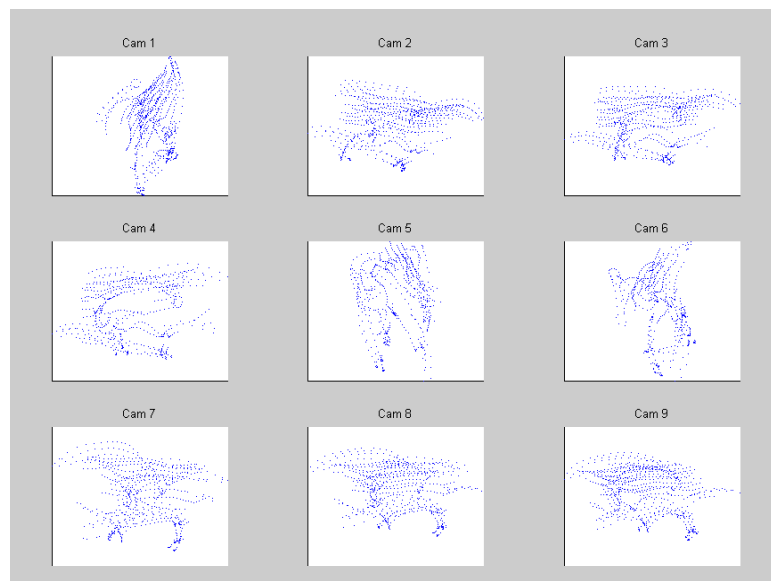
19

1) Creation of 2D strings

2) Matching 2D strings

Epipolarity constraint

3D strings

20

3) Condensation of 3D strings

4) Joining 3D strings

TVC 1

TVC 2
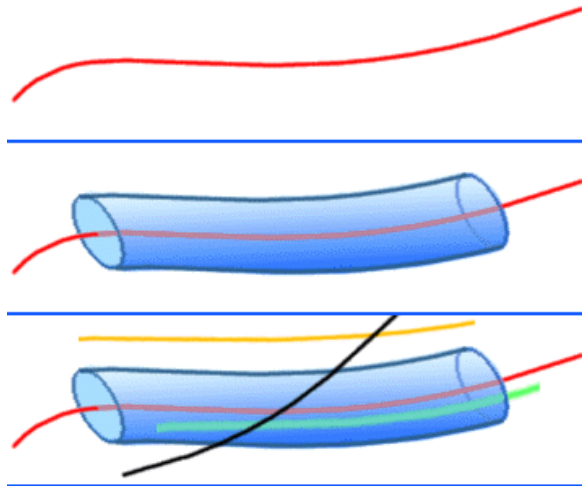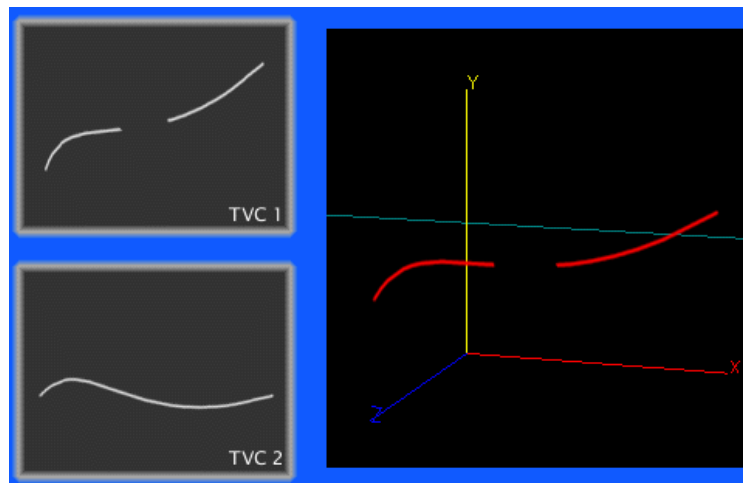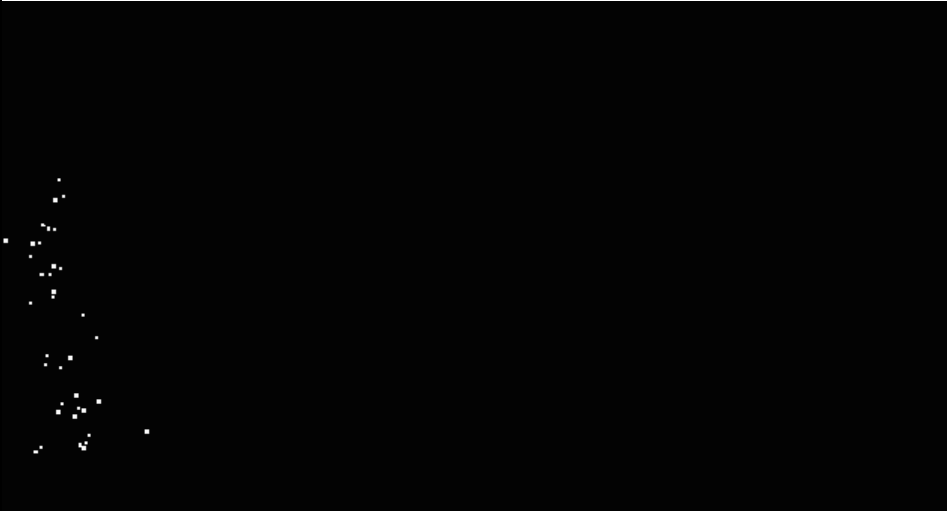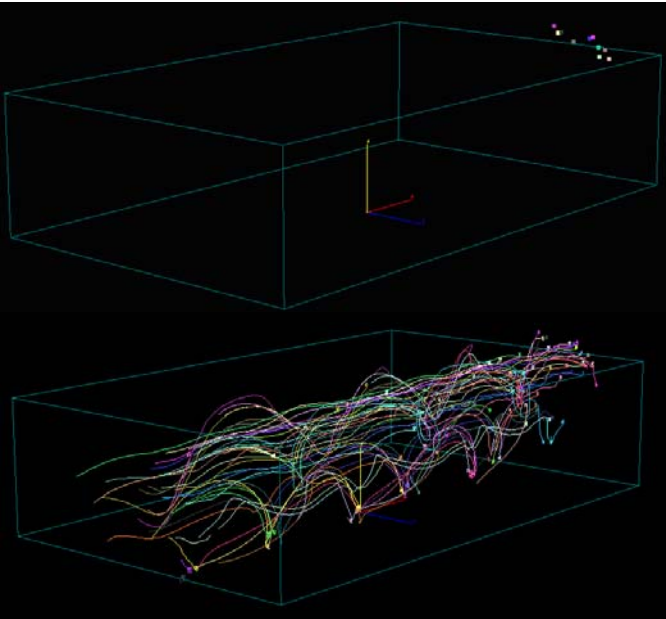
3D strings

