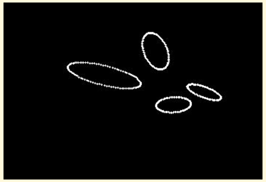
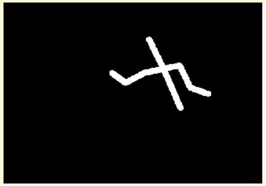
Home Page
Title Page
Page 1 of 35
Go Back
Full Screen
Close
Quit

Markerless Motion Capture using Monocular Videos: A Case Study for Bharatanatyam



by
Vishal Mamania
http://www.cse.iitb.ac.in/∼vishalm
under the guidance of
Prof. Sharat Chandran
http://www.cse.iitb.ac.in/∼sharat

July, 2004

Copyright ©2004 Vishal Mamania

Overview

- Introduction
- Our Approach
- Design of the System
- Results of Implementation
- Conclusions

Copyright ©2004 Vishal Mamania

Video Processing using Computer Vision

- Video contains a lot of data – in space as well as time

Video Processing using Computer Vision

- Video contains a lot of data – in space as well as time
- One frame $= 320 \times 240 = 76,800$ pixels

Video Processing using Computer Vision

- Video contains a lot of data – in space as well as time
- One frame $= 320 \times 240 = 76,800$ pixels
- Considering 30 frames per second, 10 second video contains $76,800 \times 30 \times 10 = 23,040,000$ pixels

Video Processing using Computer Vision

- Video contains a lot of data – in space as well as time
- One frame $= 320 \times 240 = 76,800$ pixels
- Considering 30 frames per second, 10 second video contains $76,800 \times 30 \times 10 = 23,040,000$ pixels
- Consider 16 million colors (24-bit) for each pixel....
- Information Overflow makes the problem difficult
- Need intelligent methods to decide
  - What is important and useful?
  - What is junk?

 Copyright ©2004 Vishal Mamania

Home Page

Title Page

Page 4 of 35

Go Back

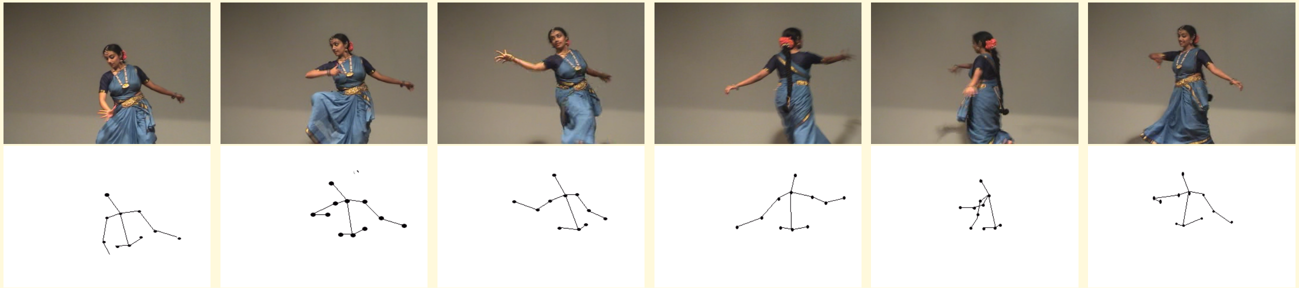
Full Screen

Close

Quit

1. Human Motion Capture

- Process of recording human body movements to get a compact representation of human skeleton and its motion
- Recovery of global position and orientation of a subject and various body parts and joints in 3D space



Copyright ©2004 Vishal Mamania

Mechanical Method of MoCap

- Exoskeleton attached to movable parts and joints of human body

- Measure the movements of corresponding parts and produce appropriate signals

- *Drawback* –
  Heavily obstructs body movements

Copyright ©2004 Vishal Mamania

Marker-based Multi-camera system of MoCap

- Done in special MoCap labs
- Dark colored clothes required
- White reflective markers attached to clothes of performer at joint locations
- These markers are viewed through multiple cameras
- *Drawback* – Highly tailored environment

Copyright ©2004 Vishal Mamania

Stages in Marker-based method

- Tracking of body parts
  - Image processing to locate the markers' positions in various viewpoints
  - Establish correspondences

Stages in Marker-based method

- Tracking of body parts
  - Image processing to locate the markers' positions in various viewpoints
  - Establish correspondences
- Reconstruction of skeleton
  - Construction of 3D structure from 2D projections
  - Establish motion parameters

Copyright ©2004 Vishal Mamania

Limitations of Marker-based Method

- **Expensive** – Requires specialized studios, multiple cameras, etc.

