Video-based Animation of People

Adrian Hilton
Professor of Computer Vision & Graphics

Visual Media Research Group
Centre for Vision, Speech & Signal Processing
University of Surrey, UK

Acknowledgements

Visual Media Research Group

Dr. Jonathan Starck

Research Fellows:

Jean-Yves Guillemaut, Joe Kilner, James Edge

PhD Students:

Aseel Turkmani, Natalaya Nadtoka,Gregor Miller
Peter Williams, Rafael Tena, Peng Huang, Chris Budd

Collaborators:

BBC Research, Sony, Snell&Wilcox, Framestore CFC,
Artem Digital, The Foundry, Hawkeye

Overview

From Video Capture to Animated Models of People

- Modelling People: model-based vs. model-free
- Surface Motion Capture (SurfCap)
- Animation from Surface Capture

Centre for Vision, Speech & Signal Processing
University of Surrey, UK

Director: Prof. Josef Kittler

12 Academic Staff
~20 Post-doctoral researchers
~35 PhD Students

4 Groups:

- Multimedia Signal Processing
- Robotics
- Medical Imaging
- Visual Media

Facilities:

- Studio with 8xHD broadcast cameras
- Motion capture
- 3D & 4D shape capture

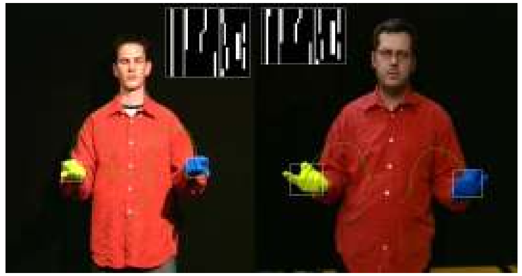
Location: Guildford, 40 minutes from London

Visual Media Research Group

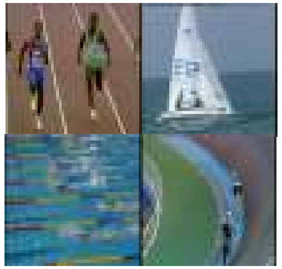
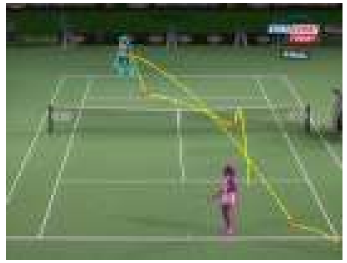
Adrian Hilton, Josef Kittler, Theo Vlachos, Richard Bowden, Jonathan Starck
~8 Post-doctoral researchers
~20 PhD students

➢ Visual Content Production
- 3D studio production
- free-viewpoint video in sports

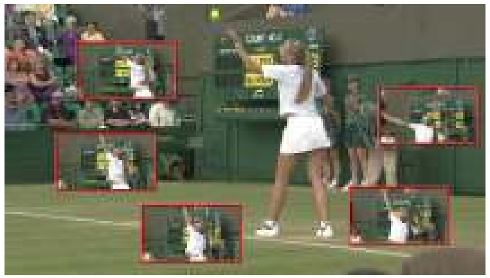
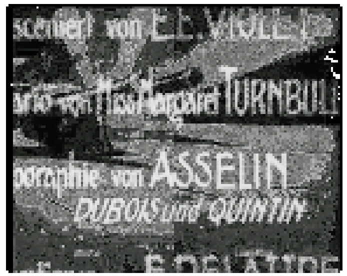

➢ Visual Interaction
- visual communication
- sign language recognition

➢ Visual Scene Understanding
- sports recognition
- sports annotation

➢ Visual Quality Evaluation
- Video coding
- Archive restoration

Modelling People

Goal: Capture animated models of real people

People are the central element of most entertainment

3D Video of People

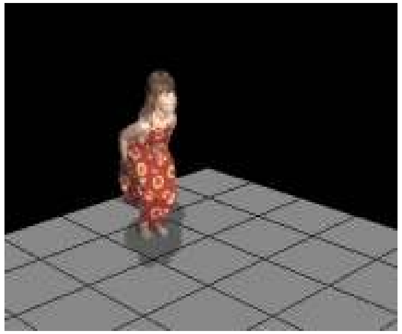
- Photo-realistic appearance
- Control of movement & viewpoint
- Reproduce dynamics of appearance and movement
  - cloth, hair, face
- Compact structured representation
- Re-lighting

Applications:

Entertainment – games, broadcast, film

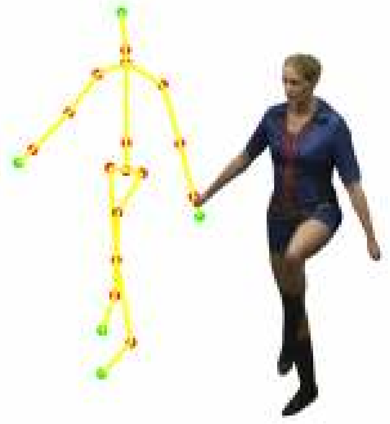
Communication – visual interaction

Capturing People – Previous Work



Model-free: Kanade'96, Matusik'00, Zitnick'04

- Virtualized reality [Kanade'96]
- Replay captured movement
- High visual quality [Zitnick'04,Starck'05]
- Unstructured representation
- **No change in movement**



Model-based: Carranza'03, Starck'03

- Structured representation
- User control of movement
- **Limited visual quality**

Goal:

- Video-quality
- Control of movement for animation
- Dynamic appearance
- Structured representation for animation pipeline

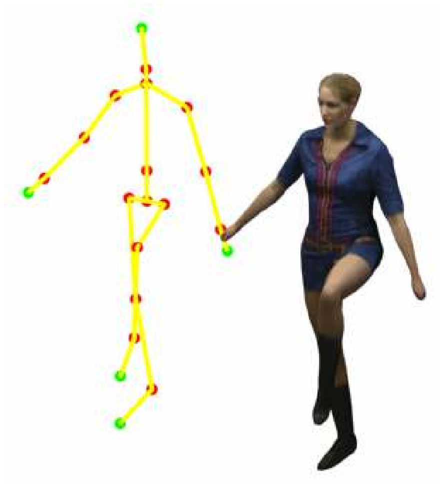
Model-based Reconstruction

Model-based Reconstruction - Methodology

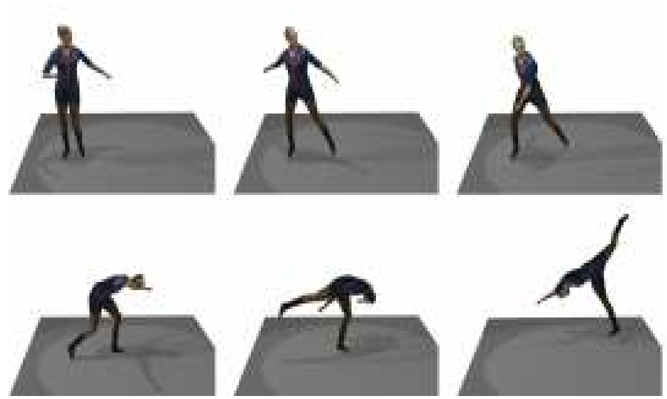
- Prior knowledge- generic functional model
  - scene geometric structure
  - kinematics constraints
  - dynamic behaviour
- Map generic model onto observations
  - shape constrained deformable surface fitting
- Layered representation of detail (shape, appearance)

Observation

Model



Animation



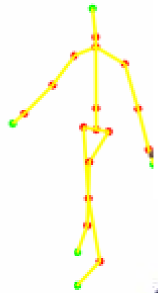
Hilton et al. Visual Computer'00
Hilton et al. 3DPVT'02
Starck & Hilton ICCV'03

Model-based Reconstruction

Input: Multiple view images or 3D data
Generic animated model

(1) Registration (Pose Estimation)

- initial alignment & feature correspondence
- automatic optimisation of model parameters



(2) Model fit to observations

- shape constrained deformable surface fitting algorithm
- shape-from-silhouette, stereo and feature matching

(3) Representation of high-resolution surface detail

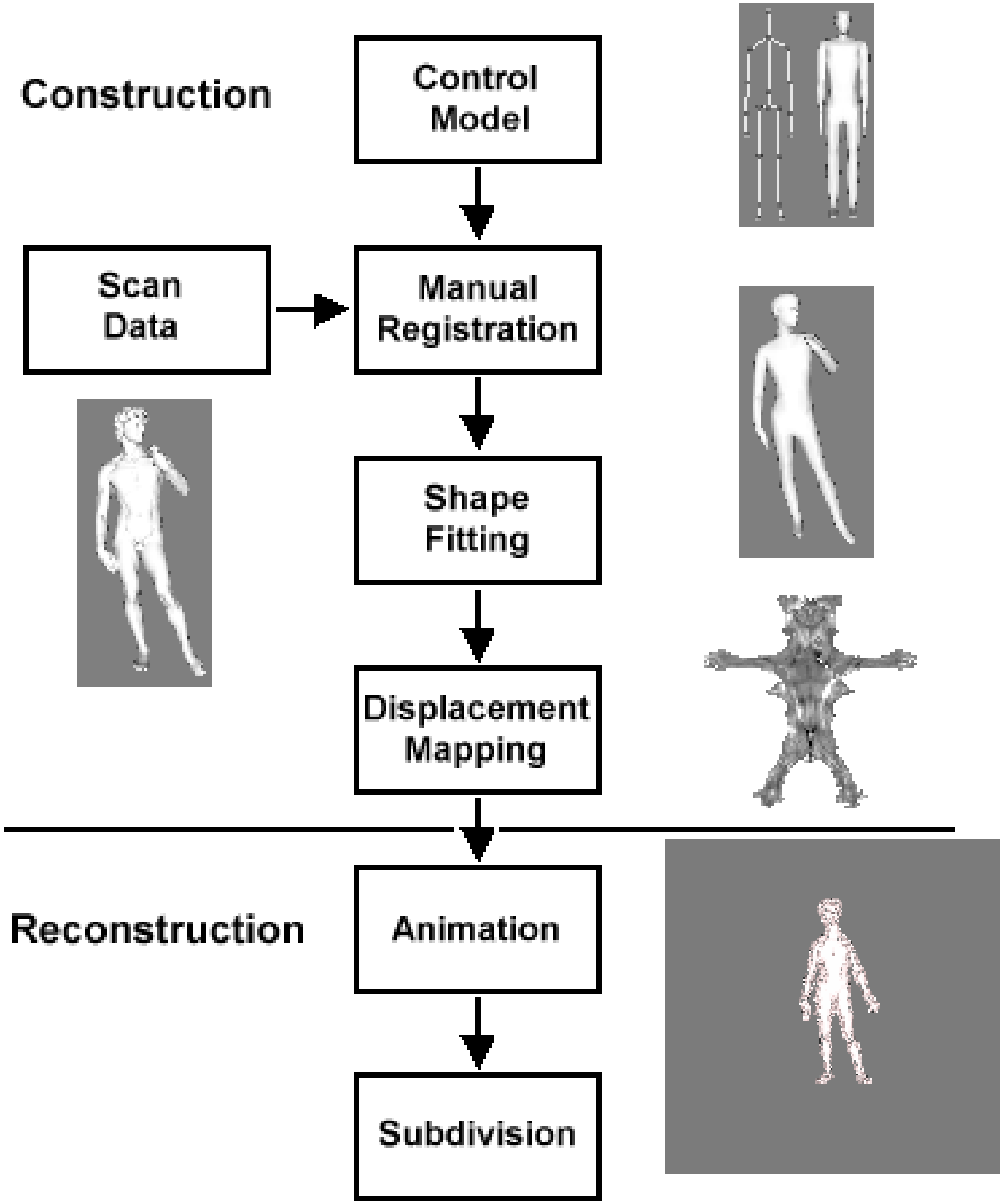
- displacement mapping
- texture mapping multiple view appearance

(4) Reconstruction

- animate control model & re-generate data in new pose

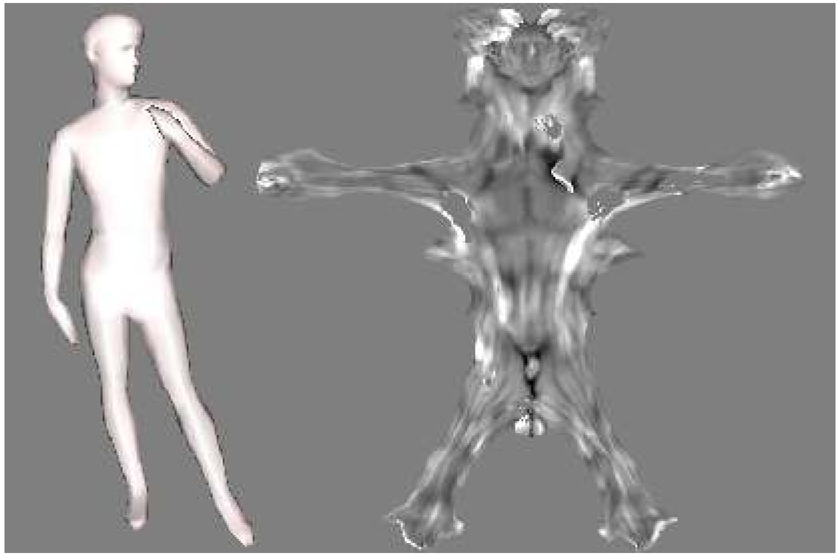
Output: Animated model of real object
(shape/appearance/motion)

