


Motion Capture
Passive markers and video-based techniques

N. Alberto Borghese
Laboratory of Applied Intelligent Systems
(AIS-Lab)
Department of Computer Science
University of Milano


Outline

Passive Markers Technology.

Low and high level processing.

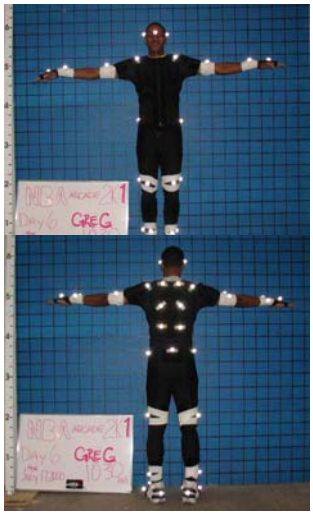
From acquisition to reconstruction.

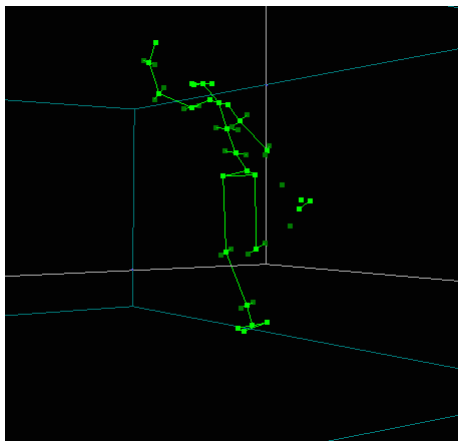
Results.

Video Based Motion Capture.

Motion Capture with passive markers

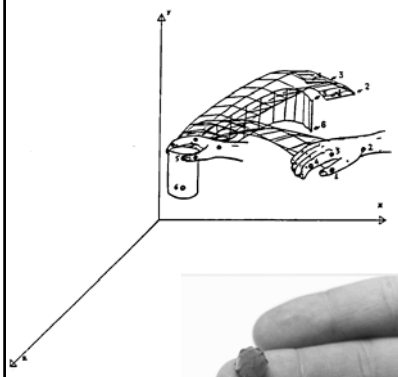
Goal: reconstruction of the 3D motion of a set of markers





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™). 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™)
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.


Outline

Passive Markers Technology.

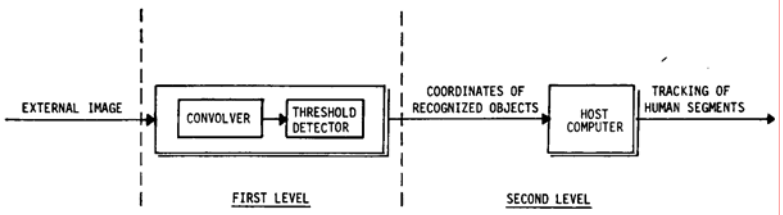
Low and high level processing.

From acquisition to reconstruction.

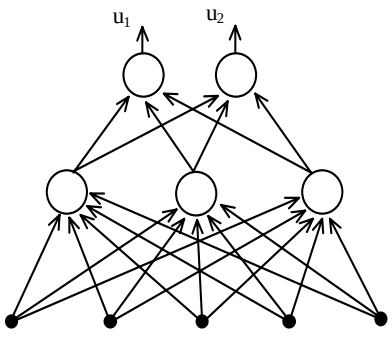
Results.

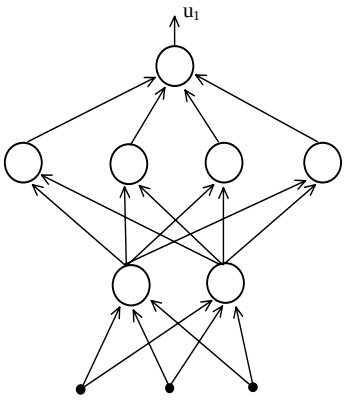
Video Based Motion Capture.

Two-levels architecture



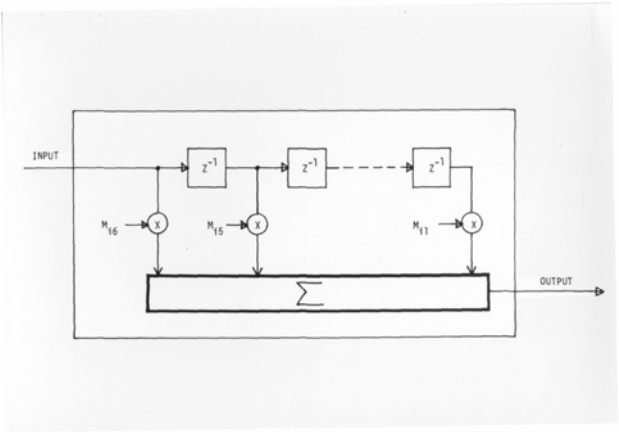
Low-level processing (low-vision)





Markers extraction through filtering

Correlation implemented by convolution (template matching or feature extraction)



Implementable with a DSP

Markers extraction through thresholding (without flash)

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

Cluster dimension.

Shape.

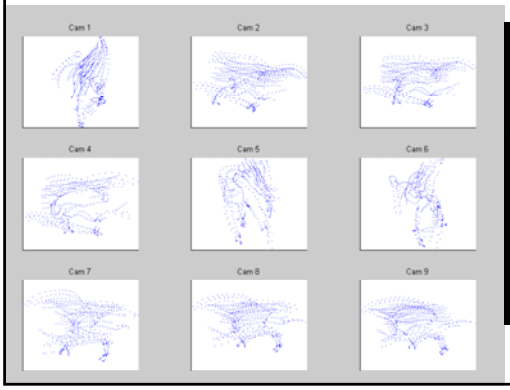
Software protection of bright target regions.