3D strings already contain motion 3D information

3D strings

22
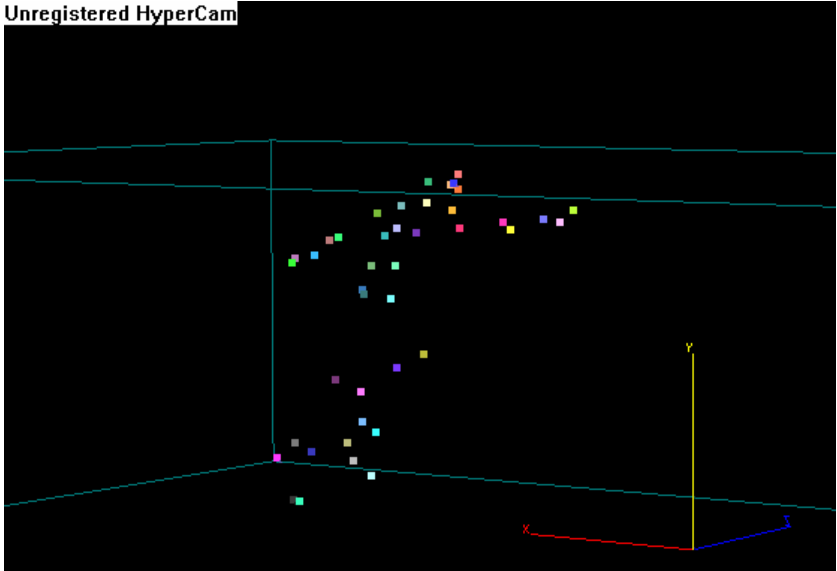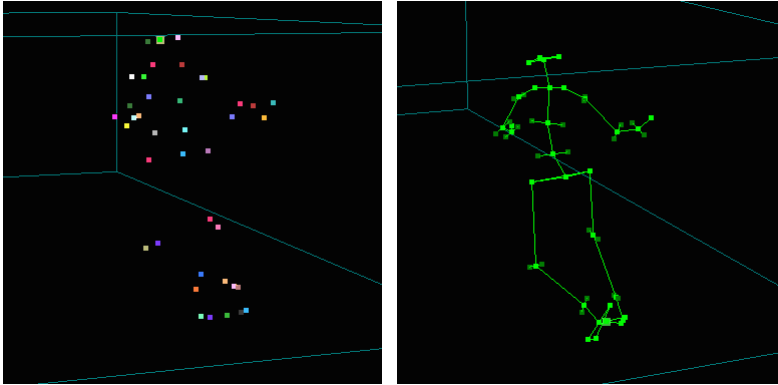
Markers Classification

5) Initial classification

Model fitting

Internal model

Reference model

What a model represents?

