COMPUCON New Zealand

copyright 2003 Compucon NZ

Detection of defined human poses for video surveillance

Final BTech451 project seminar

Author: Xu He, Zhengping Wang, Bok-suk Shin, and Reinhard Klette

Sponsored by TN Chan from Compucon NZ.
Academic supervisor: Reinhard Klette
Course coordinator: Dr S. Manoharan

Present by Xu He

ivcnz

Compucon NZ

- Compucon New Zealand was established in 1992 (registered as Modern Technology NZ Limited in Auckland)
- First New Zealand PC maker to design and produce Dual Socket file servers with RAID in 1998.
- Compucon NZ is a computing system manufacturer and a digital technology system integrator.
- The company has performed as an Industry Patner for University of Auckland since 2002.

Aim and motivation

- Traditional video surveillance system record all the information for potential later use. This consumes a huge amount of storage space.
- Modern commercial video surveillance systems are capable of recording video only when there is any movement detected. This towards reducing the bandwidth of data transmission and video storage.
- Motion detection results are still inaccurate to some degree, which may lead to false detection.

Data collection

ACM-1511 IPVS camera

Selectable MPEG-4, MJPEG

8 fps at 1280x1024 SXGA resolution

Indoor environment

Monocular camera

NVR system

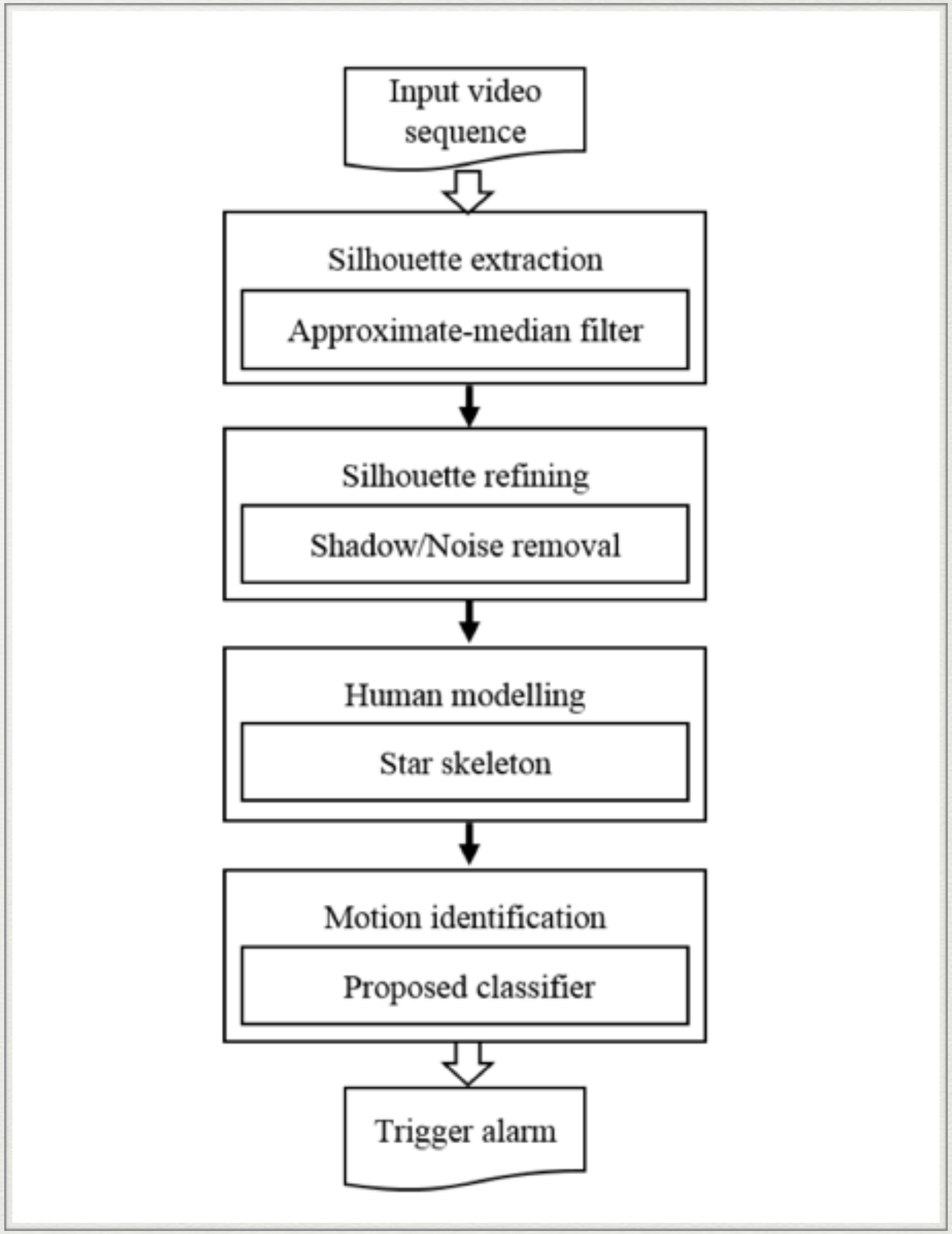
Transfer data over Ethernet

Steps of proposed system

Start with extracting a“fairly” accurate human silhouette by using approximate median filtering[1].

Matching a human model by using star skeletonization[2].

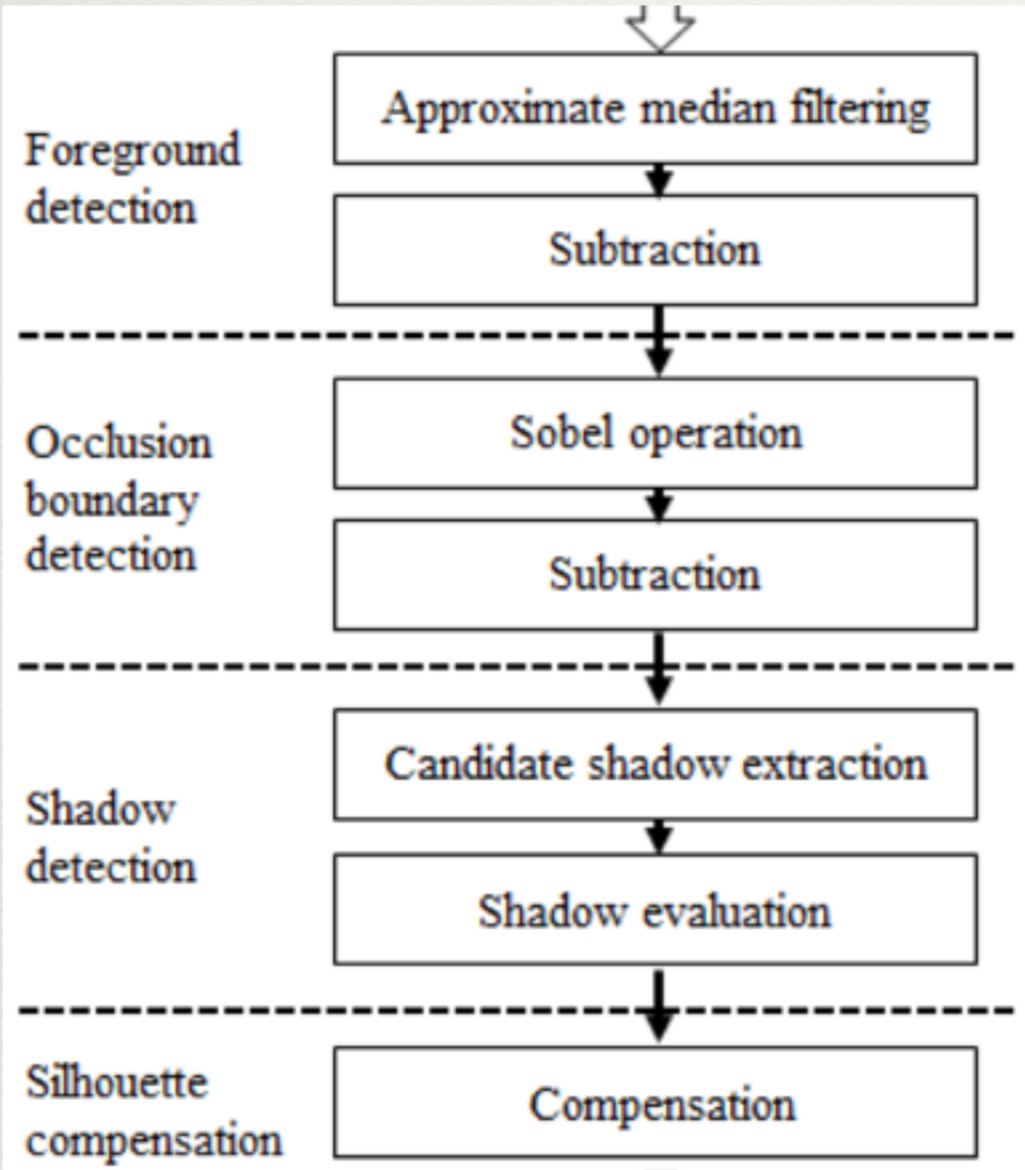
We show how star skeletons can be used to identify human poses characterised by raised hands.



