Shape Outlier Detection Using Pose Preserving Dynamic Shape Models

Chan-Su Lee and Ahmed Elgammal

Rutgers, The State University of New Jersey
Department of Computer Science

Outline

- Introduction
  - Shape Outlier Detection for Visual Surveillance
  - Previous works
- Dynamic Shape Models
  - Kinematics Manifold Embedding
  - Decomposable Generative Models
- Outlier Detection
  - Shape Normalization
  - Hole filling
  - Outlier Detection
  - Iterative Estimation of Shape Style and Outlier with Hole Filling
- Experimental Results
  - Outlier Detection in Fixed View
  - Outlier Detection in Continuous View Variations
- Conclusions & Future Works

Visual Surveillance System

- Smart video surveillance system
  - Requires fast, reliable and robust algorithms for moving object detection, tracking, and activity analysis

Block based approach

- Why shape outlier detection in visual surveillance ?
  - To monitor interactions between people and objects
  - To detect unusual event such as depositing an object, exchanging bags, or removing an object
  - Abnormal action detection

Decomposable Nonlinear Dynamic Shape Model

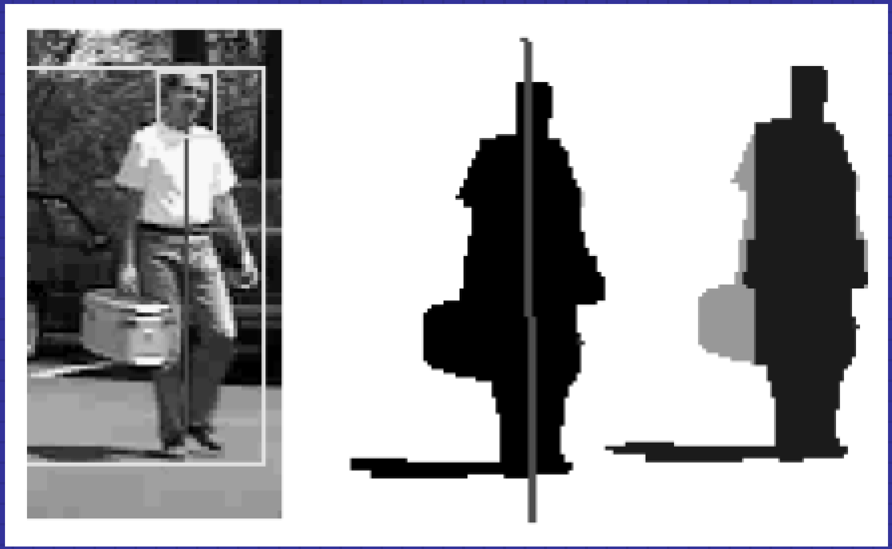
Previous Works

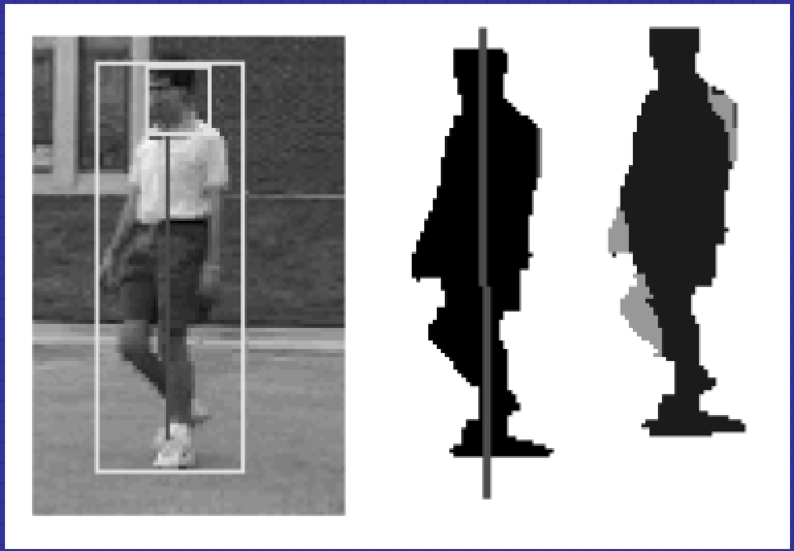
- Static shape based approaches

  [Haritaoglu et al, ICCV 1999]

  – Static shape analysis

  – Carrying object detection based on symmetric analysis & temporal analysis





Previous Works

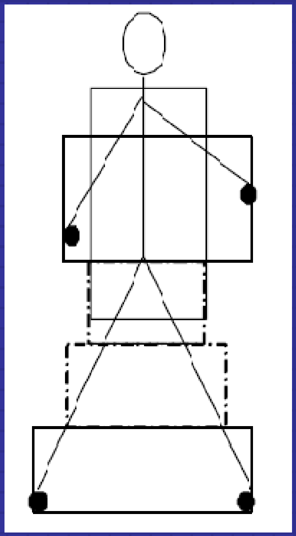
- Motion-based Recognition

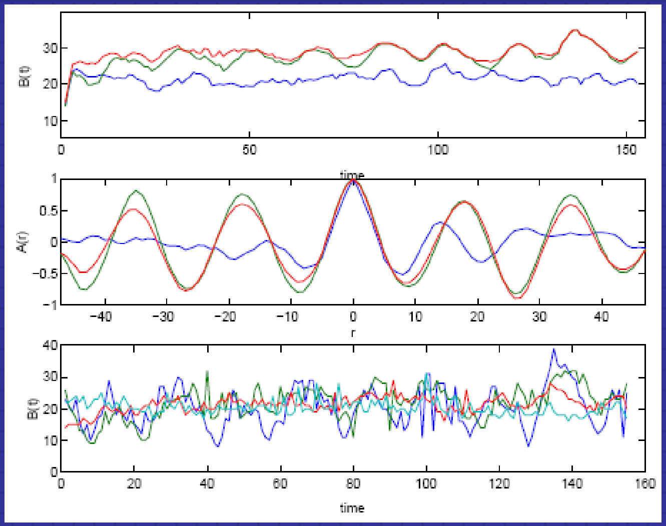
  [BenAdelkader et al, FGR 2002]

  – Subdivision of body silhouette

  – Periodicity of body part motion as constraints

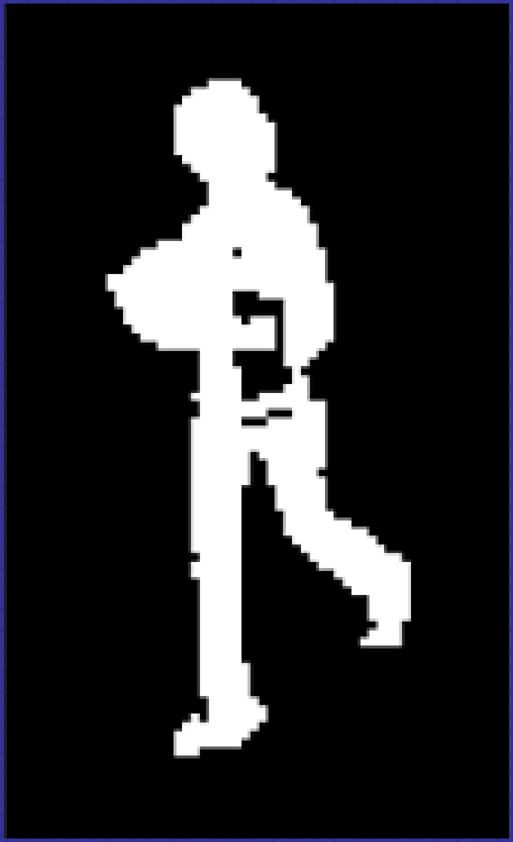
  – Pendulum-like motion of legs





Can we detect carrying object in a single image?

