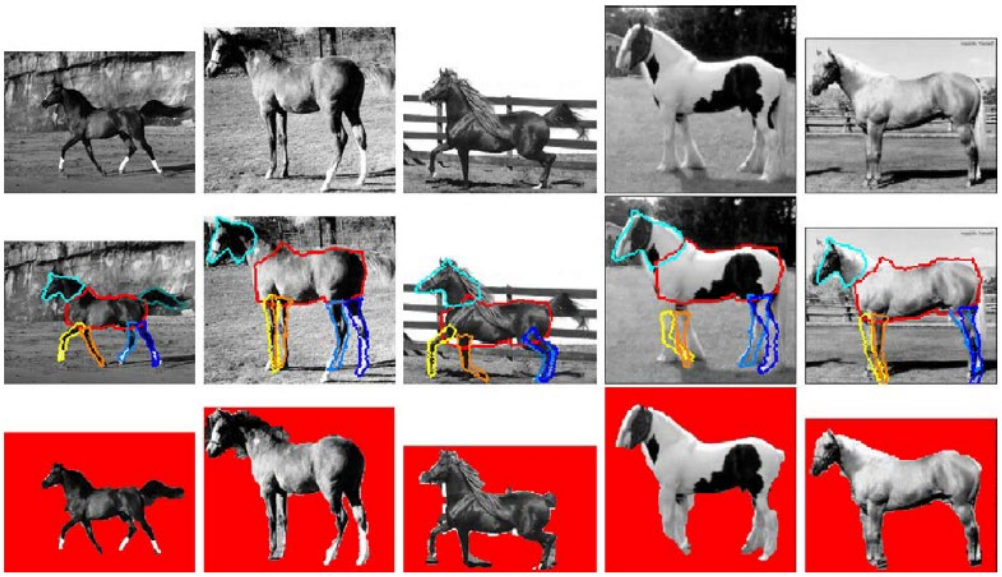
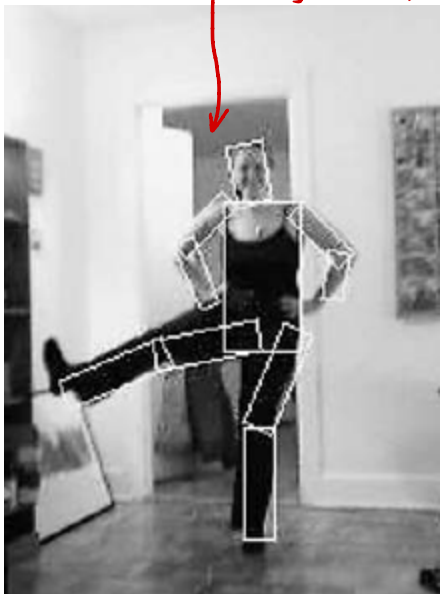
Part-based Matching

(Model-based Matching)



devide into different parts



References:

1. Fischler and Elschlager, The Representation and Matching of Pictorial Structures, IEEE Transactions on Computer, 22(1):67-92, Jan, 1973

2. Felzenszwalb and Huttenlocher, Pictorial Structures for Object Recognition, International Journal of Computer Vision, 61(1):55-79, Jan, 2005

3. Wu and Chung, A Novel Framework for Segmentation of Deep Brain Structures based on Markov Dependence Tree, NeuroImage, 46:1027-1036, 2009

4. Wu, Cai and Chung, POSIT: Part-based Object Segmentation without Intensive Training, Pattern Recognition, 43:676-684, 2010

5. Erol, Bebis, Nicolescu, Boyle and Twombly, Vision-based Hand Pose Estimation: A Review, Computer Vision and Image Understanding, 108:52-73, 2007

Representation and Matching of Pictorial Structures

http://en.wikipedia.org/wiki/Part-based_models

The goal of object detection

- Given some description of a visual object, the goal is to find that object in an actual photograph.

Notes:

1. The object might be simple
2. The description can be linguistic (based on languages), pictorial (based on pictures), etc
3. The actual photograph = Sensed scene = 2D array of gray-level values
4. Object being sought = Reference = Template object

A generic form of the non-rigid registration

1. Assume that the reference/template object is an image on a transparent rubber sheet.
2. We move this sheet over the sensed image by pulling or pushing on the rubber sheet to get the best possible alignment between the template object and the sensed image.
3. We evaluate by looking at
   a. How good a correspondence we were able to obtain?
   b. How much pushing and pulling we had to exert to obtain it?

Problem:

- An object might be composed of a number of parts. Each part might want to retain its internal shape.

Pictorial Structure Representation

1. Template object is composed of a number of parts.
   a. The rubber sheet is replaced by a reference image (template object) which is composed of a number of rigid pieces (parts).
   b. A rigid piece (part) of the reference image (template object) corresponds to a single coherent entity in the reference image.

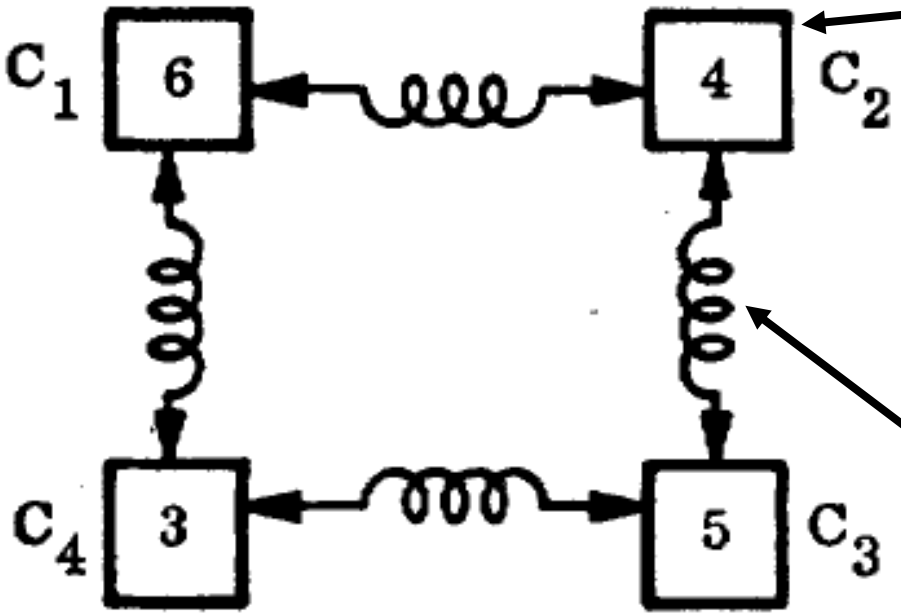
2. There are connections between parts.
   a. The rigid pieces are held together by “springs”.
   b. The springs joining the rigid pieces (parts) serve both to constrain relative movement and to measure the “cost” of the movement by how much they are “stretched”.

Fischler and Elschlager

Pictorial Structure Representation

Advantages

- Part distortion can be bounded. Internal shape can be retained.
- Relative positions between parts might be bounded.
- Object is a combination of parts, $C_1 - C_4$, with connections having spring property.