Limitations of Marker-based Method

- **Expensive** – Requires specialized studios, multiple cameras, etc.
- **Intrusive** – Require that clothes be of different (dark) color; markers placed

Limitations of Marker-based Method

- **Expensive** – Requires specialized studios, multiple cameras, etc.
- **Intrusive** – Require that clothes be of different (dark) color; markers placed
- **No Live Shows** – Not possible for live performances

Limitations of Marker-based Method

- **Expensive** – Requires specialized studios, multiple cameras, etc.
- **Intrusive** – Require that clothes be of different (dark) color; markers placed
- **No Live Shows** – Not possible for live performances
- **No Archives** – Not possible to use archive videos

Limitations of Marker-based Method

- **Expensive** – Requires specialized studios, multiple cameras, etc.
- **Intrusive** – Require that clothes be of different (dark) color; markers placed
- **No Live Shows** – Not possible for live performances
- **No Archives** – Not possible to use archive videos
- **In-house** – Can't capture outside studio

Limitations of Marker-based Method

- **Expensive** – Requires specialized studios, multiple cameras, etc.
- **Intrusive** – Require that clothes be of different (dark) color; markers placed
- **No Live Shows** – Not possible for live performances
- **No Archives** – Not possible to use archive videos
- **In-house** – Can't capture outside studio

- Solution – *Markerless Motion Capture using a single camera*

Copyright ©2004 Vishal Mamania

2. Markerless Monocular MoCap

- Develop algorithms for obtaining motion capture data in a generalized environment

2. Markerless Monocular MoCap

- Develop algorithms for obtaining motion capture data in a generalized environment
  - No special studio required
  - Using a single camera
  - No artificial aids like markers, calipers
  - No restriction on clothes

2. Markerless Monocular MoCap

- Develop algorithms for obtaining motion capture data in a generalized environment
  - No special studio required
  - Using a single camera
  - No artificial aids like markers, calipers
  - No restriction on clothes
- Using Bharatanatyam as an example

Copyright ©2004 Vishal Mamania

New Challenges

- Image processing (Tracking) becomes difficult
- A lot of clutter needs to be removed

Copyright ©2004 Vishal Mamania

New Challenges (Contd.)

- Reconstruction too becomes difficult
- Recover the depth information lost during recording
- Creating 3D from 2D !!
- Depth values have to be valid and consistent with each other

Copyright ©2004 Vishal Mamania

Our approach

- Tracking
  - Use Domain-specific knowledge
  - Information about traditional dress of Bharatanatyam

Our approach

- Tracking
  - Use Domain-specific knowledge
  - Information about traditional dress of Bharatanatyam
- Reconstruction
  - Given a 2D projection, no. of possible 3D poses is finite.
  - For n links (limbs), max 2^n poses possible.
  - Many poses impossible to achieve physically. Discard them.
  - Build a weighted graph of valid poses and find minimum weight path across the sequence. This gives the smoothest motion sequence.

Copyright ©2004 Vishal Mamania

Human Motion Capture
Markerless Monocular . . .
Phase I: Tracking
Phase II: Reconstruction
Conclusions

Home Page

Title Page

Page 13 of 35

Go Back

Full Screen

Close

Quit

What others have done

- Silhouette shape analysis
- Multiple camera voxel data
- Using motion library
- Factorization

3. Phase I: Tracking

- What we have is a grid of pixels
- What we want is the locations of joints
- Points to be considered
  - Projection Model
  - Human Model
  - Key Feature Tracking
  - Bodyparts Labeling
  - Locating Endpoints (Joints)

Copyright ©2004 Vishal Mamania

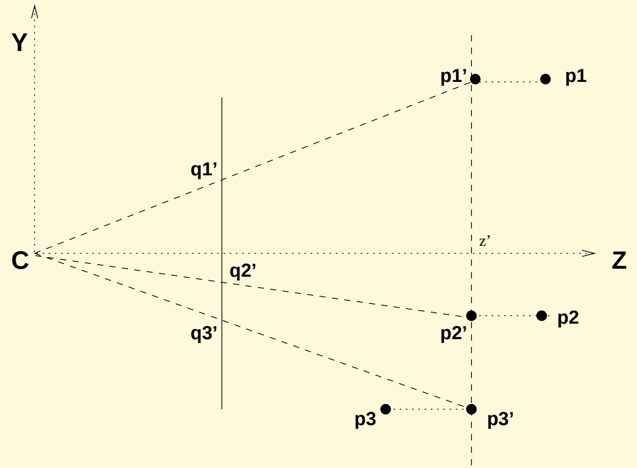
Assumptions

- Only a single person (dancer) in scene
- Dancer always in the view of camera
- Background is static
- No camera motion
- Lighting changes are limited
- Distance between dancer and camera is large