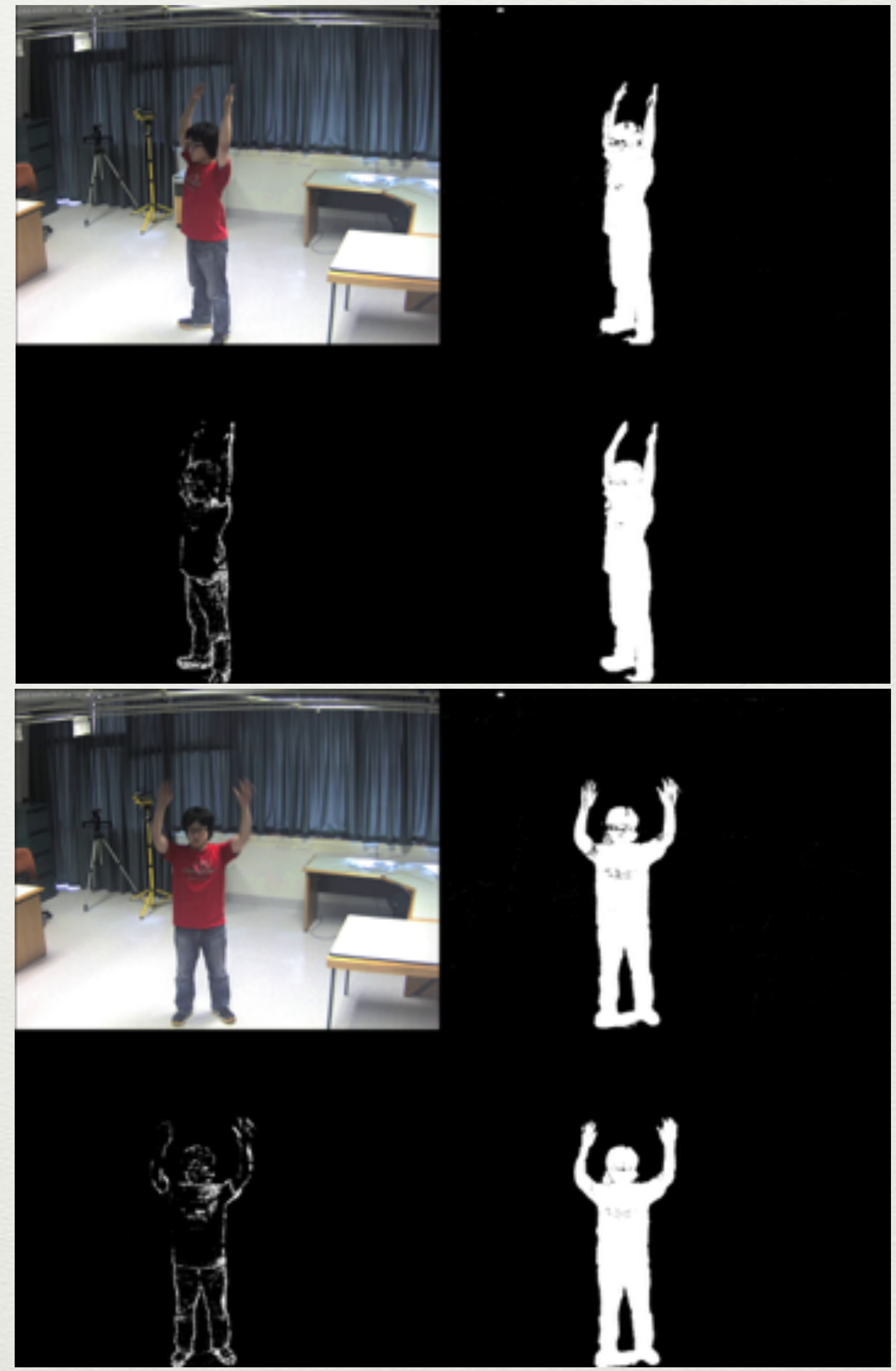
[1] Z. P. Wang, B.-S. Shin, and R. Klette. Accurate silhouette extraction of a person in video data by shadow evaluation. J. Computer Theory Engineering, 6:476–483, 2014.