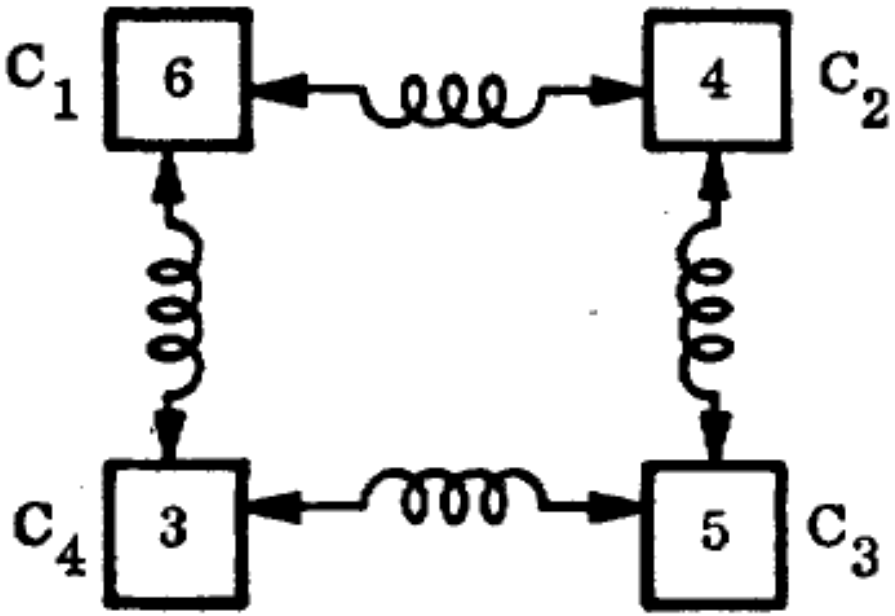
Energy for describing the mismatch of a template part against the image.

$$M_i(p_i)$$

Energy for describing the relative positions between parts

$$\sum_{i \neq j} T_{i,j}(p_i, p_j)$$

Fischler and Elschlager



$$E(P) = \sum_{i} M_i\left(p_i\right) + \sum_{i}\sum_{i \neq j} T_{i,j}\left(p_i, p_j\right)$$