Copyright ©2004 Vishal Mamania

Scaled Orthographic Projection

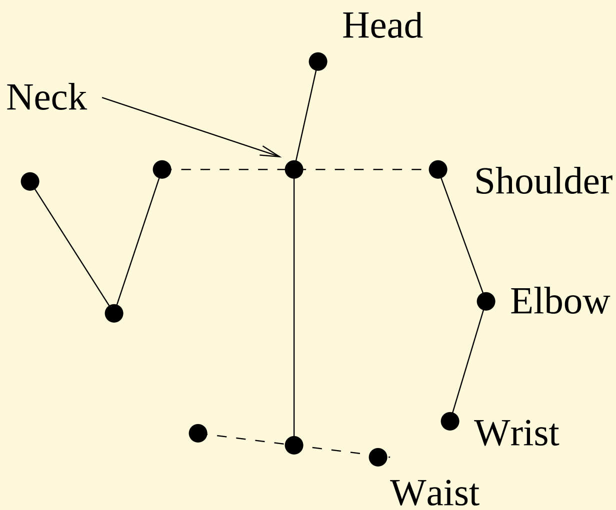
- A simple approximation to perspective projection
- Can be used
  - When the range of depth values of a scene is small compared to distance from camera
  - Distance between object and camera is large compared to size of object



 Copyright ©2004 Vishal Mamania

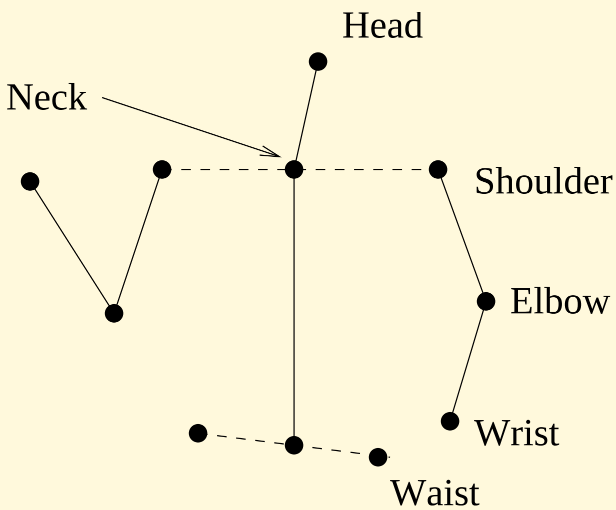
Human Model and Key features

- We use stick-figure representation
- Represent joints as points, bones as lines



Human Model and Key features

- We use stick-figure representation
- Represent joints as points, bones as lines



- Key features are the points to be tracked across the sequence
- All joints in the stick-figure
- Head, Shoulder, Elbow, Wrist, Waist

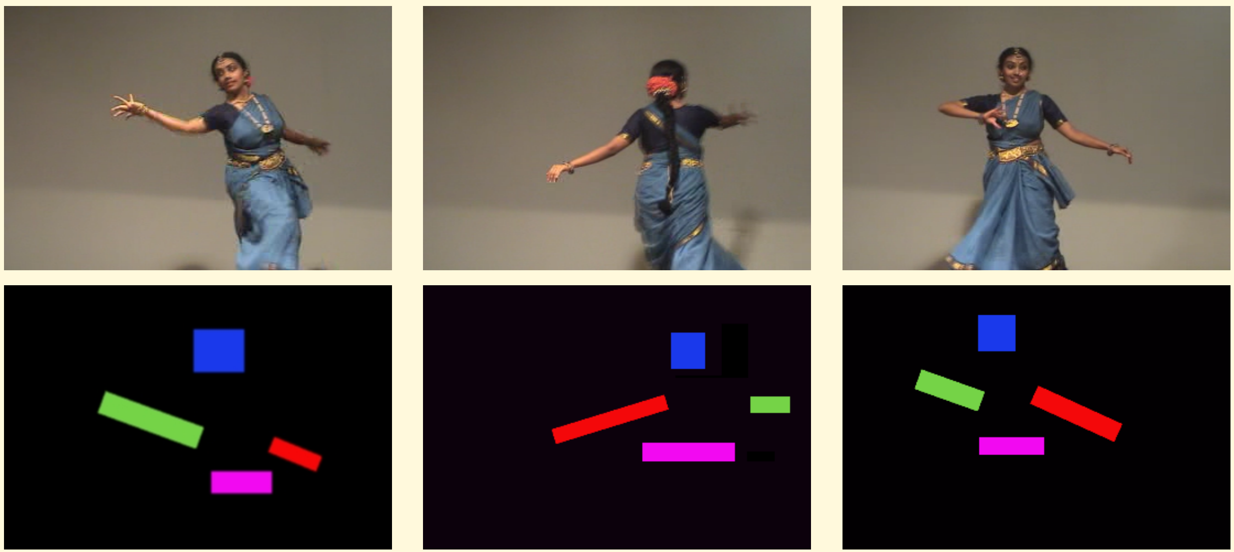
Copyright ©2004 Vishal Mamania

Feature Tracking

We need to track the features across the sequence and mark them accordingly

Feature Tracking

We need to track the features across the sequence and mark them accordingly



Copyright ©2004 Vishal Mamania

Skin color model

- We use *skin color model* to detect these features
- According to skin color model, normalized color components of skins of people of different races, genders, complexion are similar
- RGB components are normalized as follows

