Stefan Krauß, Juliane Hüttl
SE Visual Computing, SoSe 2011, HU-Berlin

PERFORMANCE CAPTURE FROM SPARSE MULTI-VIEW VIDEO

Uses of Motion/Performance Capture

- movies
- games, virtual environments
- biomechanics, sports science, medicine
- robot control
- human-computer interaction

Types of Motion/Performance Capture


electromagnetic

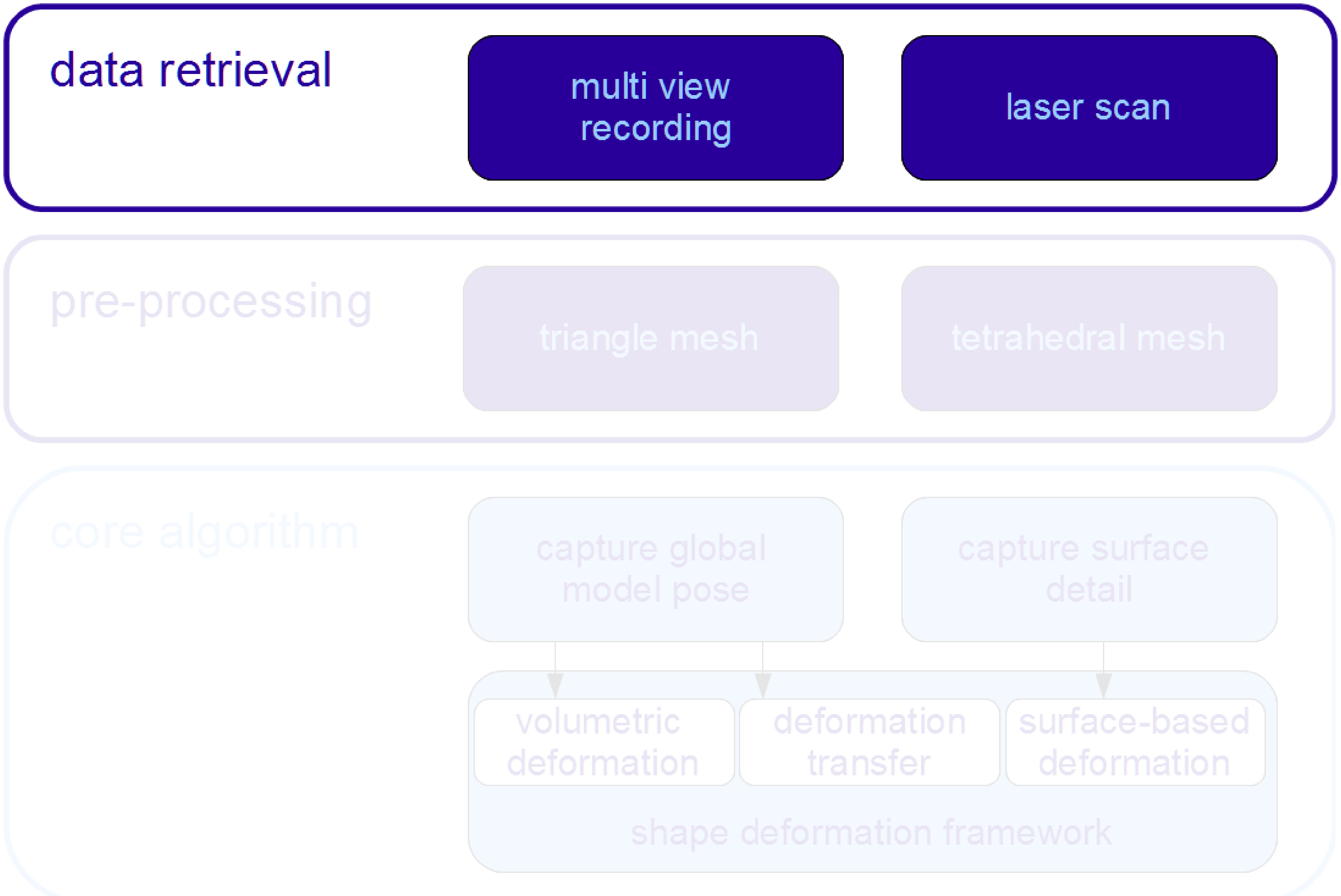

inertial


optical marker-based


electromechanical


optical markerless

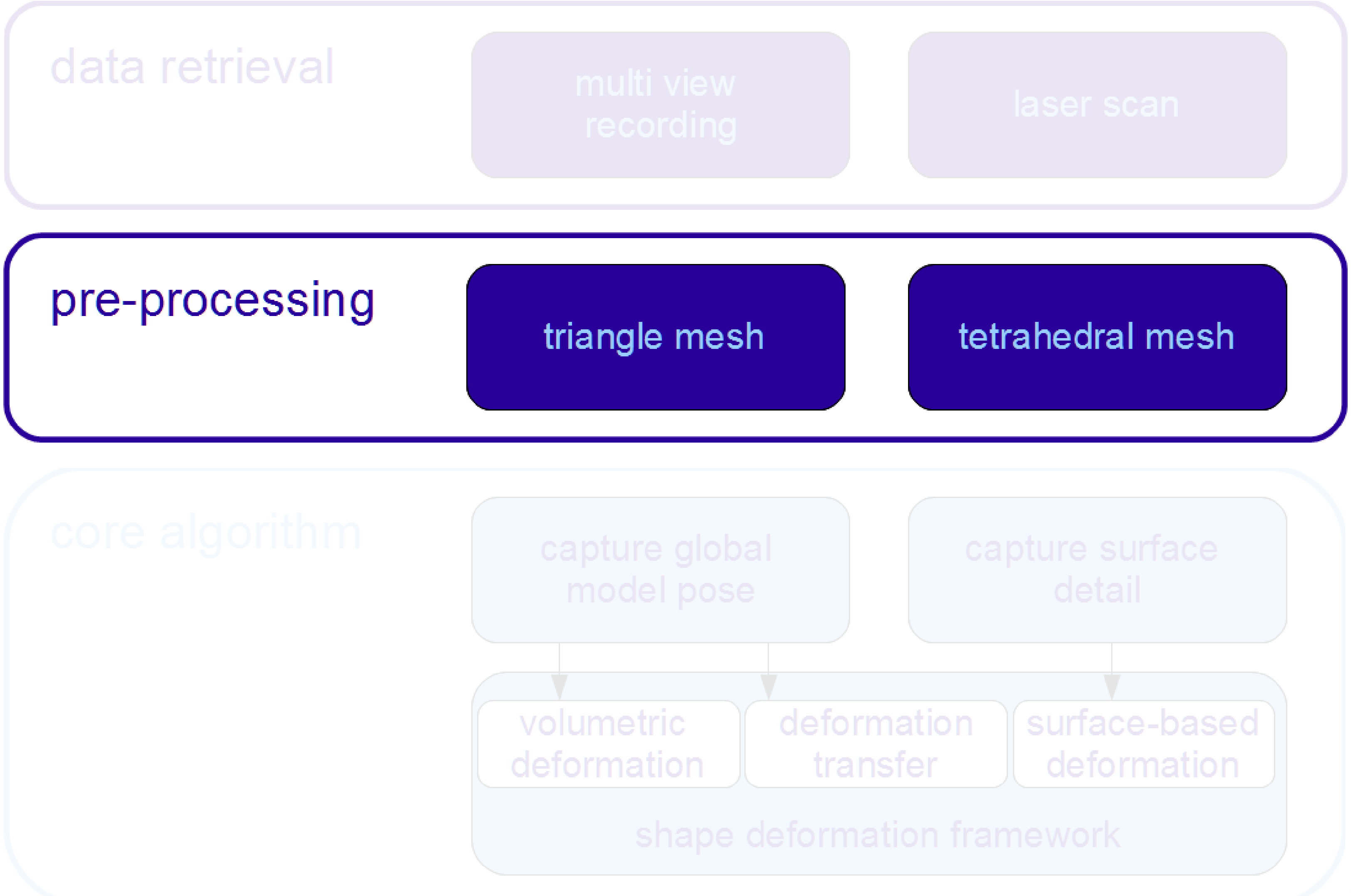
Processing Steps



Data Retrieval

- full body laser scan
- 8 cameras , 24fps , 1004x1004 pixels , placed in circular configuration