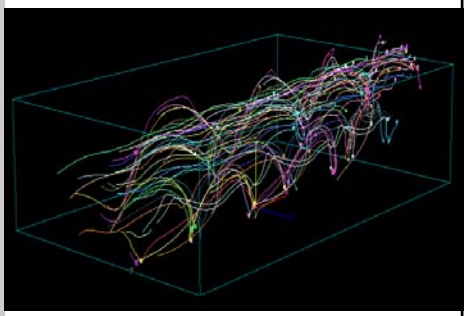
High-level processing

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.



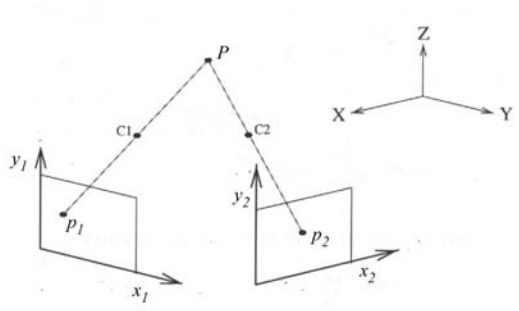


The difficulties in data processing

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.

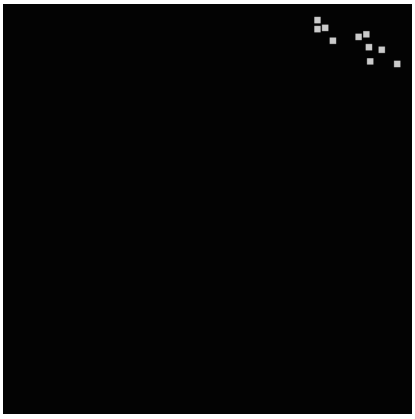
From 2D to 3D

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.

Outline

Passive Markers Technology.

Low and high level processing.

From acquisition to reconstruction.

Results.

Video Based Motion Capture.

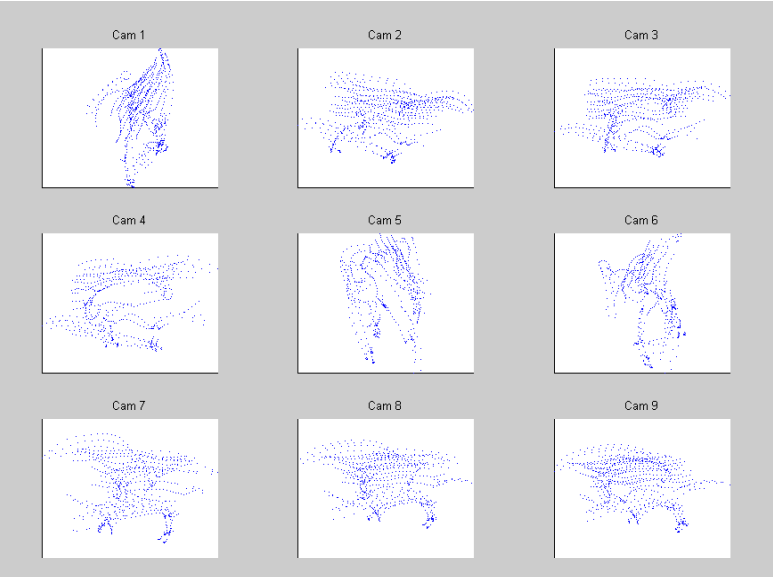
Tracking processing steps

ACQUISITION OF 2D POINTS

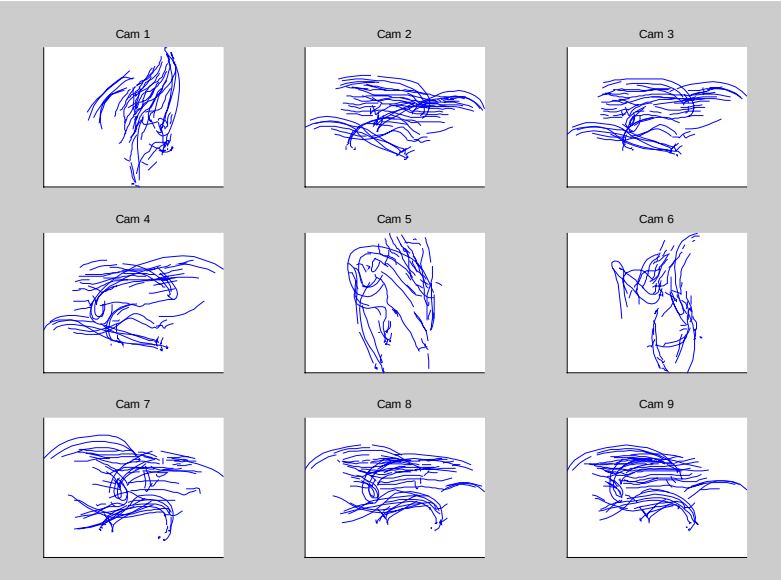
TRACKING:
1) From 2D points to 2D strings.
2) Pairing 2D strings with the epipolar constraint to create 3D strings.
3) Condensation of 3D strings.
4) Joining 3D strings.