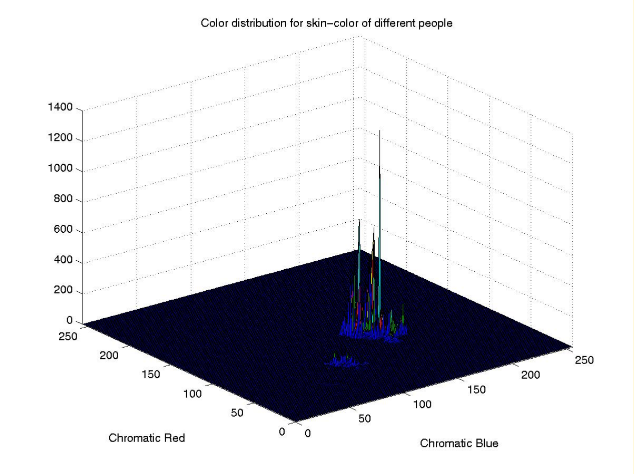
$$r = \frac{R}{(R+G+B)};\; b = \frac{B}{(R+G+B)}$$

Skin color model

- We use *skin color model* to detect these features
- According to skin color model, normalized color components of skins of people of different races, genders, complexion are similar
- RGB components are normalized as follows

$$r = \frac{R}{(R+G+B)};\ b = \frac{B}{(R+G+B)}$$

- Plot color histogram of r and b from distribution of skin color of different people.
- This histogram is clustered



Skin Color Model (Contd.)

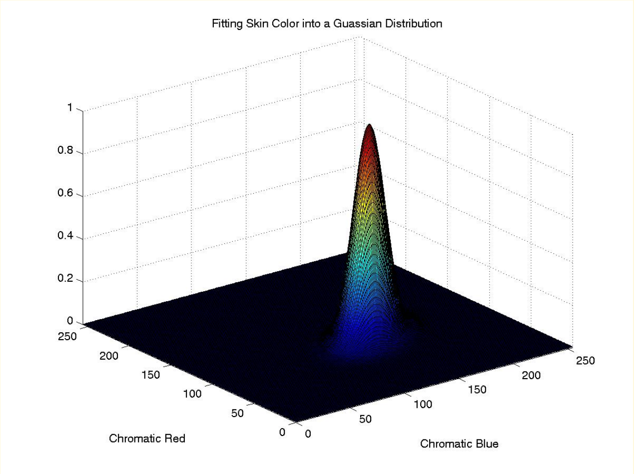
We approximate the color distribution to a Gaussian model $G(m, C)$ with

mean, $m = E\{x\}$, where $x = \begin{bmatrix} r \\ b \end{bmatrix}$

covariance, $C = E\{(x - m)(x - m)^T\}$

Skin Color Model (Contd.)

We approximate the color distribution to a Gaussian model $G(m,C)$ with



mean, $m = E\{x\}$, where $x = \begin{bmatrix} r \\ b \end{bmatrix}$
covariance, $C = E\{(x-m)(x-m)^T\}$

- Likelihood of each pixel belonging to skin is given as

$$\text{likelihood} = P(r,b) = exp[-0.5(x-m)^T C^{-1}(x-m)]$$

- This number is thresholded to get the skin regions

 Copyright ©2004 Vishal Mamania

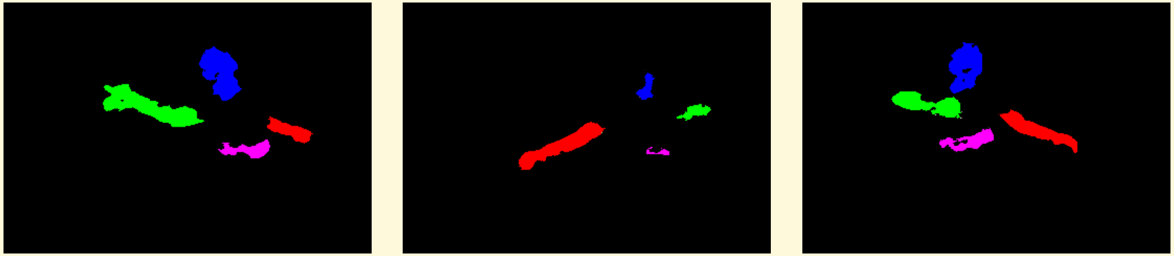
Results of skin detection

Note that the golden belt in the waist region is also detected as skin color region.