Markered subject   Modello 3D   Modello a stick   Modello hidden

24
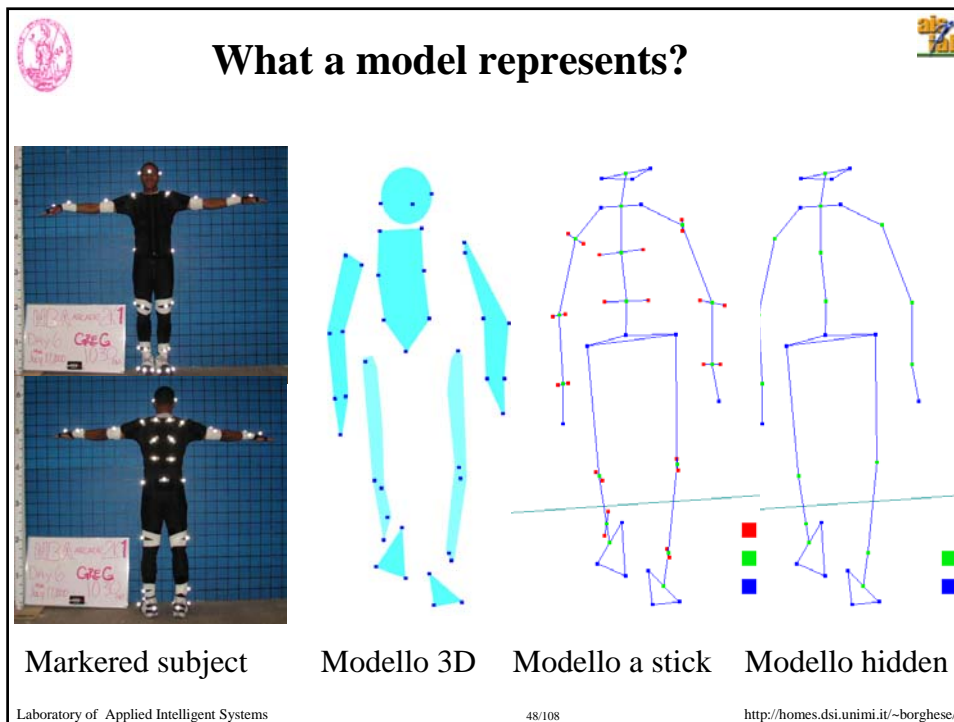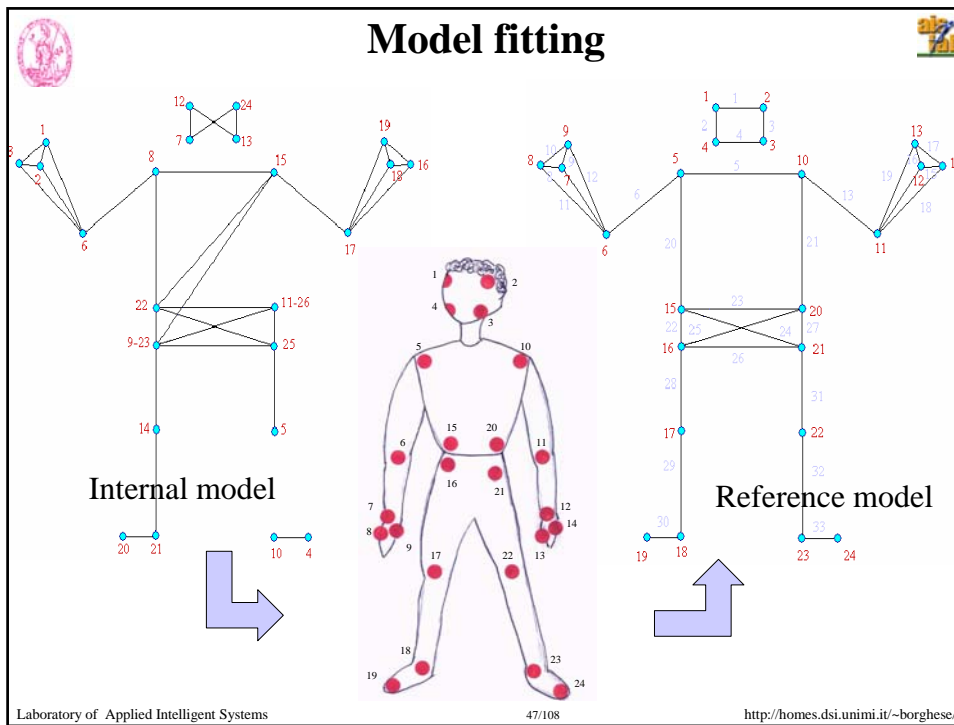
6) Classification extension

3D strings are automatically extended in this phase.

Frame 100   Frame 101   Frame 102 - before   Frame 102 - after
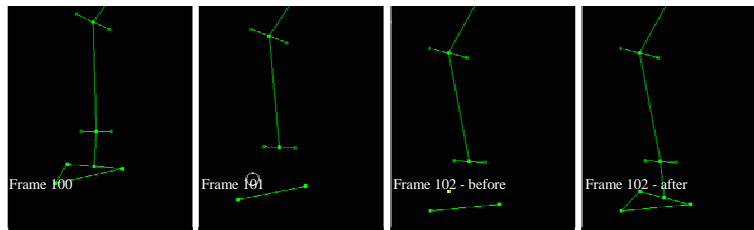
Two strings are joined on the base of:
- Smooth motion.
- Model checking (a dynamic prioirty is coded in the number of links).
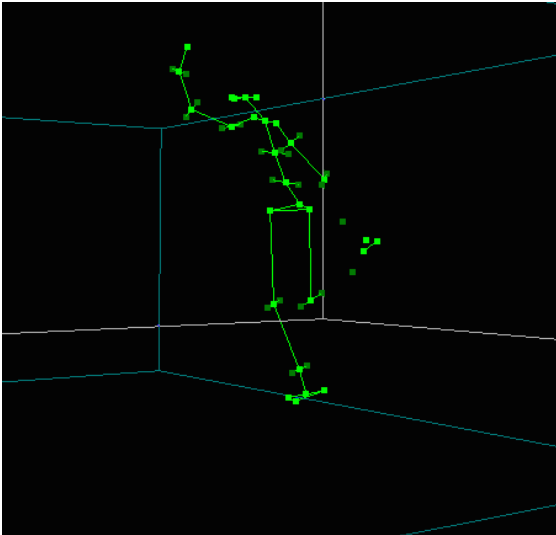
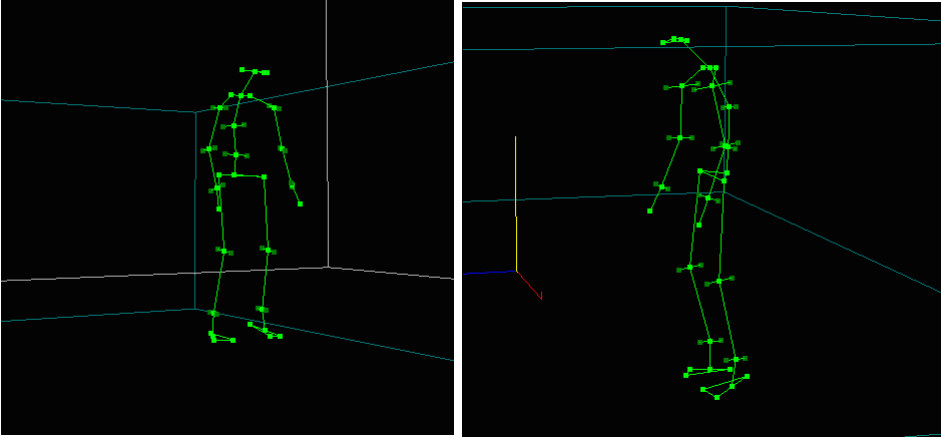# Calibration is a pre-requisite

Camera calibration
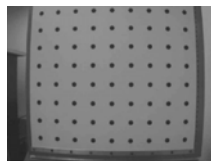
Set-up calibration
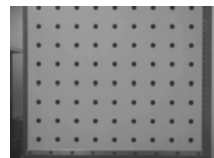
Excellent for special effects, not so good for measurements….
Cameras are not metric.