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

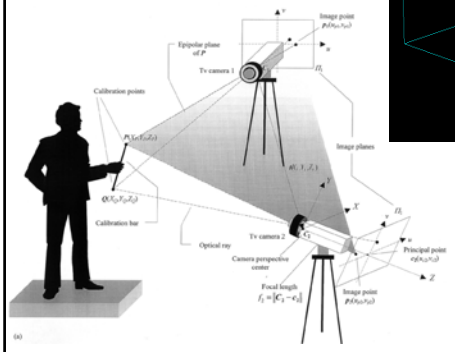


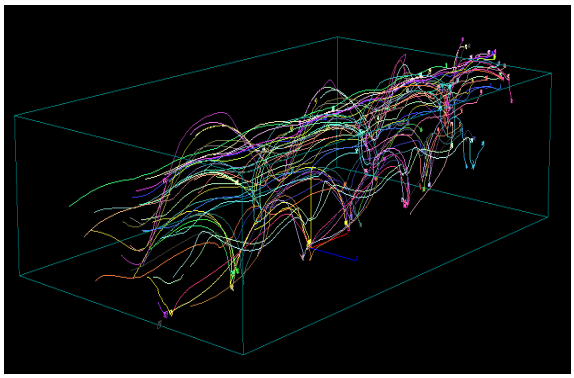
1) Creation of 2D strings



2) Matching 2D strings

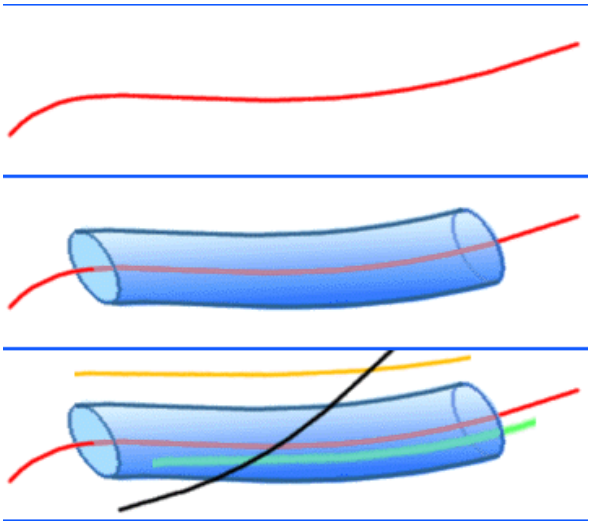
Epipolarity constraint



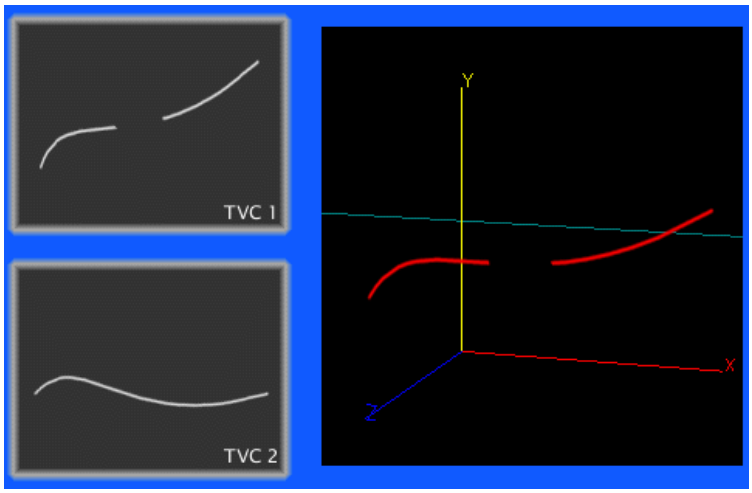


3D strings

3) Condensation of 3D strings



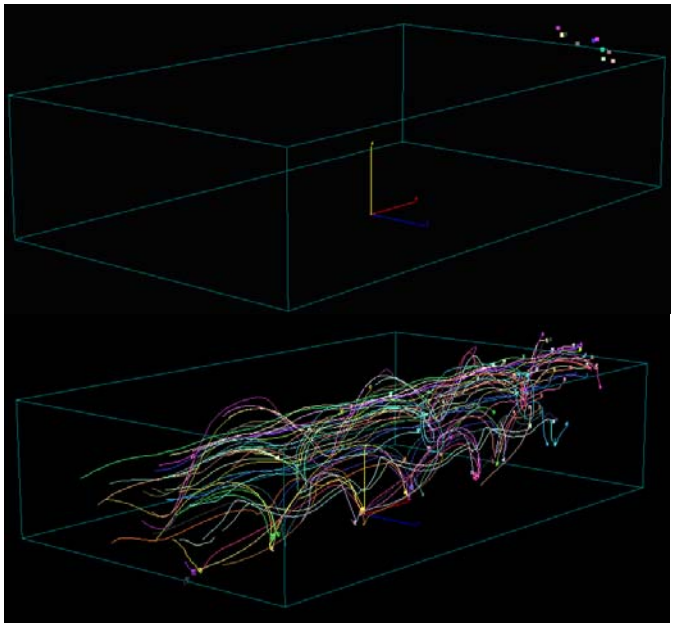