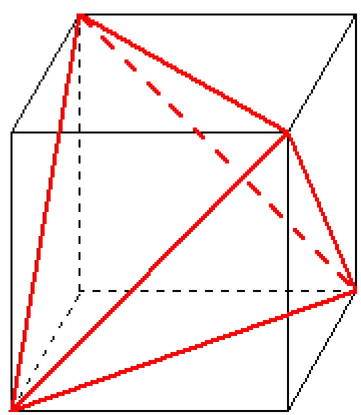
Processing Steps



Pre-Processing

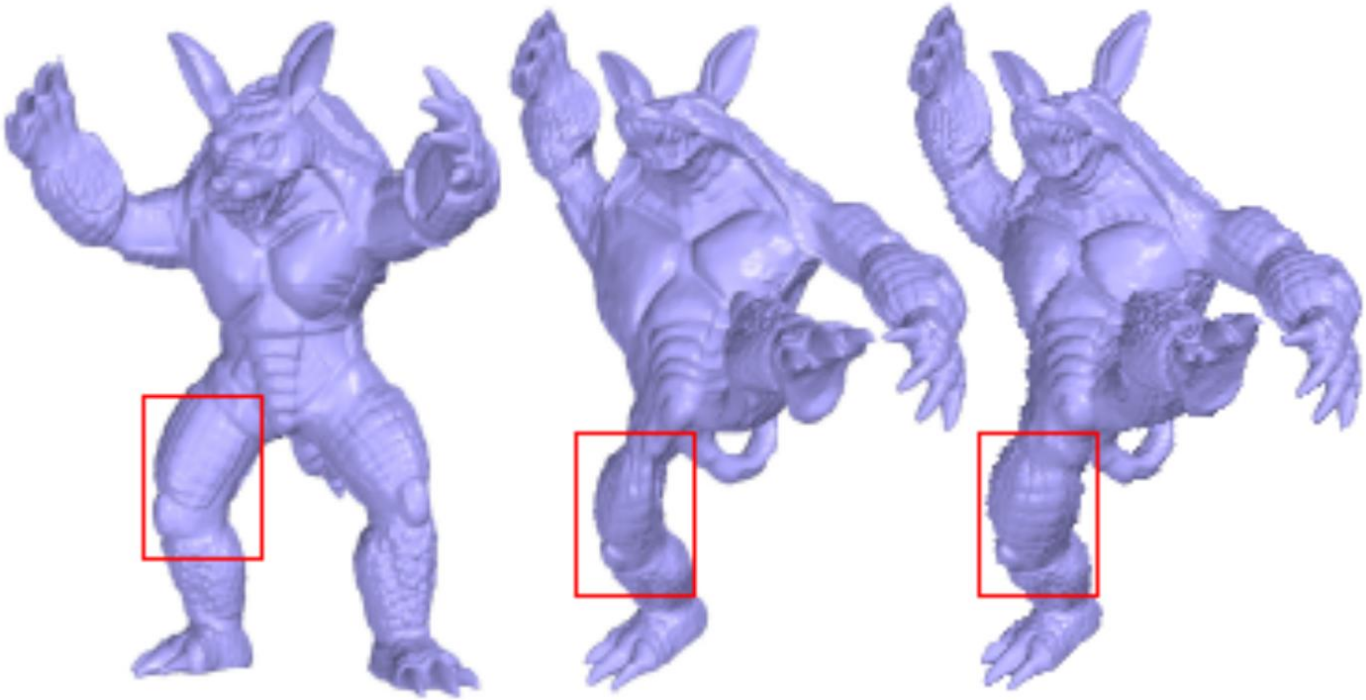
- color-based background subtraction for silhouette of model
- generate triangle mesh via Poisson surface reconstruction
- create tetrahedral version of surface scan