 Copyright ©2004 Vishal Mamania

Labeling body parts

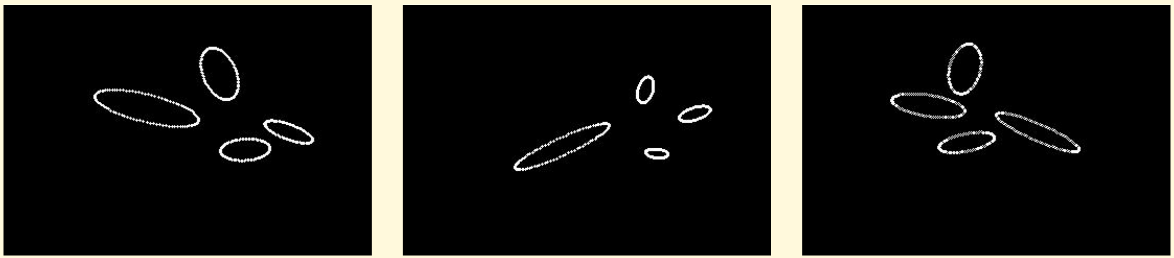
- After morphological operations, different blobs of skin colors are formed. Only large blobs are maintained
- Blob labels are initialized in the first frame
- These are tracked using the motion factor and proximity to previous frame's blobs
- Blobs may get merged or broken



 Copyright ©2004 Vishal Mamania

Fitting ellipses to blobs

- Ellipses fitted around the boundary of blobs, using algebraic fit algorithm
- Endpoints of major axis are endjoints of limbs



Copyright ©2004 Vishal Mamania

Human Motion Capture
Markerless Monocular . . .
Phase I: Tracking
Phase II: Reconstruction
Conclusions

Home Page

Title Page

Page 24 of 35

Go Back

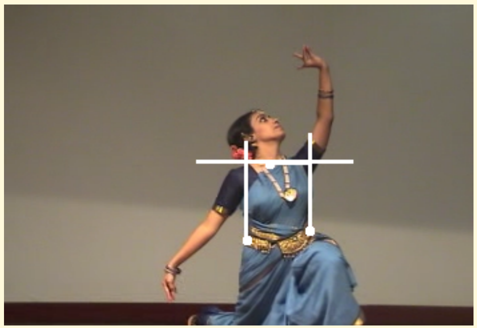
Full Screen

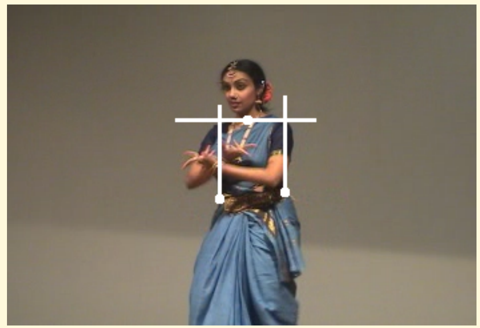
Close

Quit

Shoulders’ Position

- *Observation* – In most cases, except when the body is tilted, the position of the shoulders is exactly above the waist region end-points and in horizontal line with the lower end of the neck.



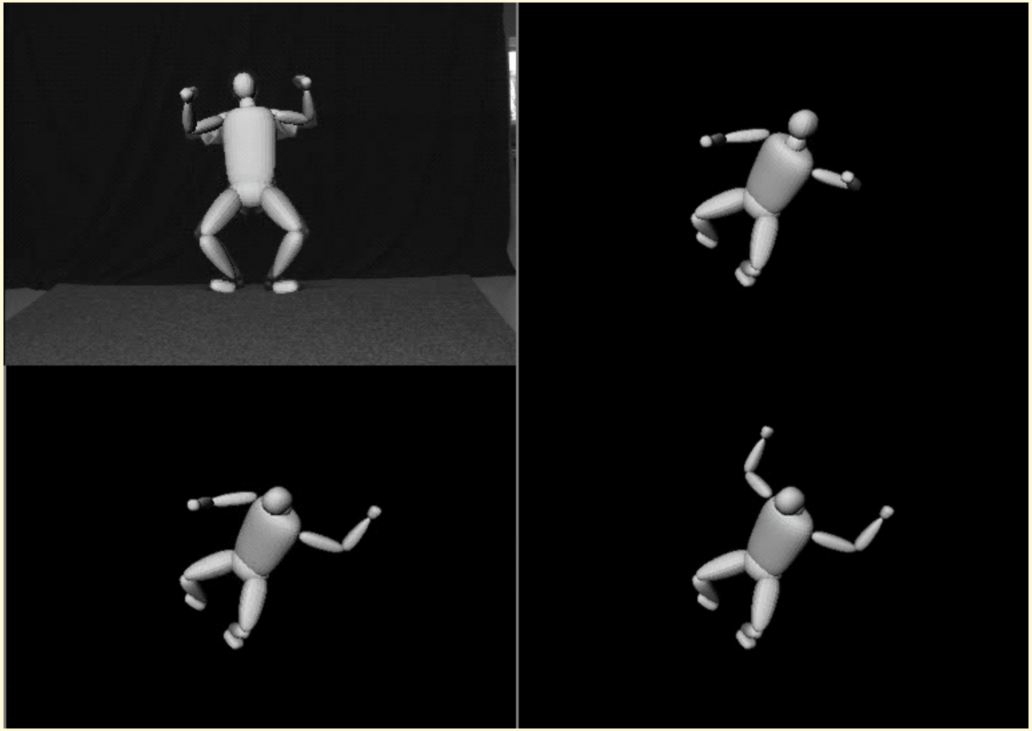


- Needs improvement.
- We currently use manual adjustment.