[2] H. Fujiyoshi and A. J. Lipton. Real-time human motion analysis by image skeletonization. In Proc. IEEE Workshop Applications Computer Vision, pages 15–21, 1998.

Silhouette extraction



this picture from below reference



Z. P. Wang, B. S. Shin, and R. Klette.

Accurate silhouette extraction of a person in video data by shadow evaluation.

In *Int. J. Computer Theory Engineering*, **6**:476–483, 2014.

Silhouette extraction

The background of the video sequence is updated by using Approximate median filter[1]

$$B(x,y,t) = \begin{cases} B(x,y,t-1)+1, \\ \qquad\qquad \text{if } I(x,y,t) > B(x,y,t-1) \\ \\ B(x,y,t-1)-1, \\ \qquad\qquad \text{if } I(x,y,t) < B(x,y,t-1) \end{cases}$$

- I(x,y,t) - the value of an image pixel at position (x,y) at time t
- B(x,y,t) - the value of a background pixel at postion(x,y) at time t
- I(x,y,0) be the initial value of B(x,y,0).

[1] Z. P. Wang, B.-S. Shin, and R. Klette. Accurate silhouette extraction of a person in video data by shadow evaluation. J. Computer Theory Engineering, 6:476–483, 2014.

Background Subtraction

$$F(x,y,t) = \begin{cases} 1 \;\; if|I(x,y,t) - B(x,y,t-1)| > \sigma_t \\ \\ 0 \;\; I(x,y,t) < otherwise \end{cases}$$

$$\sigma_t \;\; = \;\; \sqrt{\left(\sum_{x=1}^{N_{cols}} \sum_{y=1}^{N_{rows}} (I(x,y,t) - \mu_t)^2 \right) / |\Omega|} \qquad (3)$$

$$\text{with } \mu_t = \left(\sum_{x=1}^{N_{cols}} \sum_{y=1}^{N_{rows}} I(x,y,t) \right) / |\Omega| \qquad (4)$$

- Sigma is the standard deviation of all the intensity value of frame at time t

[1] Z. P. Wang, B.-S. Shin, and R. Klette. Accurate silhouette extraction of a person in video data by shadow evaluation. J. Computer Theory Engineering, 6:476–483, 2014.

Foreground extraction

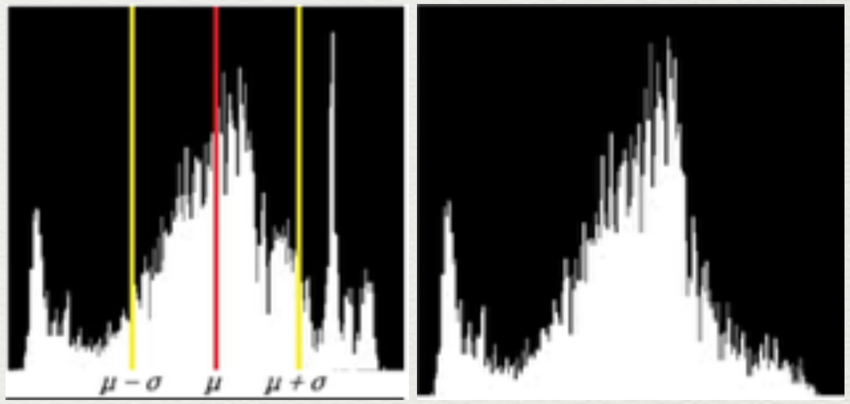
- Using the Sobel operator as a simple and robust edge estimator on the subtracted background image for obtaining raw occlusion boundaries of a person
- Subtracting the background boundaries from the raw occlusion boundaries (of a person) in order to extract the true occlusion border of a person.
- Filling the true occlusion border to obtain the foreground mask ,also called the silhouette.

[1] Z. P. Wang, B.-S. Shin, and R. Klette. Accurate silhouette extraction of a person in video data by shadow evaluation. J. Computer Theory Engineering, 6:476–483, 2014.