Delaunay tetrahedralization

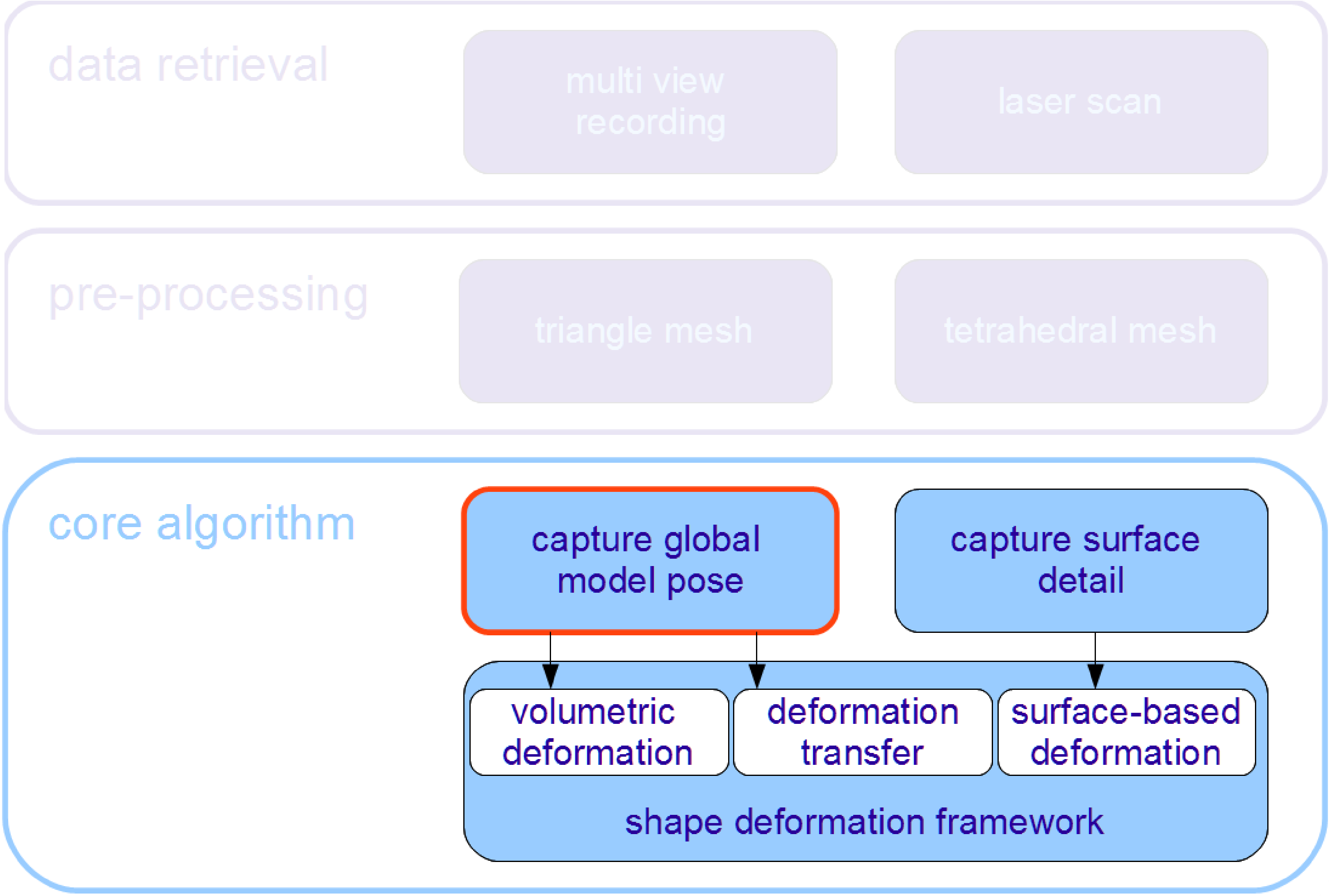


Visual Computing
Department of Computer Science

Deformation Problem on Triangle Meshes



Processing Steps



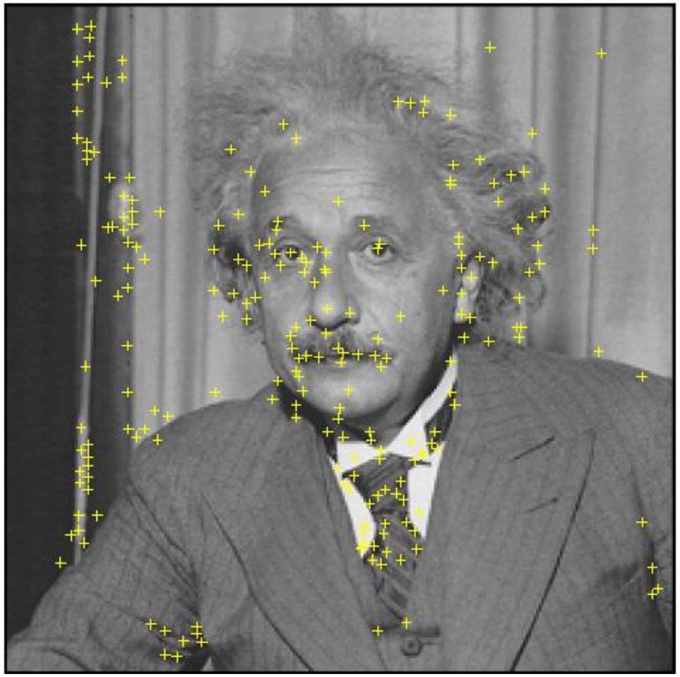
Capturing the Global Model Pose

- recover global pose of tetrahedral mesh for each time step of video
- subdivide surface of model into 100-200 regions
- compute deformation constraints from each pair of subsequent time steps

1. Pose Initialization from Image Features
2. Refining the Pose using Silhouette Rims
3. Optimizing Key Handle positions

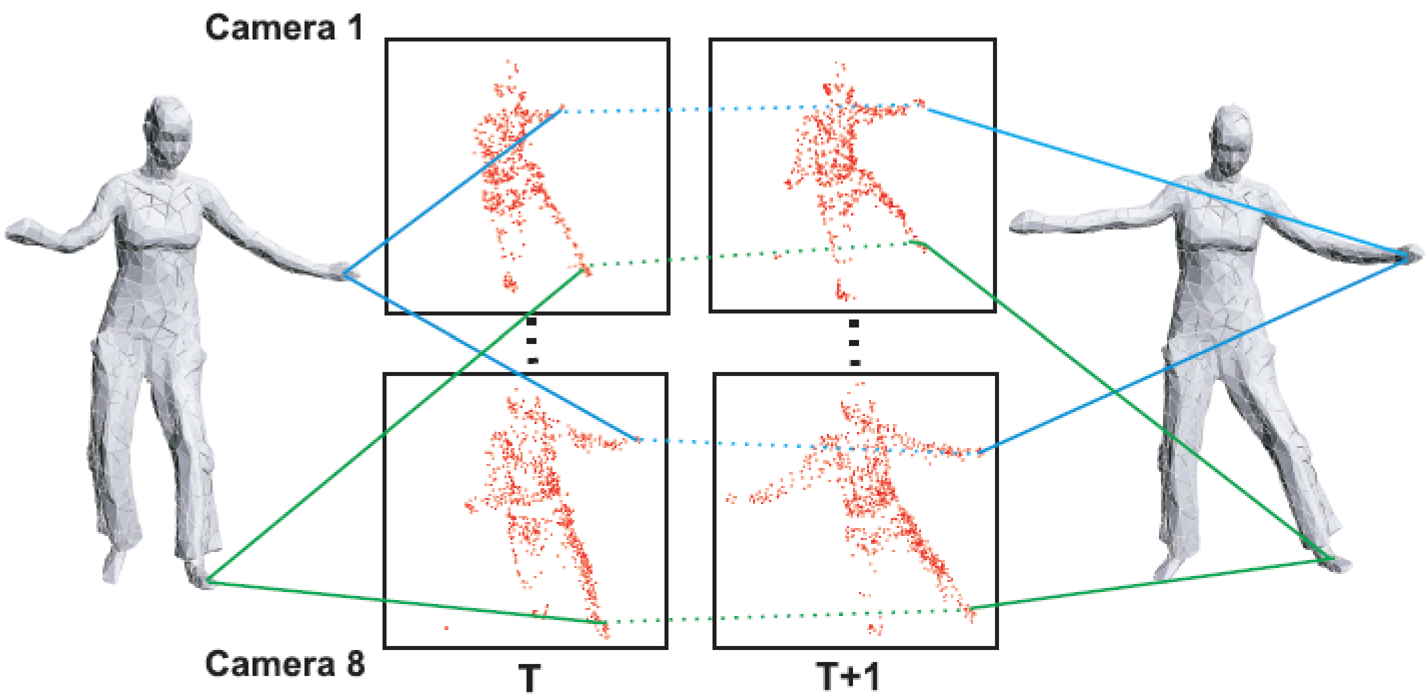
Pose Initialization from Image Features

- SIFT – scale-invariant feature transform
- features largely invariant under
  - illumination
  - out-of-plane rotation
  - image noise
  - small changes in viewpoint
  - scaling
- reliable correspondence finding for fast scene motion



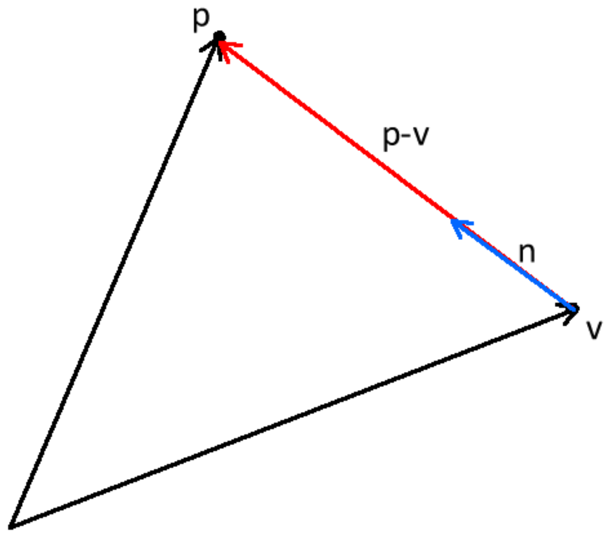
Pose Initialization from Image Features

- associate vertices of tetrahedral mesh with descriptors at time t
- compare intersection point of reprojected rays with vertex
- find corresponding association at time $t + 1$