Framework: Model-based representation of 3D scans



Michelangelo's David

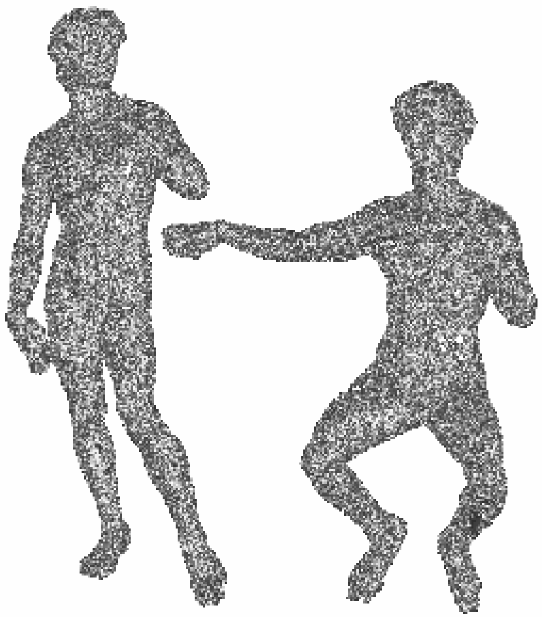
Data
>100Mb



Model + Displacement Map

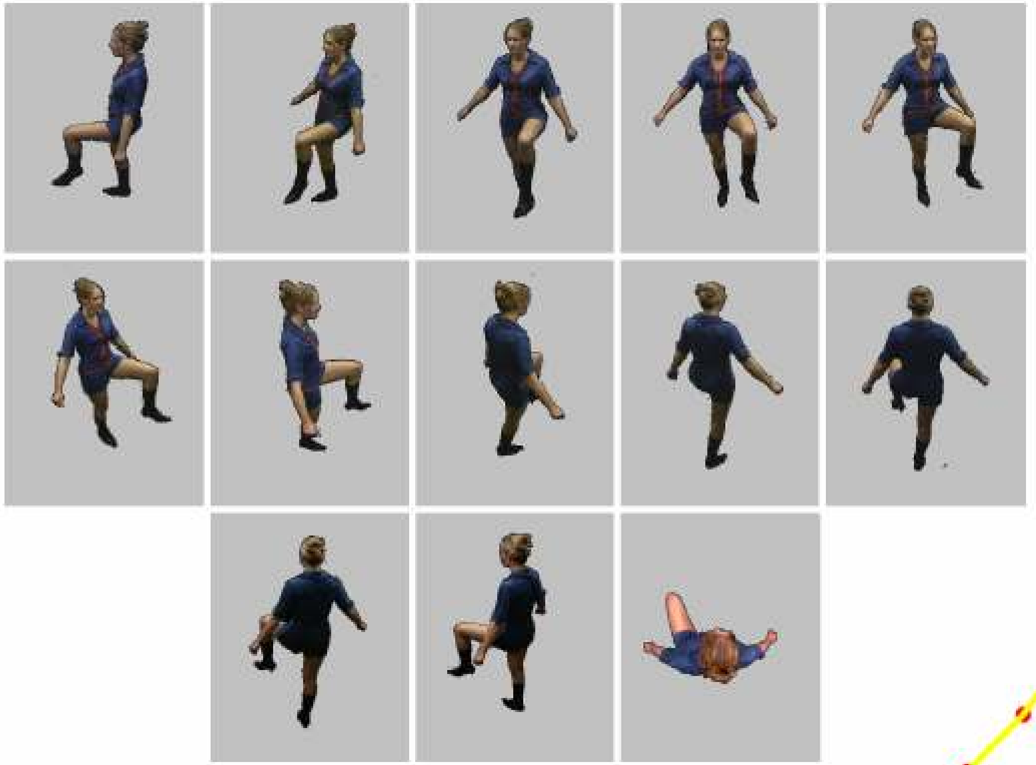
Reconstruction
<1Mb

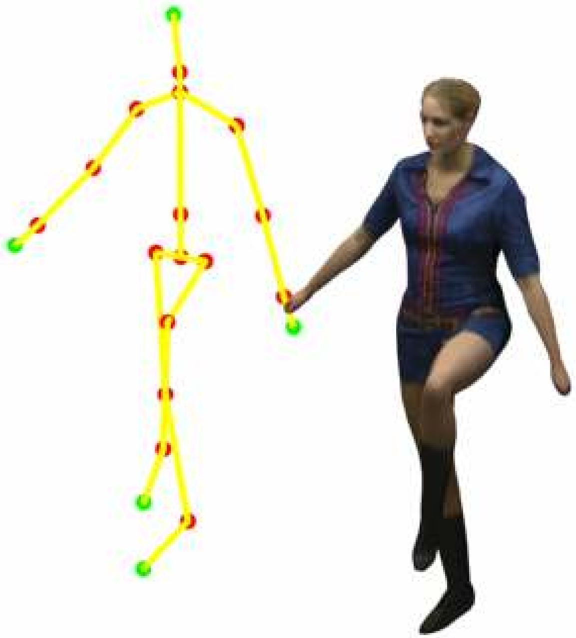
Data Courtesy Stanford Computer Graphics Lab.



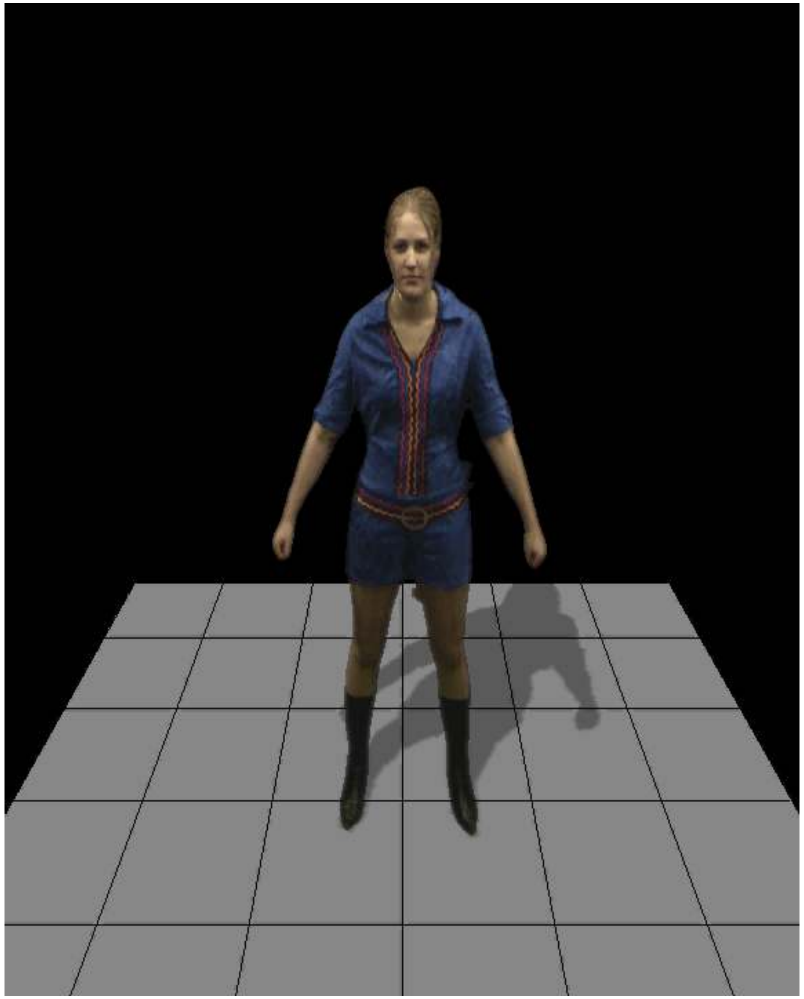
'Animated Statues' MVA'02

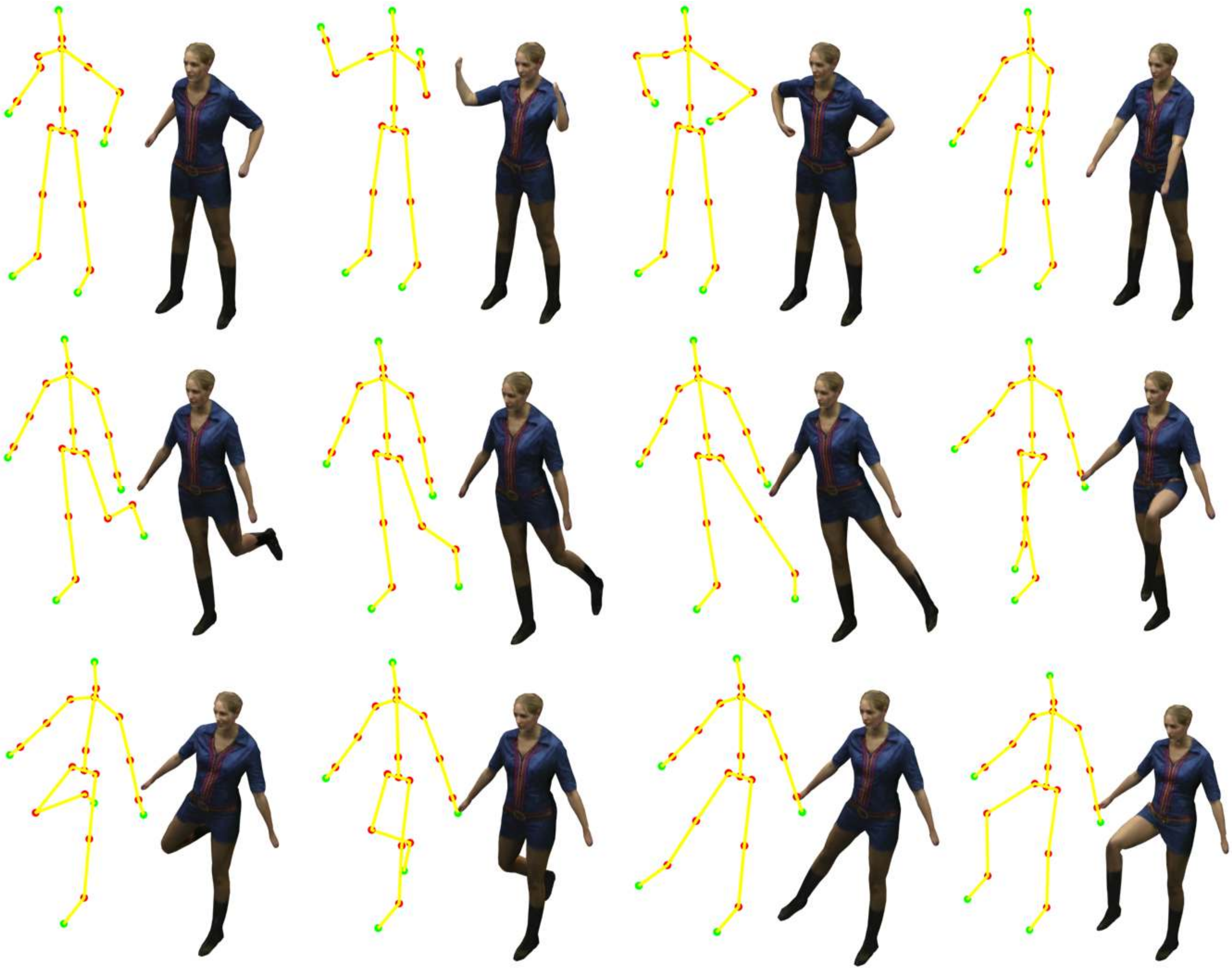
Model-based Reconstruction: Multi-camera studio