Silhouette refining

- Shadow removal - candidate shadow detection and shadow evaluation.
    - Most of shadow pixels is far away from the top of the normal distribution of horizontal histogram.

$$S(x,y,t)=\begin{cases}1 & \text{if } F(x,y,t)=1 \\ & \quad \text{and } |I(x,y,t)-\mu|>\sigma \\ 0 & \text{otherwise}\end{cases}$$



- Noise removal - Morphological erosion.
    - large blob removed by coating total number of pixels.
    - Set a fixed threshold 800 (approx. one third of total number of border pixels of a person)
- Reducing noise features and connecting isolated body part
    - applying dilation twice followed by an erosion.

[1] Z. P. Wang, B.-S. Shin, and R. Klette. Accurate silhouette extraction of a person in video data by shadow evaluation. J. Computer Theory Engineering, 6:476–483, 2014.

Star skeletonization

There are five steps:

Step 1: Trace the border of the extracted silhouette.

Step 2: Find the centre of gravity of the target border.

$$x_c = \frac{1}{N}\sum_{i=1}^{N} x_i, \quad y_c = \frac{1}{N}\sum_{i=1}^{N} y_i$$

Step 3: Define a distance function D(i) which is the distance between the centre of gravity and each border pixel point.

$$D(i) = \sqrt{(x_i - x_c)^2 + (y_i - y_c)^2}$$

H. Fujiyoshi and A. J. Lipton. Real-time human motion analysis by image skeletonization. In Proc. IEEE Workshop Applications Computer Vision, pages 15–21, 1998.

Star skeletonization (cont.)

Step 4: Smooth the distance signal D(i) by applying a low-pass filter in the frequency domain. The low pass filter has a cutoff- threshold **a** for filtering out the high-frequency components. Let **D**(u) be the Fourier transform of D(i). We set,

$$\mathbf{D}(u) = 0 \ \text{ if } \ |u| \geq a \cdot N$$

- u as the frequency coordinate
- N as the to total number of border pixels
- a = 0.0004 as the threshold for our experiment.
    - 0.015 suggested in the paper[1] but with a much lower resolution.
    - The lower **a**, the more maxima will be found.

[1] H. Fujiyoshi and A. J. Lipton. Real-time human motion analysis by image skeletonization. In Proc. IEEE Workshop Applications Computer Vision, pages 15–21, 1998.

Star skeletonization (cont.)