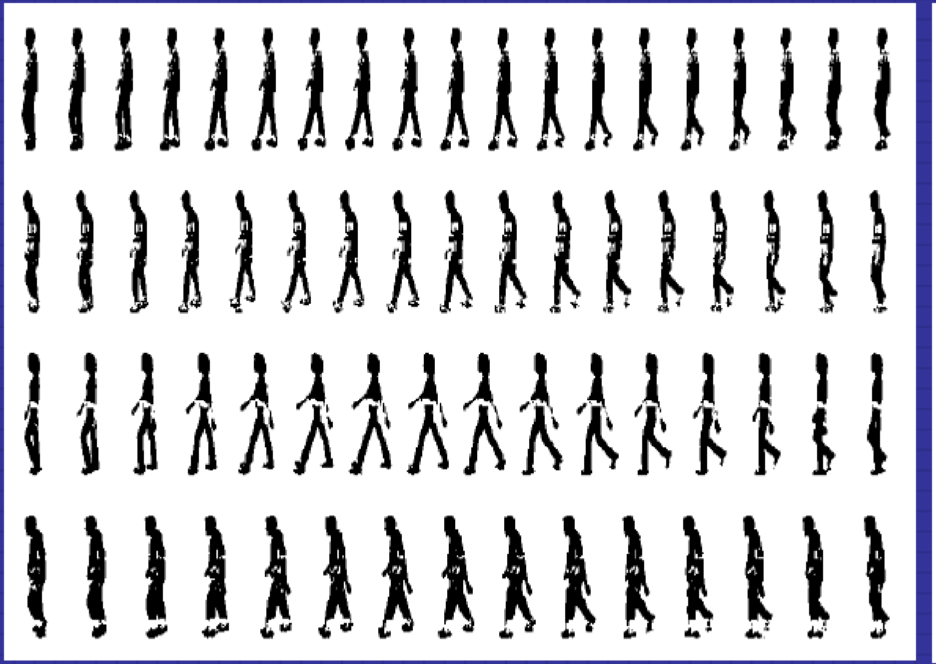
- People can detect carrying object even a single foreground shape image
  - People know possible shape of normal walking in different views in different people



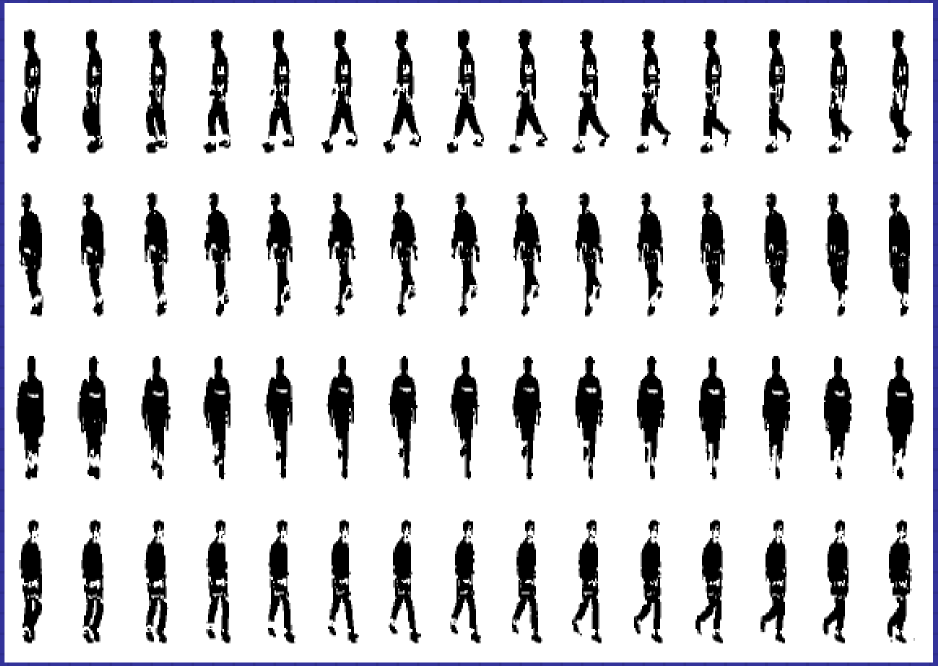
Dynamic Shape Models

Dynamic Shape Deformations

- Shape Deformations in Gait
  - Temporal variations (Body configuration)
  - Different in different people, in different view, etc.



Walking sequence in different people



Walking sequence in different view

Dynamic Shape Models

- Learning nonlinear generative models

$$y_t = \gamma_o(x_t; a_1, a_2, \cdots, a_n)$$

  - Representation of configuration space
    - Compact, and low dimensional
    - Dynamic characteristics + time invariant factors
  - Learn nonlinear mapping
    - Capture nonlinearity in body configuration and observed data
  - Factorize static parameters

Generative Model for Gait

A generative model for walking silhouettes
for different people from different views

- Time invariant
- Person invariant
- Characterizes the view

View Parameter

v

Representation of the motion phase

x_t

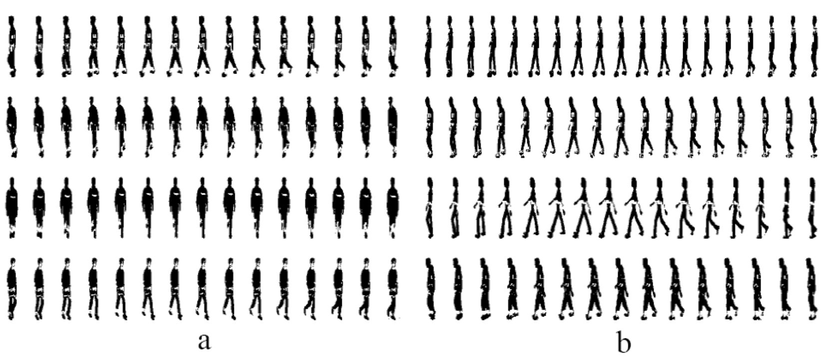
- Function of time
- Invariant to person
- Invariant to view
- Characterizes the motion phase: body configuration

s

Person Parameter

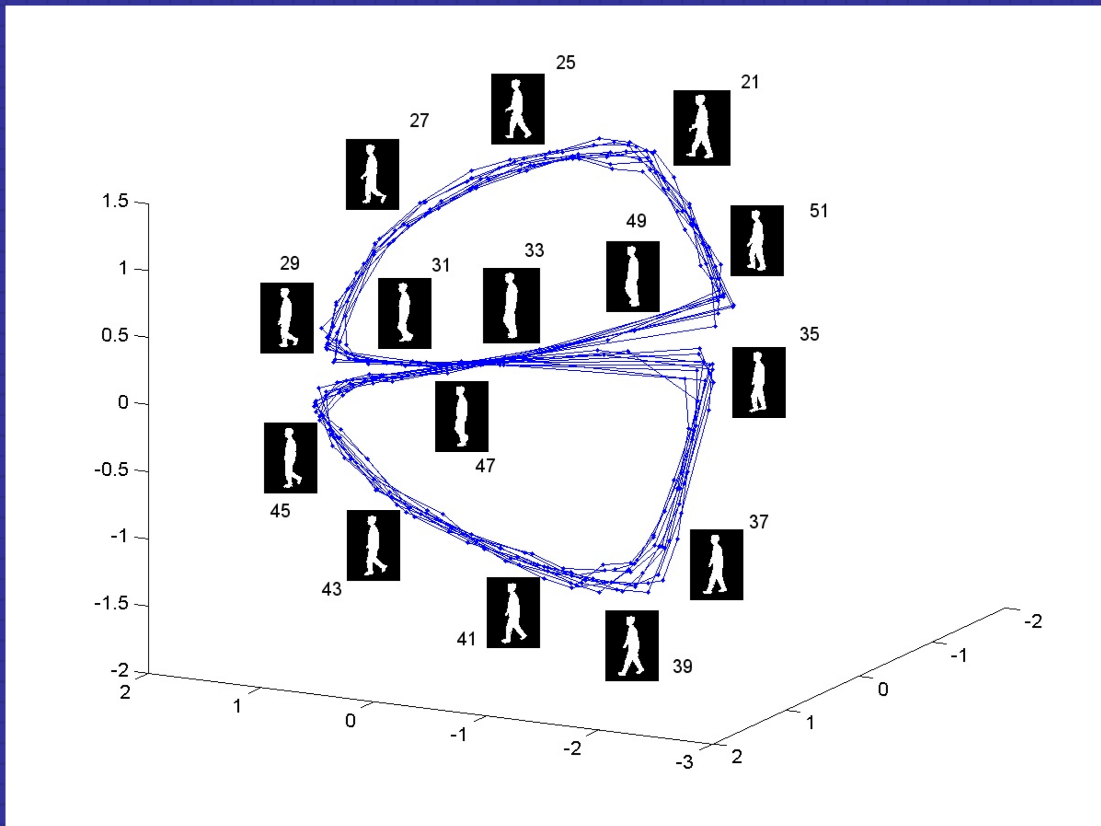
- Time invariant
- view invariant
- Characterizes the person shape