each part

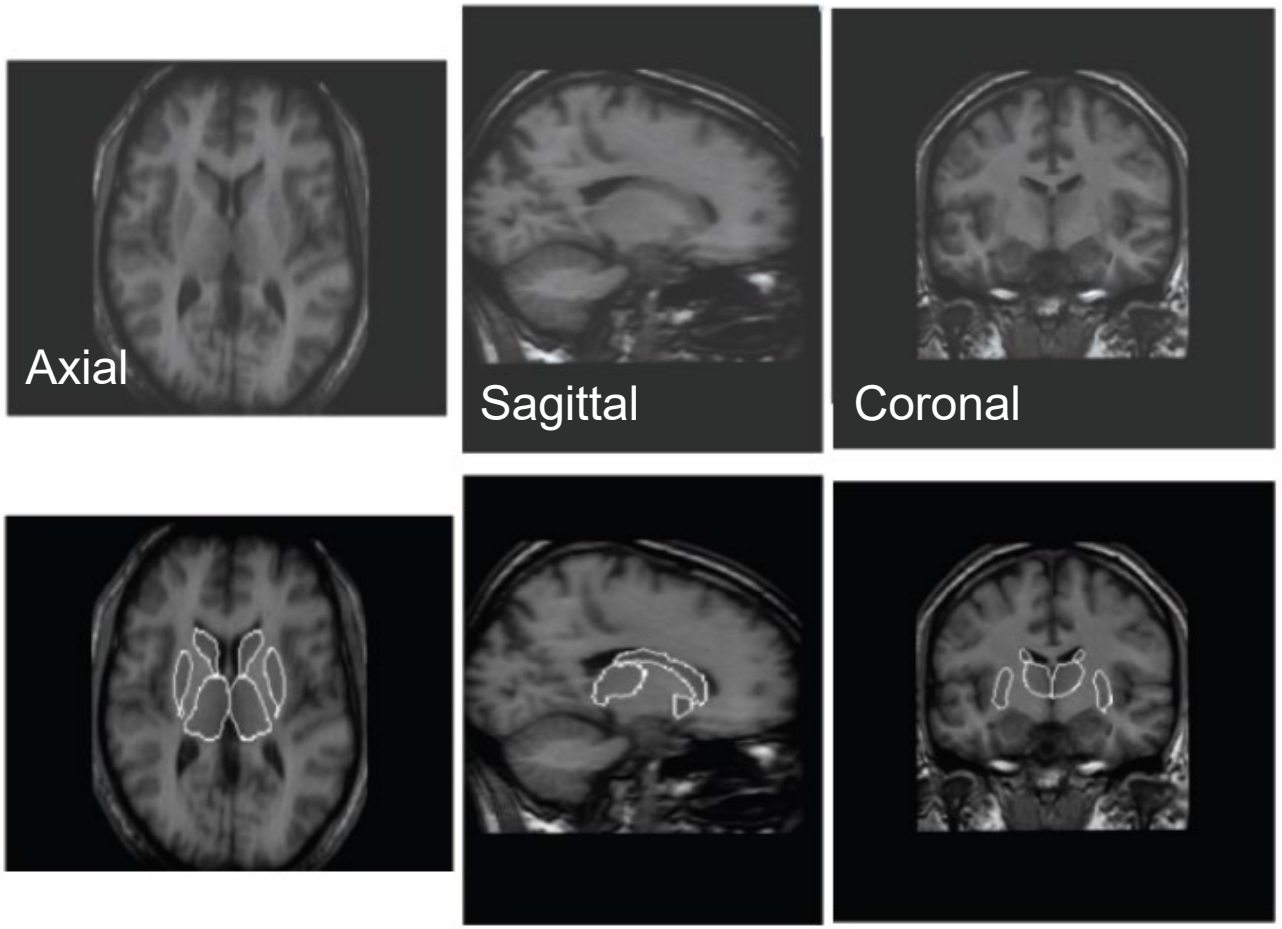
relationship

Fischler and Elschlager

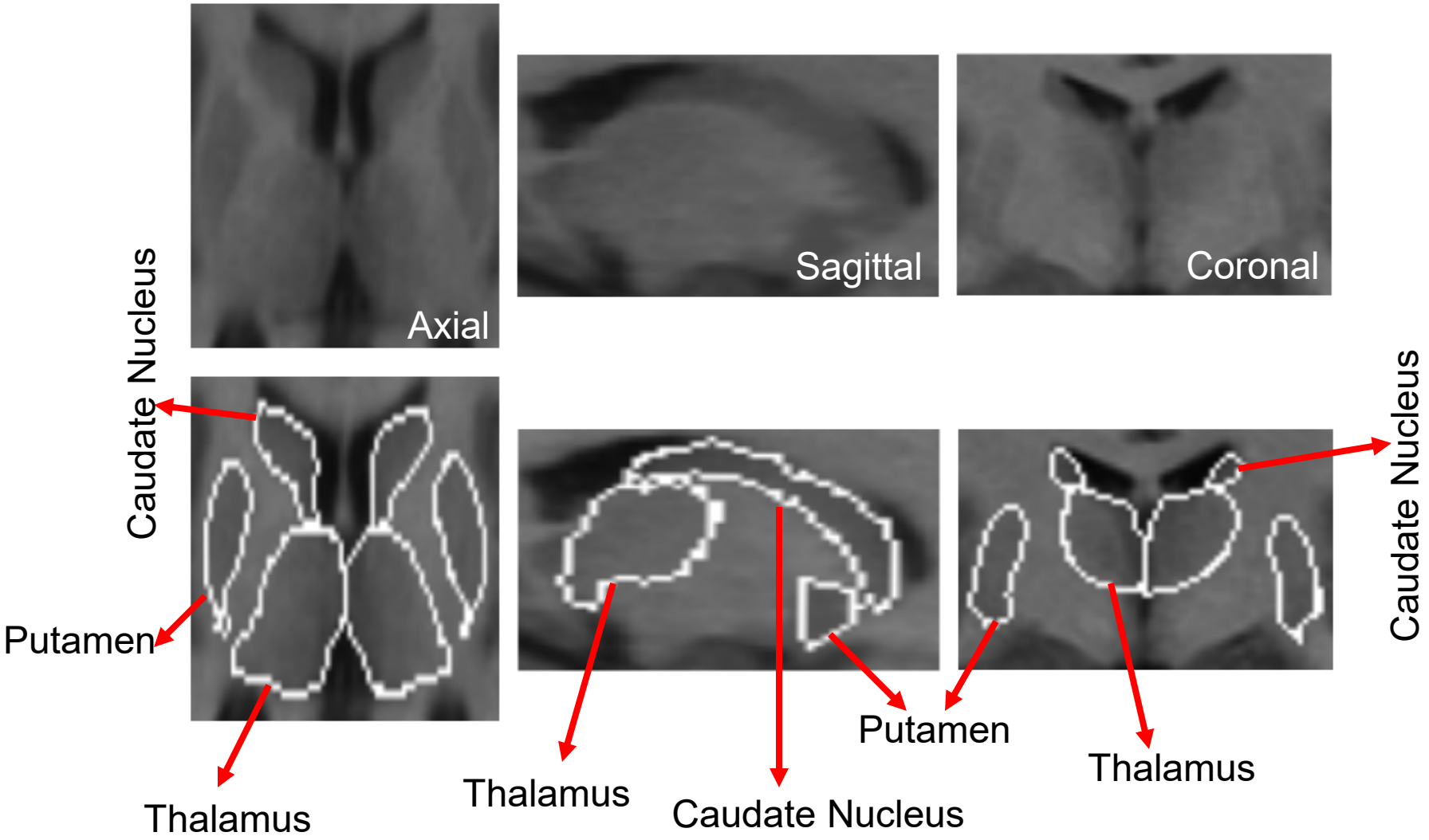
Detection of Deep Brain Structures

Deep Brain Structures

- Caudate nucleus, putamen and thalamus play important roles in human brain functioning.

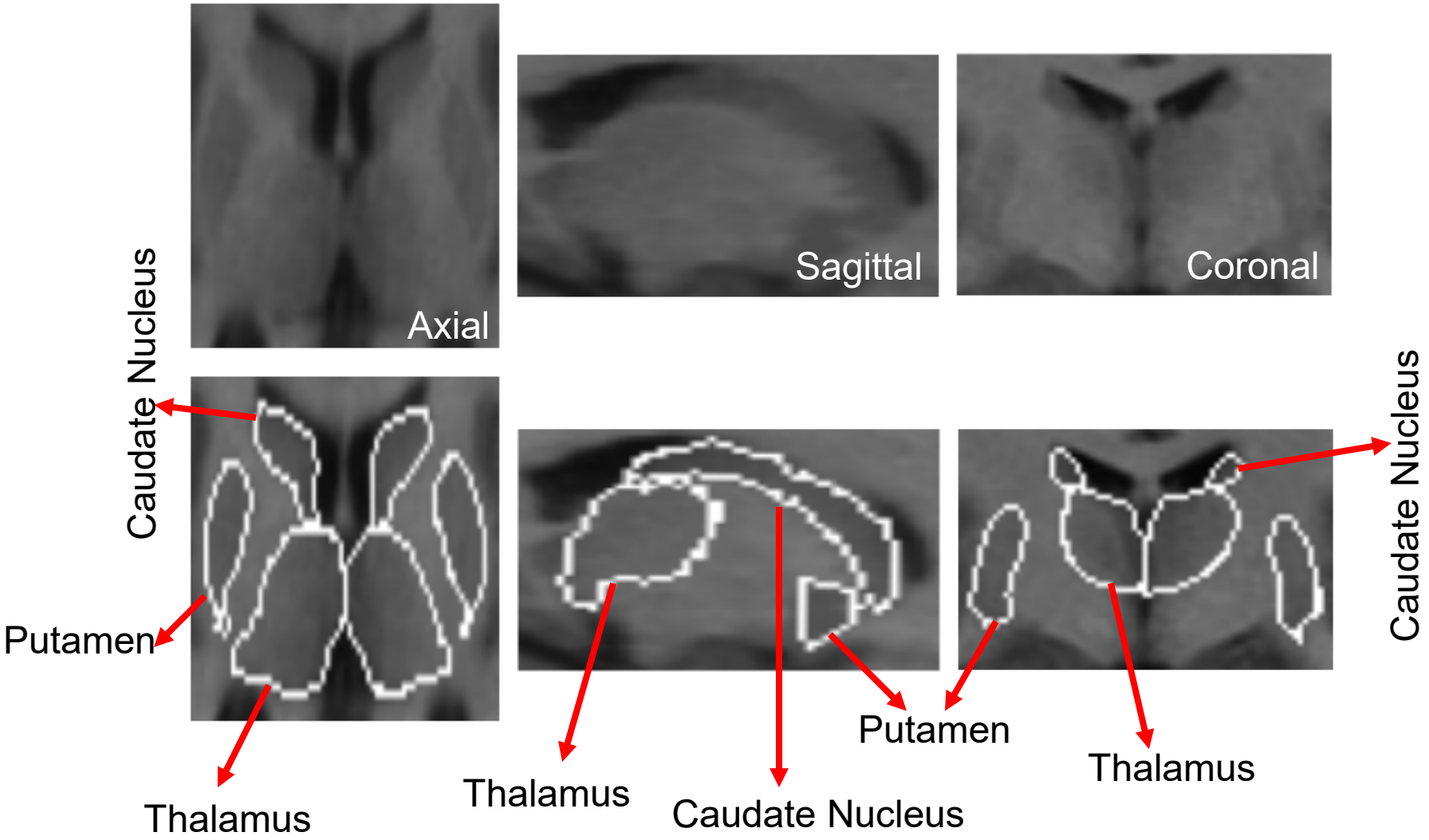


Close-up axial, sagittal and coronal views



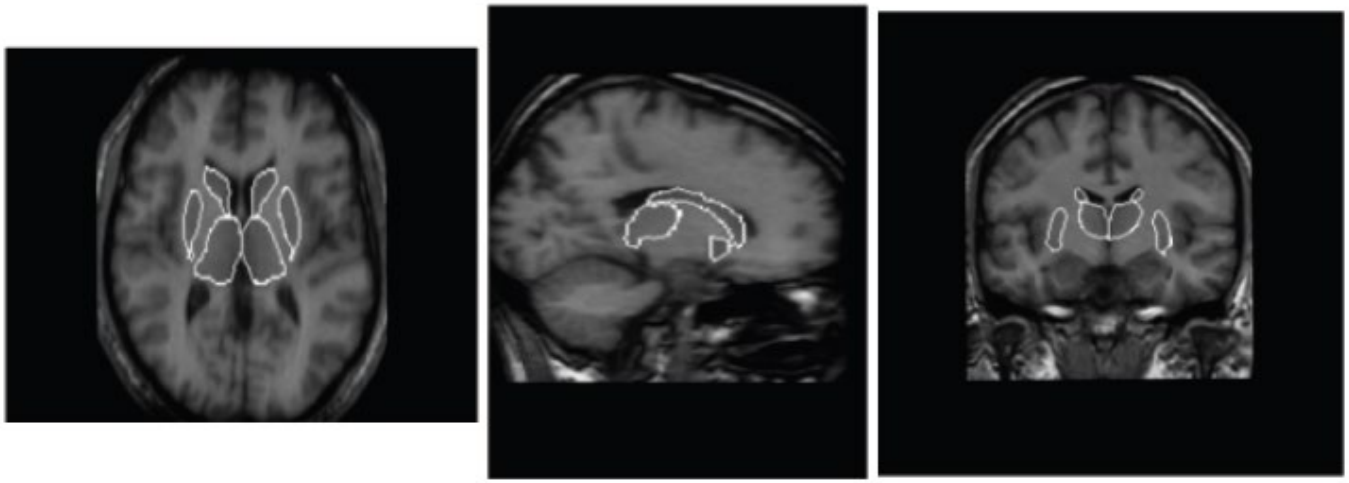
To segment them, there are several challenges…

Challenge #1: Boundaries are blurry



Active contour model may fail to find missing or weak boundaries.