Model-based Reconstruction 13 views





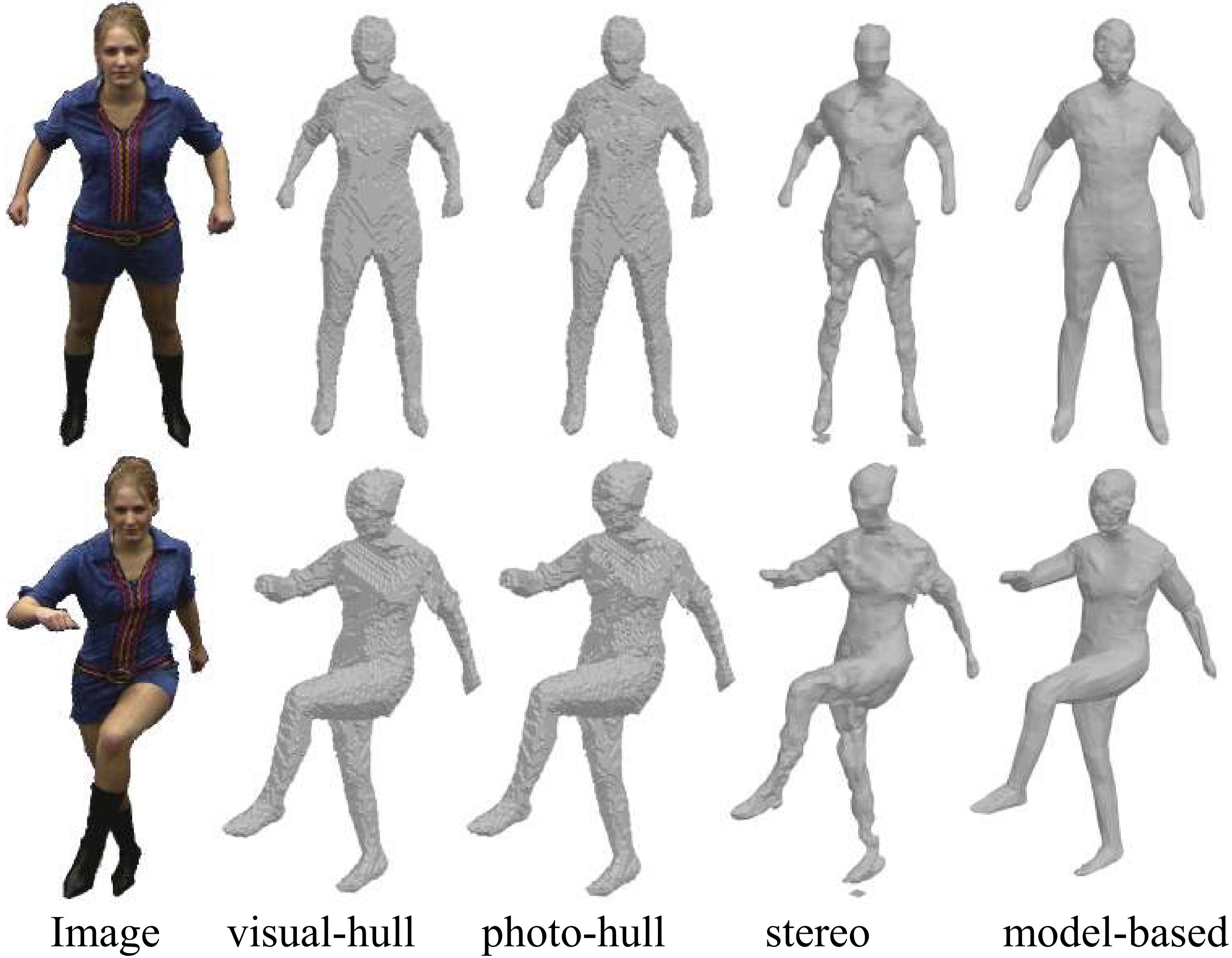
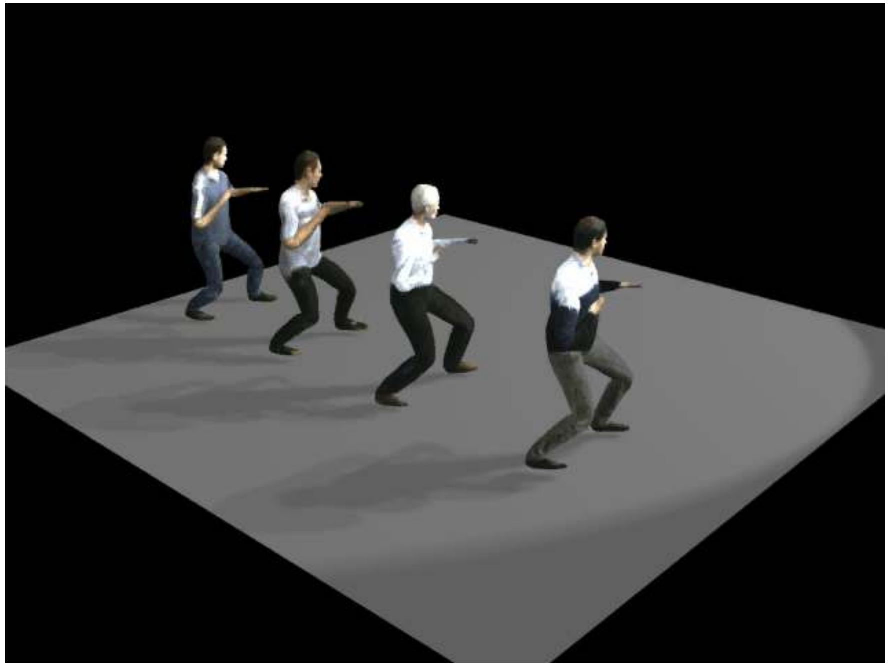


Image
visual-hull
photo-hull
stereo
model-based

Image visual-hull photo-hull stereo model-based

Model-based Reconstruction



[Starck&Hilton ICCV'03]

Limitations:

- shape detail face/hands/hair/clothing
- not video quality

Problem: Generic model over simplifies shape

Model-free Reconstruction

Model-free Reconstruction

No prior knowledge of scene

- remove limitations of model-based approach
- no animation structure

Multi-view reconstruction

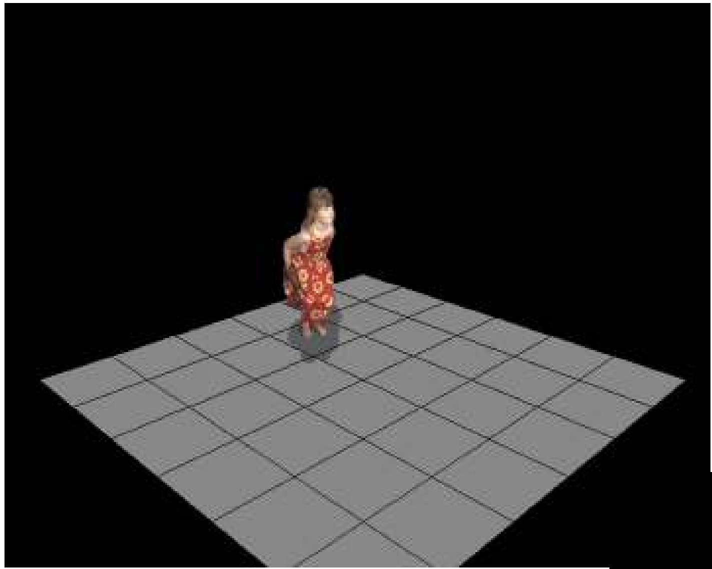
- same framework as model-based
- initial shape from visual hull
- optimise visual-hull surface for stereo re-projection
- wide-baseline stereo matching
- regularise with surface smoothness (membrane, thin-plate)

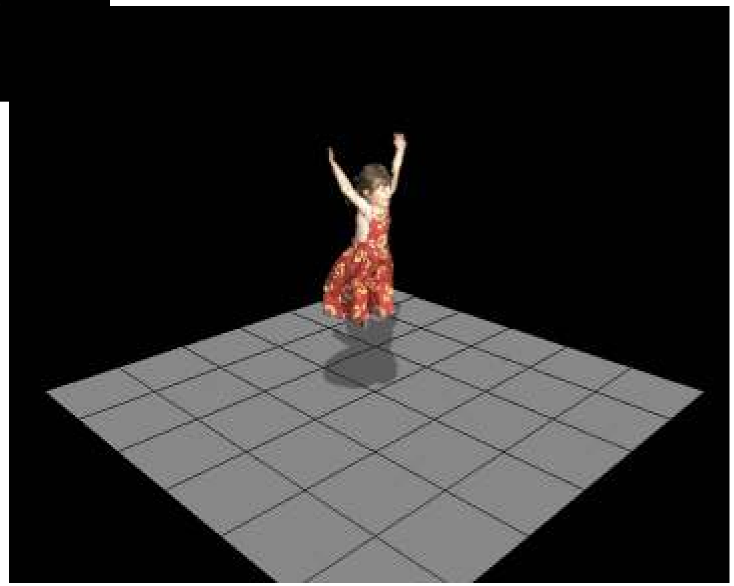
View-dependent optimisation of shape

- minimise error for stereo re-projection
- robust to calibration error
- changes in appearance with viewpoint

Hilton&Starck 3DPVT'04
Starck&Hilton Graphical Models'05

Model-free reconstruction: 8 cameras





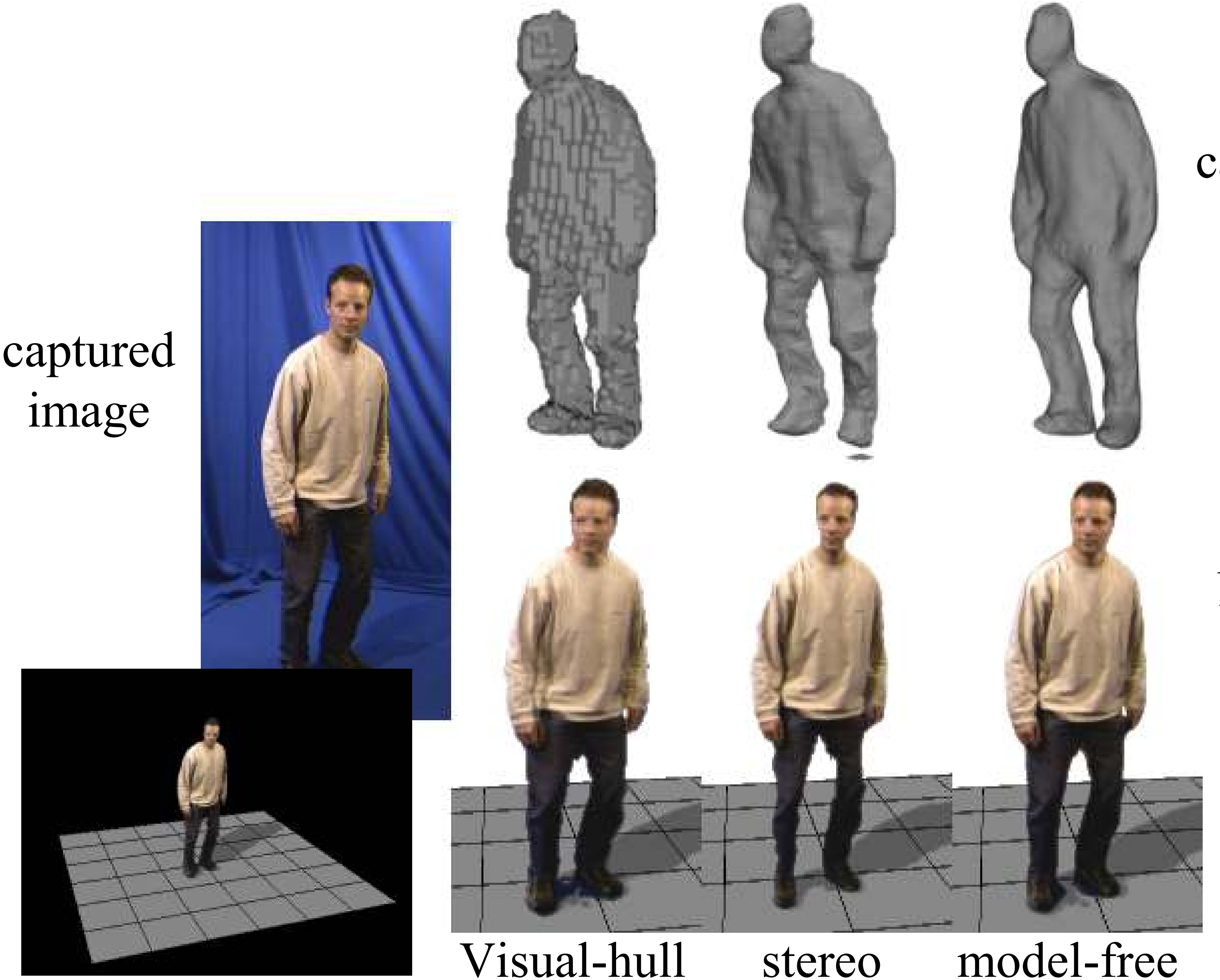
Model-free reconstruction

Pre-computation of video-based representation

Interactive user-control of viewpoint for free-viewpoint video

Miller, Hilton, Starck CVMP'05

Visual-hull vs. Model-free



Model-free reconstruction in sports

iView project: BBC R&D, S&W, Hawkeye

Objective: Free-viewpoint video for production of live sports events

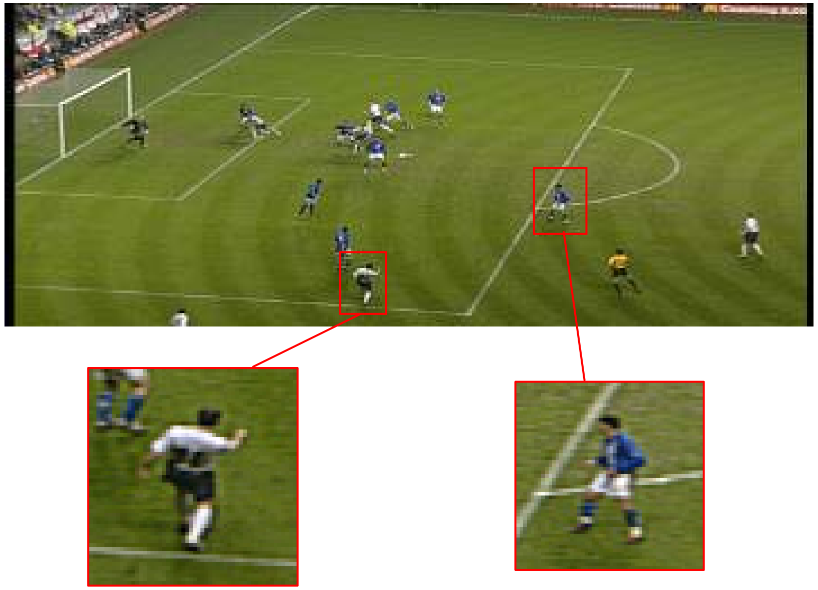
Methodology: Transfer model-free reconstruction from studio to outdoor