4) Joining 3D strings



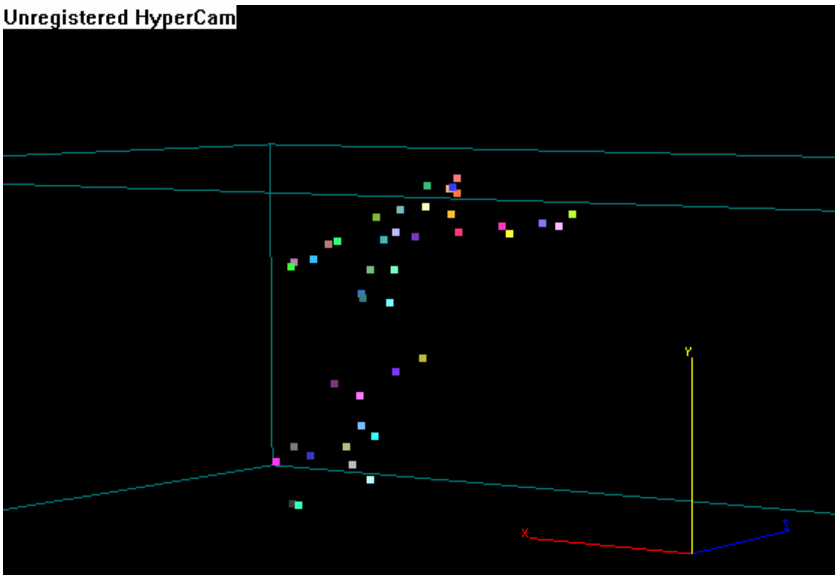
3D strings

3D strings already contain motion 3D information

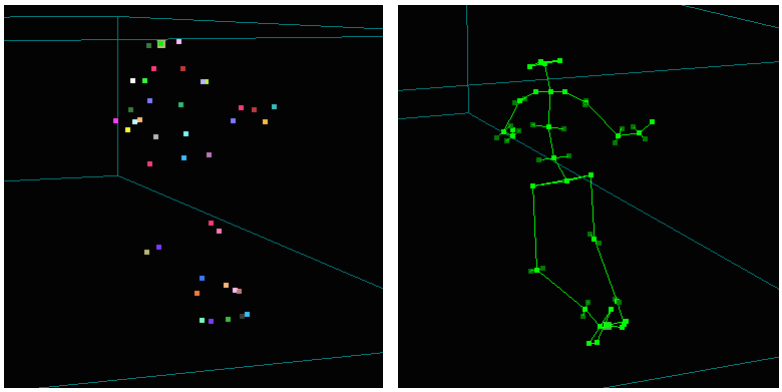
3D strings



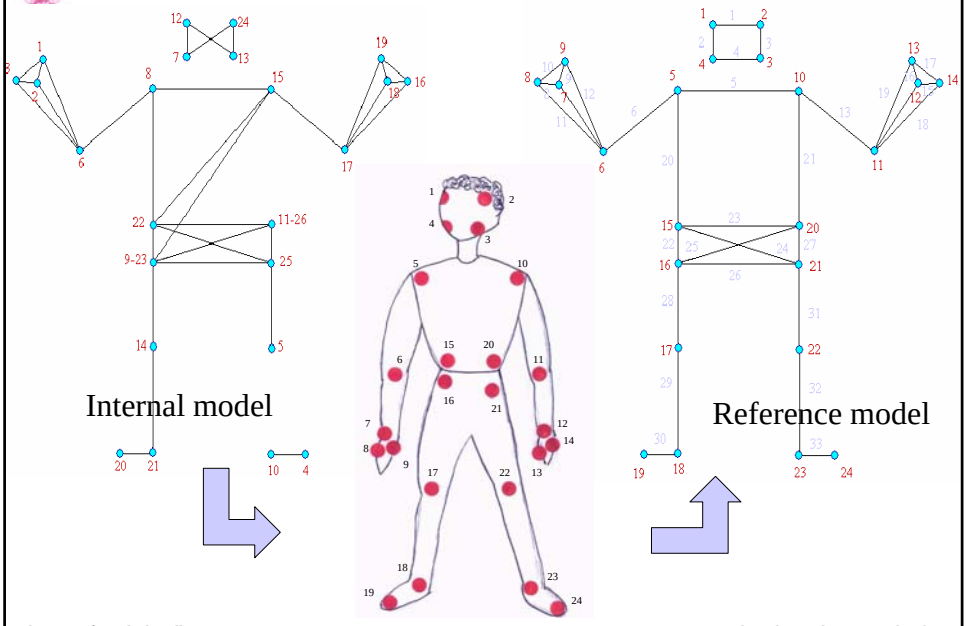
Markers Classification



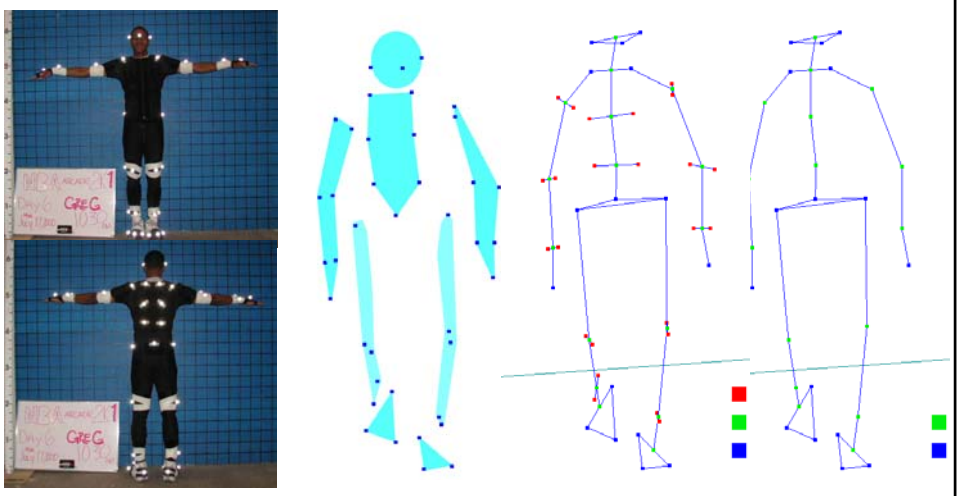
5) Initial classification



Model fitting



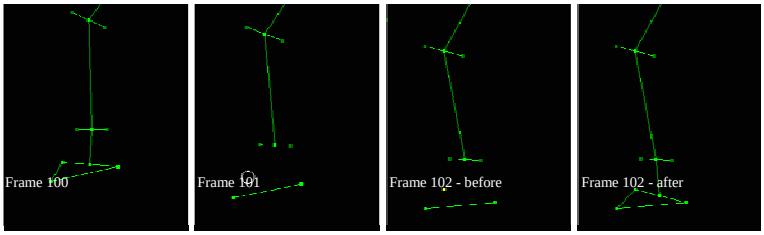
What a model represents?