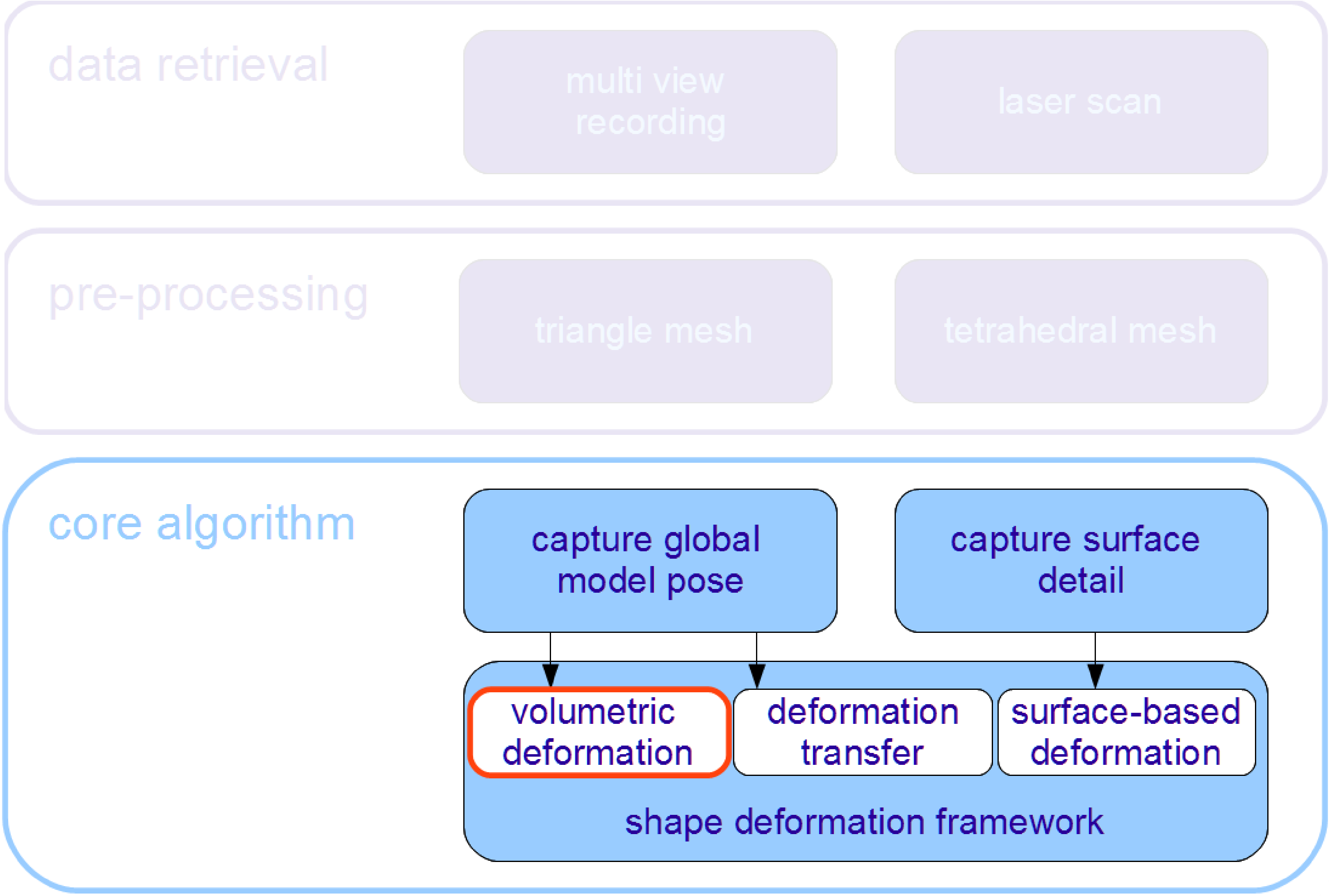


Pose Initialization from Image Features

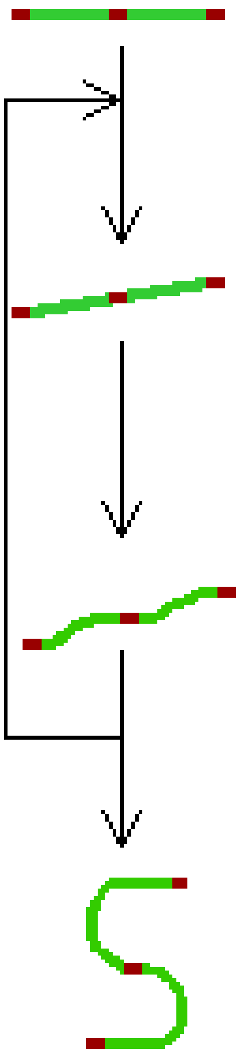
- each vertex found is candidate for deformation handle
- find one best handle for each region:
  normal n of vertex v most collinear with difference vector of intersection point p and vertex v
- define intermediate target position
- step-wise deformation until silhouette overlap error is minimal



Processing Steps



Volumetric Deformation



1. Linear Laplacian deformation

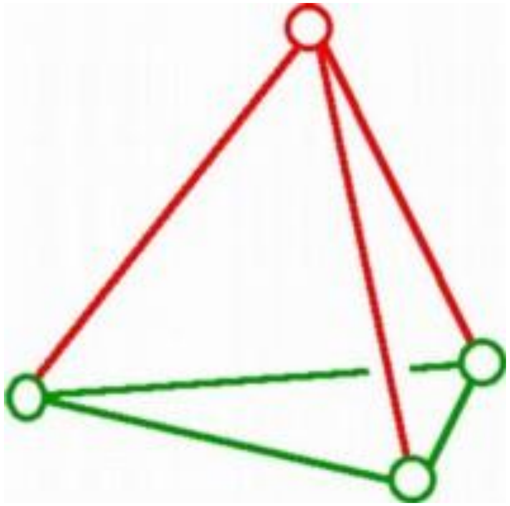
2. Rotation extraction

3. Update

Iterate until non-rigid deformation is minimized

Linear Laplacian Deformation

- calculate gradient operator matrix G for each vertex of every tetrahedron
- construct Laplacian operator from G and volumina D (and constraints)
- solve Laplacian System:
- $Lv = \delta$ with $L = G^T D G$ and $\delta = G^T D g$
- calculate new vertex-positions



Rotation Extraction and Update

- extract transformation matrix for each tetrahedron
- split transformation matrix in rigid and non-rigid part
- iterate until non-rigid-part is minimum
- update differential

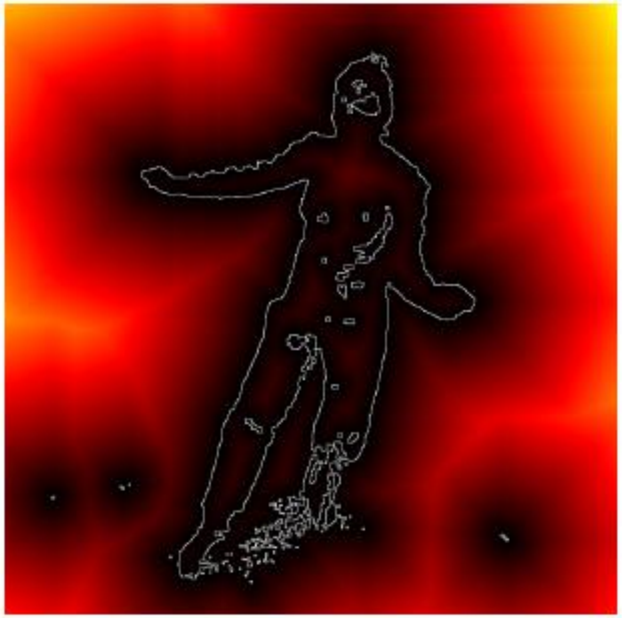
Capturing the Global Model Pose

- recover global pose of tetrahedral mesh for each time step of video
- subdivide surface of model into 100-200 regions
- compute deformation constraints from each pair of subsequent time steps