Risultati: run
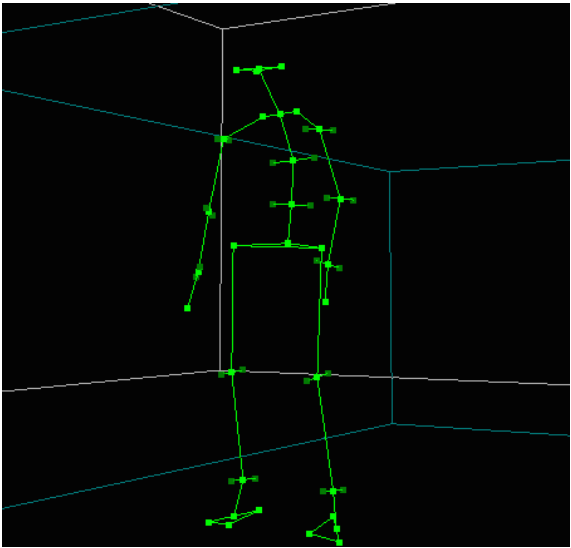
Risultati: escape

26

Risultati: head_turn

Risultati: fall_run

27
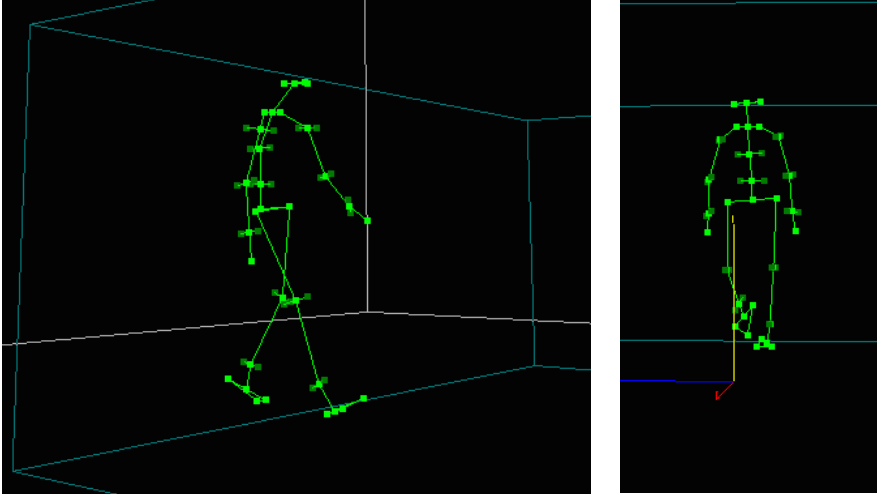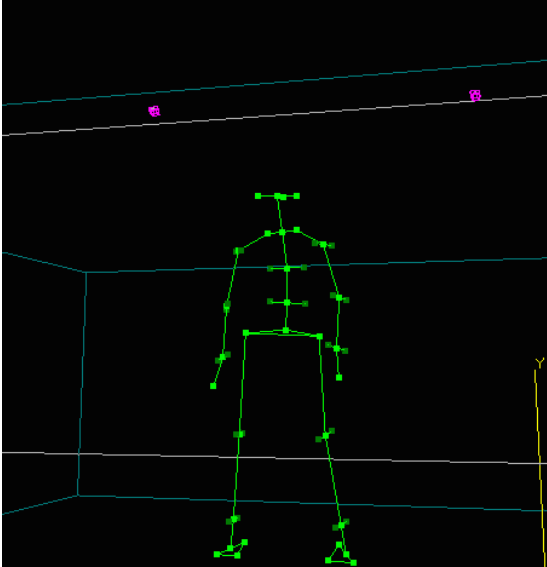
Risultati: walk

Risultati: roll

28

A motion capture system

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.

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.
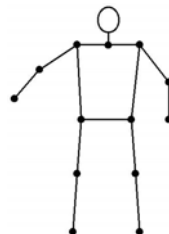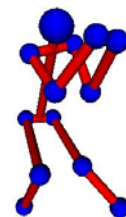
A photogrammetric approach

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

Computer vision techniques

Silhouette (-> Skeleton)
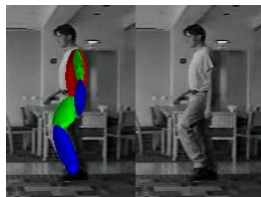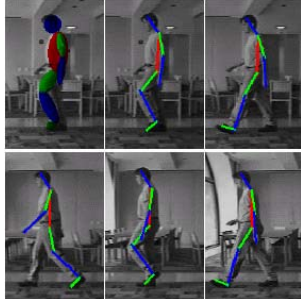
Set of difficult problems:

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/

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.)
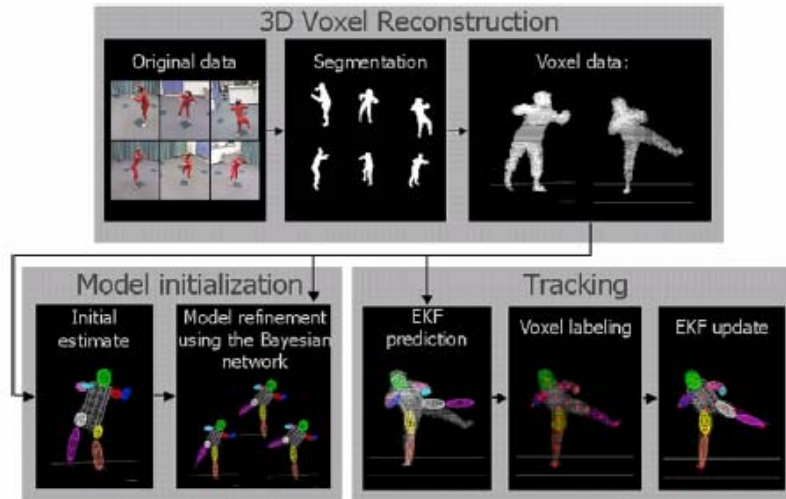
Voxel reconstruction