Copyright ©2004 Vishal Mamania

4. Phase II: Reconstruction

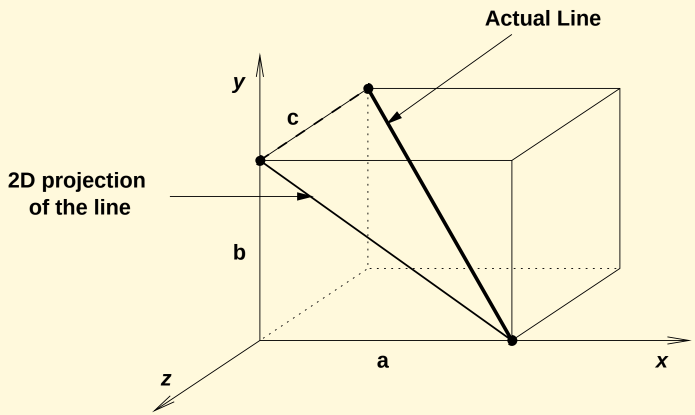
- What we have is 2D projections of joints in all frames
- What we want is 3D positions of joints in all frames



 Copyright ©2004 Vishal Mamania July 8, 2004

Calculating depth

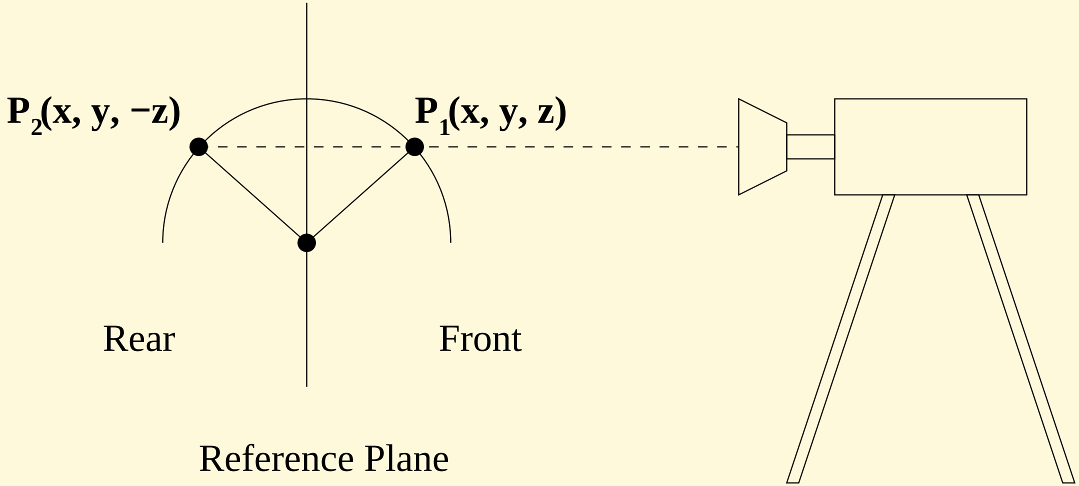
- *Observation* – Given a sufficiently long sequence, each link becomes parallel or nearly parallel at least once.
- Using anthropometric data to adjust the length values
- Given the 2D length & 3D length, depth of link can be calculated using basic trigonometry.



 Copyright ©2004 Vishal Mamania

Reflective Orthographic Ambiguity

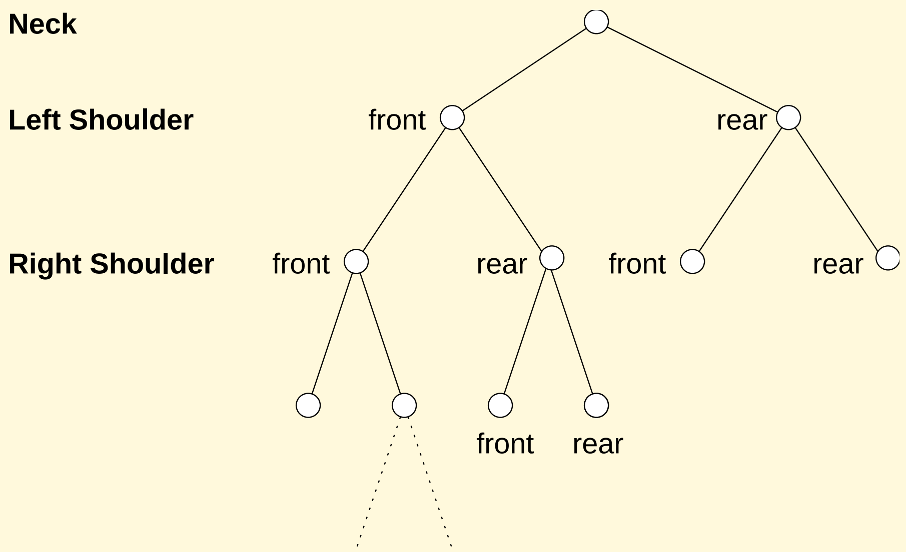
- For each link, there are two possibilities of z-values.
- One endpoint can be in front or in rear of the other endpoint.