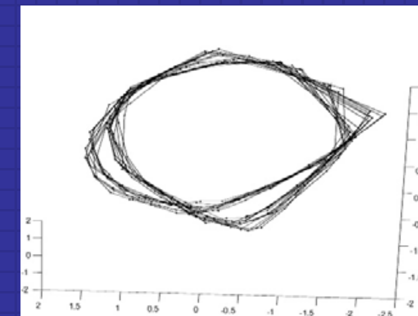
$$y_t = \gamma(x_t\,; v, s)$$

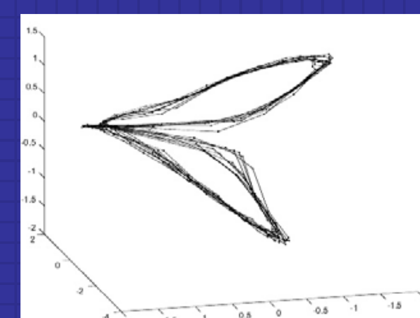


Embedding the Gait Manifold

[Elgammal, A, & Lee, C.-S. CVPR 2004a]



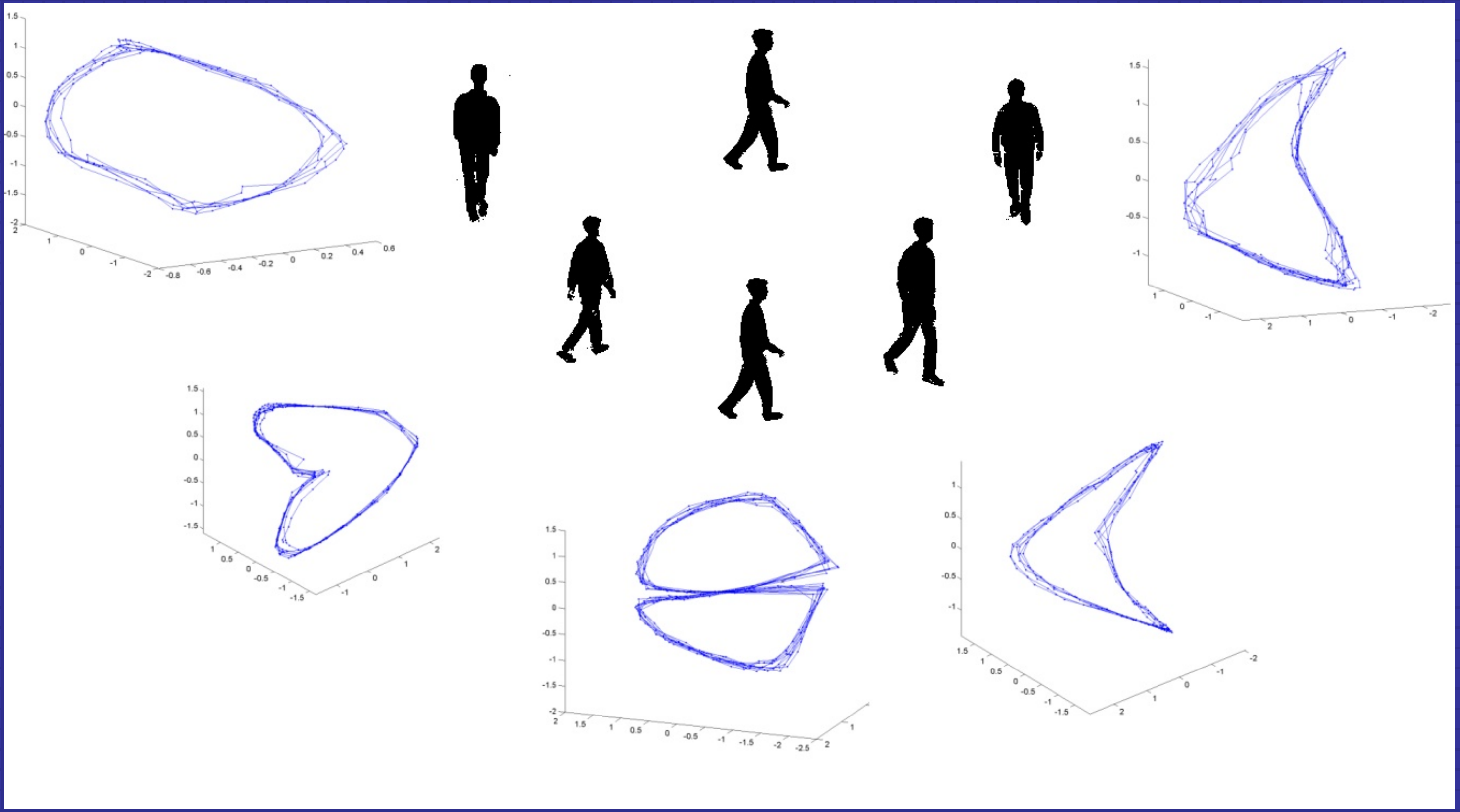




- ❑ Walking cycle:300frames
- ❑ No temporal information.
- ❑ Obtain embedding that shows body configuration

Embedding Gait Manifolds in Different View

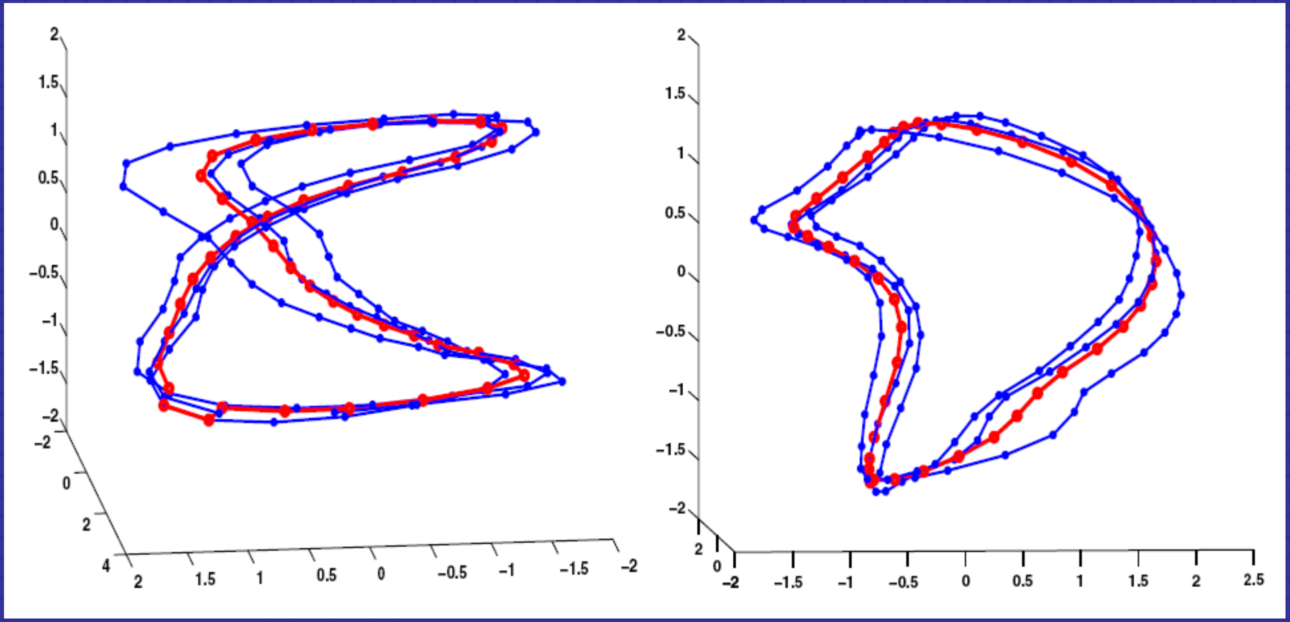
- Manifold twists differently depending on the view point, the body shape, clothing, etc.