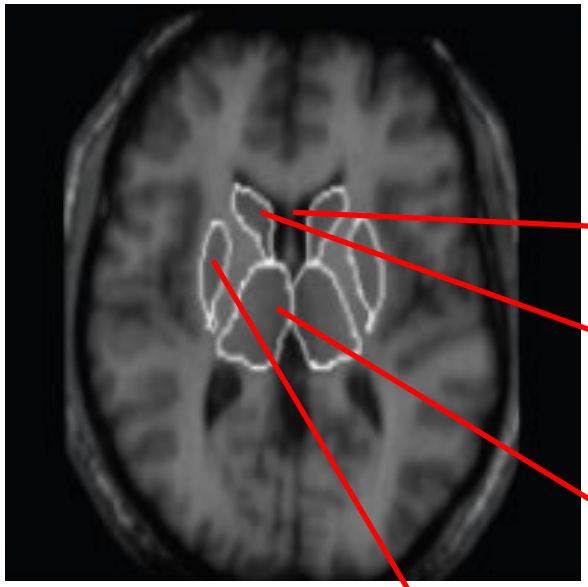
Challenge #2: Small volume size and nearby structures



When segmenting these small structures with blurry boundaries, atlas-based methods may be affected by the larger structures with clear boundaries.

Nearby structures are larger and they have clearer boundaries. Active contours may be attracted to the nearby structures if there are no shape constraints.

- Part-based presentation
  - Each structure is viewed as a part of a larger complex object
  - Each part is a binary map and an intensity map of the structure
  - Each part can non-linearly, individually, locally deform



Ventricles

Left Caudate Nucleus

Right Caudate Nucleus

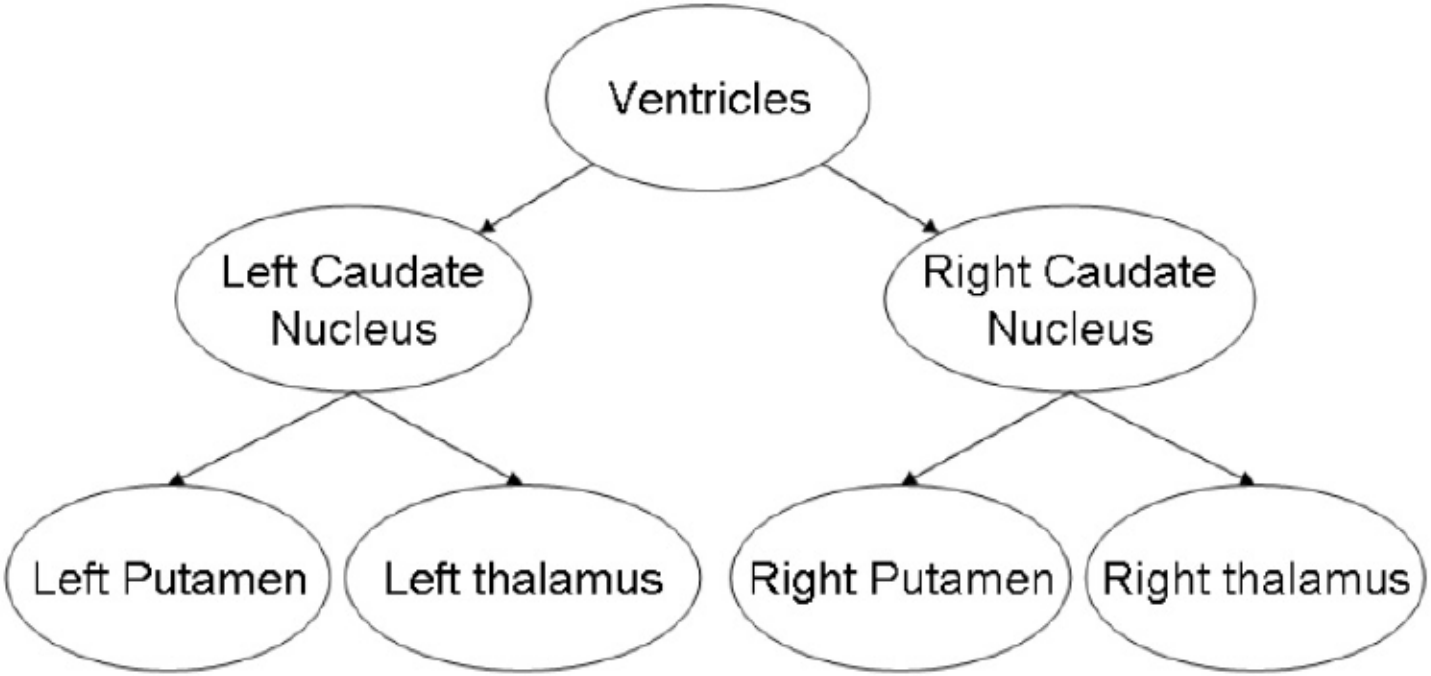
Left Putamen

Left thalamus

Right Putamen

Right thalamus

- Markov dependence tree (MDT) representation
  - A tree is used to model the non-linear, inter-part interactions
  - “Ventricles” part is the tree root
  - The current status is characterized by the MDT total energy

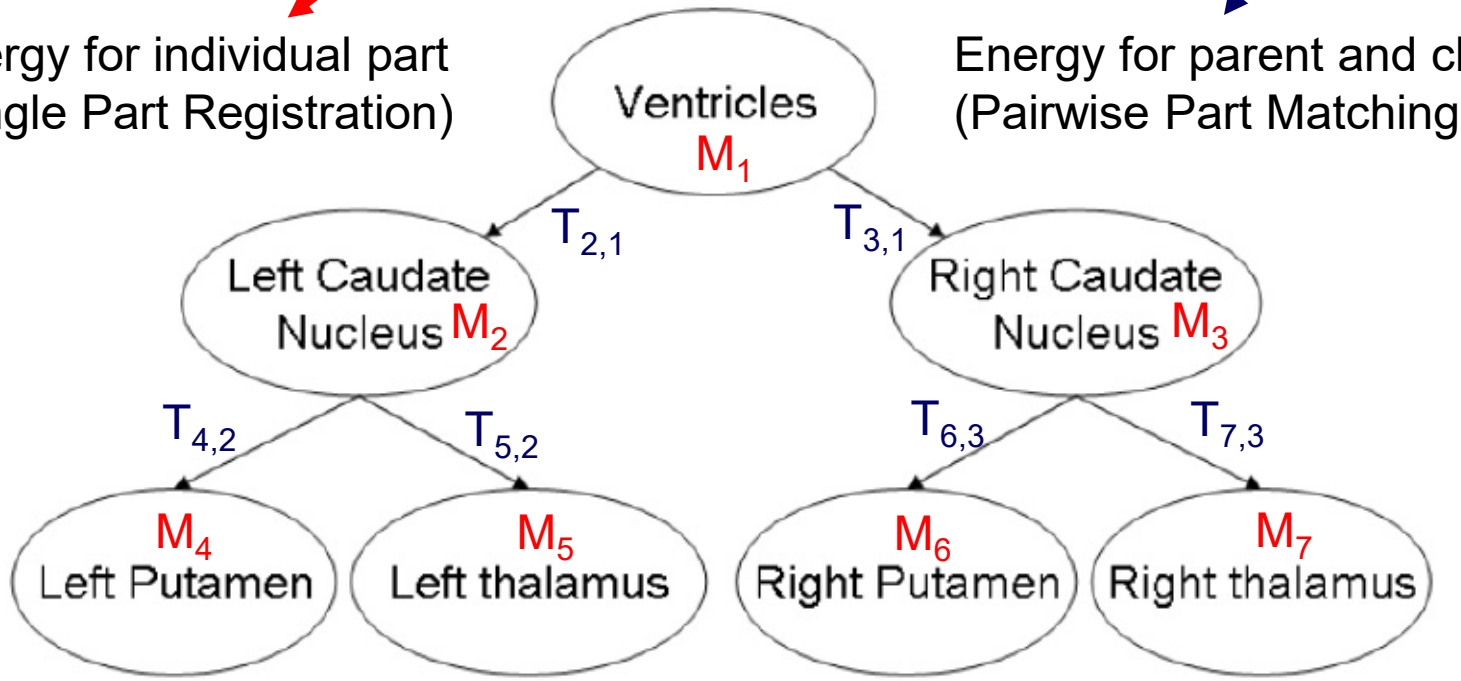


Markov Dependence Tree (MDT) Total Energy

$$E(P) = \sum_i M_i(p_i) + \sum_i \sum_{j=\text{parent of } i} T_{i,j}(p_i|p_j)$$