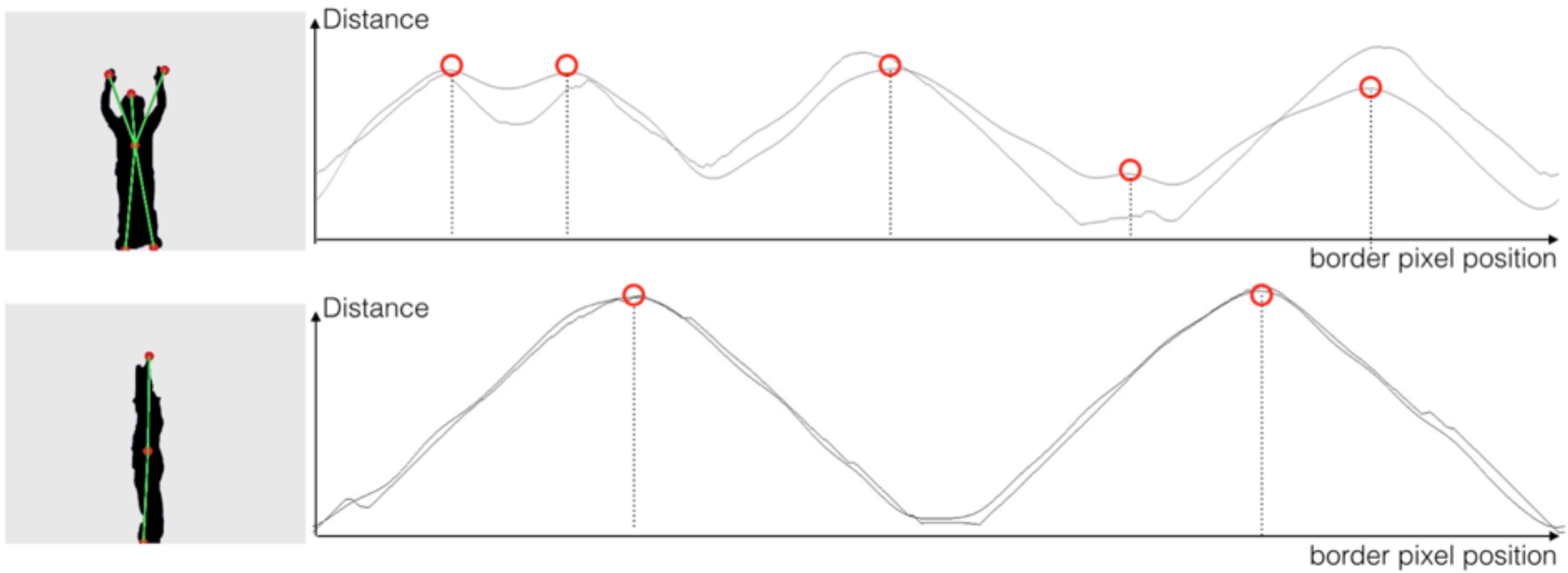


Figure 4: *Right:* **Original distance signal** D **and smoothed distance signal** $\hat{D}$ **defined by a low pass in the Fourier domain.** *Left:* **Calculated skeletons.**

Step 5: Taken all local maxima in the filtered signal **D**(u)

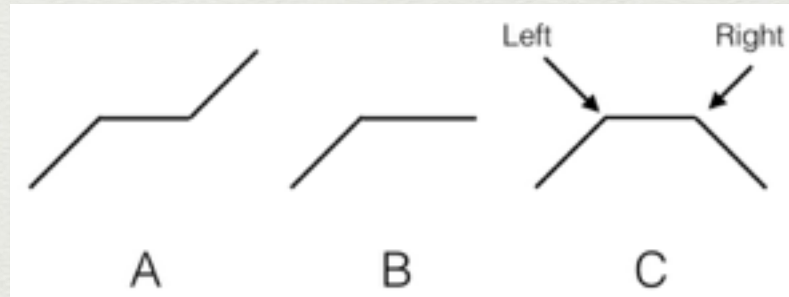


Figure 3: Three point sequences. A and B are not a maximal situation. C is a maximal situation.