[Elgammal, A, & Lee, C.-S. CVPR 2004b]

Kinematics Manifold Embedding

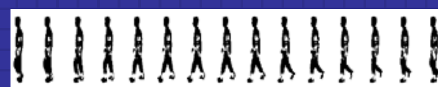
- Representation of body configuration in low dimensional space
  - Applying nonlinear dimensionality reduction for motion capture data
  - Invariant in different views



Multilinear Decomposition

$$y_t^{sv} = B^{sv}\psi(x_t^c)$$

$$y_t^{12} = B^{12}\psi(x_t^c)$$



$$y_t^{11} = B^{11}\psi(x_t^c)$$

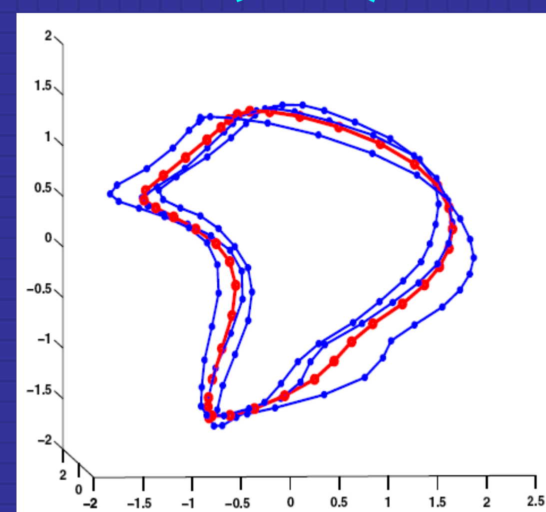
$$y_t^s = \mathcal{A} \times b^s \times c^v \times \psi(x_t^c)$$

mode-n tensor product

$c^v : K \times 1$ View vector

$b^s : J \times 1$ Style vector

$\mathcal{A} : d \times N_s \times N_v \times J$ Fourth-order tensor



Kinematics Manifold Representation

Estimation of Parameters

- To synthesize new gait shape, we need to know states of shape images (body configuration, view type, person style)

$$E(x_t, v, s) = \left\| y_t - C \times v \times s \times \Psi(x_t) \right\|$$

$$\sum_{k=1}^{K_v} \beta_k v^k \qquad \sum_{k=1}^{K_s} \alpha_k s^k$$

- Estimation of configuration for the known style and view factors is a nonlinear 1-dimensional search problem
- Obtain style(view type) conditional class probability by assuming a Gaussian density around the mean of the style classes(view classes)

$$p(y \mid x, s^k, v) \approx N(C \times v \times s^k \times \psi(x), \Sigma^{s^k})$$

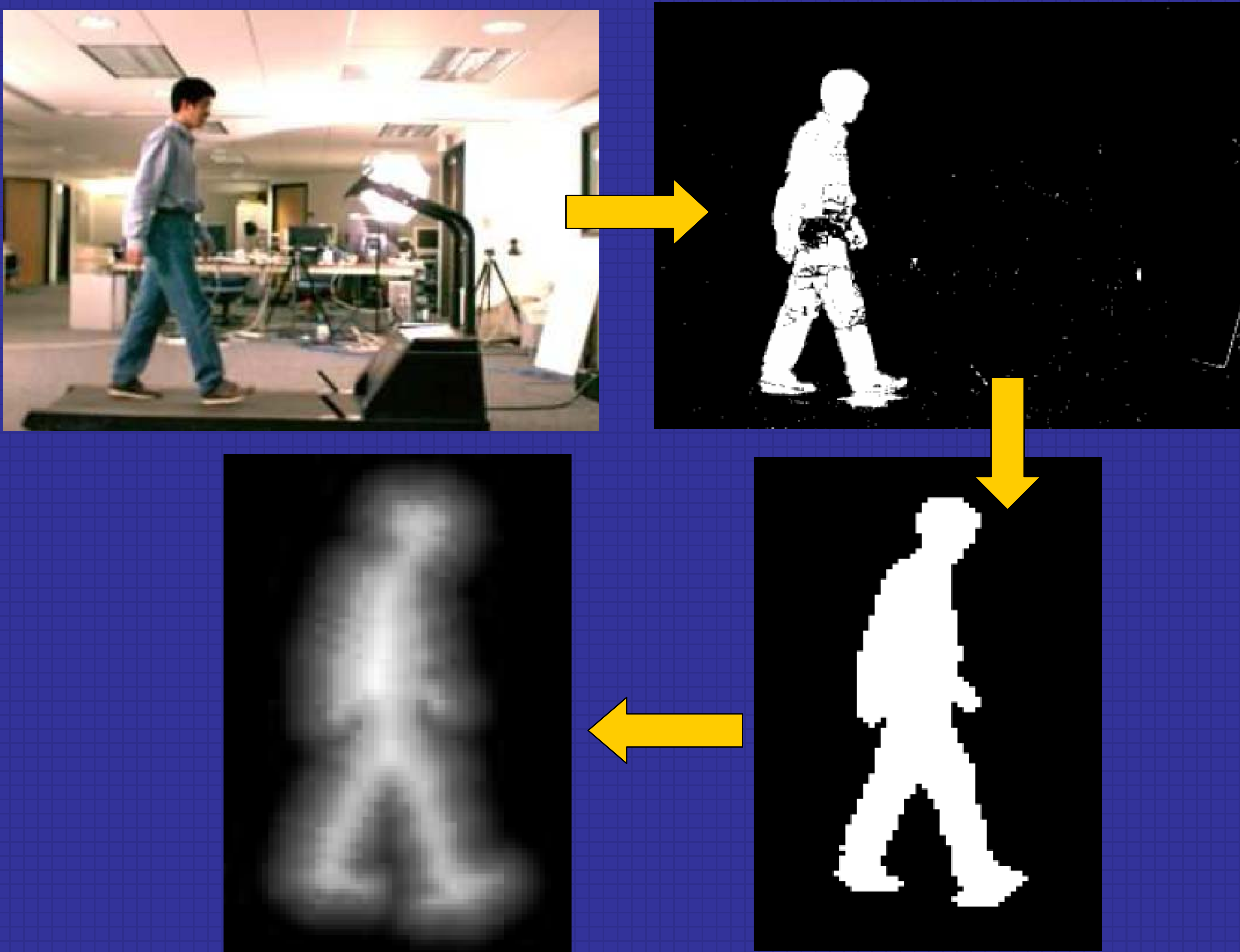
$$\Rightarrow \quad \alpha_k \propto p(s^k \mid y, x, v) \qquad \beta_k \propto p(v^k \mid y, x, s)$$

Iterative Estimation with Annealing

- This setting favors an iterative procedure
- However, wrong estimation of any of the factors would lead to wrong estimation of the others
- Avoid hard decision: at the beginning weights are forced to be close to uniform weights. The weights are gradually become discriminative thereafter
- Deterministic Annealing-like procedure: adaptive view and style class variances

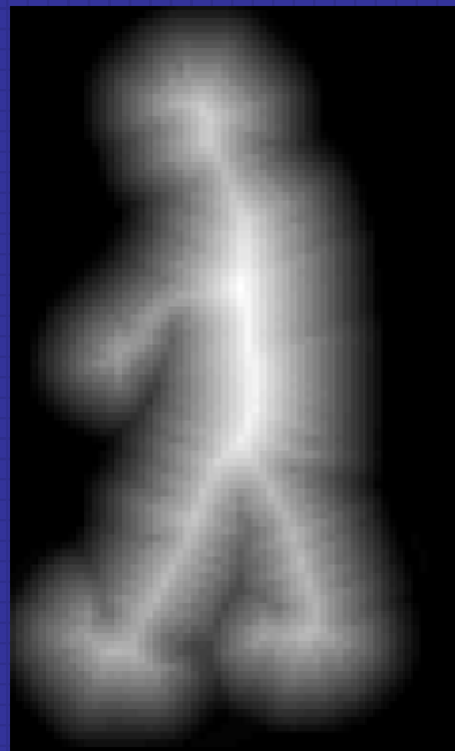
$$\boxed{\sum^{v} = T_v \sigma_v^2 I} \qquad \boxed{\sum^{s} = T_s \sigma_s^2 I}$$