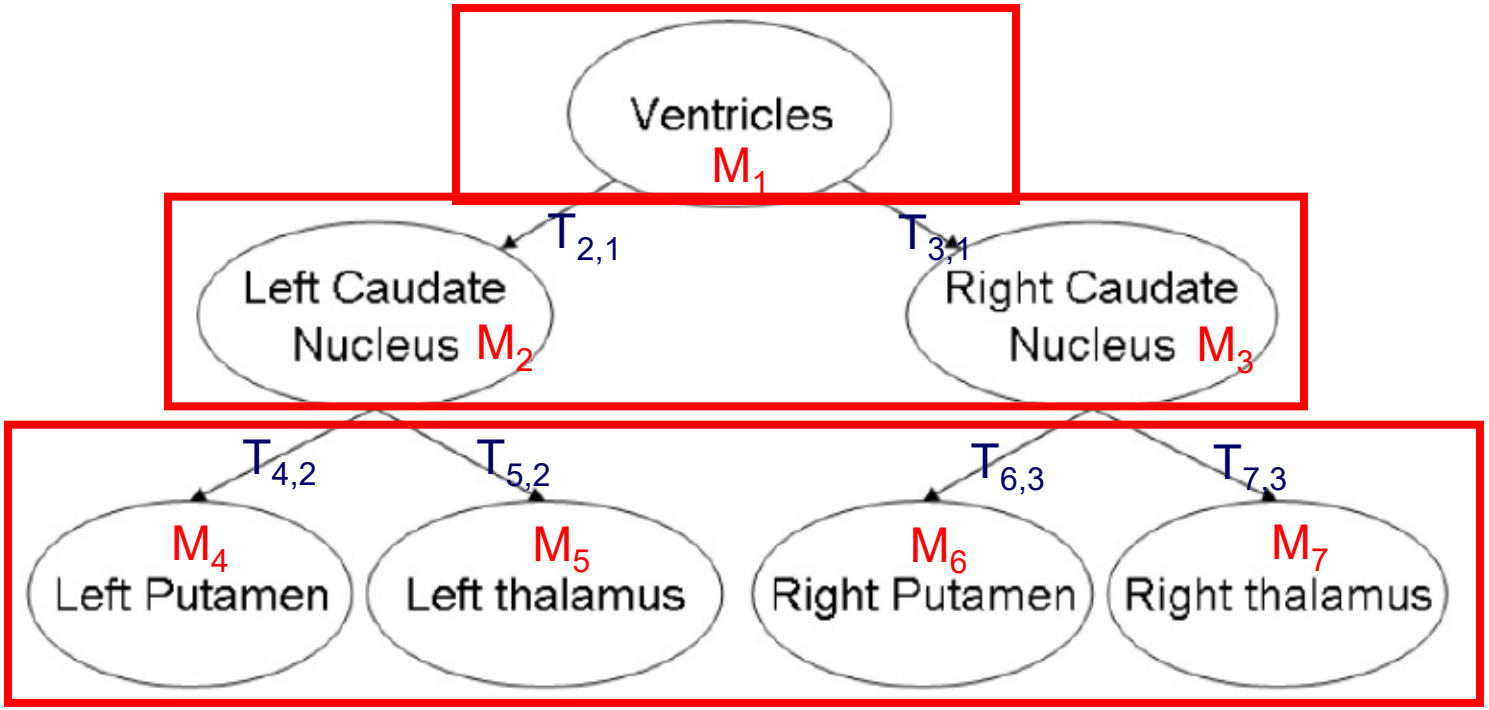
Energy for individual part
(Single Part Registration)

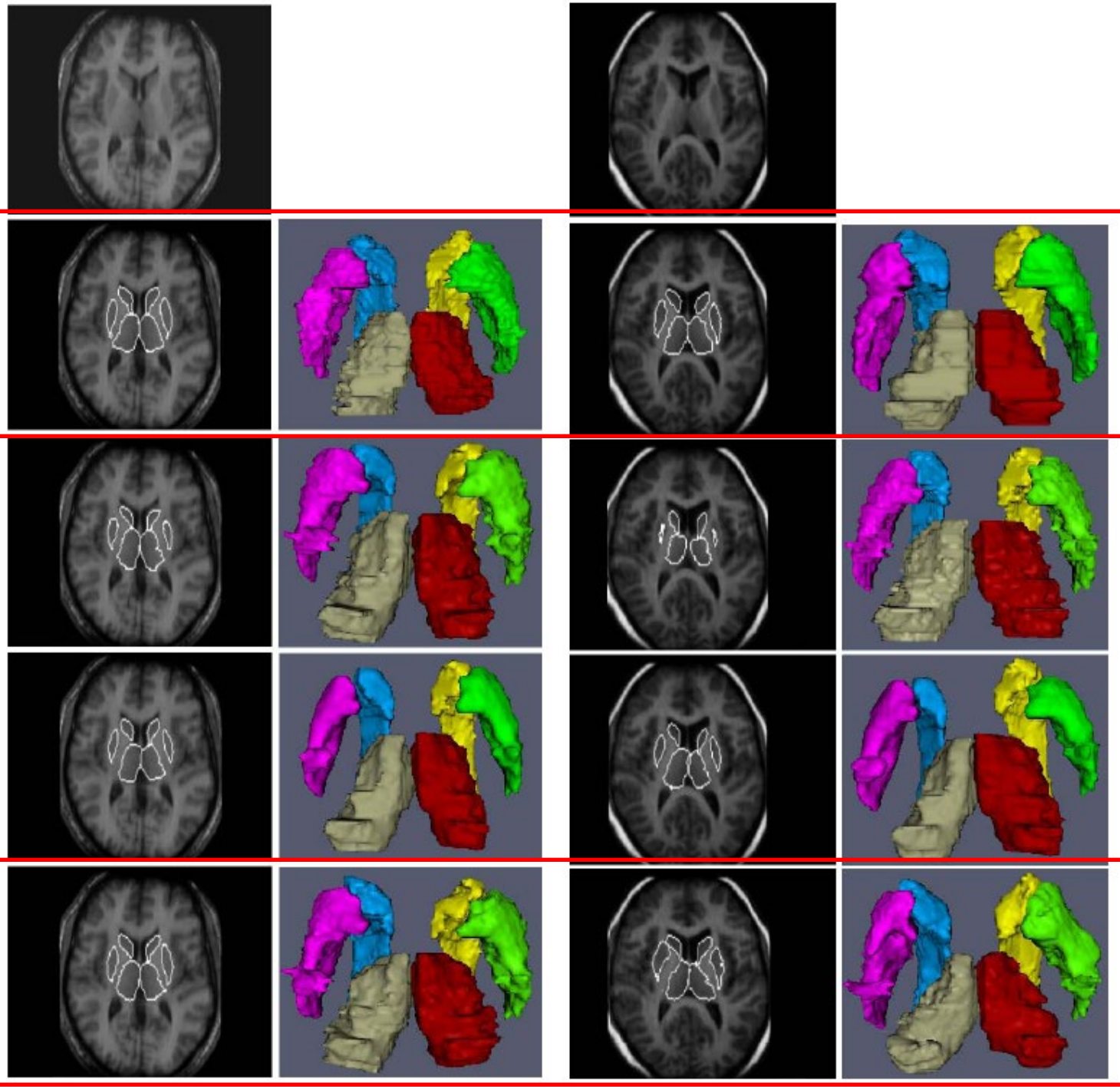
Energy for parent and child parts
(Pairwise Part Matching)