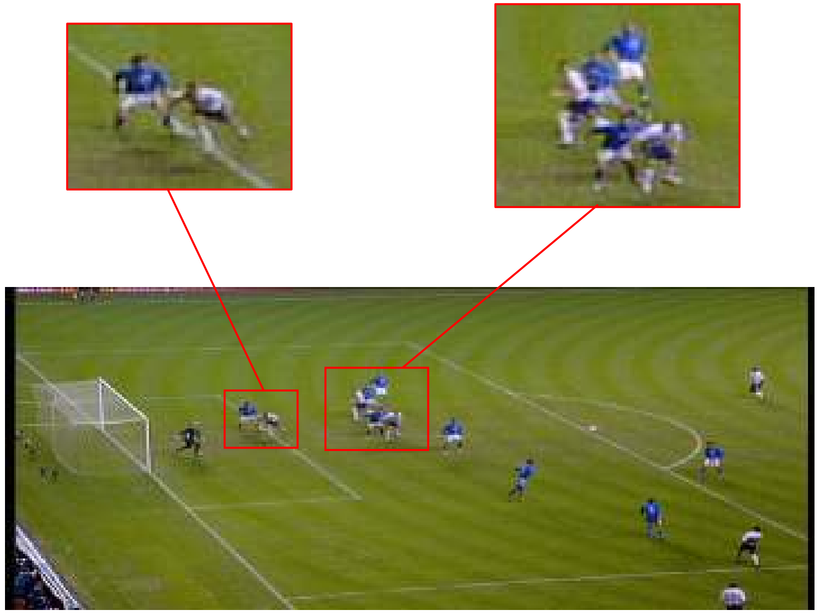
Challenges:

- large areas (football pitch)
- uncontrolled illumination
- wide-baseline views (8-12 match cameras)
- broadcast quality

Funding: DTI Technology Programme/EPSRC 2006-2009

Why is it challenging?



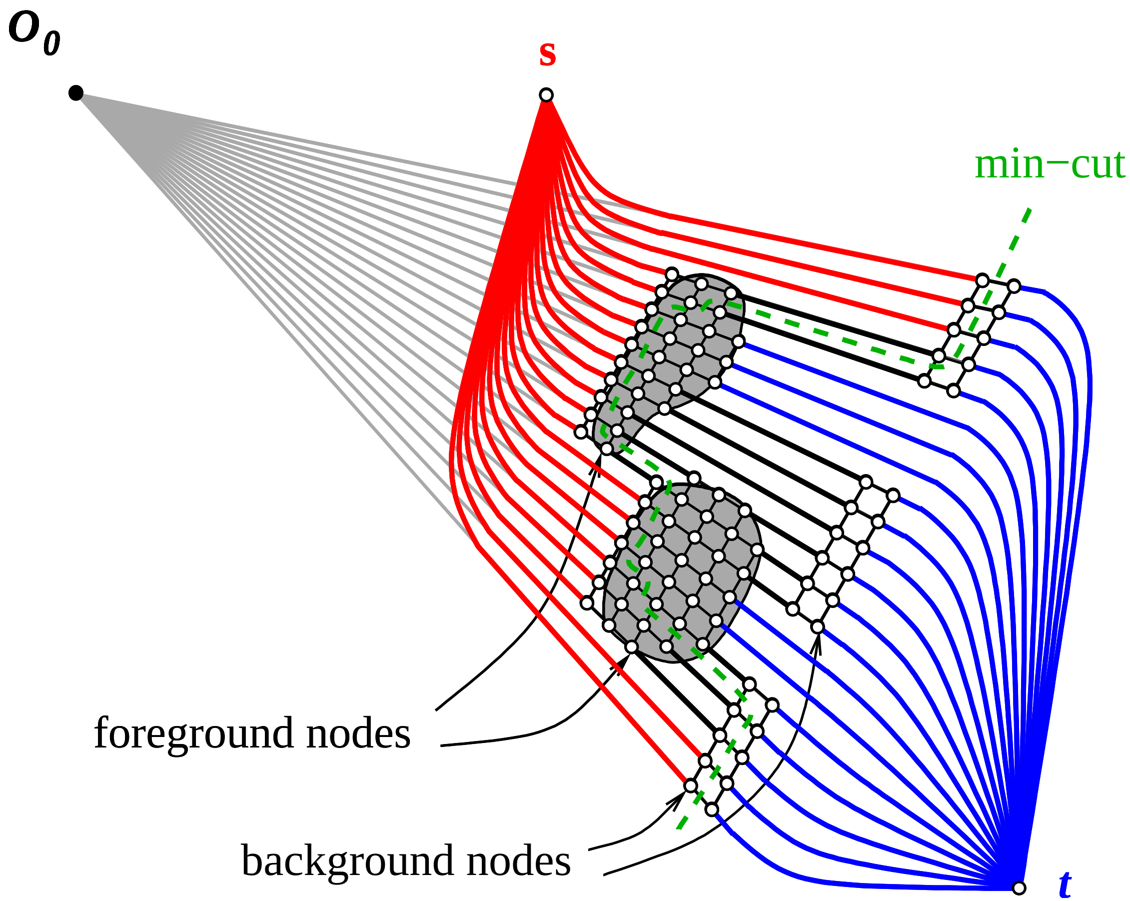


- Small player size in the images
- Occlusions
- Image blur

- **Matting errors**
- **Camera calibration errors**

Model-free matting + reconstruction

View-dependent graph-cut

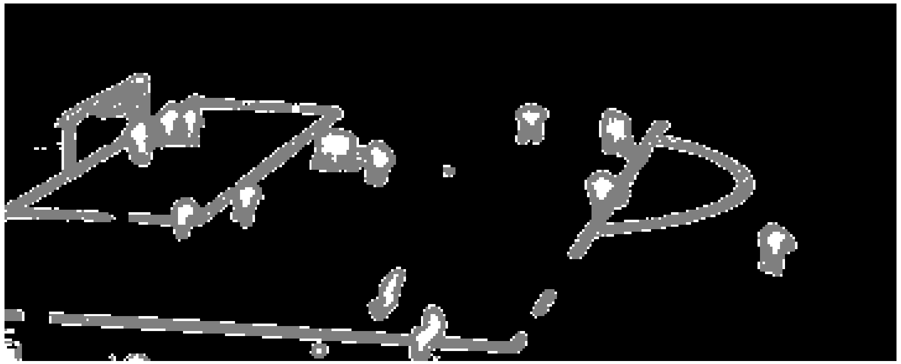


[Guillemaut 3DIM'07]

Reconstruction and matting results

Input reference image

Background image



Trimap

Depth map

Model-free reconstruction in football

Free-viewpoint 3D Video

Goal-view 3D Video

Production trials - BBC Sport

England European cup qualifiers - Wembley, Nov'07

Rugby 6 Nations - Twickenham, Feb'08

Sports Production

Production trials - BBC Sport

England European cup qualifiers - Wembley, Nov'07

Rugby 6 Nations - Twickenham, Feb'08

Sequence 1

Sequence 2

Model-free reconstruction in rugby

Model-free reconstruction in sports

Input camera views

Output goal-keepers view

[Kilner CVMP'06]

Model-based vs. Model Free

Model-based

+ structured animation model
+ robust to visual ambiguities of visual-hull & stereo
- restricted shape: hair, hands
- loss of shape detail: face, hands
- loss of appearance detail: view-dependant

Model-free

+ free-viewpoint video
+ view-dependent optimisation
+ no prior assumptions on scene structure
+ high visual quality approaching captured video
- no structure
- no control of movement

Animation from Surface Capture

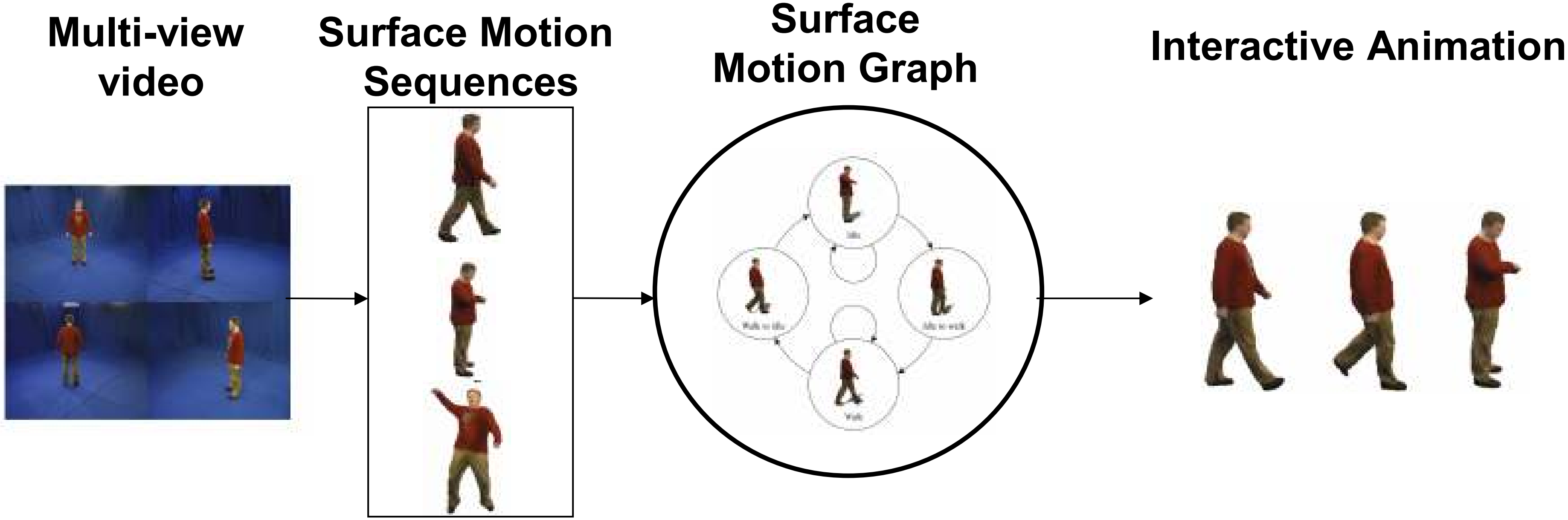
Animation from Surface Motion Capture

Surface Motion Capture

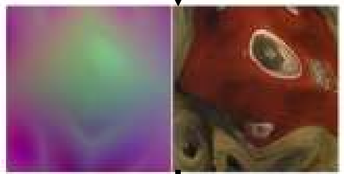
- model-free reconstruction
- representation with consistent temporal structure

Surface Motion Graph

- Example-based animation
- User control of movement



Overview



1. Surface motion capture

- Multiple view video capture of multiple motions
- Visual reconstruction using

→unstructured 3D mesh sequences

2. Structured Representation

- Geometry images (genus-N)
- Spherical mapping to 2D domain

→ temporally consistent mesh structure

3. Surface motion graph

- Spherical matching for motion transitions
- Wide-timeframe surface matching of sequences

→graph of surface motion sequences

4. Interactive animation

- User input of character motion
- User control of viewpoint

→ user controlled character animation

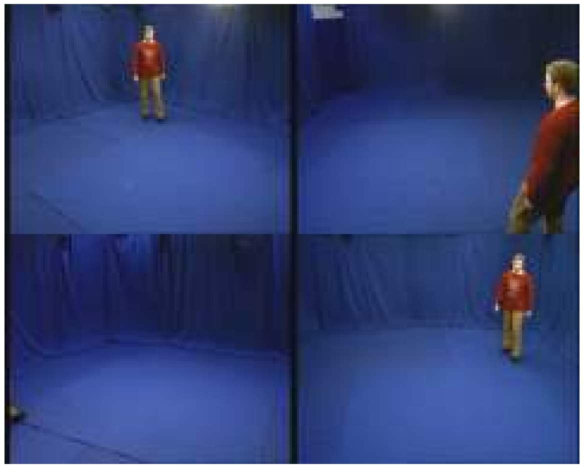
Starck, Miller, Hilton SCA'05
Starck & Hilton IEEE CGA'07

Surface Motion Capture

Video Capture
- 8-10 cameras
- chroma-key studio
- capture volume 2.5m^3

Reconstruction
- Global surface optimisation with graph-cuts
  stereo, silhouette rims, features