1. Pose Initialization from Image Features
2. Refining the Pose using Silhouette Rims
3. Optimizing Key Handle positions

Refining the Pose using Silhouette Rims

- find rim vertices
- compute distance field
- value from distance field at projected location defines displacement length



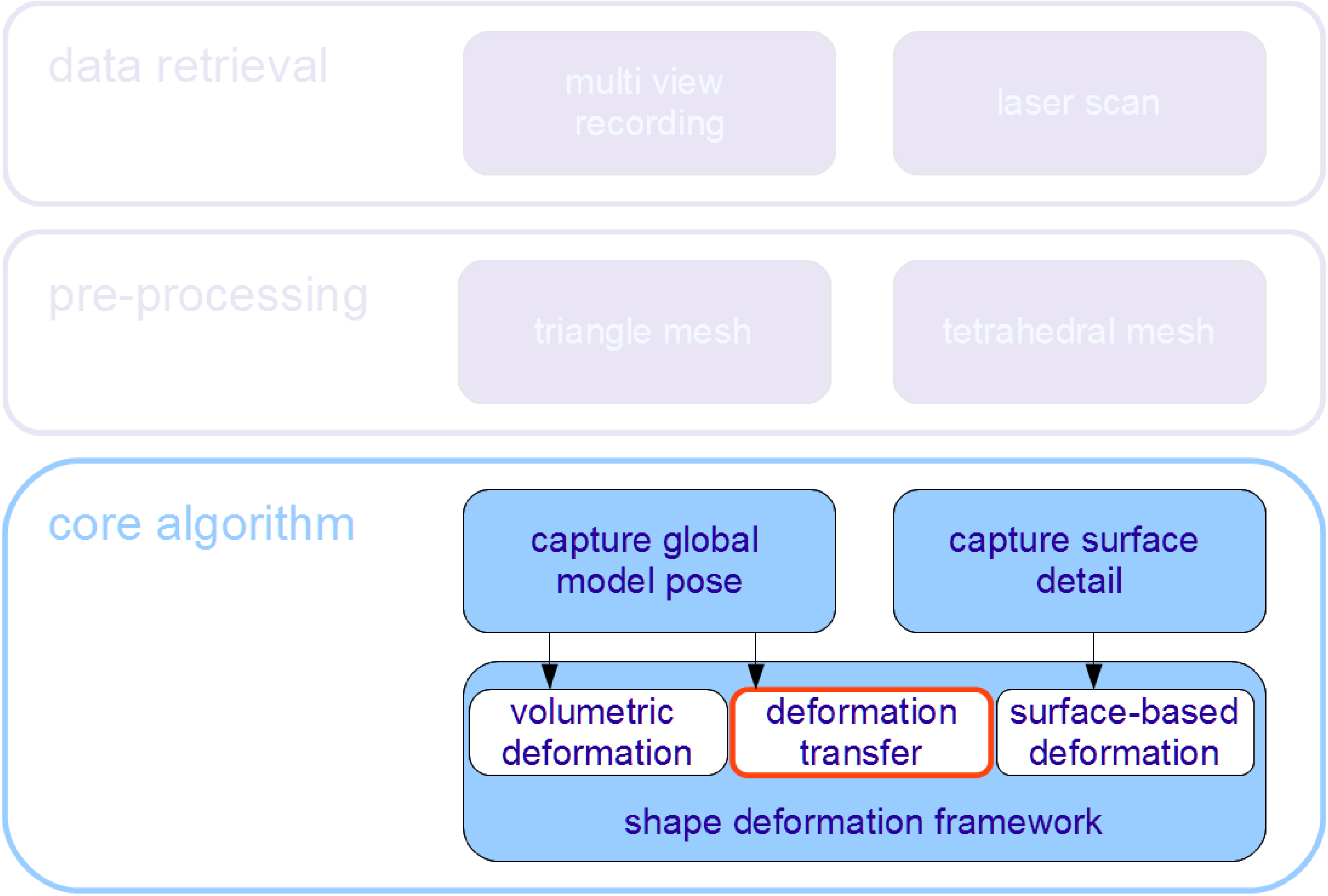
Optimizing Key Handle positions

- find optimal multi-view silhouette overlap
- choose 15-25 key vertices manually
- optimize vertex positions until tetrahedral deformation produces optimal silhouette overlap (based on overlap error)