Copyright ©2004 Vishal Mamania

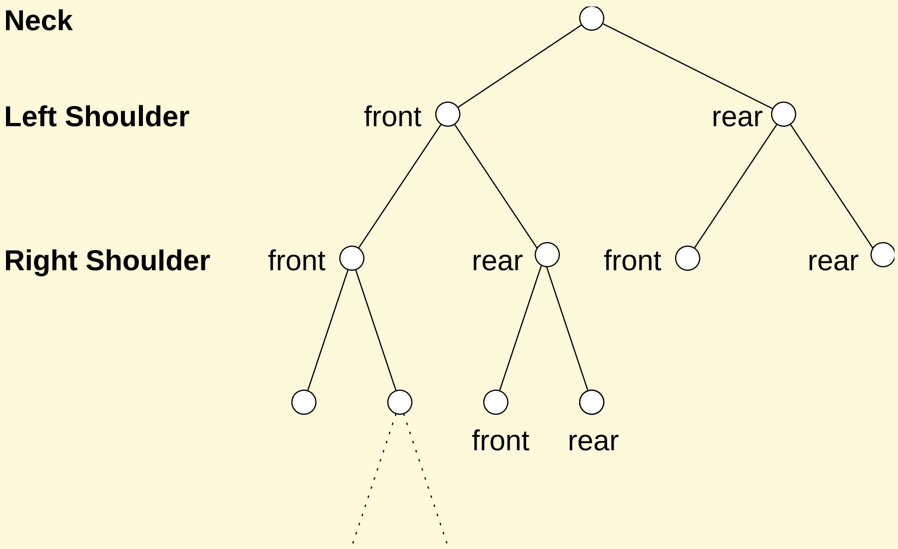
Pose Generation

- For each link, two possibilities $\Longrightarrow 2^n$ possibilities for n links.



Pose Generation

- For each link, two possibilities $\Longrightarrow 2^n$ possibilities for n links.



- Not all of 2^n poses are physically attainable.
- Need to put constraints to filter out impossible poses.

Copyright ©2004 Vishal Mamania

Body Constraints

- Joint Angle Limits
  - Each joint of body has a maximum and a minimum limit of angle of bend.

Body Constraints

- Joint Angle Limits
  – Each joint of body has a maximum and a minimum limit of angle of bend.
- Collision Constraints
  – One body part cannot penetrate through another part.
  – Find distance between different links, they should be less that sum of corresponding radii.

 Copyright ©2004 Vishal Mamania

Graph Formulation

- What we have is a set of valid poses for each frame.

Graph Formulation

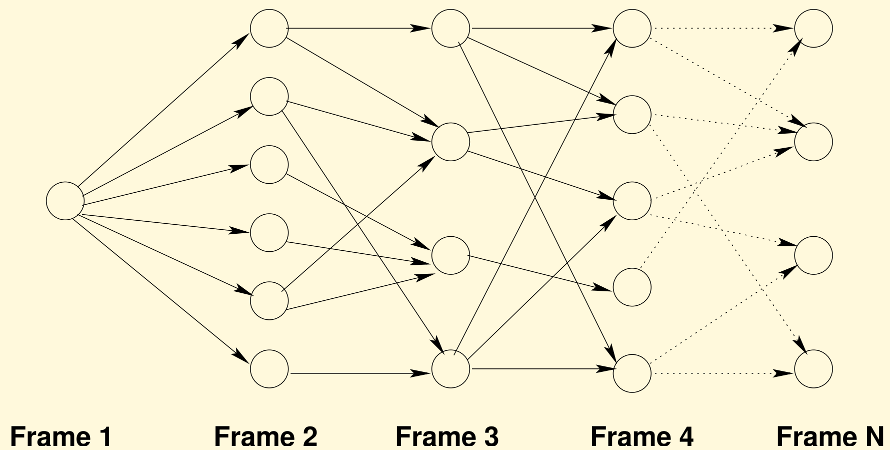
- What we have is a set of valid poses for each frame.
- What we want is a valid pose sequence across frames.