Outlier Detection



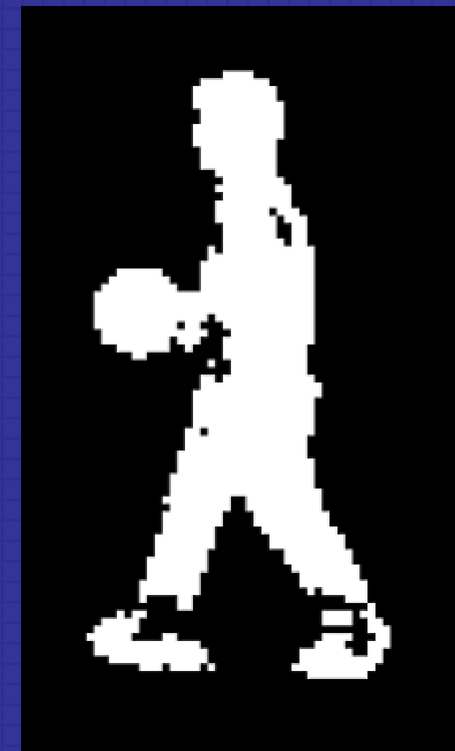
Hole Filling

Normalized Input



Estimated Shape

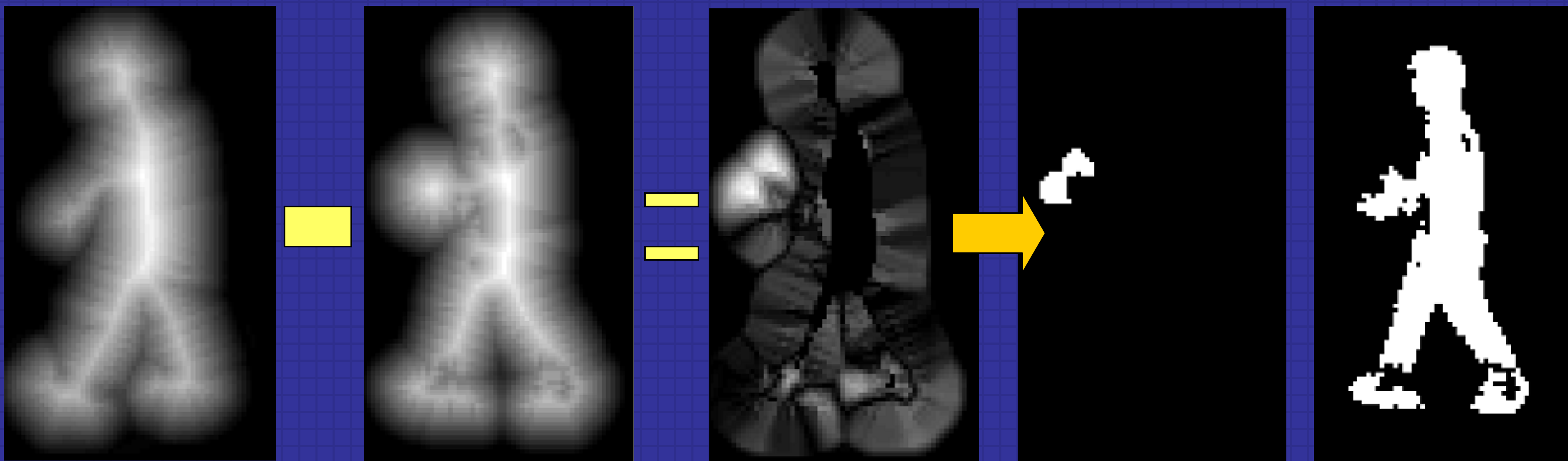
Generated Mask



Hole filled Silhouette

$$h(x)_{hole\ mask} = \begin{cases} 1 & d_c(x) \geq d_c^{TH_{hole}} \\ 0 & otherwise \end{cases}$$

Shape Outlier Detection



$$O(x)_{outlier\ mask} = \begin{cases} 1 & \left\| z_c(x) - z_c^{est}(x) \right\| > e_c^{TH_{outlier}} \\ 0 & otherwise \end{cases}$$

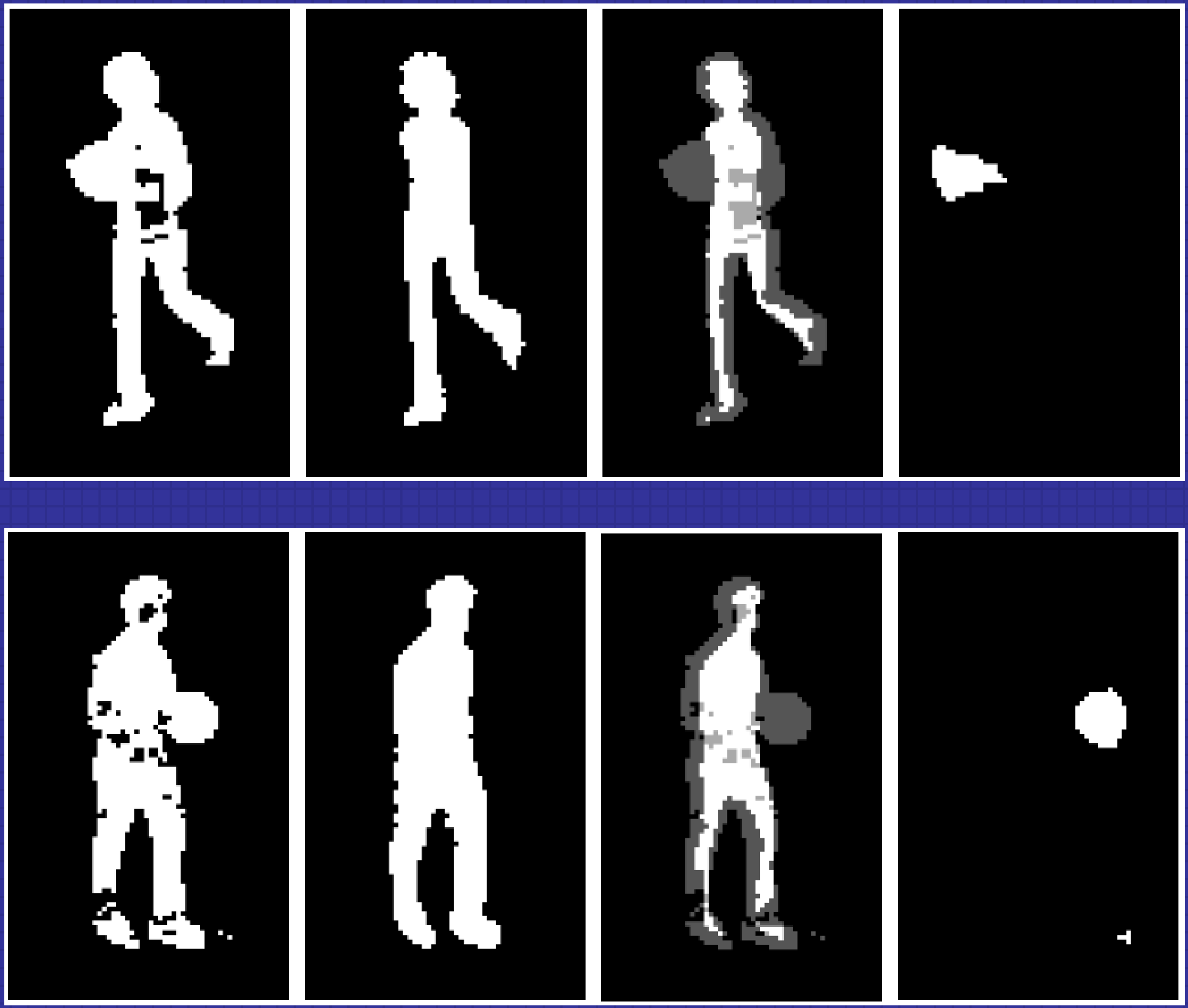
$$y_{outlier}(x) = \mathrm{z}\left(\mathrm{bin}\left(z_c(x) \otimes z_c^{est}(x)\right) \bullet O(x)_{outlier\ mask}\right)$$