model and silhouette overlap before optimization

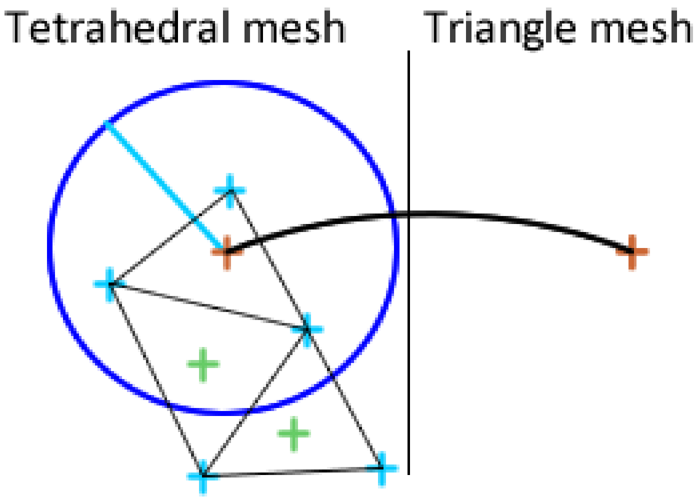
after key vertex optimization

Processing Steps

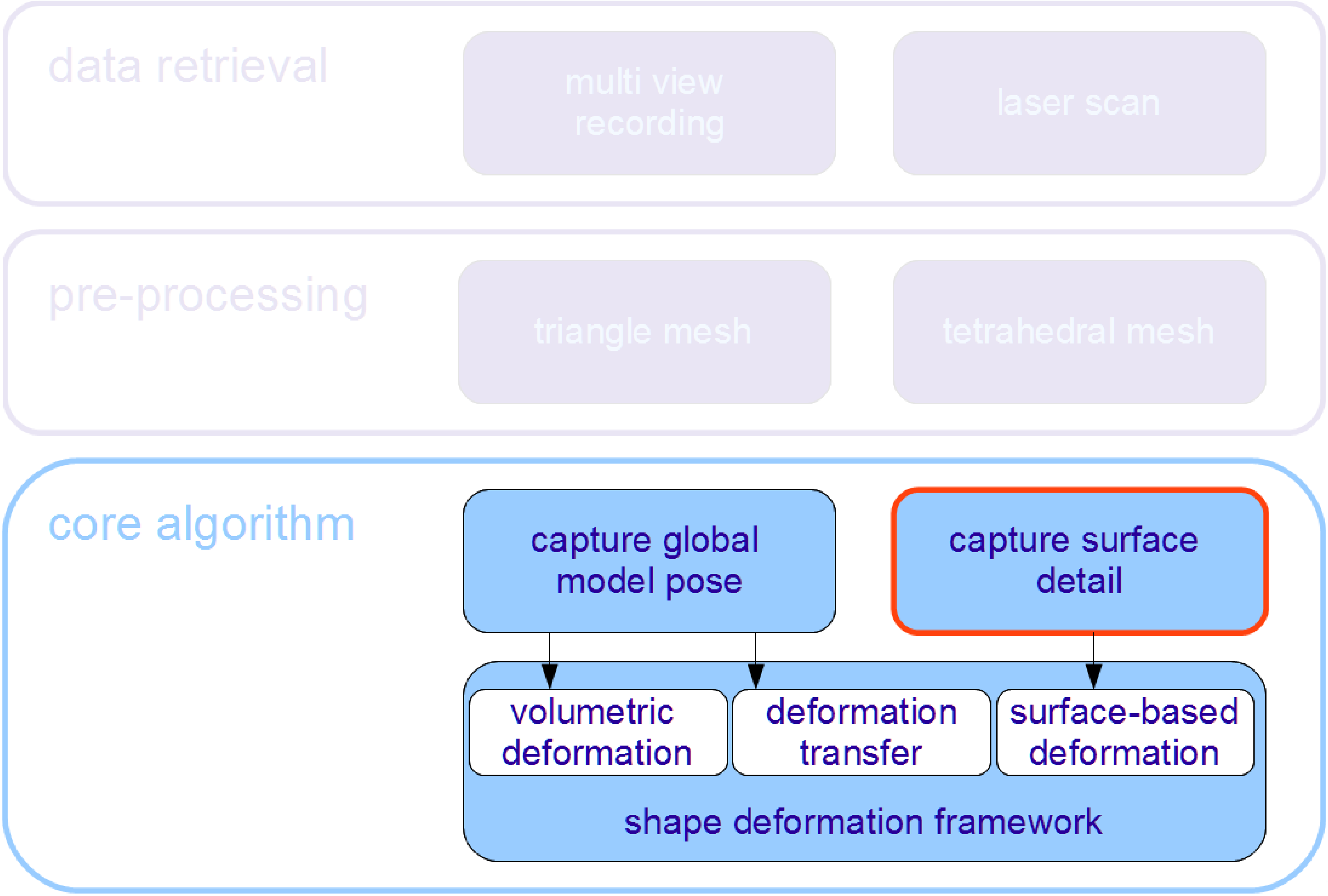


Deformation Transfer

- match vertices of input triangle mesh to input tetrahedral mesh
- vertices of triangle mesh are linear combination of tetrahedral mesh
- combine linear coefficients in matrix B
- $T'_{tri} = T'_{tet} * B$



Processing Steps



Capturing Surface Detail

- map tetrahedral mesh to triangle mesh
- capture shape detail at each time step

1. Adaptation along Silhouette Contours
2. Model-Guided Multi-View Stereo

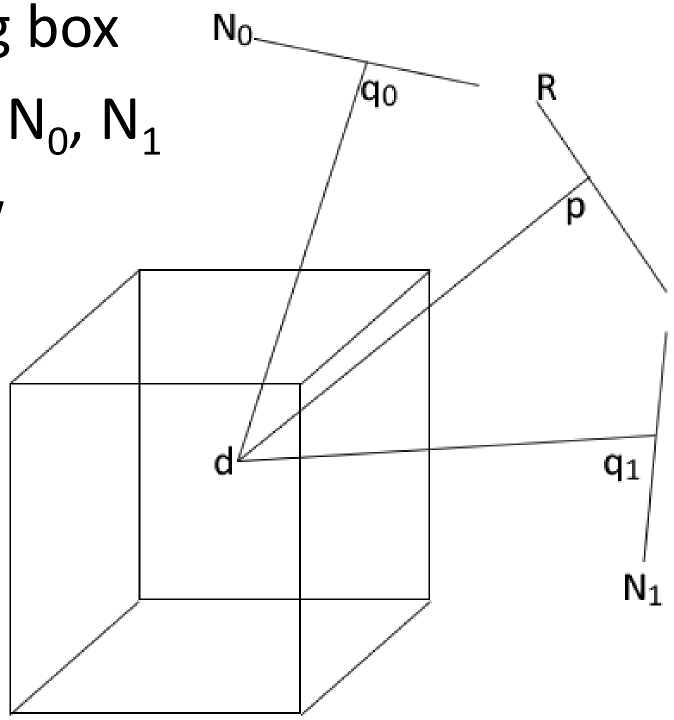
Adaptation along Silhouette Contours

- calculate rim vertices
- find closest 2D point on silhouette boundary for each rim vertex
- compute image gradients at input silhouette points and reprojected model contour image
- if distance between gradients smaller threshold, add as constraint

Model-Guided Multi-View Stereo

- multi-view stereo method by [Goesele et al. 2006]

- project pixel p of view R into bounding box
- reproject point to neighbouring views N_0, N_1
- compute cross-correlation for window around p and corresponding window around q_0, q_1
- high score for valid depth d