- 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.
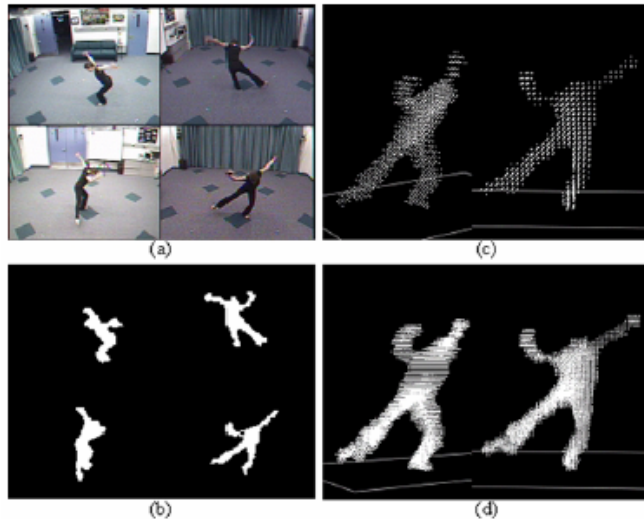
c) 50mm   d) 25mm

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.
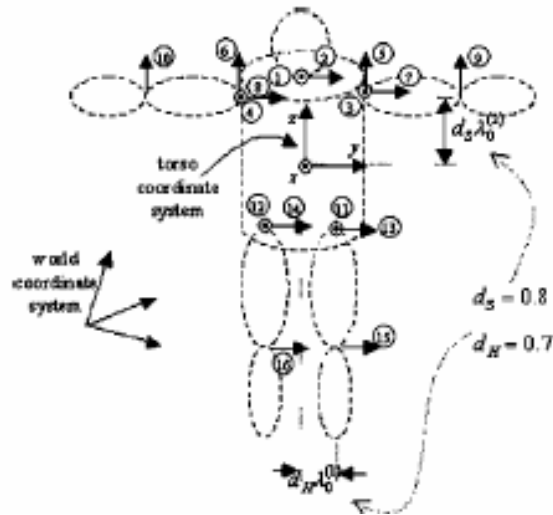
Model acquisition

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
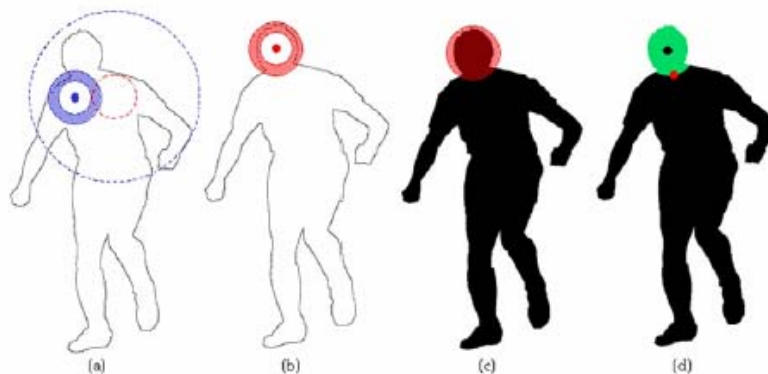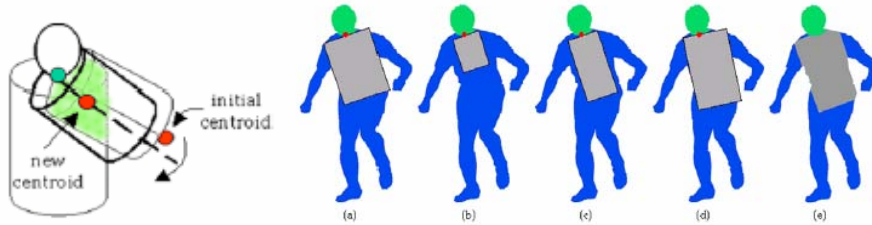

Spherical template with minimum and maximum radius.

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.
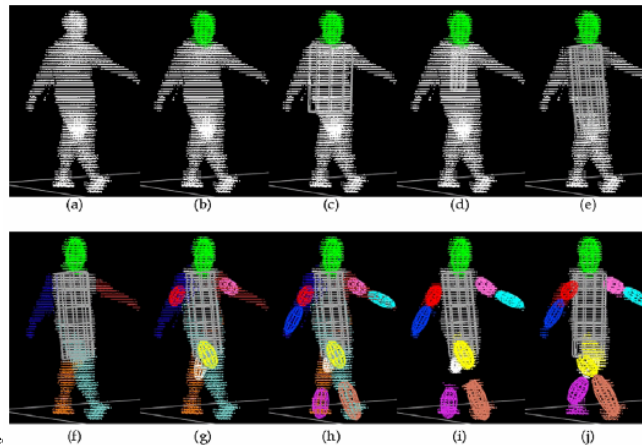
Tracking

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.
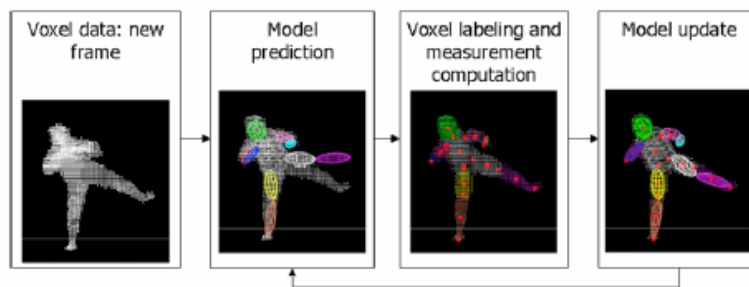
Results: stepping (640 x 480, 10Hz)
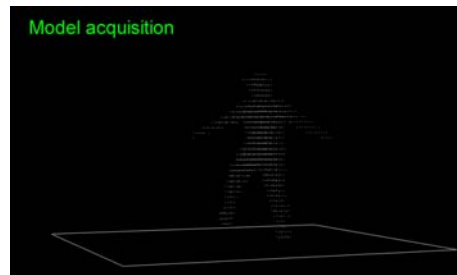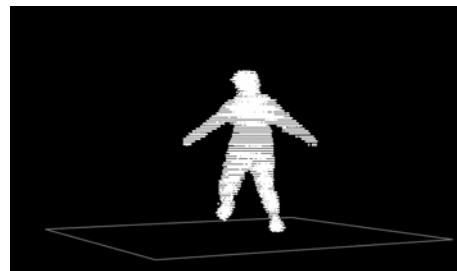
Model acquisition