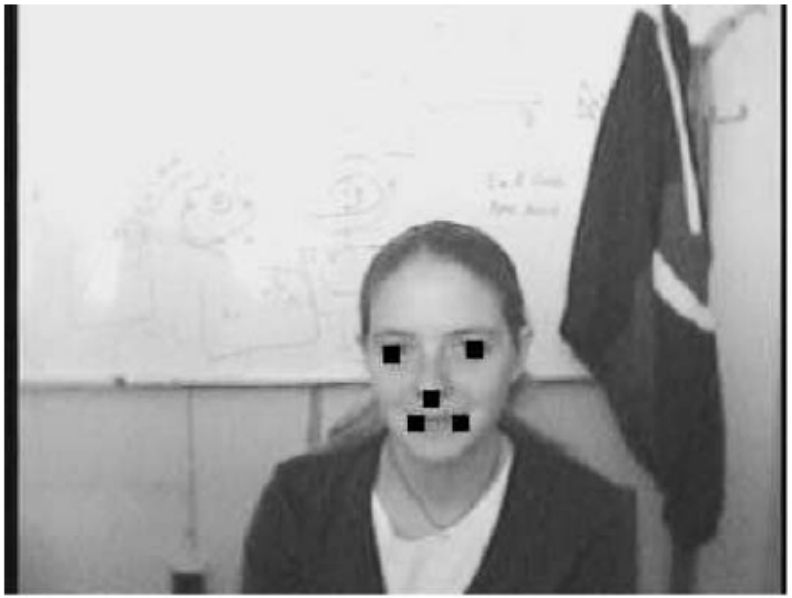
Energy Minimization Strategy

- Total energy is minimized by finding the optimal transformations for all parts, similarity transform followed by affine transform
- Top-to-down energy minimization strategy is used
- It finds the optimal transformations sequentially

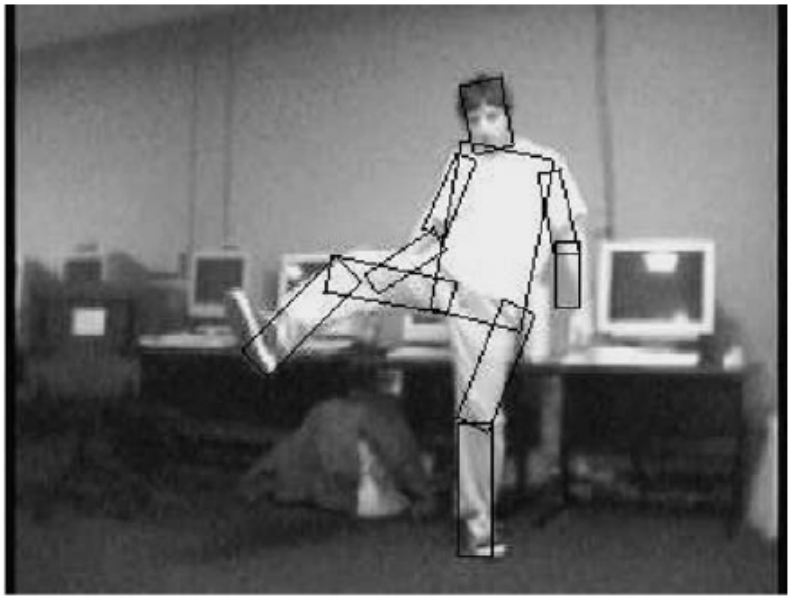



Original
Ground truth
Direct B-spline FFD
Interim results
Final results

Detection of Face and Human Body



(a)



(b)

Figure 1. Sample results for detection of a face (a); and a human body (b). Each image shows the globally best location for the corresponding object, as computed by our algorithms. The object models were learned from training examples.

$$L^* = \arg\min_L \left(\sum_{i=1}^{n} m_i(l_i) + \sum_{(v_i, v_j) \in E} d_{ij}(l_i, l_j) \right)$$

http://cs.brown.edu/~pff/index.html Felzenszwalb and Huttenlocher

Figure 3. Three examples from the first training set showing the locations of the labeled features and the structure of the learned model.

Figure 5. Matching results on occluded faces. The top row shows some input images and the bottom row shows the corresponding matching results. The MAP estimate was a good match when the faces had up to two of five parts occluded and incorrect when three parts were occluded.

Figure 4. Matching results.

Figure 6. Matching results on an image with multiple faces. See text for description.

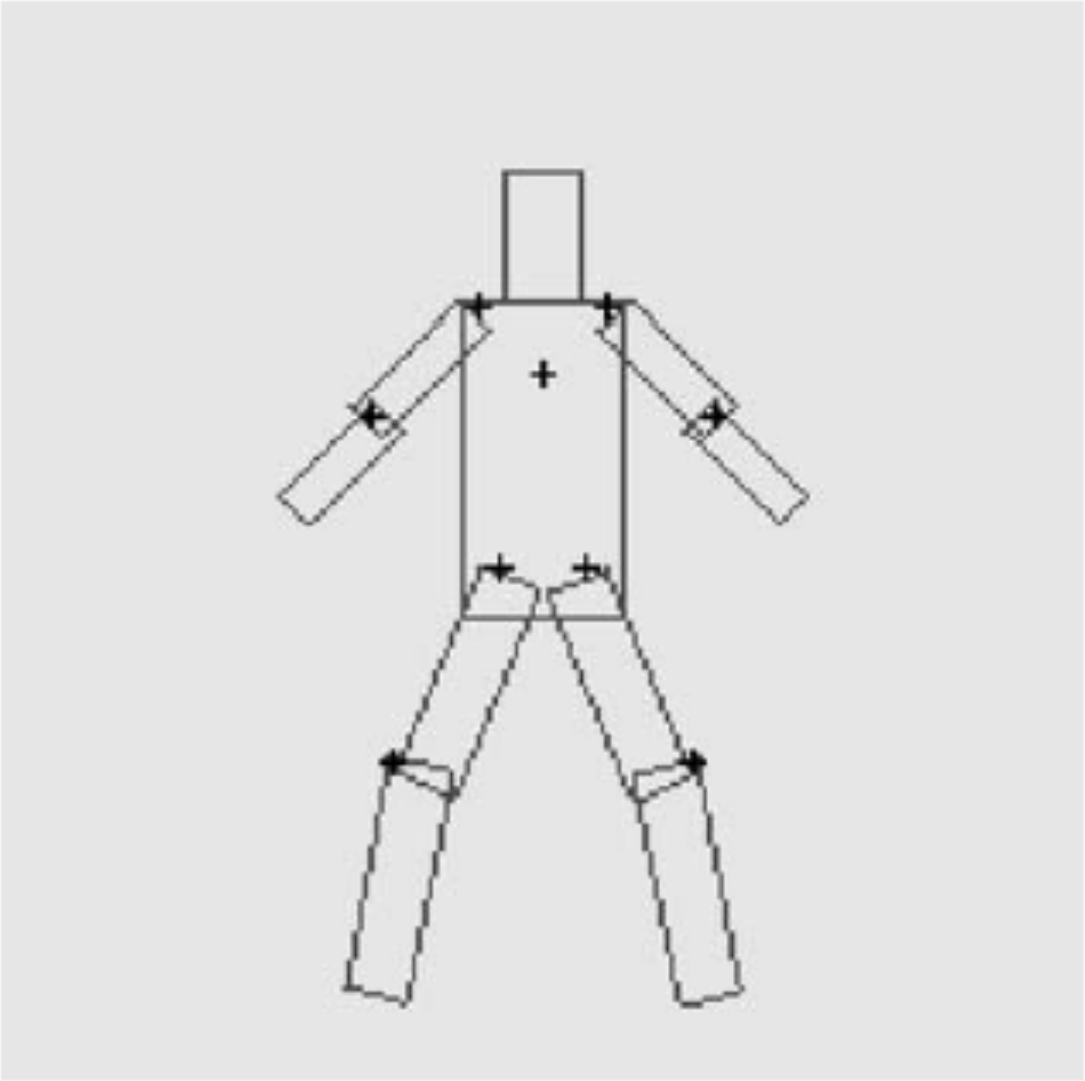


Figure 11. Human body model learned from example configurations.