Multiple view video



Capture:
walk,run,jump,kick…

Reconstruction

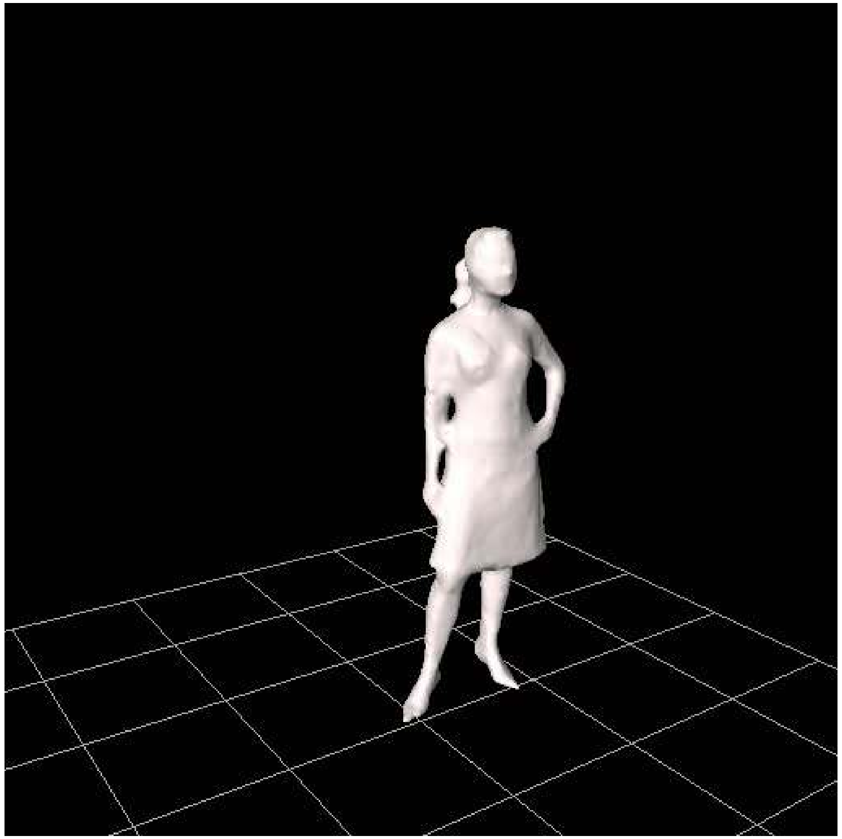
Unstructured 3D Mesh

Starck&Hilton IEEE CGA'07

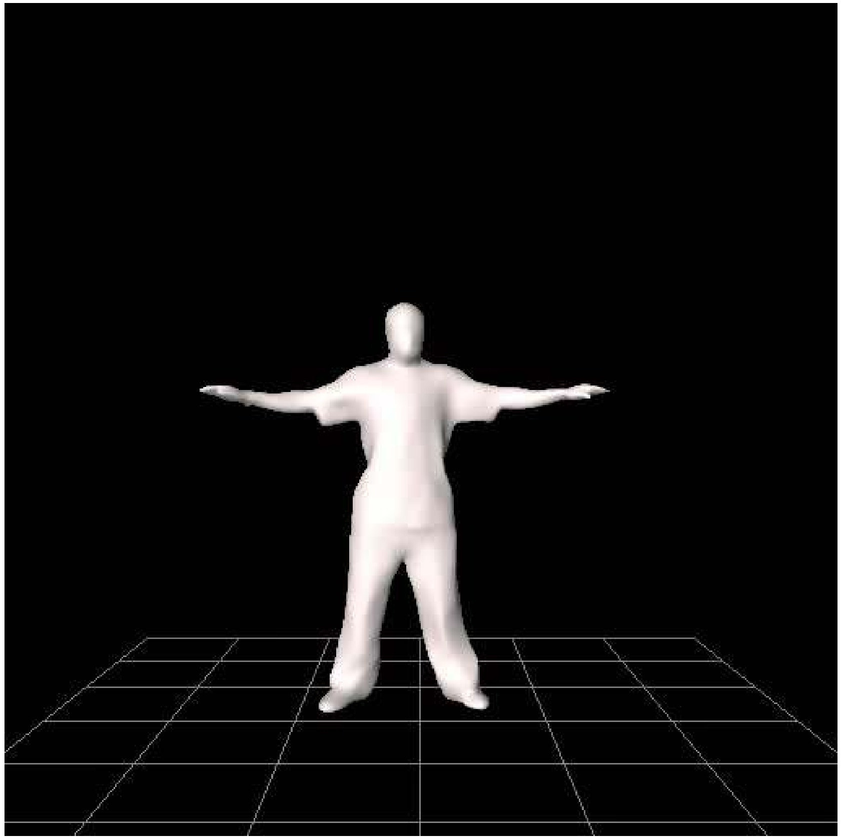
Surface Motion Capture

3D surface sequence

- unstructured raw mesh sequence
- temporally consistent structure



Roxanne



JP

Genus-N Geometry Video

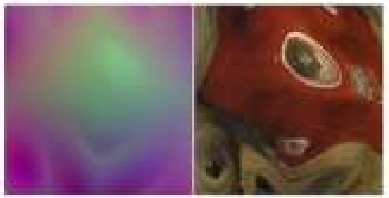
Genus-0 (spherical) geometry images [Praun&Hoppe'03]

Input: 3D mesh sequence *M(t)* of genus-N

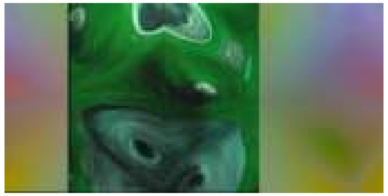
1. Cut Mesh to genus-0 [Steiner&Fischer'02]

2. Progressive mesh decimation: [Hoppe'96]
M → tetrahedra *T*

3. Spherical mapping: *M* → unit sphere *S*
i. *T* → *S*
ii. Progressive embedding *M* → *S*
Optimise embedding to minimise sampling density



4. Image mapping: *S* → image *I*
i. Adaptive resampling of *S* with a regular subdivision
Optimise resampling density [Zhou et al.'04]
ii. Unfold subdivision mesh *S* → *I*



Output: Mapping of mesh to geometry video *M(t)* → *I(t)*
Mesh with constant connectivity, no correspondence

Surface Motion Graphs

Motion graphs [Kovar,Gleicher,Pighin’02]

- Marker-based motion capture (MoCap)
- Skeletal motion
- Example based animation from mocap sequences
- Transitions between similar skeletal pose & motion

Surface Motion Graphs

- Example based animation from surface motion sequences
- Transition between similar surface shape & movement

Problem: Surface-to-surface motion transitions

Surface-to-surface Motion Transitions

Surface Motion Capture

- walk, run, turn, kick, idle

Transitions

Different motions: walk-run, run-walk, walk-kick, walk-idle
Cyclic motions: walk-walk, run-run

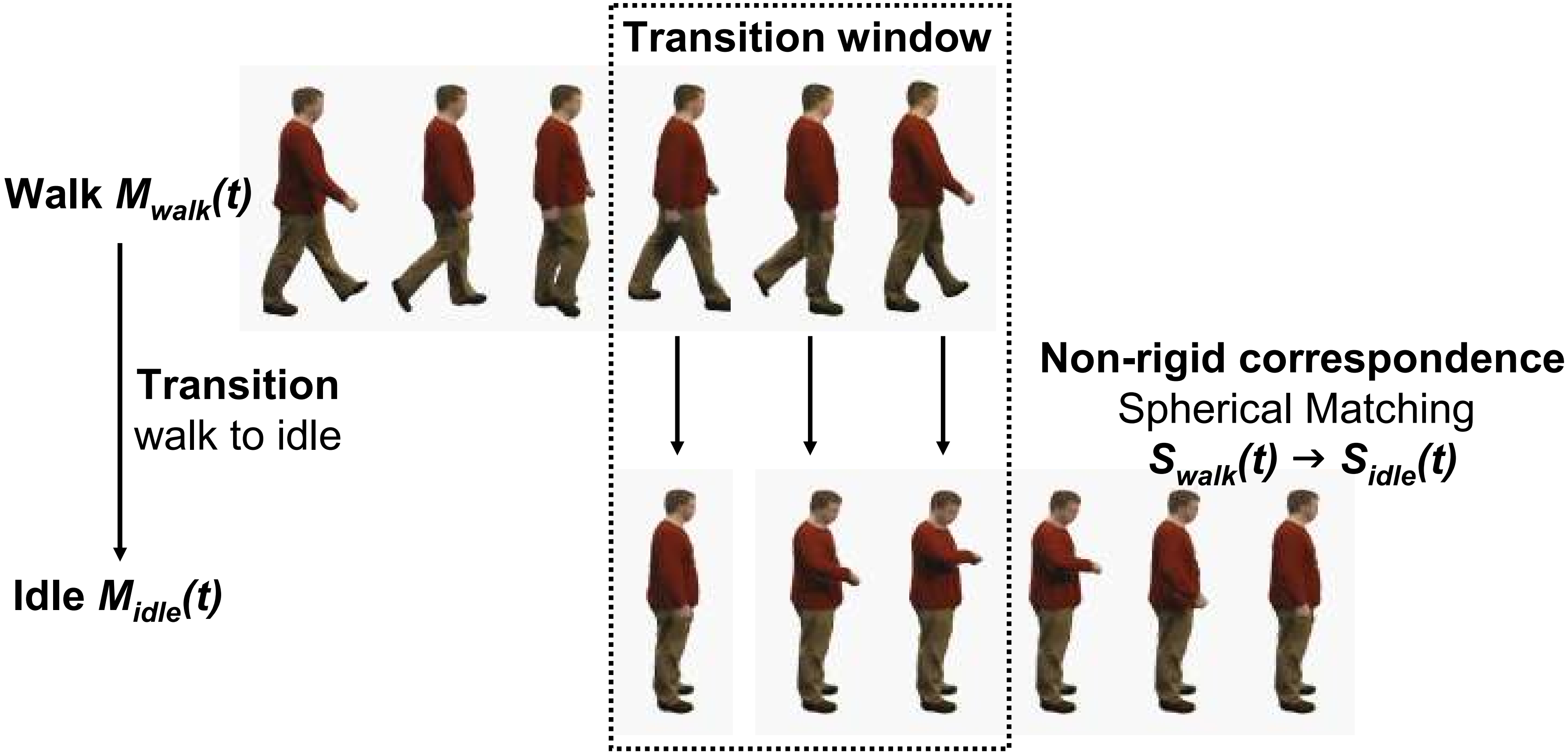
Surface-to-surface sequence matching

1. Identify frames with similar poses/movements
2. Estimate correspondence in shape and appearance
3. Transition by morphing between sequences

Surface Motion Transitions

Transition between motion $M_{walk}(t) \rightarrow M_{idle}(t)$

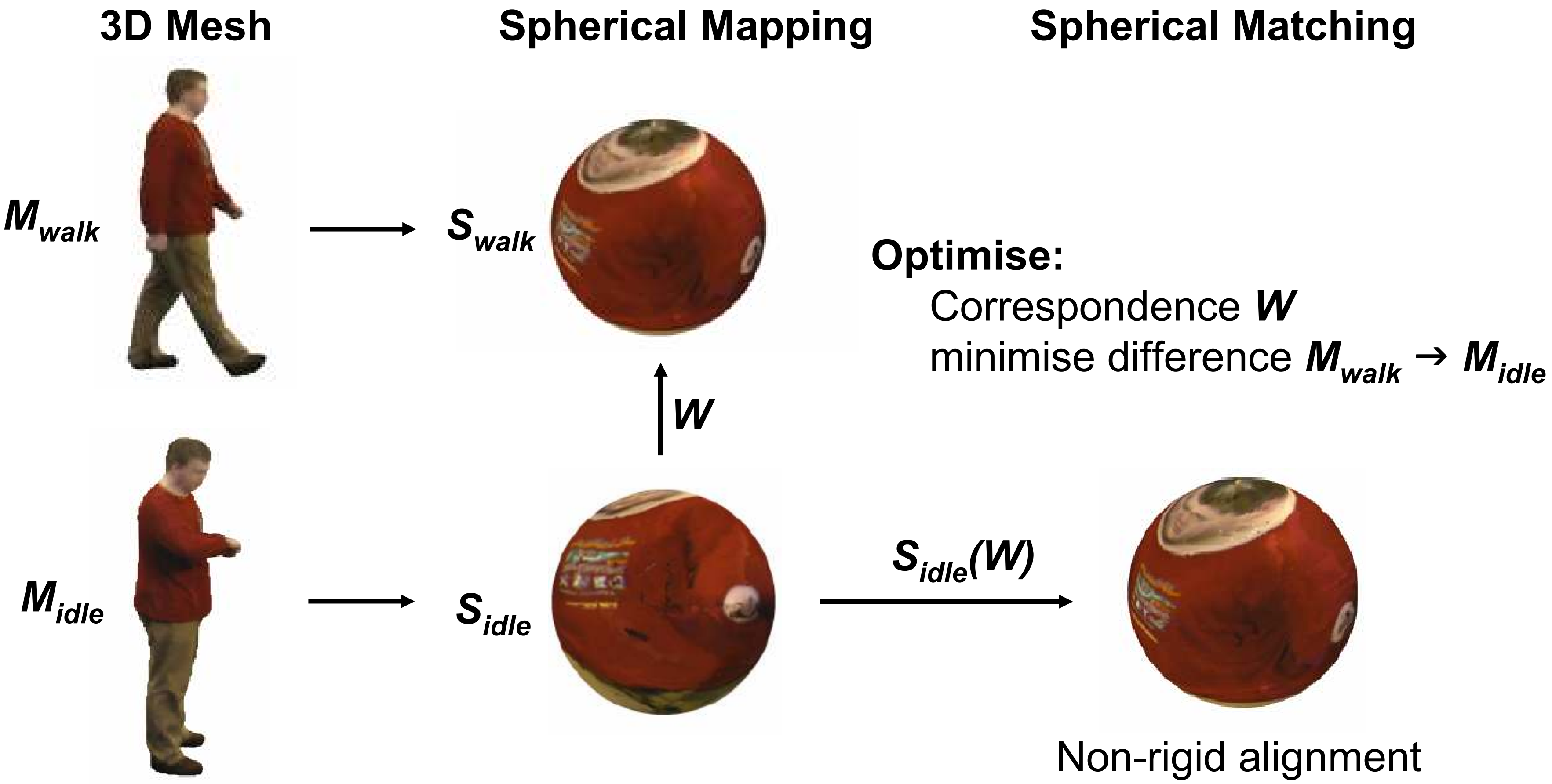
- Non-rigid correspondence
- Blend between geometry videos