Detection of raised hands

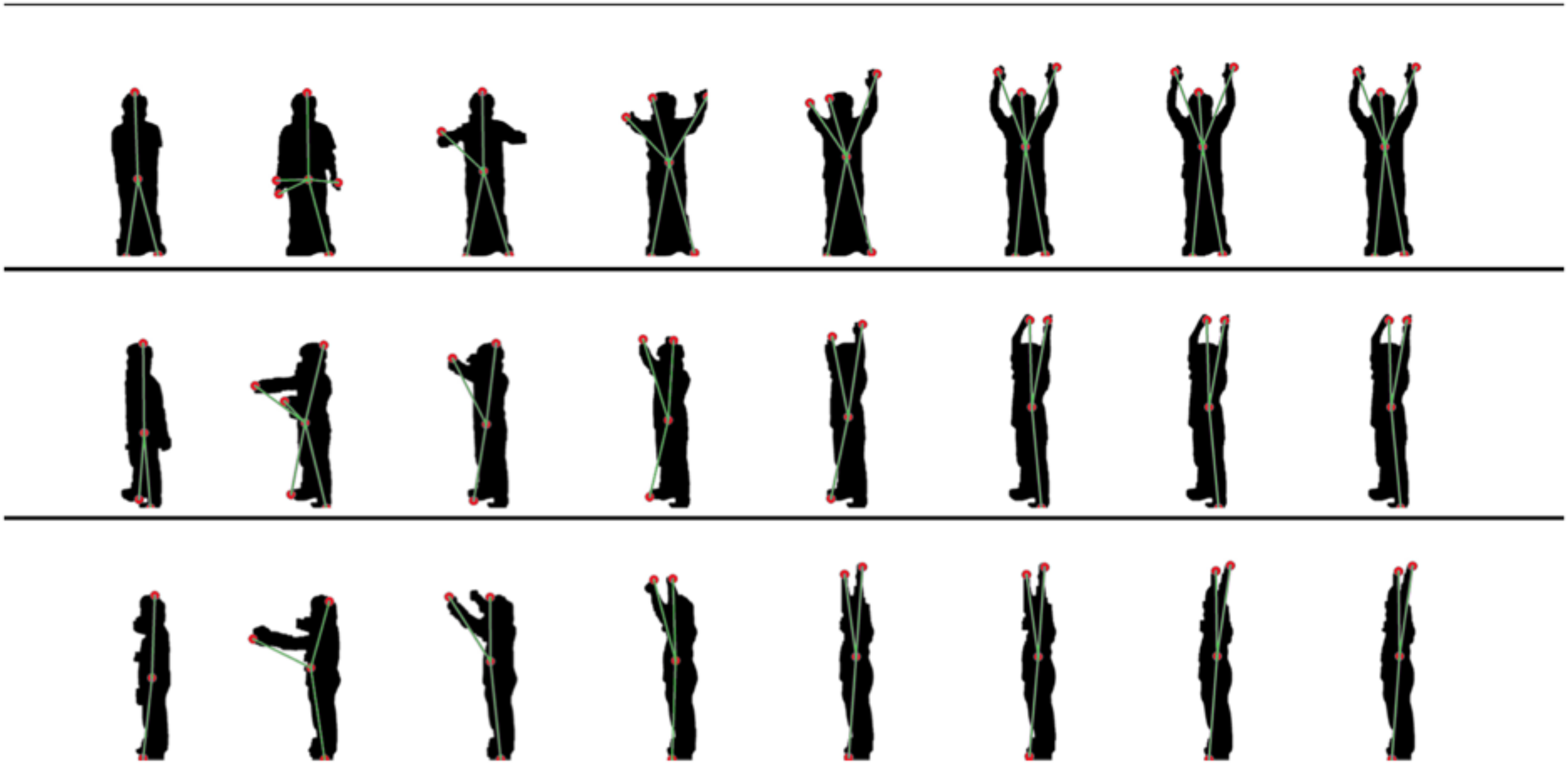
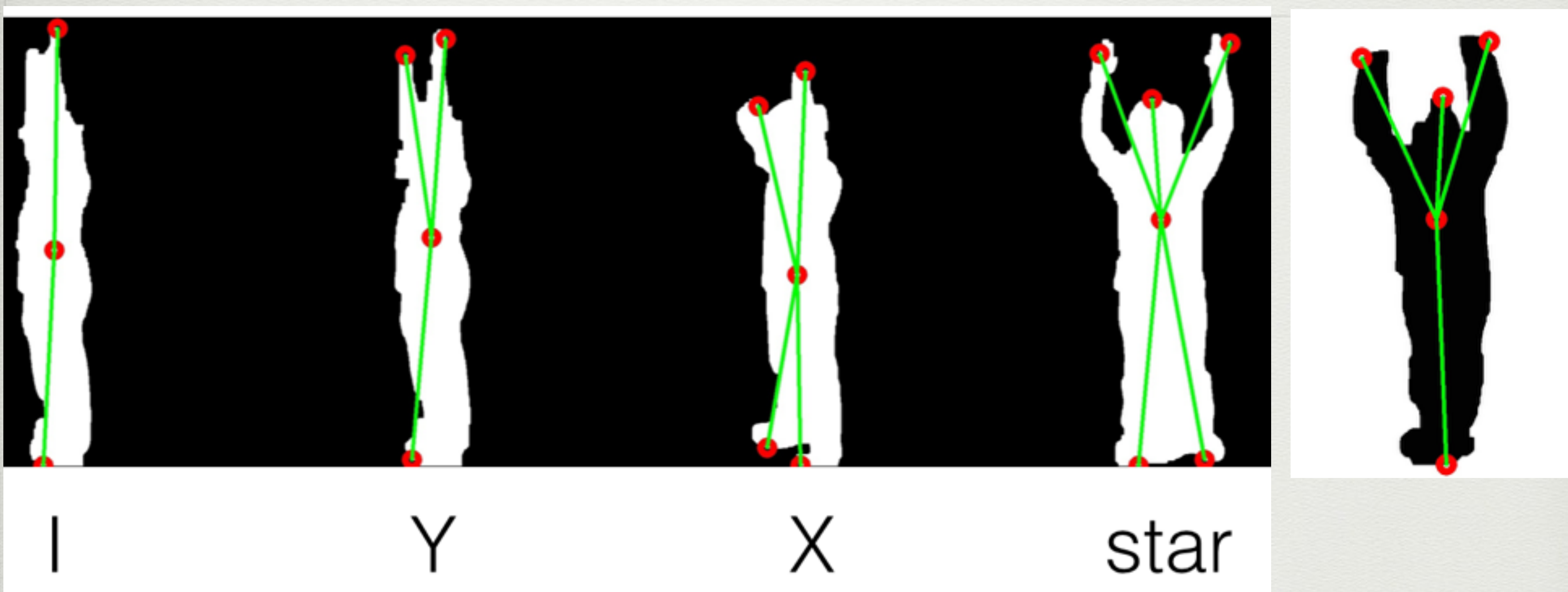


Figure 6: Different states of a person which raises the hands. *Top Row:* **Front view.** *Second Row:* **Half-side view.** *Bottom Row:* **Side view.**

Detection of raised hands



$$y_{\text{hands}} > 0.8 \cdot H \ \& \ y_{\text{feet}} < 0.2 \cdot H$$

In all cases the common feature is that the positions of the two hands are on top of the head position at some stage.