Markered subject Modello 3D Modello a stick Modello hidden

6) Classification extension

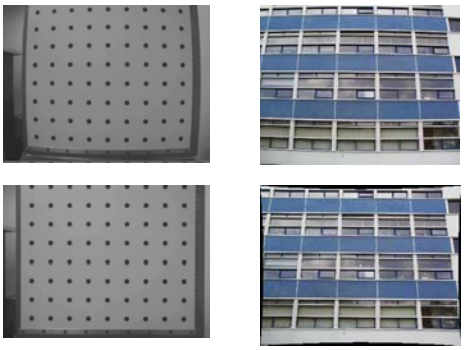
3D strings are automatically extended in this phase.



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.


Outline

Passive Markers Technology.

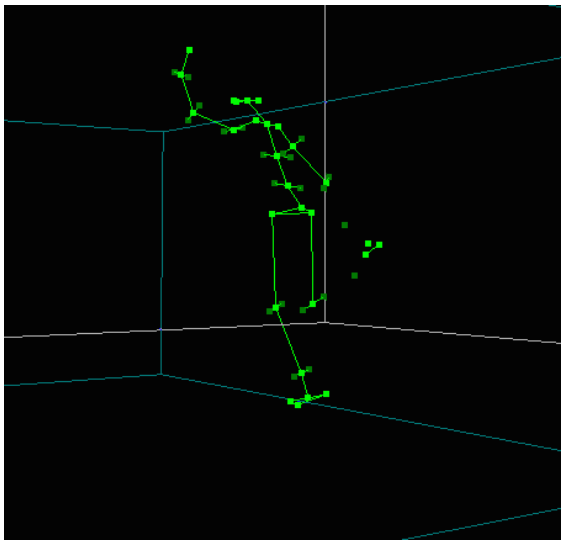
Low and high level processing.

From acquisition to reconstruction.

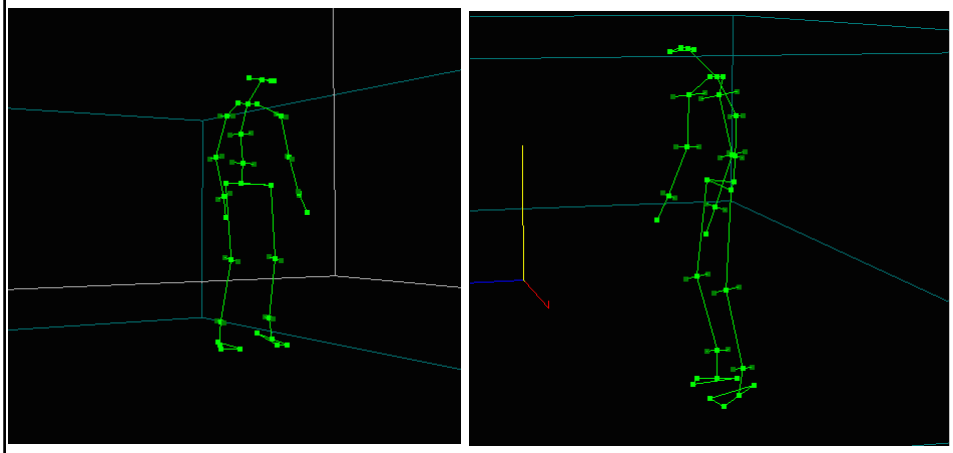
Results.

Video Based Motion Capture.