Spherical Matching of Genus-0 Surfaces

Non-rigid correspondence: $M_{walk}(t) \rightarrow M_{idle}(t)$

- Match in 2D spherical domain $S_{walk}(t) \rightarrow S_{idle}(t)$
- Guarantee 1:1 correspondence everywhere
- Preserve the spherical embedding



Starck&Hilton ICCV'05

Multi-resolution Spherical Matching

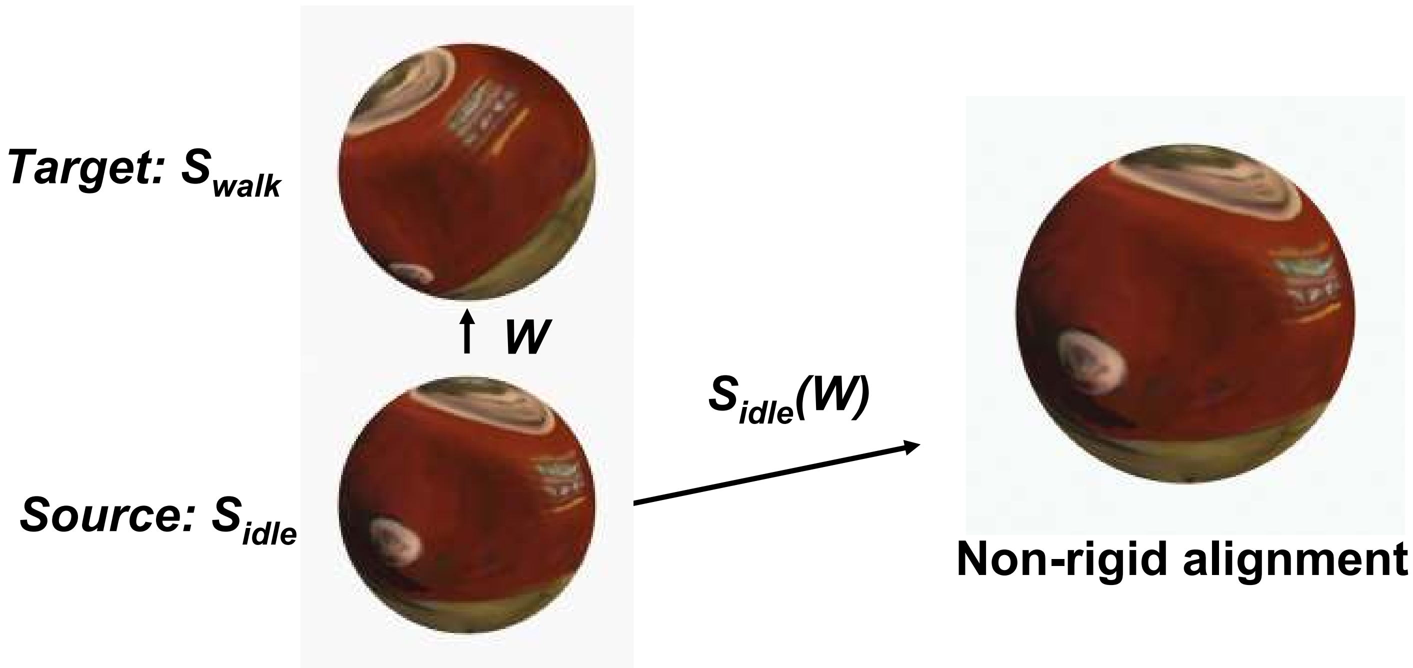
Energy minimisation:

Optimise correspondence $\boldsymbol{W}$ which minimises difference $\boldsymbol{M_{walk}} \rightarrow \boldsymbol{M_{idle}}$

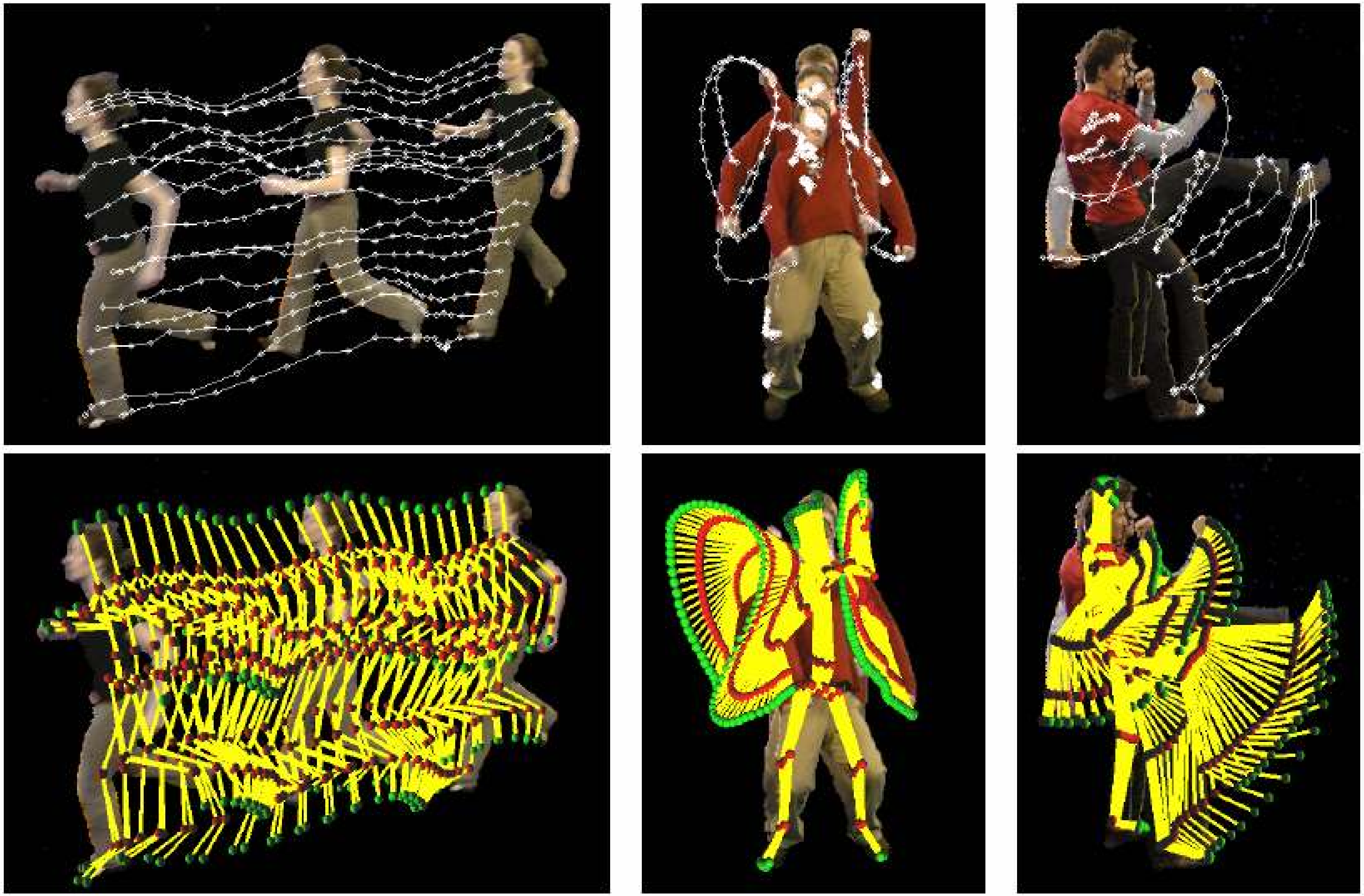
Difference: Shape, Color, Normal

Robust optimisation in spherical domain $\boldsymbol{S_{walk}} \rightarrow \boldsymbol{S_{idle}}$

- coarse-to-fine: mesh subdivision
- multi-resolution: many-to-many to one-to-one correspondence $\boldsymbol{W}$



Spherical Matching Results

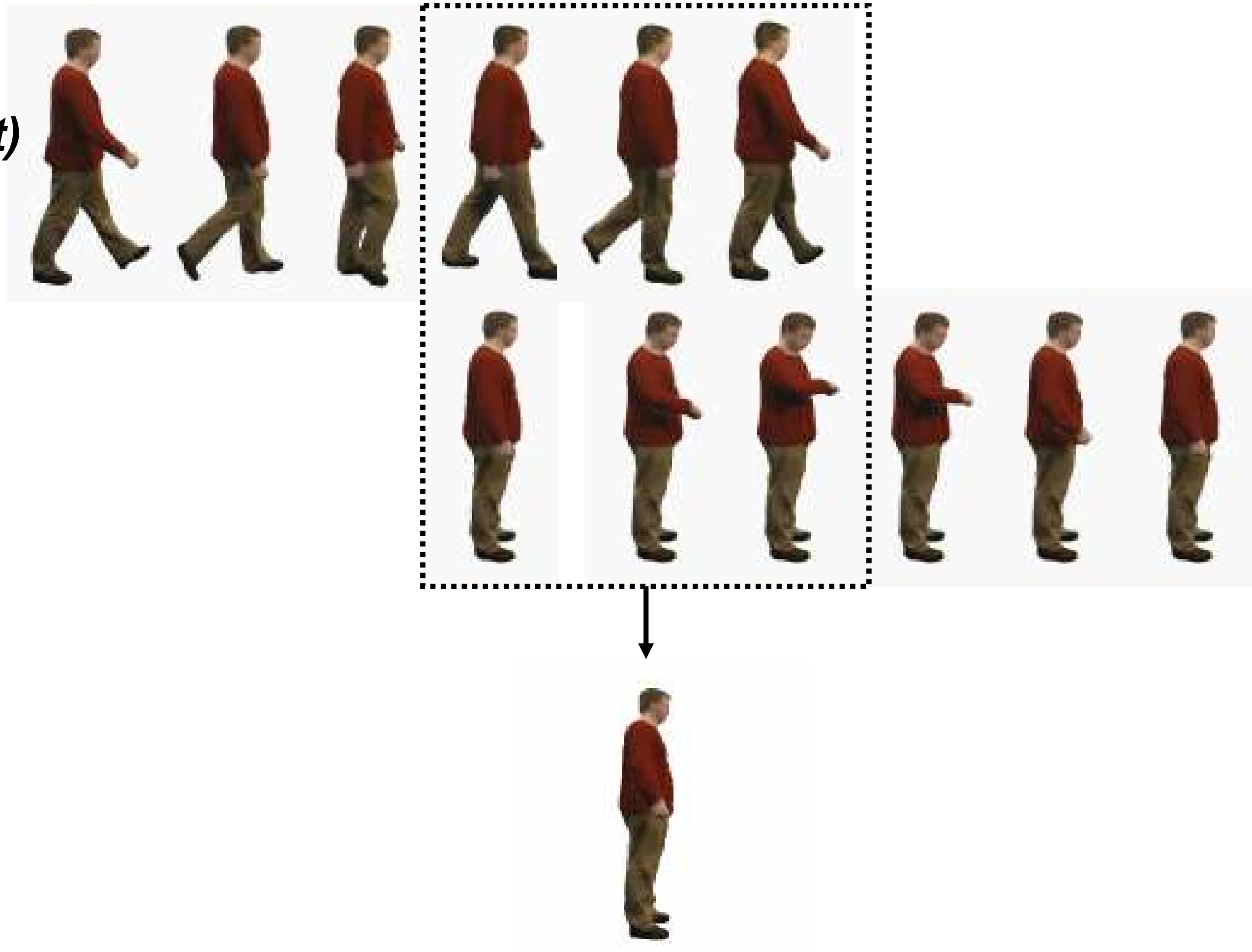


Surface Motion Transitions

Walk $M_{walk}(t)$

Idle $M_{idle}(t)$

Transition $M_{walk\text{-}idle}(t)$



Transition

- Spherical matching $S_{walk}(t) \rightarrow S_{idle}(t)$ over 5 frame window
- Linear blend of shape & color