Results: dancing
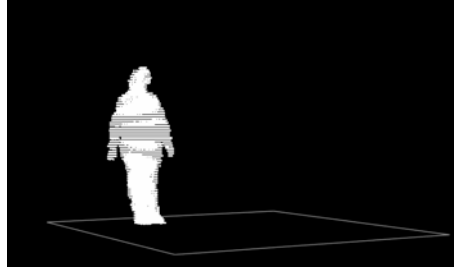
Model acquisition
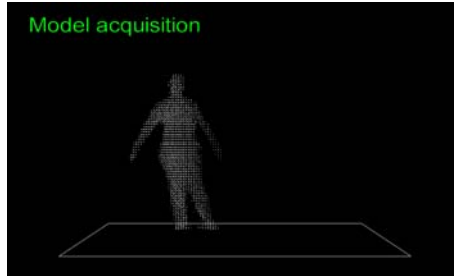
Results: jumping

Model acquisition

38
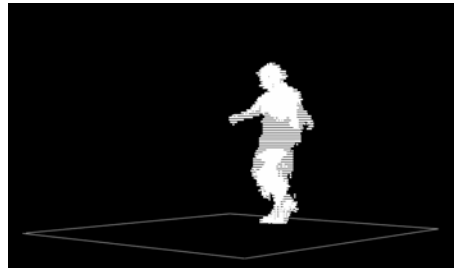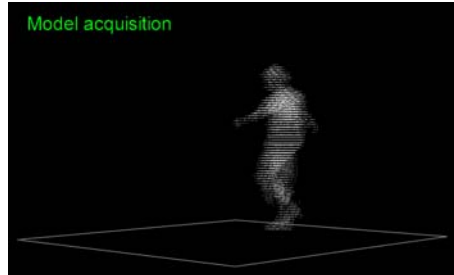
Results: cartwheel
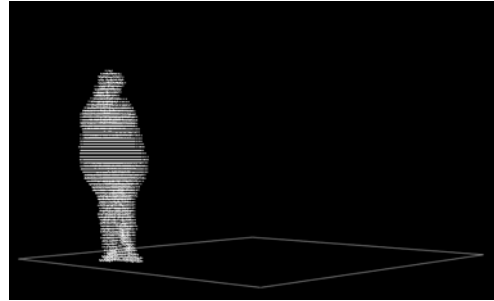
Model acquisition
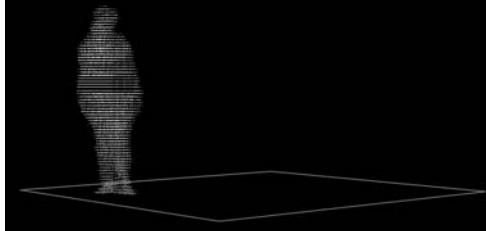
Motion capture for animation

- 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)
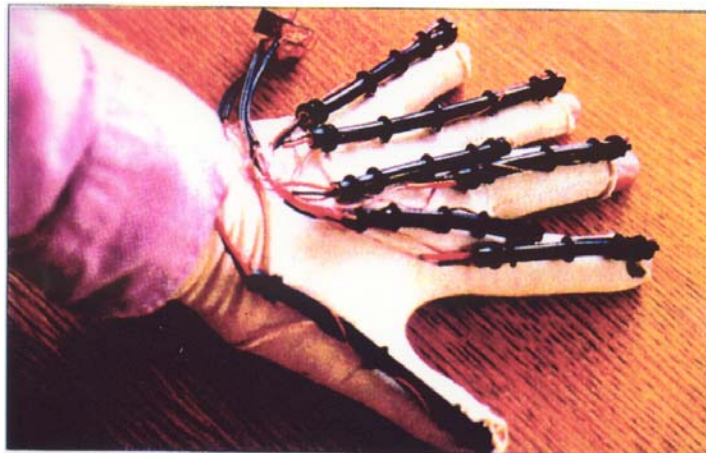
MIT glove (1977)

Digital Data Entry Glove (1983)
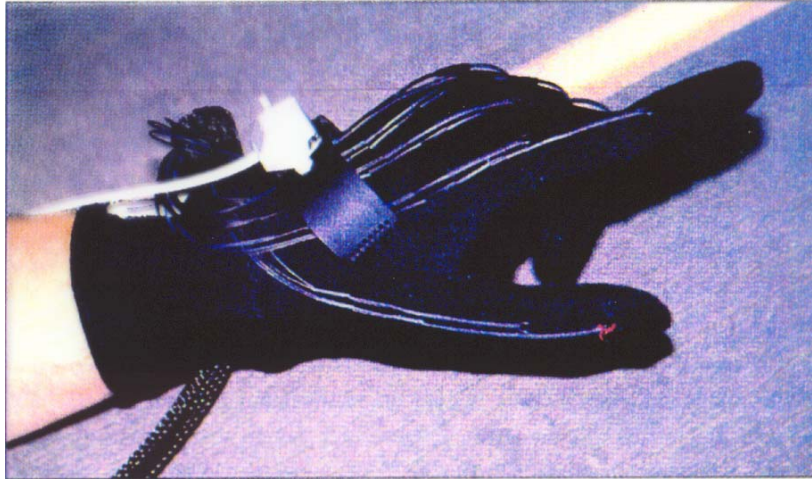
Data Glove (1987)