Iterative Estimation

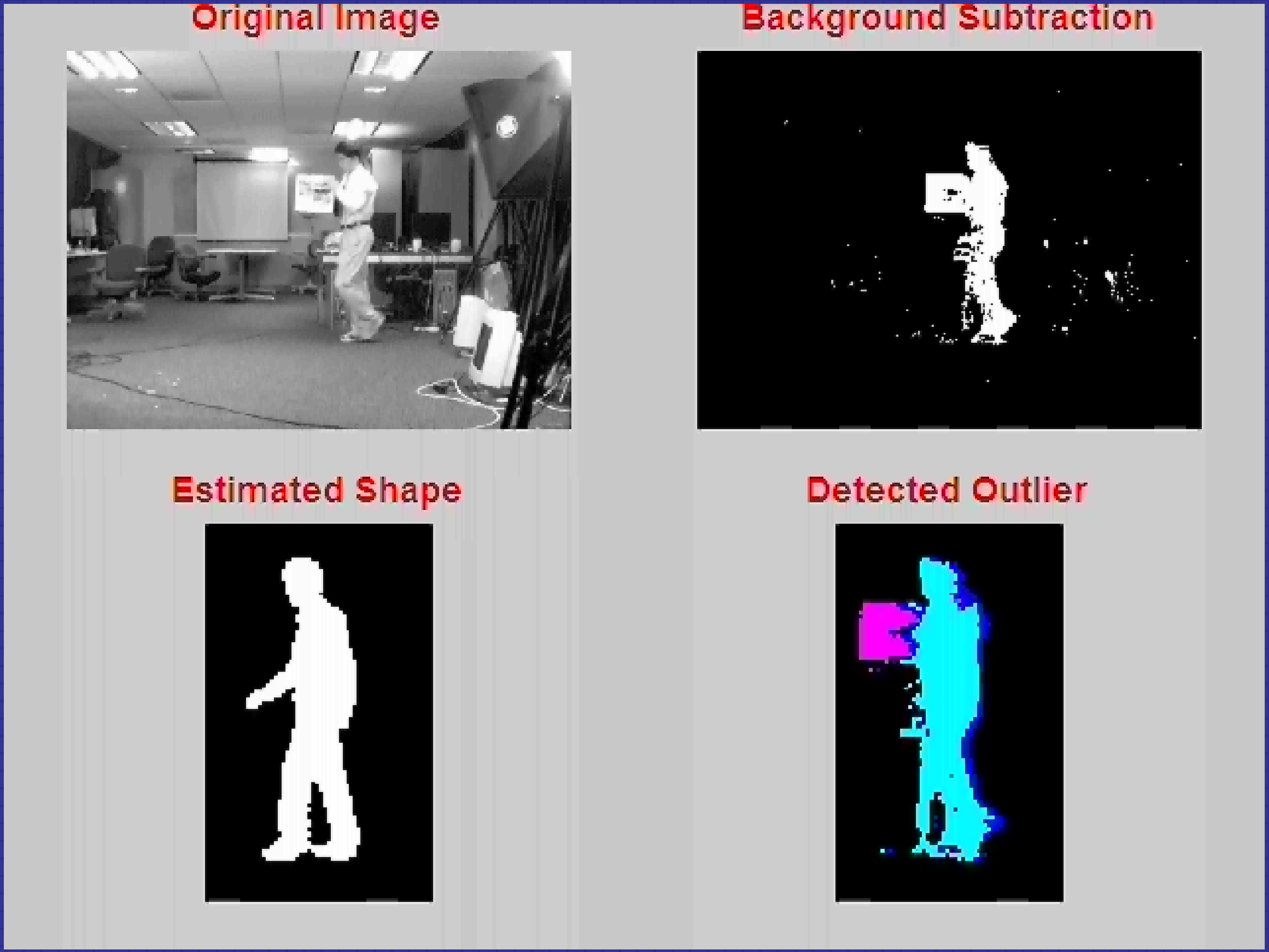
- Input
  - Image shape y_i estimated view v , estimated style s
- Iteration
  - Generate N configuration samples based on estimated view and style $y_i^{sp}, i = 1, \cdots, N_{sp}$
  - Generate hole filling masks from sample $h_i = h_{hole\ mask}(y_i^{sp})$
  - Estimate best fitting configuration sample with hole filling masks
  - Update input silhouette with hole filling
  - Estimate outlier from hole filled sample
  - Remove outlier
- Update
  - Reduce hole threshold value $d_c^{TH_{hole}}$, outlier threshold value $e_c^{TH_{outlier}}$