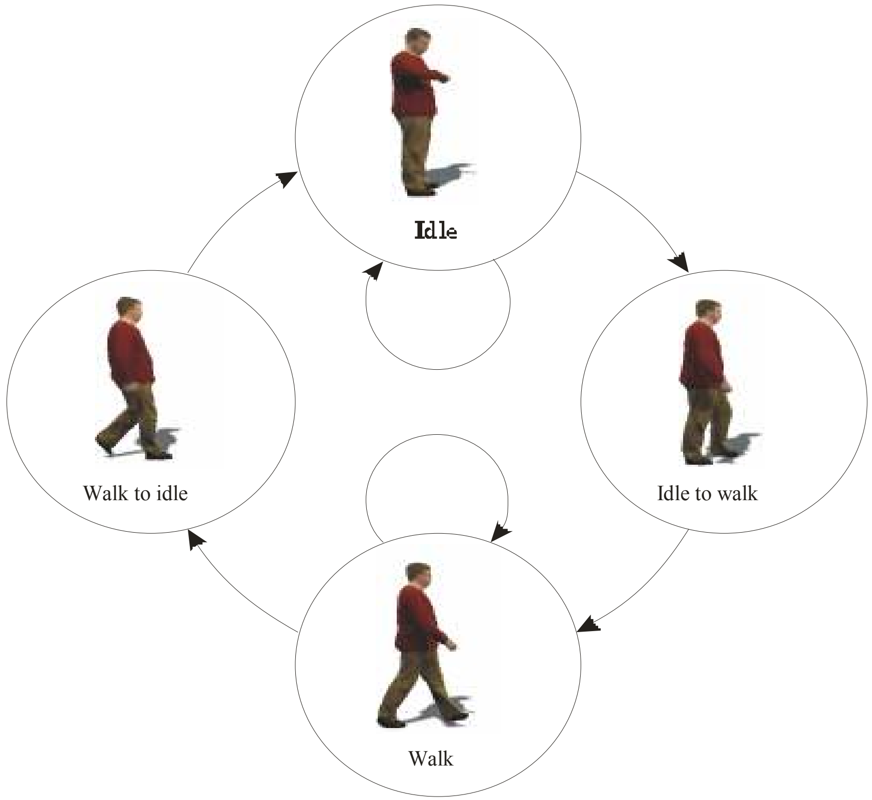
Surface Motion Graph of Geometry Videos

Geometry videos: walk, run, turn, kick, idle

Pre-computed motion transitions:

walk-idle,walk-run, run-walk, walk-kick

cyclic transitions: walk-walk, run-run



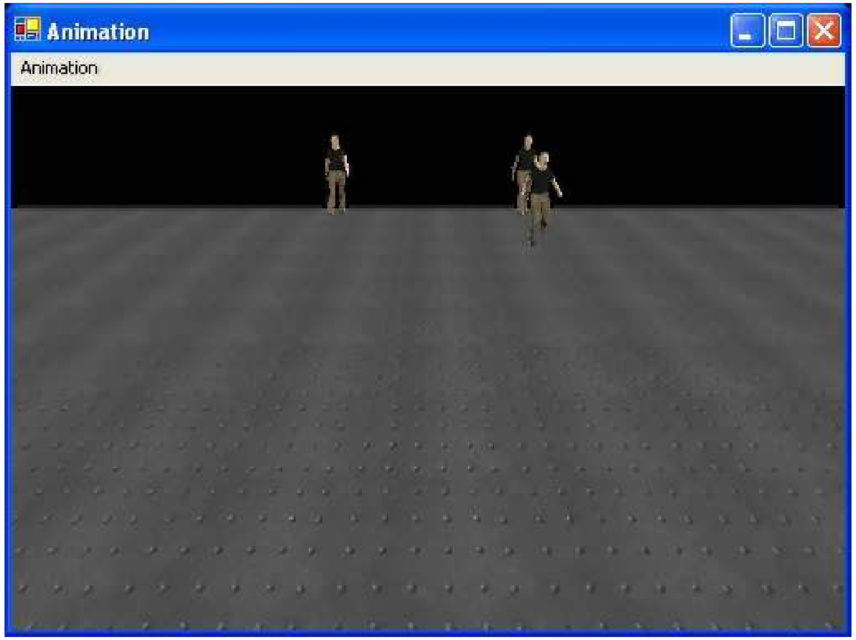
Surface Motion Capture Animation - Results

Geometry image resolution: 256^2 (24 bit)
Rendering performance: 88 frames/second on nVidia 6600 GT

Surface Motion Capture Animation - Results

Geometry image resolution: 256^2 (24 bit)
Rendering performance: 88 frames/second on nVidia 6600 GT

Surface Motion Capture Animation - Results

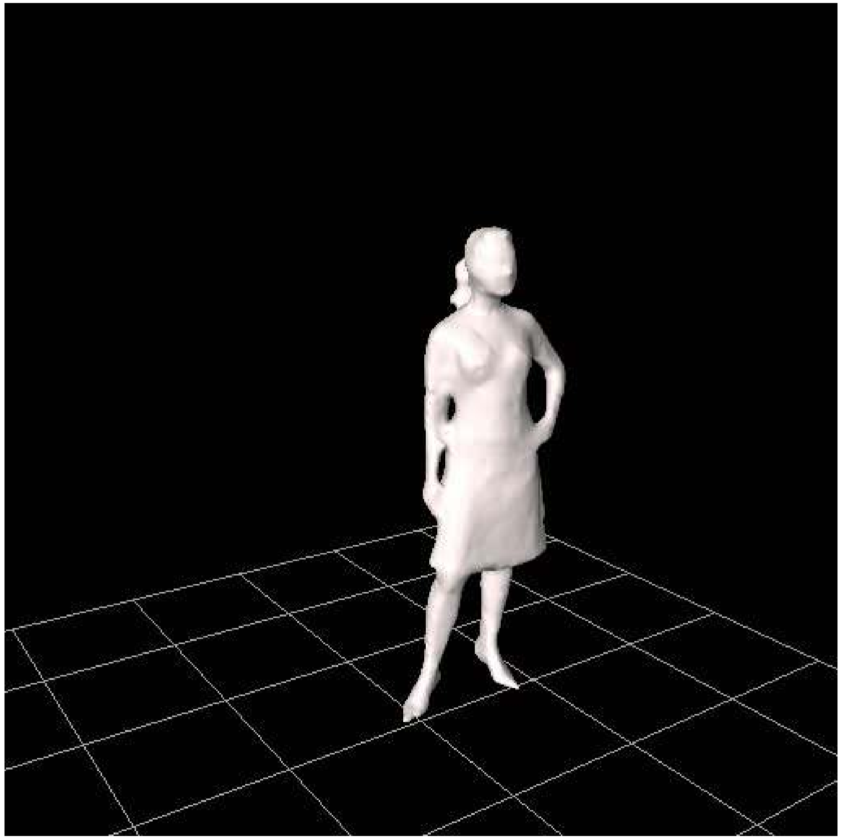


Geometry image resolution: 256^2 (24 bit)
Rendering performance: 30 frames/second on nVidia 6600 GT

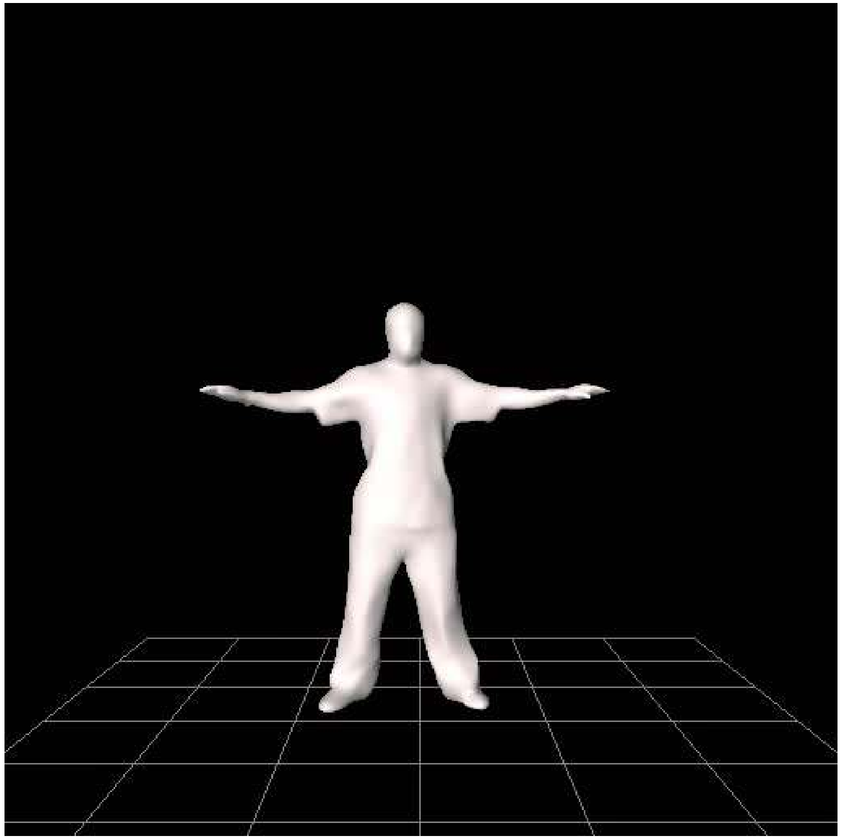
Surface Motion Processing

Temporal Processing

- Low-pass filter to reduce reconstruction noise
- temporal editing of shape or appearance
- temporally consistent representation with correspondence



Roxanne



JP

Surface Matching for Genus-N Surfaces

Multi-resolution non-rigid surface matching

- Progressive mesh hierarchy for coarse-to-fine matching
- Global rigid alignment at lowest resolution
- Robust multi-point matching [Chui&Rangarajan 2003]
- Annealed refinement of surface-to-surface correspondence
- Multi-resolution matching range based on mesh hierarchy
- Minimise difference in shape and appearance

High-Resolution non-rigid surface matching

- refinement of multi-resolution matching to sub-pixel accuracy
- normalised cross-correlation of surface using input images
- hard assignment of corresponce

Output: Dense non-rigid surface-to-surface correspondence

Wide-timeframe Surface Matching

Match surfaces from different motions
- gross non-rigid deformation
- changes in topology

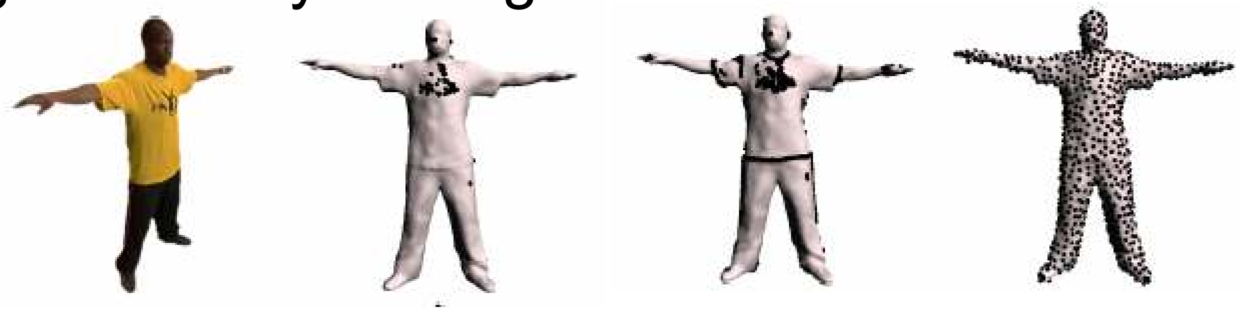
Wide-timeframe Surface Matching

Deformation invariant feature matching

- shape + appearance

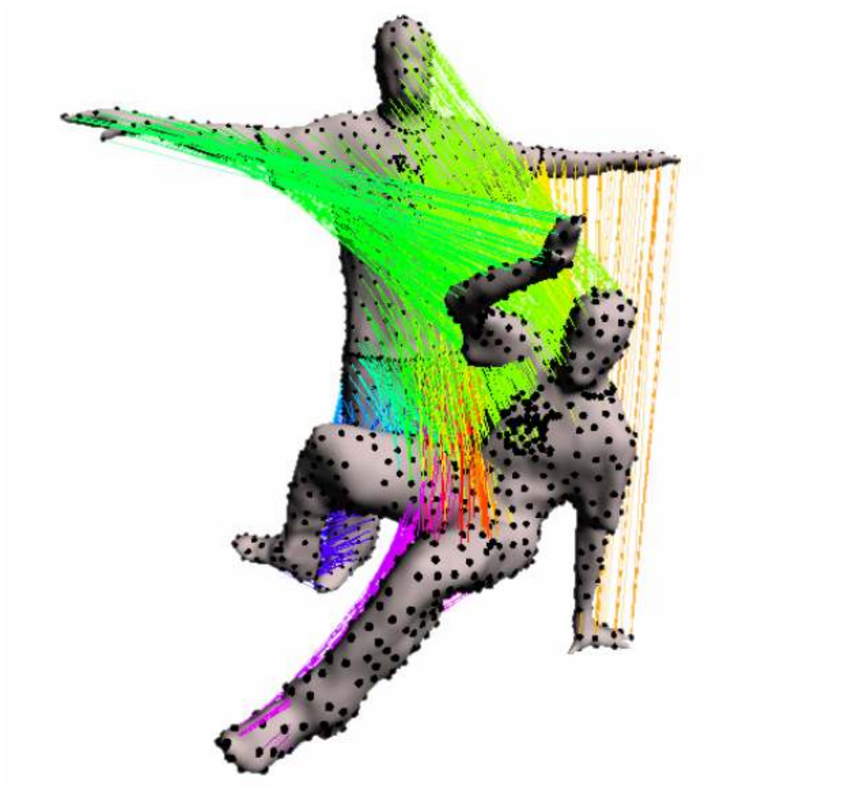
MAP-MRF correspondence labelling

- sparse to dense labelling (corners, edges, regions)
- belief propagation
- cross-validation for consistent matching
- regularised by non-rigid deformation