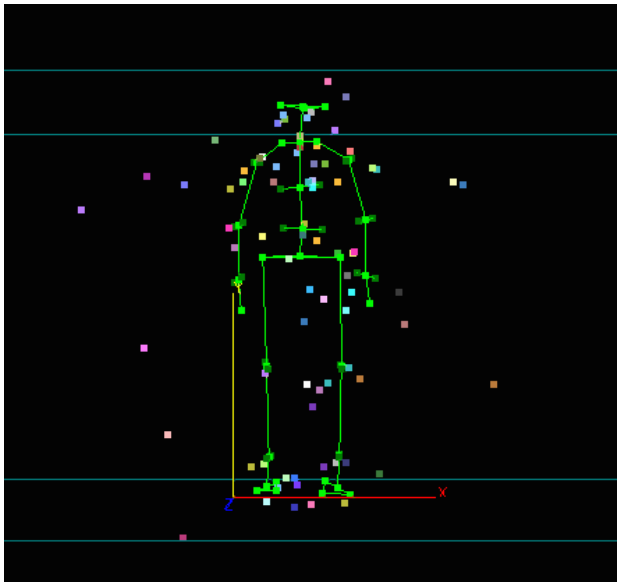

Risultati: run



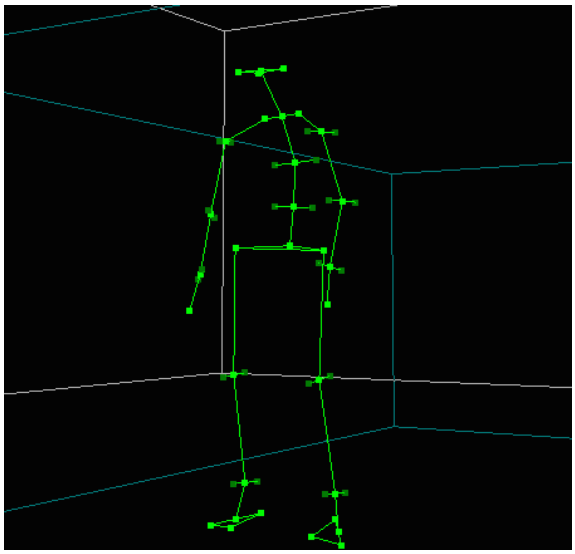
Risultati: escape



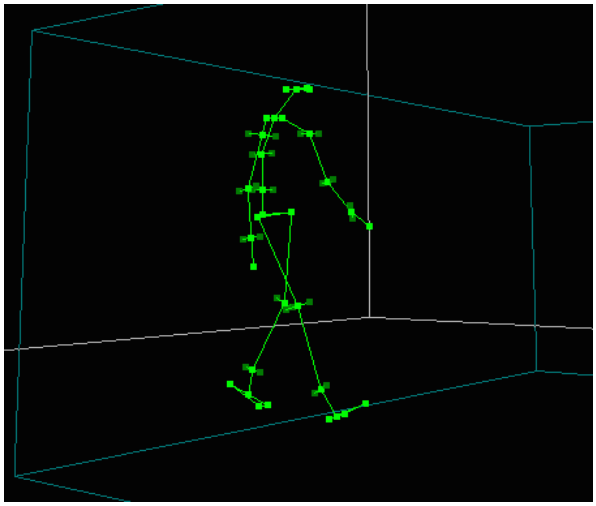
Risultati: head_turn

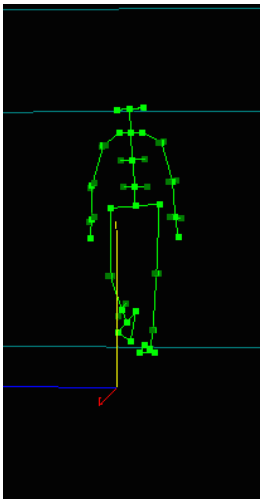


Risultati: fall_run

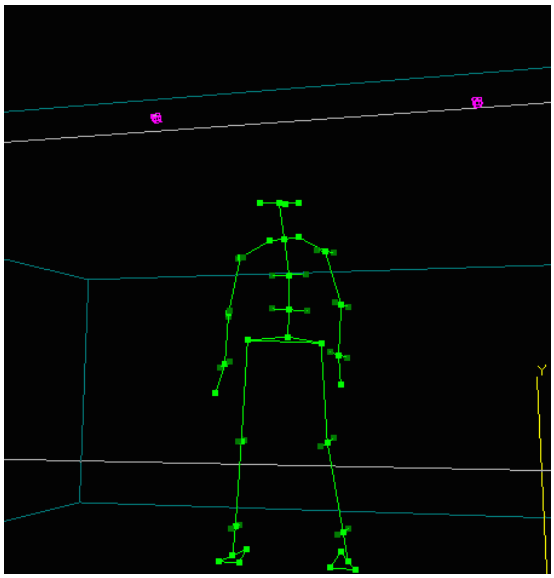


Risultati: walk





Risultati: roll



Outline

Passive Markers Technology.

Low and high level processing.

From acquisition to reconstruction.

Results.

Video Based Motion Capture.

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.

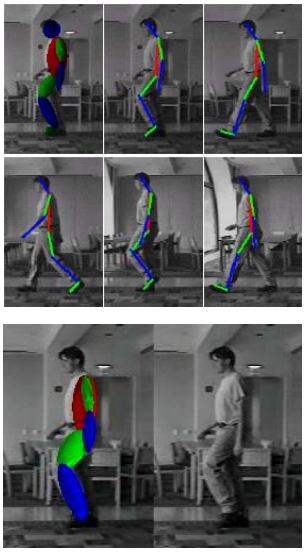
A photogrammetric approach

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/

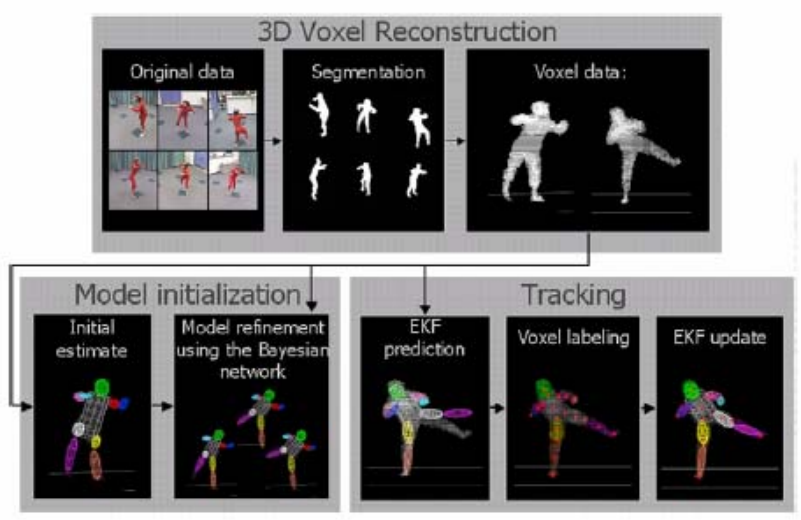
The volumetric approach: a possible solution

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.