Power Glove (1990)

43

Cyber Glove (1995)
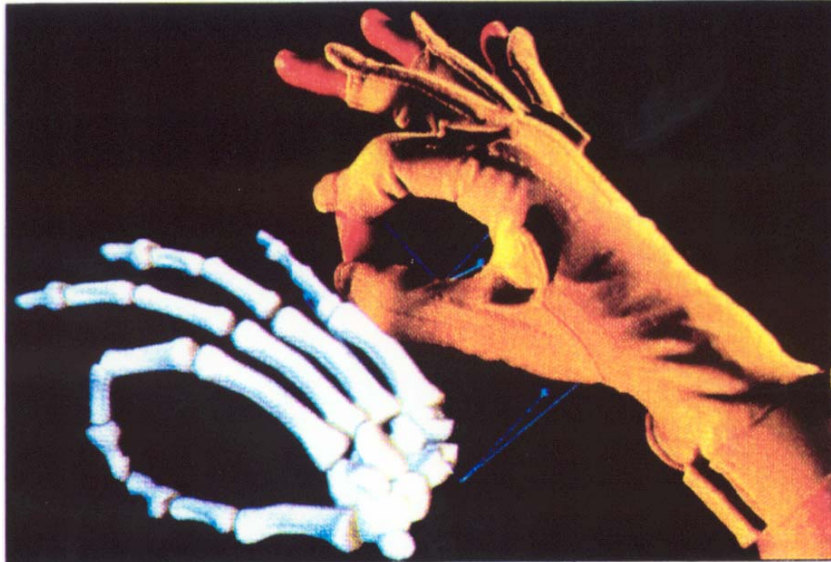
Calibration

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).
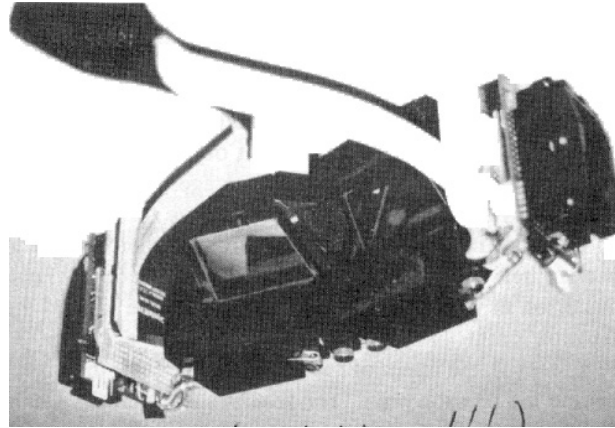
Vision based eye trackers

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

World generators

Integrated systems for 3D CAD and Animation:
- Maya (ex-Alias/Wavefront)
- XSI (ex-Softimage)
- 3D Studio Max.

- 3D Structure.
- Colour and Texture
- Motion (animation)
- Rendering (lights, shadows)

Camera motion, trasparencies….

3D structure

Solid modeling
- 3D geometric solids: cubes, cylinders, cones…
- Superquadrics (Terzopoulos and Metaxas, 1991): global parameters + local parameters.
- Revolution surfaces.
- NURBS (Piegle, 1993). CAD, high interactivity.
- Subdivision surfaces (Schroeder, 1999).

Finite element models
- It is a class per sé. Local modeling. Mechnical modeling.
- Largely used for animation in medicine (facial animation, deformation of tissue during surgery). Multi-layer modeling.

Autoscan - 1997
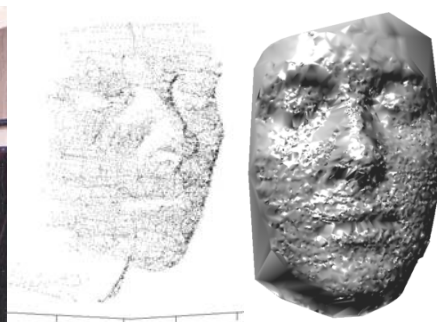
- Scansione manuale attraverso puntatore laser.
- Guida alla scansione dal feed-back su monitor.
- Flessibilità nel set-up e portabilità.
- Acquisizione spot laser in tempo reale a 100 Hz. (max 100 punti /sec)
  La triangolazione diretta dei punti pone dei problemi per la presenza di rumore.

3D structure from data

Linear approximation (mesh):
- Delauney triangulation (Watson, 1981; Fang and Piegl, 1992). Direct tessellation (no filtering).
- Alpha shapes, Ball Pivoting (Bernardini et al., 2000), Power Crust (Amenta, 2002). Post processing to regularize a Delauney tessellation.

•Surface fitting to range data
- Snakes (Kass et al., 1988). Energy based approach. Best curves.
- Kohonen maps (1990).
- Radial Basis Functions Networks (Poggio and Girosi, 1995; Borghese and Ferrari, 1998).

.....

Models from range data

Cyberware whole body scanner, WB4

Models from range data (II)

Cyberware smaller model 3030

3D structure from range data (III)

Polhemus hand held laser scanner

49

Models from range data (IV)

Digibot II.
•Platform rotates
•Scanner line translates.