[Starck & Hilton ICCV'07]

Wide-timeframe Surface Matching



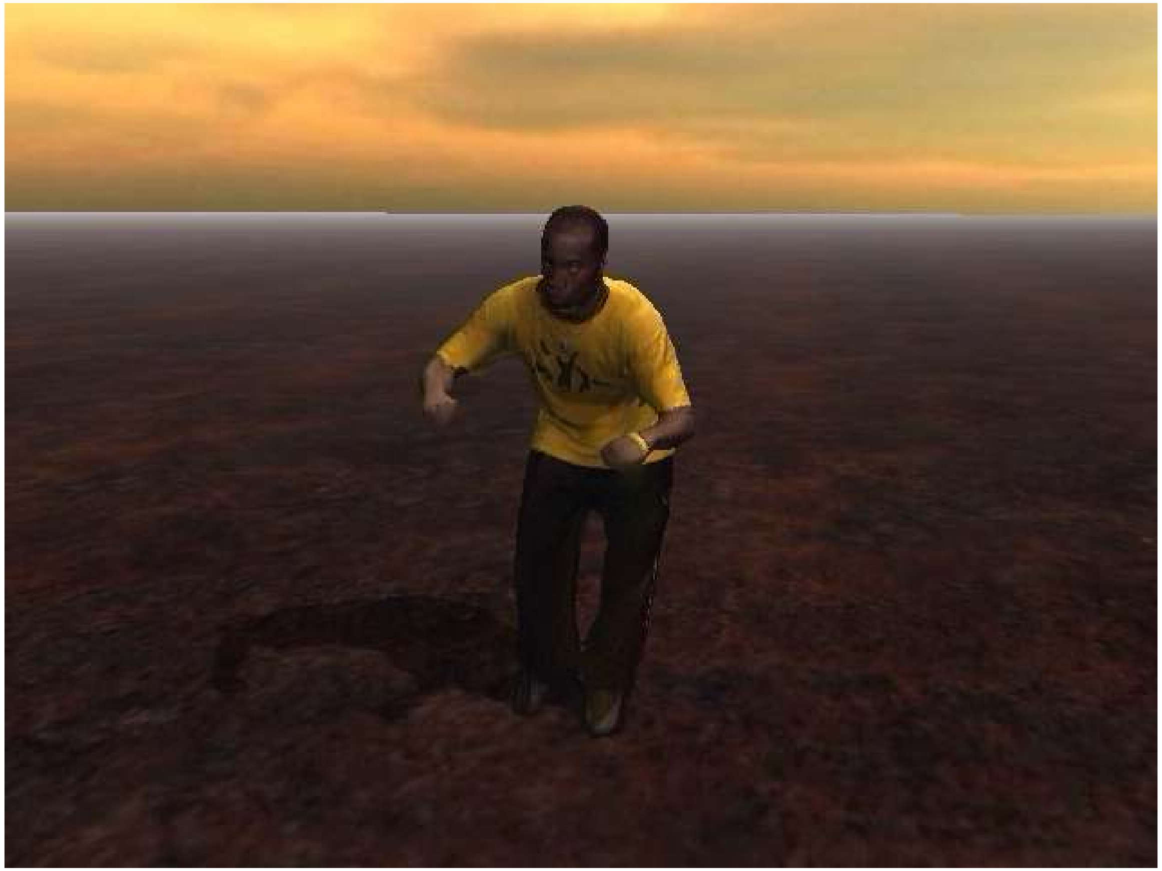
Wide-timeframe Surface Matching

Correspondence labelling for wide-timeframe
free-form surface matching

Paper ID 434

[Starck & Hilton, ICCV'07]

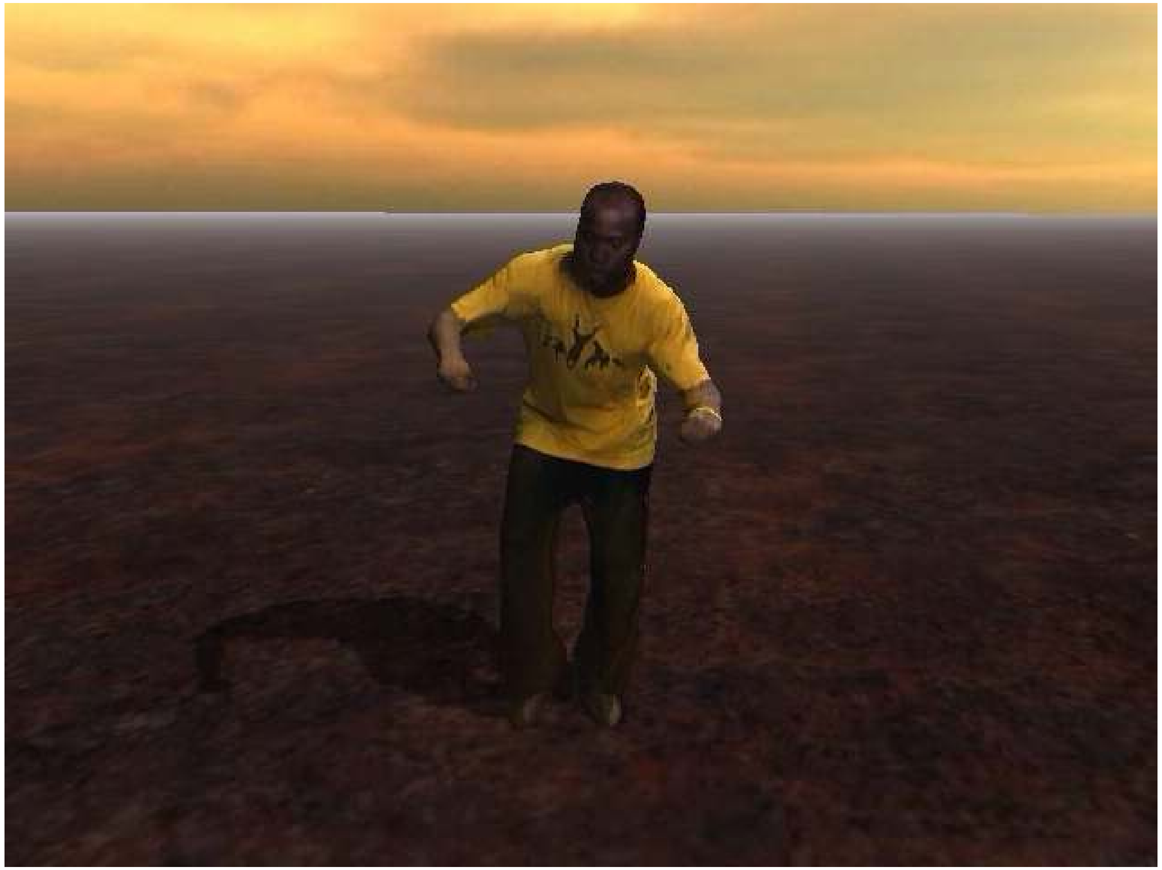
Animation from Surface Motion Capture



Game Demo JP

Starck&Hilton IEEE CGA'07

Surface Motion Capture



Starck&Hilton IEEE CGA'07

Animation from Surface Capture

Starck&Hilton Jan'06

Animation from Surface Capture

Starck&Hilton IEEE CGA'07

Animation from Surface Capture

Starck&Hilton IEEE CGA'07

Performance

Pre-processing:

- **Capture time per person:** 1hr. + studio setup
- **Reconstruction time:** 15s/frame
- **Progressive mesh construction:** 8 min/frame
- **Spherical parameterisation:** 6 min/frame
- **Geometry image construction:** 30s/frame
- **Spherical matching:** 3 min/frame
- **Total:** 18min/frame

Representation: $512\text{x}256^2$ 24bit geometry+texture image

- **Captured video (10 views):** 12MB/frame
- **Geometry video:** 0.4MB/frame uncompressed
- **Walk sequence (23 frames):** 9MB uncompressed
- **Motion graph (walk,turn,run,idle):** 37MB uncompressed

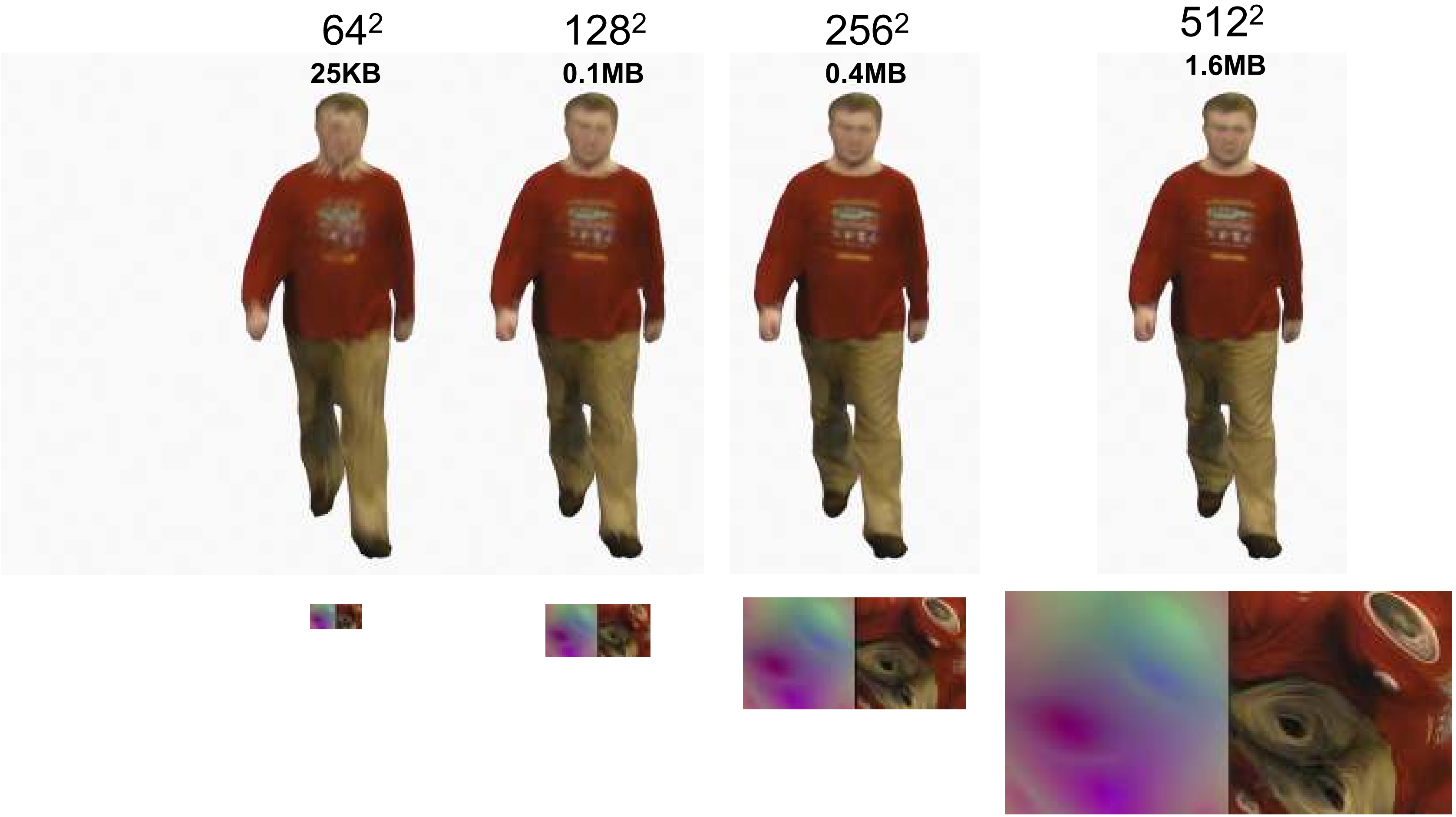
Rendering: nVidia 6600 GT (128MB graphics RAM)

- **one person:** 88fps
- **3 people:** 30fps with texture transfer

Representation Cost

256^2x2 24bit geometry+texture image

- **Captured video (10 views):** 12MB/frame
- **Geometry video:** 0.4MB/frame uncompressed
- **Walk sequence (23 frames):** 9MB uncompressed
- **Motion graph (walk,turn,run,idle):** 37MB uncompressed



Conclusions – Video-based Animation

Model-based

- Structured representation
- Limited visual quality

Model-free

- High visual quality
- No animation control (unstructured)

Conclusion: Model-free reconstruction (no restrictions on shape)
+ Model-based structure (animation)

Surface Motion Capture

- High visual quality (model-free)
- Structured representation (geometry videos)
- Animation control 'Surface Motion Graphs'
- Surface matching for transitions & temporal processing

➢ **Realistic animation of people**

Future work – 3D Video Production

- Integration into animation production pipeline
  - standard animation controls (rigging)
  - editing motion
  - editing appearance

- Geometric and appearance detail
  - face
  - hands
  - hair

- Re-lighting

Video-based Animation of People

Adrian Hilton & Jonathan Starck

Visual Media Research Group
University of Surrey, UK
http://www.ee.surrey.ac.uk/CVSSP/VMRG/surfcap.html