Felzenszwalb and Huttenlocher

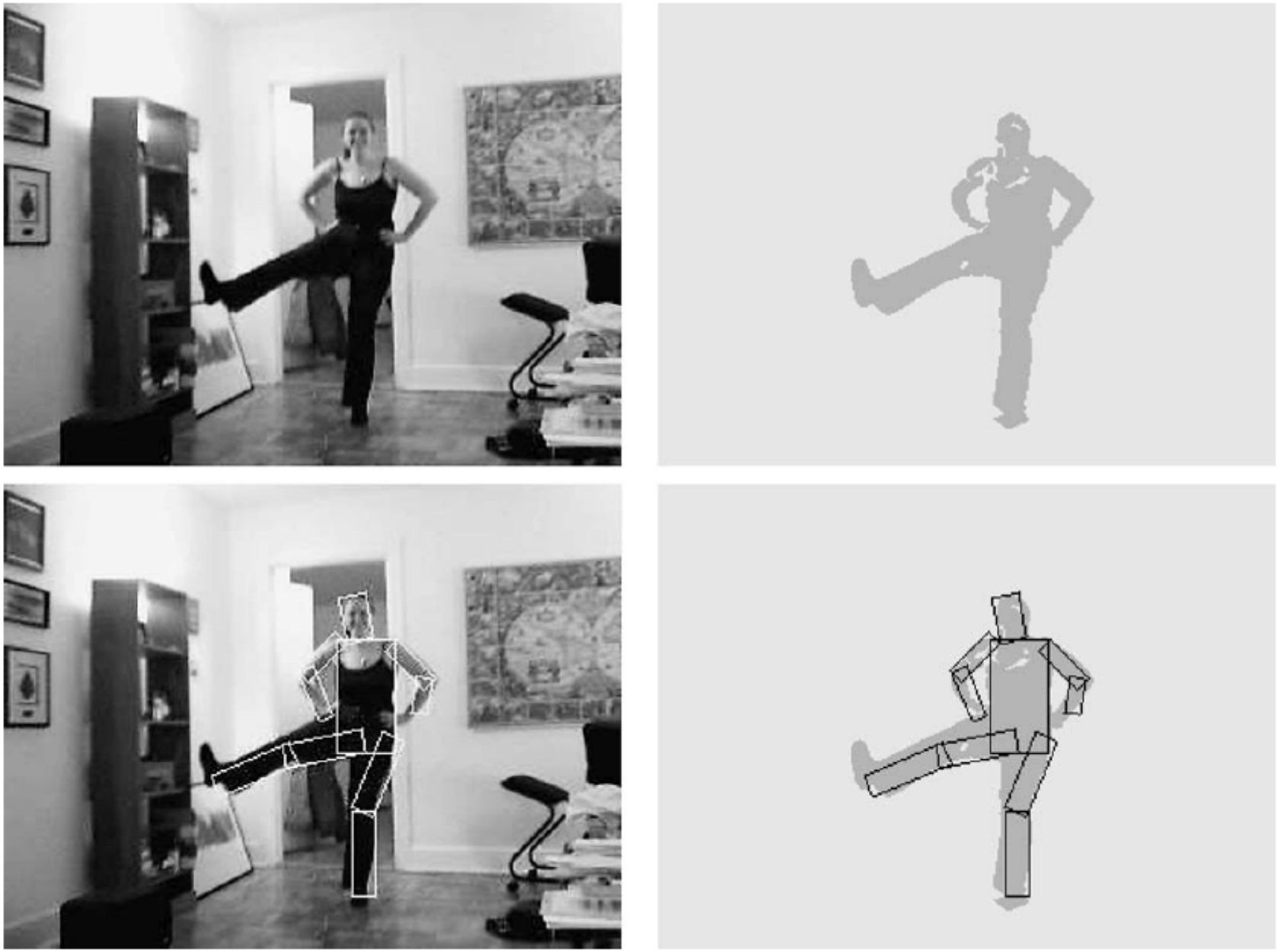


Figure 8. Input image, binary image obtained by background subtraction, and matching result superimposed on both images.

Felzenszwalb and Huttenlocher

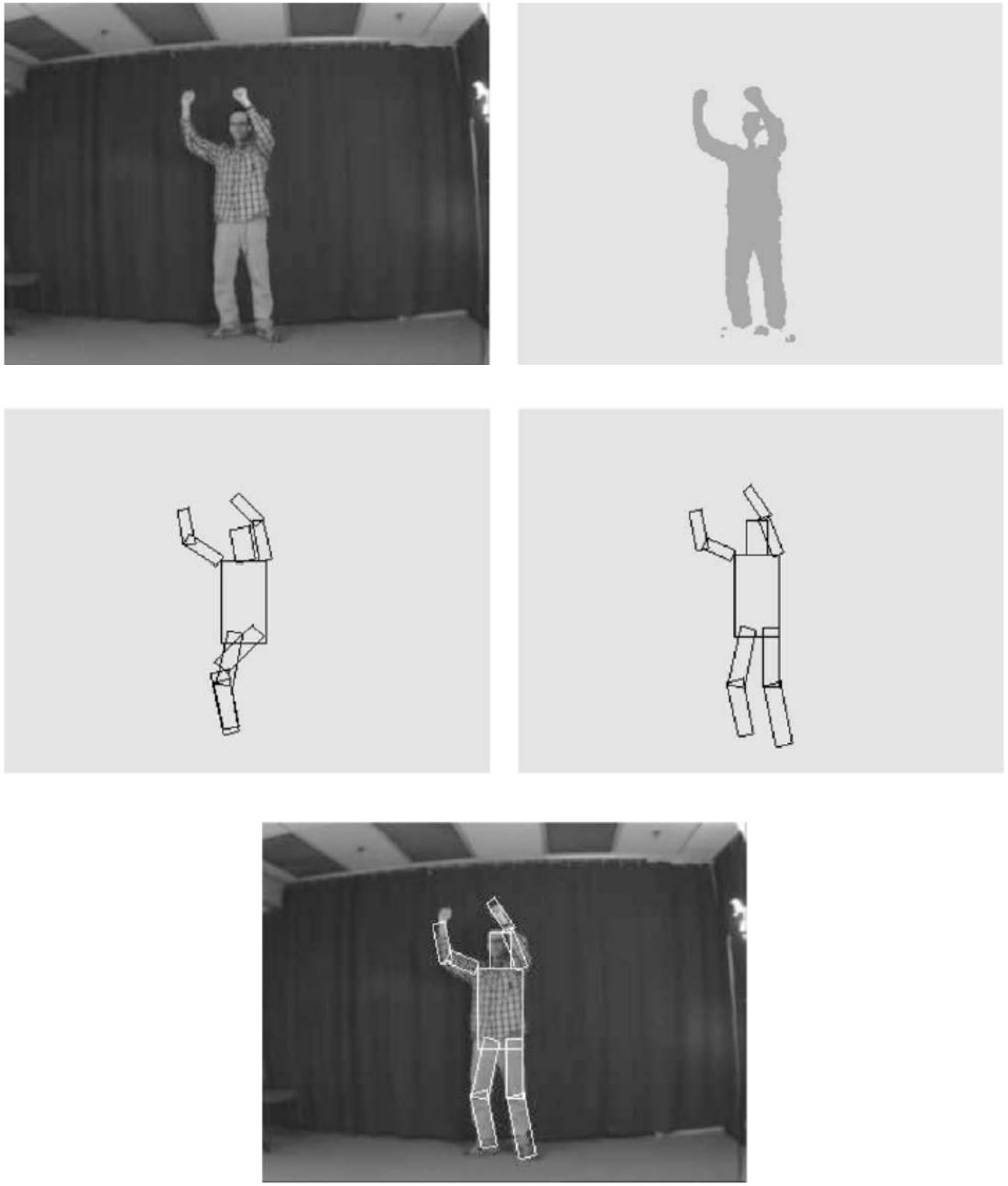


Figure 12. Input image, binary image, random samples from the posterior distribution of configurations, and best result selected using the Chamfer distance.