Experimental Results

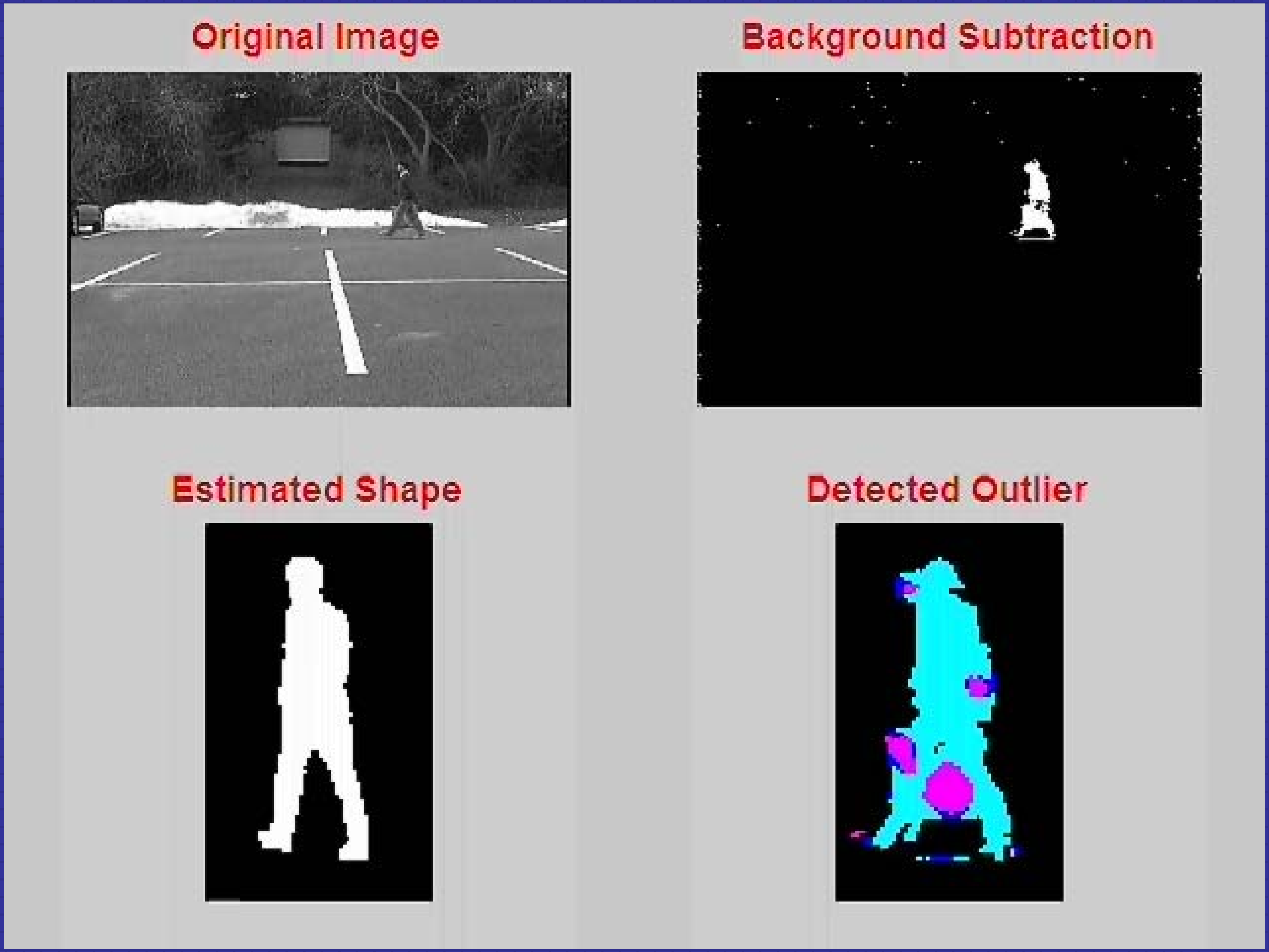
Outlier Detection in Fixed Views



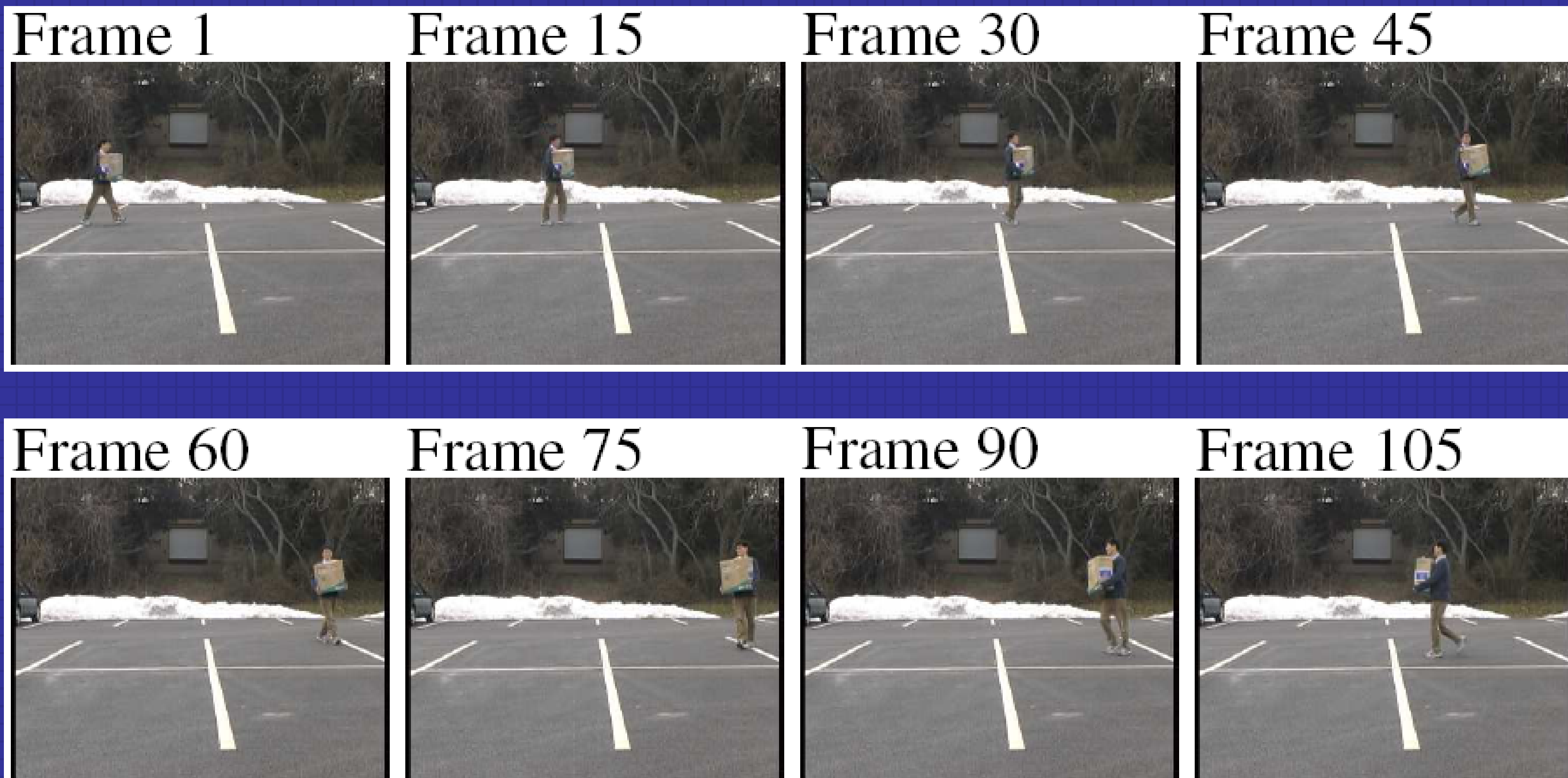
Outlier Detection in Fixed Views



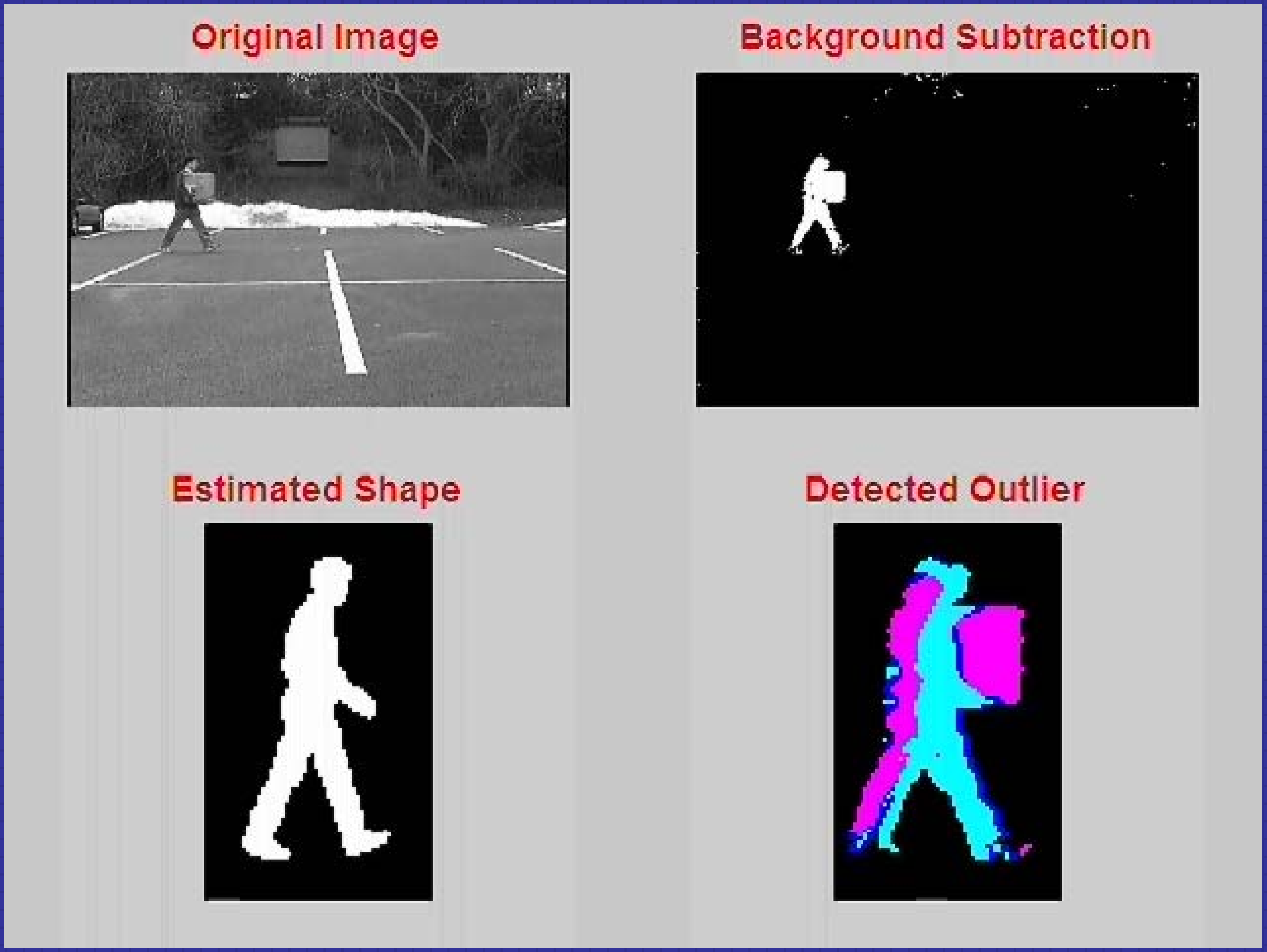
Outlier Detection in Fixed View



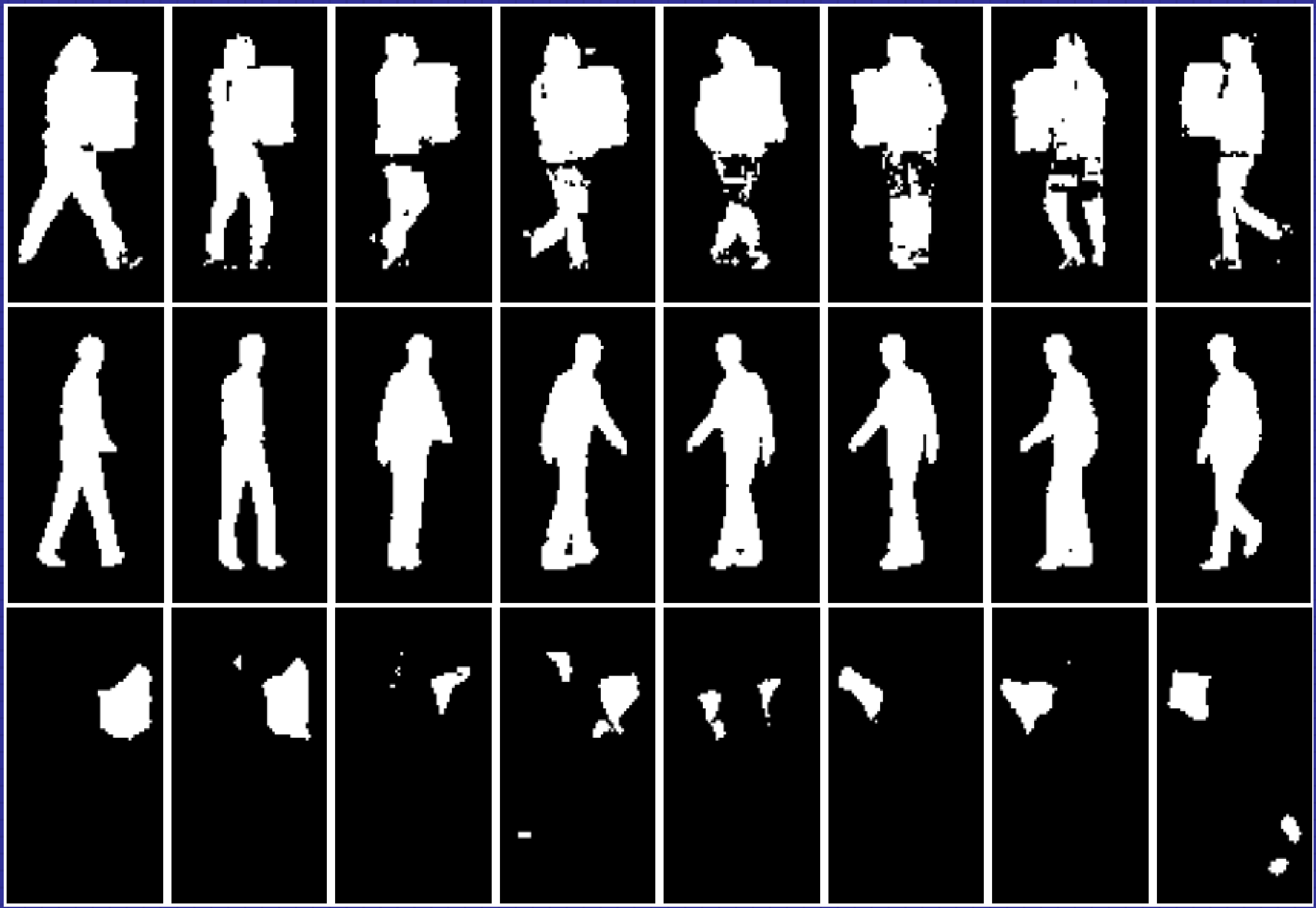
Outlier Detection in Continuous View Variations



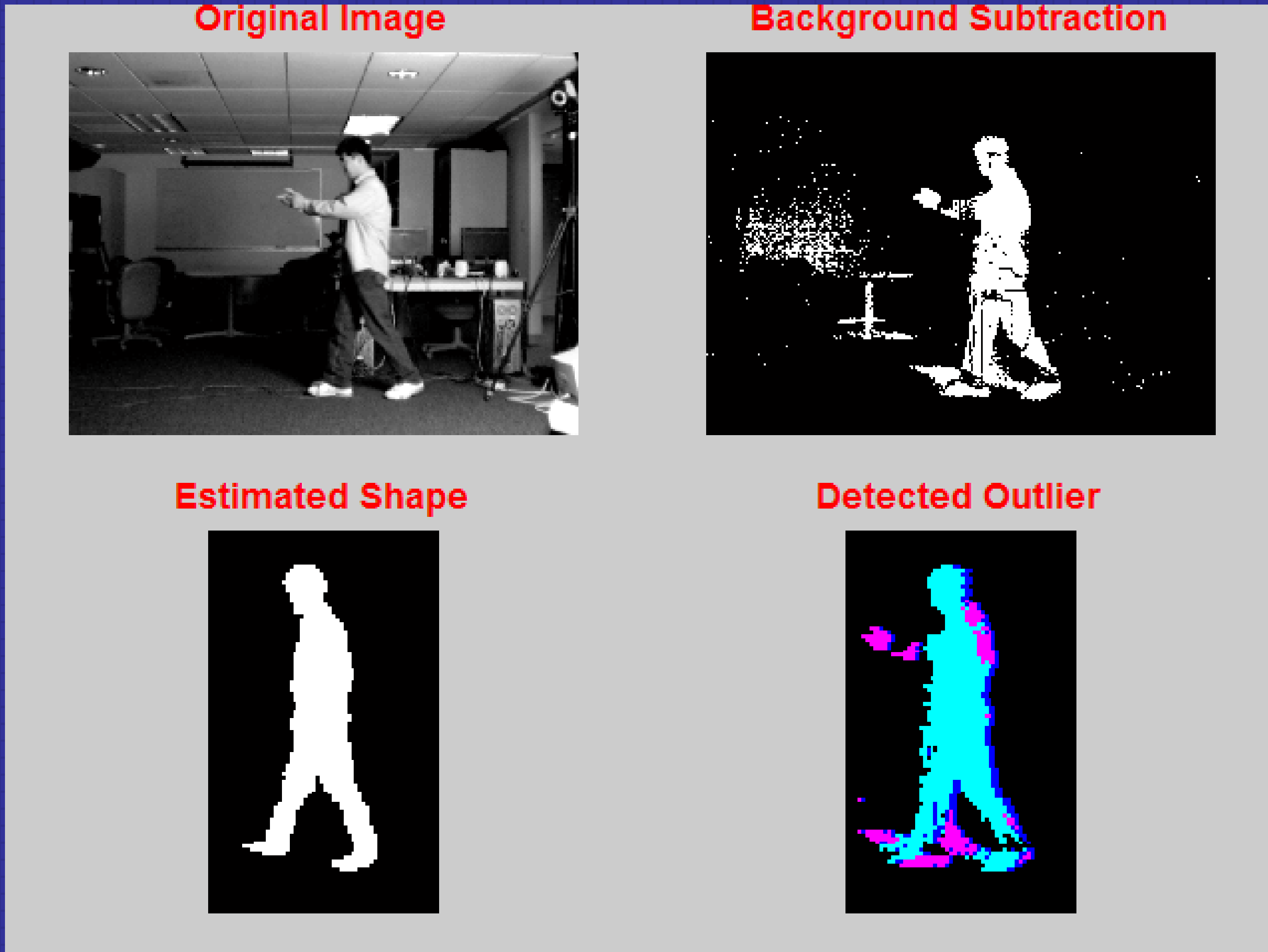
Outlier Detection in Continuous View Variations



Outlier Detection in Continuous View Variations



Shadow and abnormal pose detection



Conclusions

- Nonlinear Decomposable dynamic shape model
  - Provides shape model in different view and people
  - Detect shape outlier/carrying object by outliner detection with hole filling
  - Gradual reduction of threshold value for outlier detection and hole filling mask to be gradual reduction of misalignment due to outlier or hole

Future Works

- Analysis of temporal characteristics
  - Analysis of sequence of outlier for the detection of high level classification of outlier
    - Carrying object / Shadow / Abnormal action / …
  - Estimation of shape models with temporal coherence

Thank you