Graph Formulation

- What we have is a set of valid poses for each frame.
- What we want is a valid pose sequence across frames.
- We create a layered graph to model this situation.
- One layer for each frame.
- Each valid pose for a frame is represented as a node in the corresponding layer.
- Edges are put between nodes in adjacent layers of transition between those poses is possible.

Copyright ©2004 Vishal Mamania

Graph Formulation (Contd.)



- Assign weights to edges
- Find minimum weight path from first to last layer, which gives optimal path

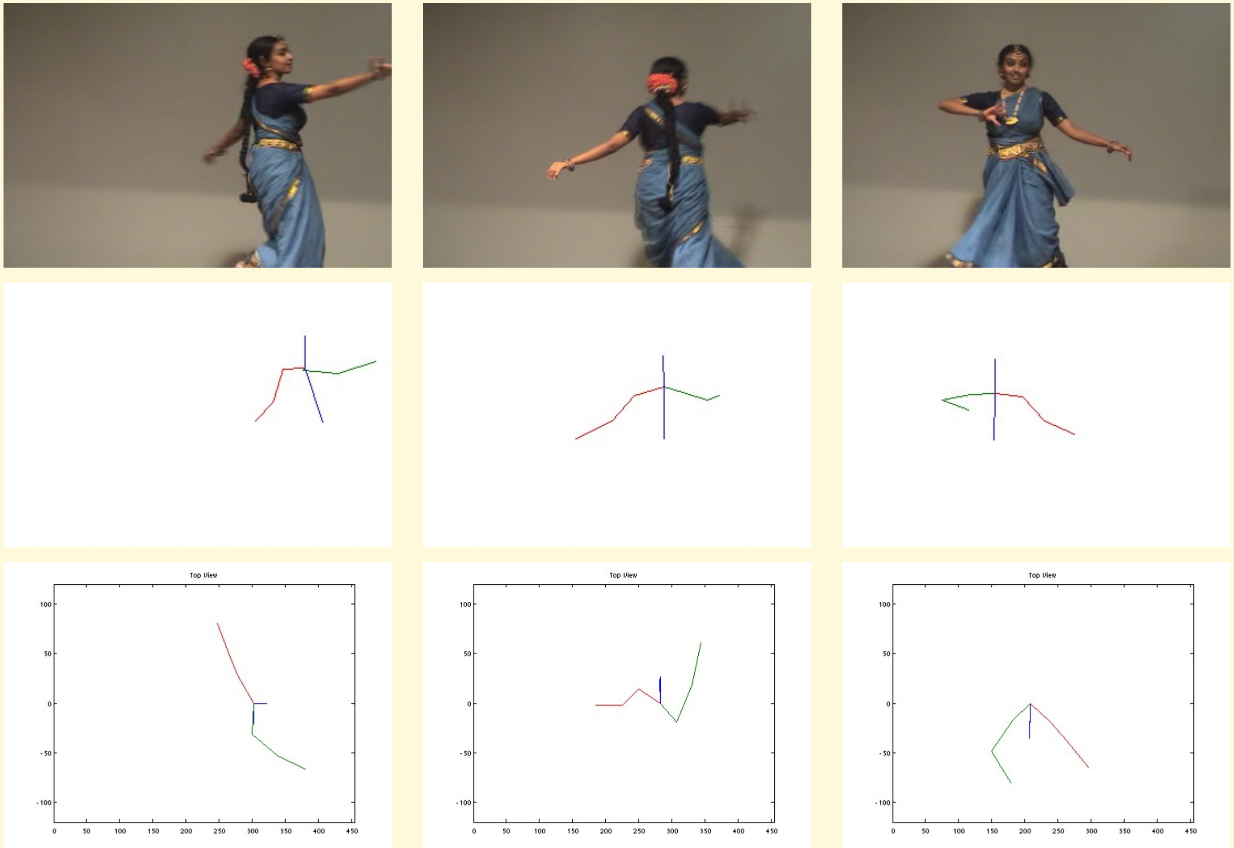
 Copyright ©2004 Vishal Mamania

Calculating Weights

- For jerk free motion, the change in angles, velocities should be as smooth as possible.
- Weight represent the difficulty of transition from one pose to another.
- Various possibilities tested
  - Change in depths of joints
  - Change in angles at joints
  - Change in velocities
  - Estimation based on velocity
- Last method produces the best results.

Copyright ©2004 Vishal Mamania

Final Results



 Copyright ©2004 Vishal Mamania

5. Conclusions

- Captured upper body motion for Bharatanatyam sequence
- Tracking done using the domain-specific knowledge
- Tracking is not completely automatic. We need some manual intervention for blobs labeling and final positions
- Reconstruction done using a graph-based approach
- Reconstruction produces accurate results in majority of frames

Copyright ©2004 Vishal Mamania

THANK YOU !!

Copyright ©2004 Vishal Mamania