Detection of raised hands

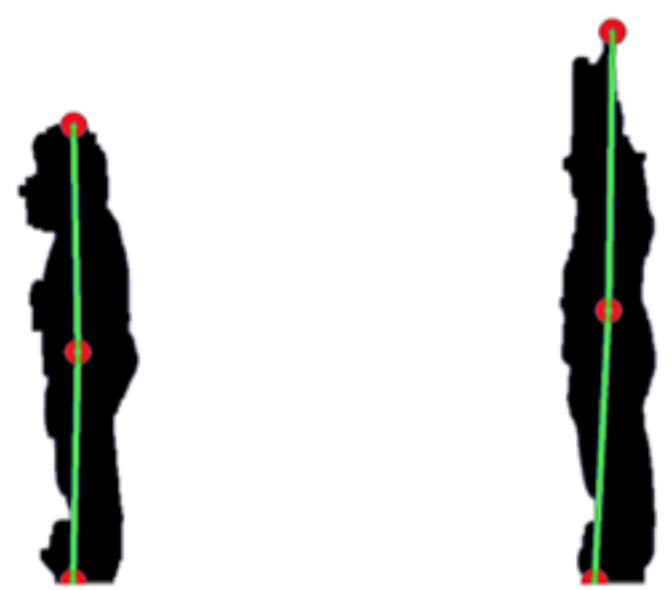


Figure 5: The two extreme cases denoted as *I-shape*. A temporal change in height indicates the raised hands. *Left:* A side view of a standing person. *Right:* A side view of a standing person having the hands up.

Results

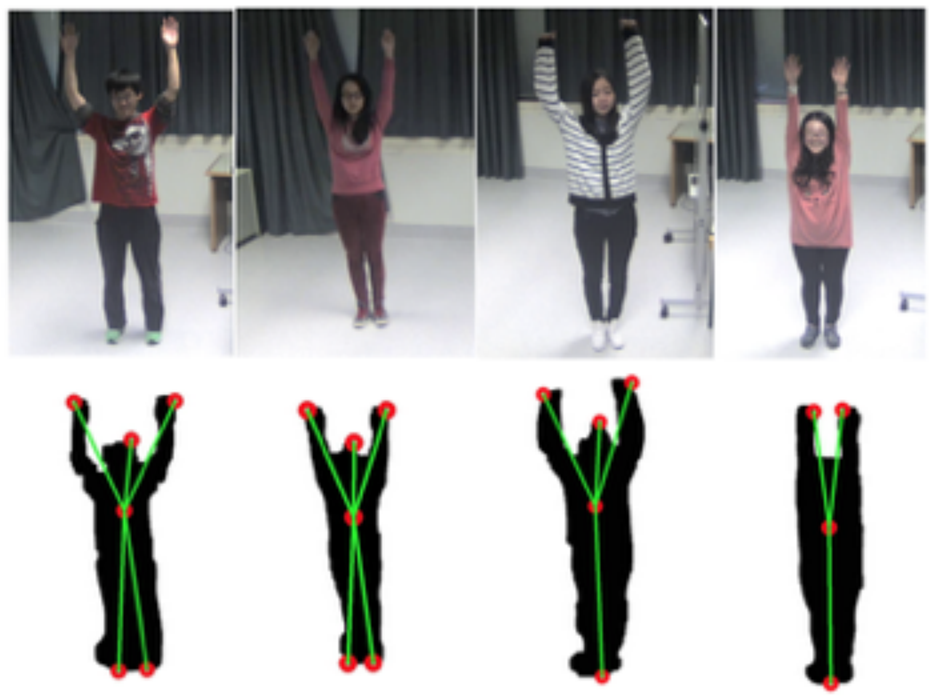


Figure 7: Samples of extracted silhouettes for test persons B to E, left to right.

Sequence	Total frames	FP	FN	Ratio
#A	172	0	4	97.67%
#B	305	2	10	96.07%
#C	287	22	4	90.94%
#D	293	4	8	95.90%
#E	243	5	4	96.30%
Total	1300	33	30	95.15%

Thank you