http:\\homes.dsi.unimi.it\~borghese\

Minolta scanner 3D

http://kmpi.konicaminolta.us/eprise/main/kmpi/content/ISD/ISD_Category_Pages/3dscanners
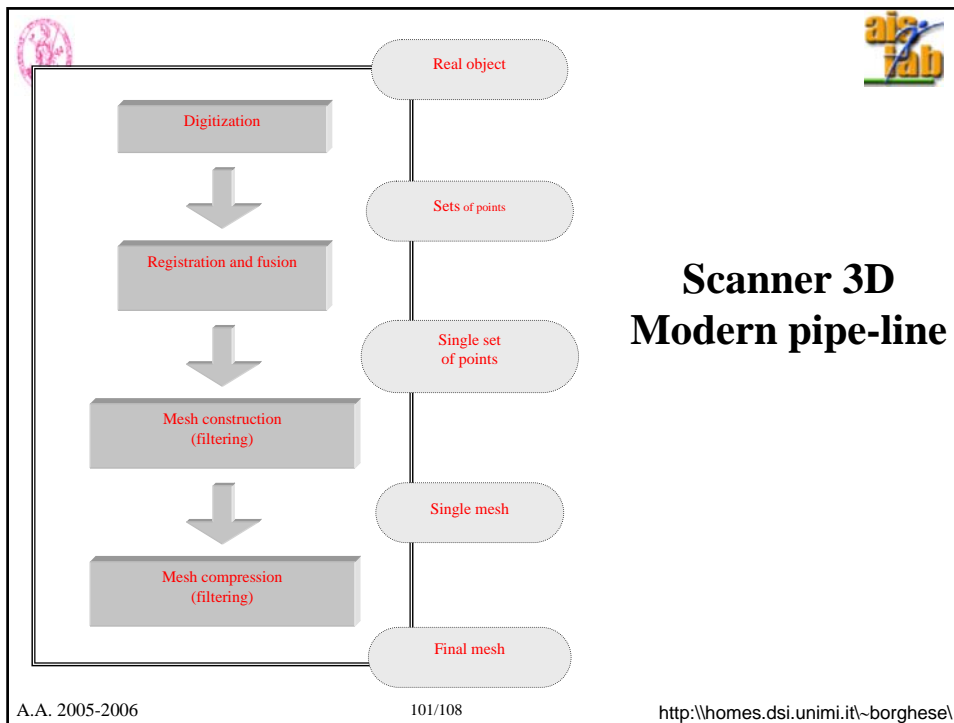
http:\\homes.dsi.unimi.it\~borghese\

Scanner 3D
Modern pipe-line

Real object

Digitization

Sets of points

Registration and fusion

Single set of points

Mesh construction (filtering)

Single mesh

Mesh compression (filtering)

Final mesh

Sommario

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

Graphical representation

Graphical engines represent triangles => Every shape is transformed into triangles.

•The models created by the scanners are ensembles of triangles (milions of).

•Much more than required by applications.

⇓

Mesh compression. Representation of the same.
geometry/pictorial attributes, with a reduced set of triangles.

LOD Modelli a dettaglio e risoluzione diversi.

.

VRML format

```
#VRML V2.0 utf8
Viewpoint {
      position 0 0 3
      orientation 0 0 1 0
      fieldOfView 0
}
DirectionalLight {
      intensity 0.2
      ambientIntensity 0.2
      color 0.9 0.9 0.9
      direction 0 -1 -1
}
Group {
children Group{
children [
Transform {
children Shape {
appearance Appearance {
material Material {
ambientIntensity 1
diffuseColor 0.9 0.9 0.9
specularColor 0 0 0
emissiveColor 0 0 0
shininess 0
transparency 0
}
}
```

```
geometry IndexedFaceSet {
coord Coordinate {
point [
      -30.180237 -231.844711 -101.136322,
      -9.759983 -198.816086 -112.282883,
...
      41.981602 -72.366501 -38.740982,
      33.281391 -76.643936 -48.074211,
]
}
color Color {
color [
  0.9 0.9 0.9,
  0.9 0.9 0.9,
...
  0.9 0.9 0.9,
  0.9 0.9 0.9,
]
}
coordIndex [
      10, 685, 970, -1,
      0, 1133, 1162, -1,
…
      263, 472, 1176, -1,
      263, 666, 1176, -1,
]
```

```
colorPerVertex TRUE
ccw TRUE
solid TRUE
creaseAngle 8
}
}
translation 0 0 0
center 0 0 0
scale 1 1 1
}
]
}
}
```

LOD model
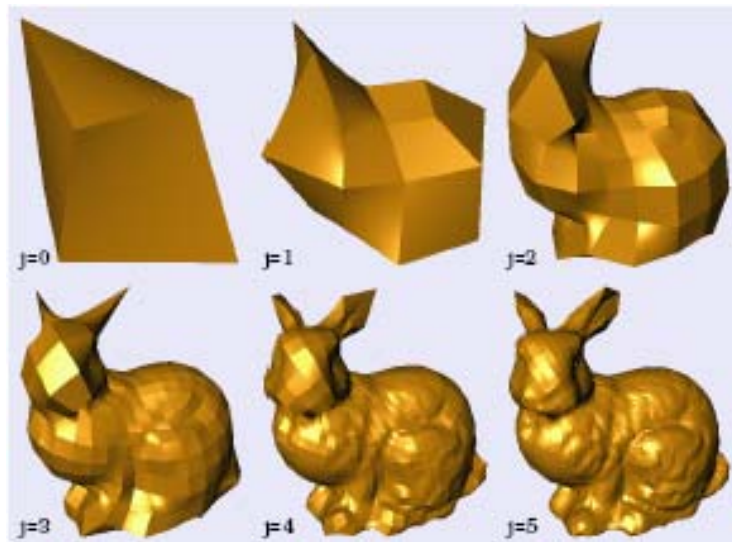
The graphical engine (visual computing)

Double buffering (for real-time visualization of 3D models) + rasterization.

Hardware acceleration of graphical operations (OpenGL, texture mapping…). GPU.

GPU programming language (cf. nVidia).

Graphical pipelining (geometry, colour, texture, shadowing, rendering….).

Parallelization.

Multiple cache levels.

Look-ahead code optimization (compiler optimization).

53

Collision detection

Computational demanding (On^{2}EF).

Use of multiresolution models.

Hierarchical detection.

Geometry semplification (axes aligned faces).

Check for common volumes.

Extraction of the faces belonging to these volumes.

Octree of the pairs of candidate faces.

Check for intersection.

Sommario

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