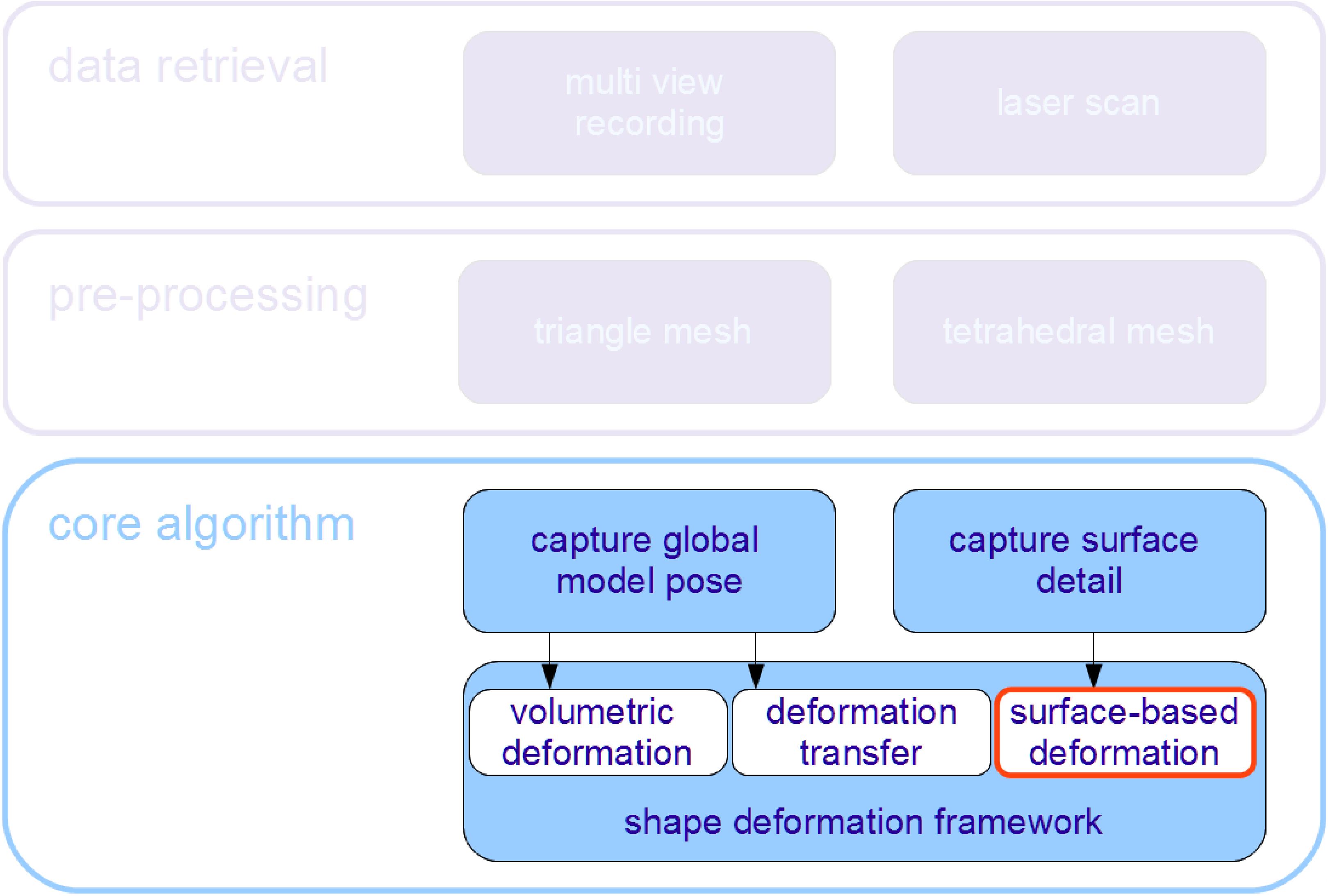


Visual Computing
Department of Computer Science

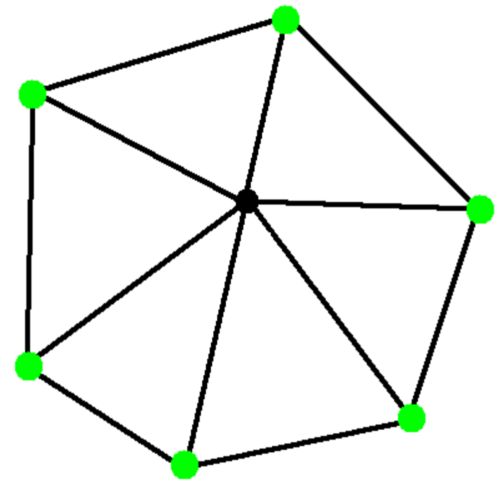
Model-Guided Multi-View Stereo

- constrain correlation search to 3D points at most ±2cm away from triangle mesh
- merge depth maps into single point cloud
- project points from triangle mesh onto point cloud
- projected points provide additional position constraints