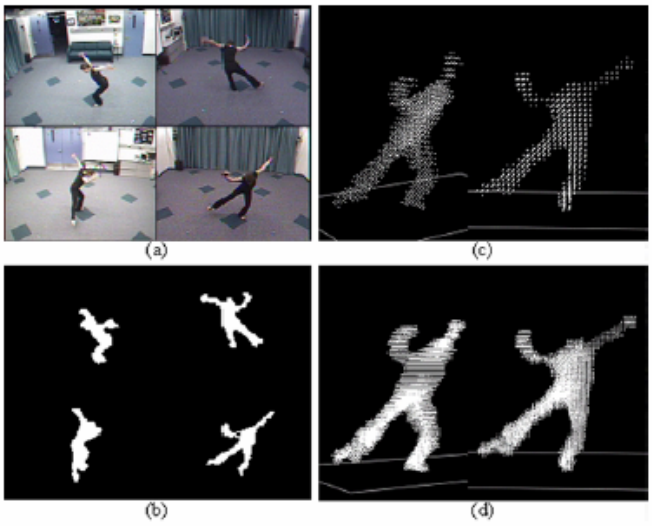
Jain et al., 3D video, Proc. VRAIS, 1994.

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.

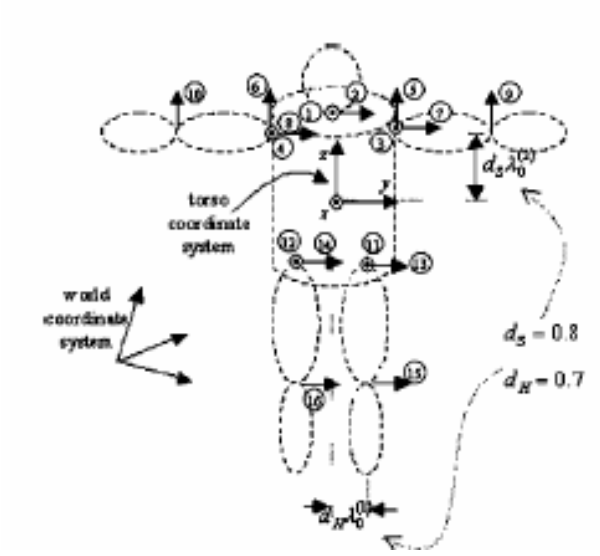


c) 50mm d) 25mm

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:

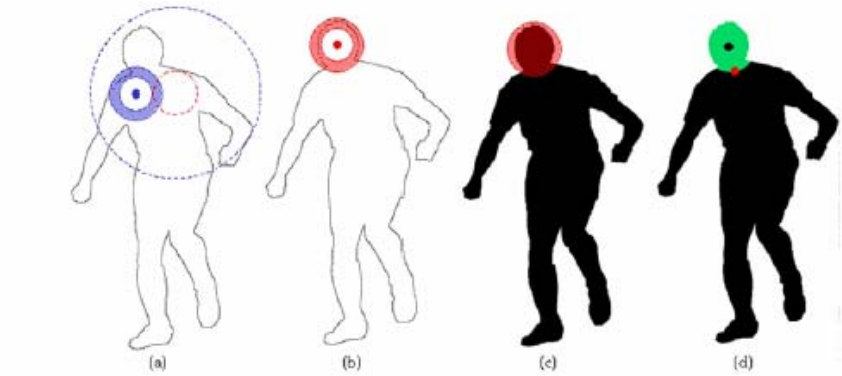
An average cylinder is attached to the neck and oriented as the centroid of the remaining voxels.

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

Every k-steps, its orientation is recalculated.

3) Limbs:

Head initial identification

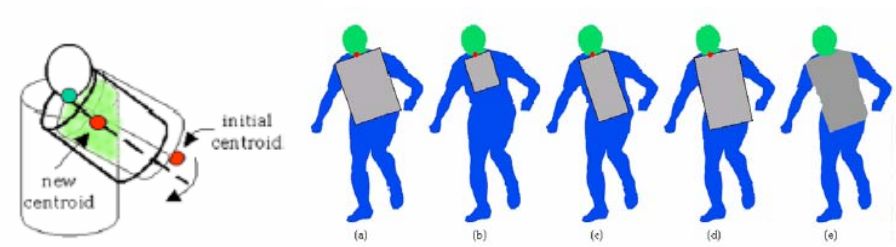


Spherical template with minimum and maximum radius.

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.

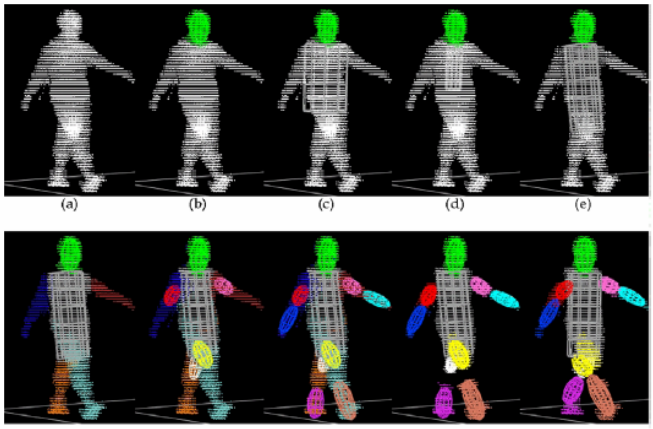
It is then shrunk and grown again until it incorporates empty voxels.