Processing Steps

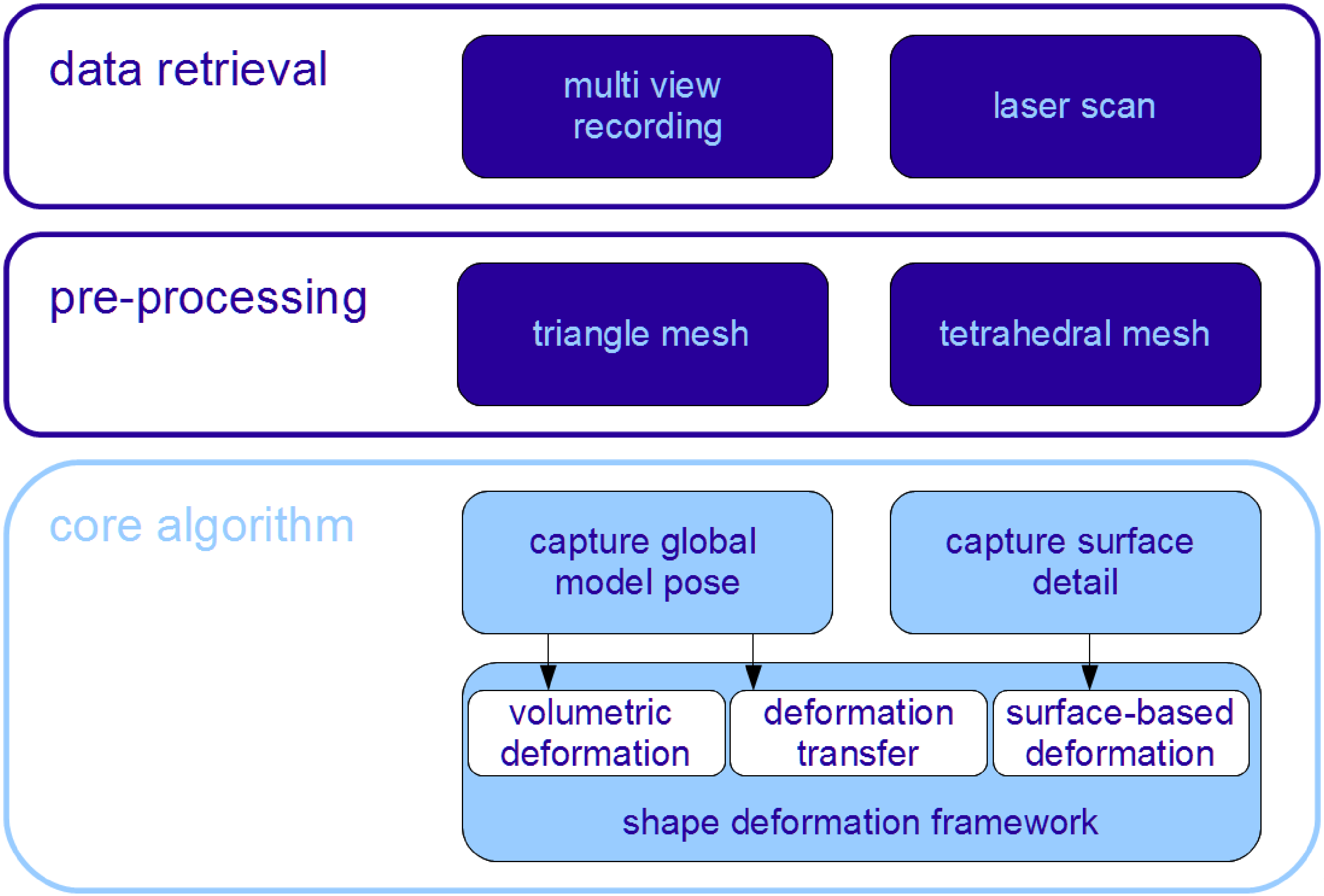


Surface-based Deformation

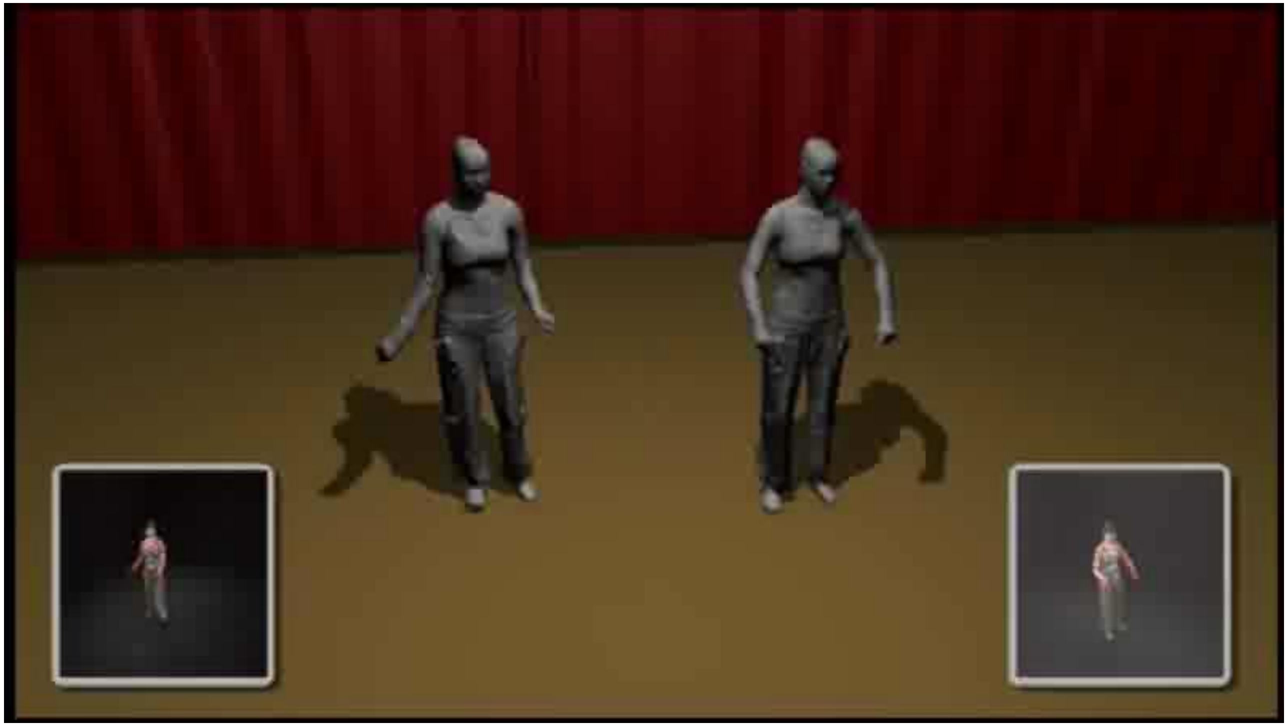


- calculate differential coordinates δ
- L is cotangent Laplacian matrix
- C is diagonal matrix, only entries for constrained vertices q
- minimize: $\{\|Lv - \delta\|^2 + \|Cv - q\|^2\}$
- new triangle mesh in v

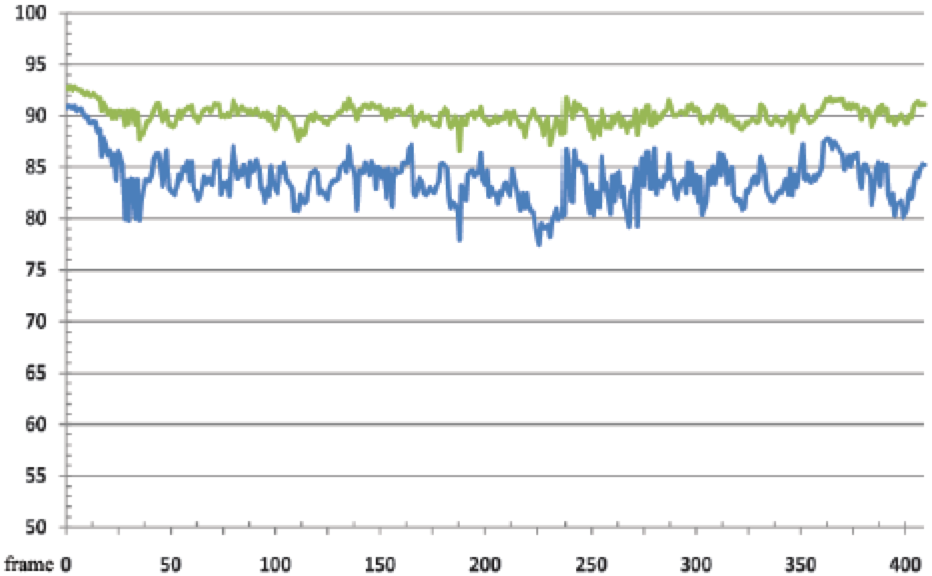
Processing Steps



Summary



Summary



silhouette overlap [%]
blue: global pose estimation
green: surface detail reconstruction

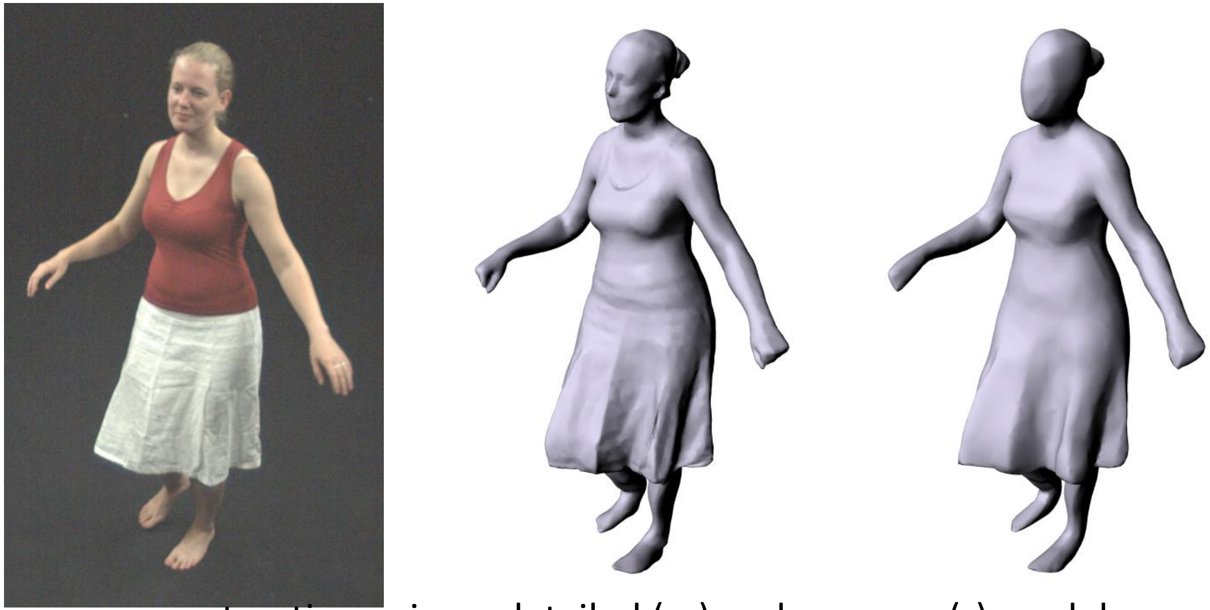
overlay between input image and model

Summary

- non-intrusive method
- all test input scenes were reconstructed fully automatically
- minimal user input:
  - mark head and foot regions to exclude from surface detail capture
  - mark 25 deformation handles for key handle optimization
- only few isolated pose errors due to motion blur or strong shadows

Summary

- limitations:
  - topological structure of input silhouette must be close to reprojected model silhouette (no outdoor scenes)
  - topology of input scanned model preserved (movement of hair cannot be captured)
  - deformation technique mimics elastic deformation → rubbery look
  - resolution limit to deformation capture (high-frequency details “baked in“)



reconstruction using a detailed (m) and a coarse (r) model

Summary

- large application range of algorithm
  - tight and loose apparel
  - fabrics with prominent texture as well as plain colors
  - simple walks and different dance styles
  - complicated self-occlusions



References

- De Aguiar et al, 2008 „Performance Capture from Sparse Multi-view Video“
- Stoll et al, 2007, „A Volumetric Approach to Interactive Shape Editing“
- Goesele et al, 2006, „Multi-view stereo revisited“
- Taku Komura, http://homepages.inf.ed.ac.uk/tkomura/cav/index.html, access: 27.05.2011