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

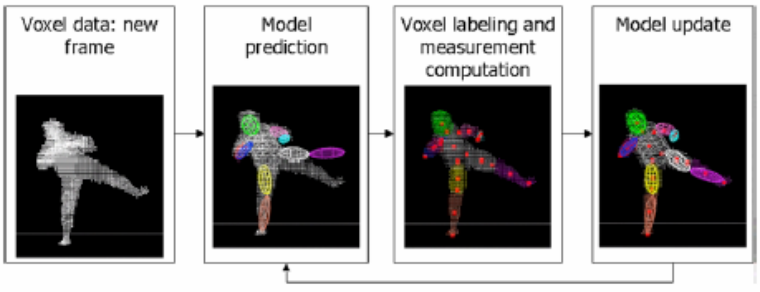
Kalman filter.

Silhouettes → Voxel reconstruction.

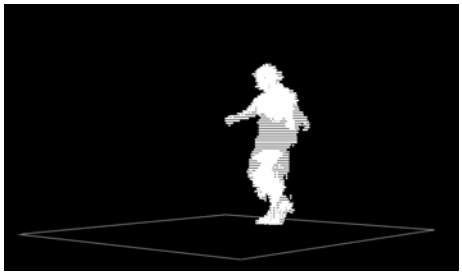
Model prediction through Kalman filter → Voxel labeling.

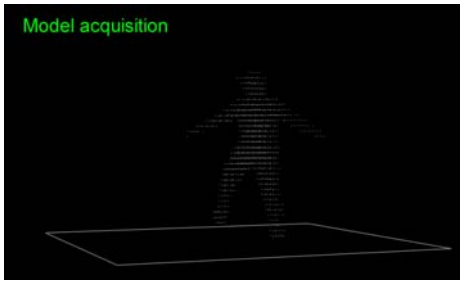
Measurement on labelled voxels of body positions (end-points and centroid of the segments).

From the difference from the measurement and the prediction, the new position of the model is determined.

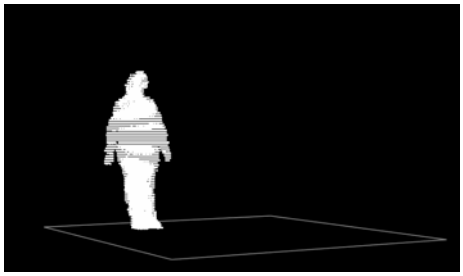


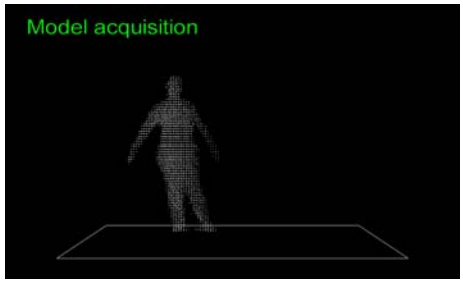
Results: stepping (640 x 480, 10Hz)



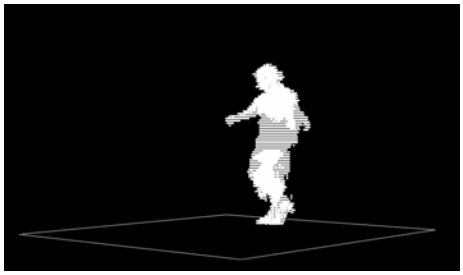


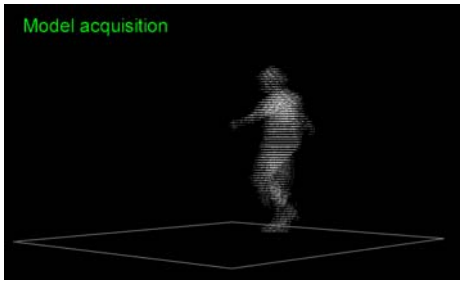
Results: dancing



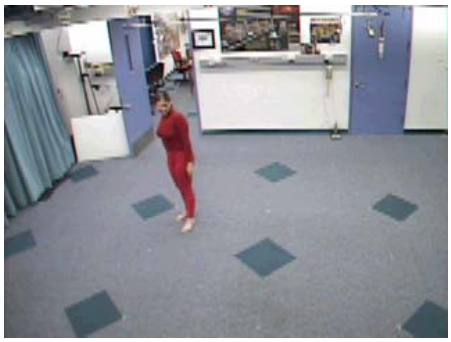


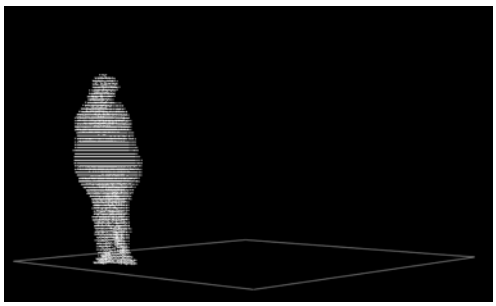
Results: jumping

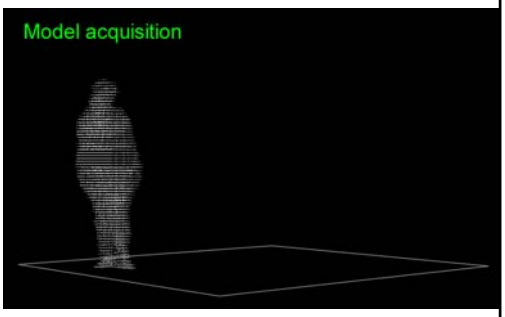




Results: cartwheel







Outline

Passive Markers Technology.

Low and high level processing.

From acquisition to reconstruction.

Results.

Video Based Motion Capture.