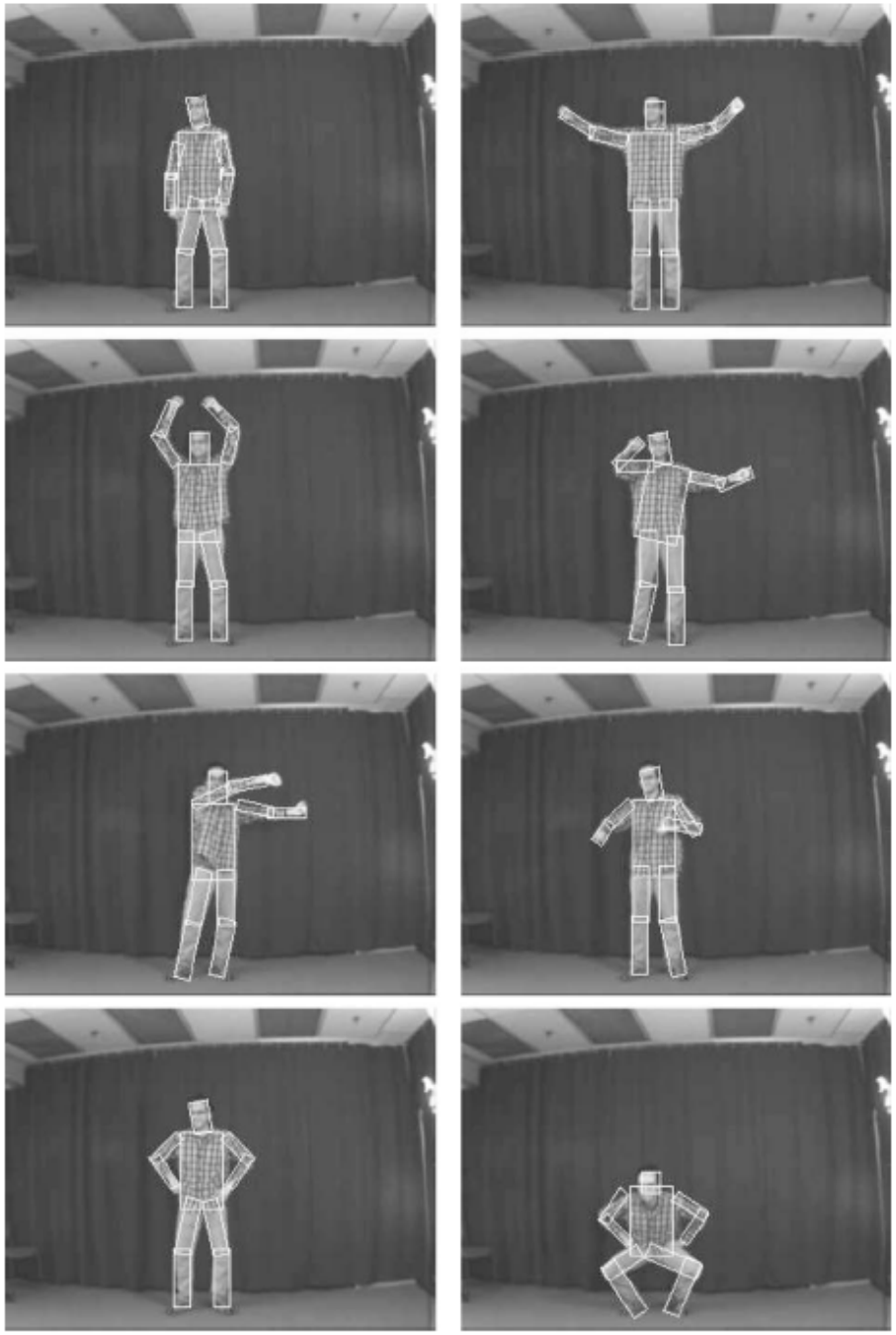


Figure 13. Matching results (sampling 200 times).

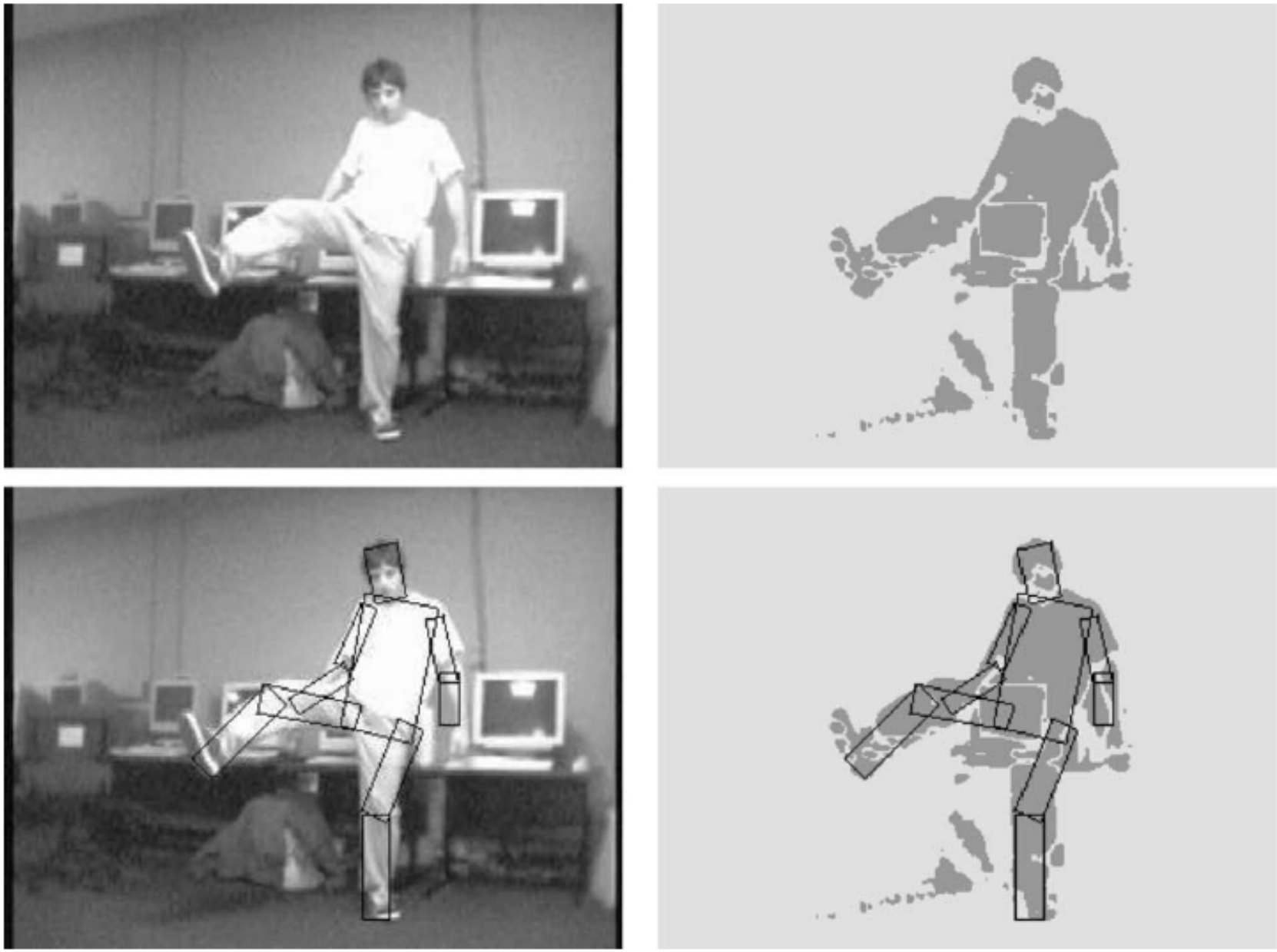


Figure 15. This example illustrates how our method works well with noisy images.

Felzenszwalb and Huttenlocher

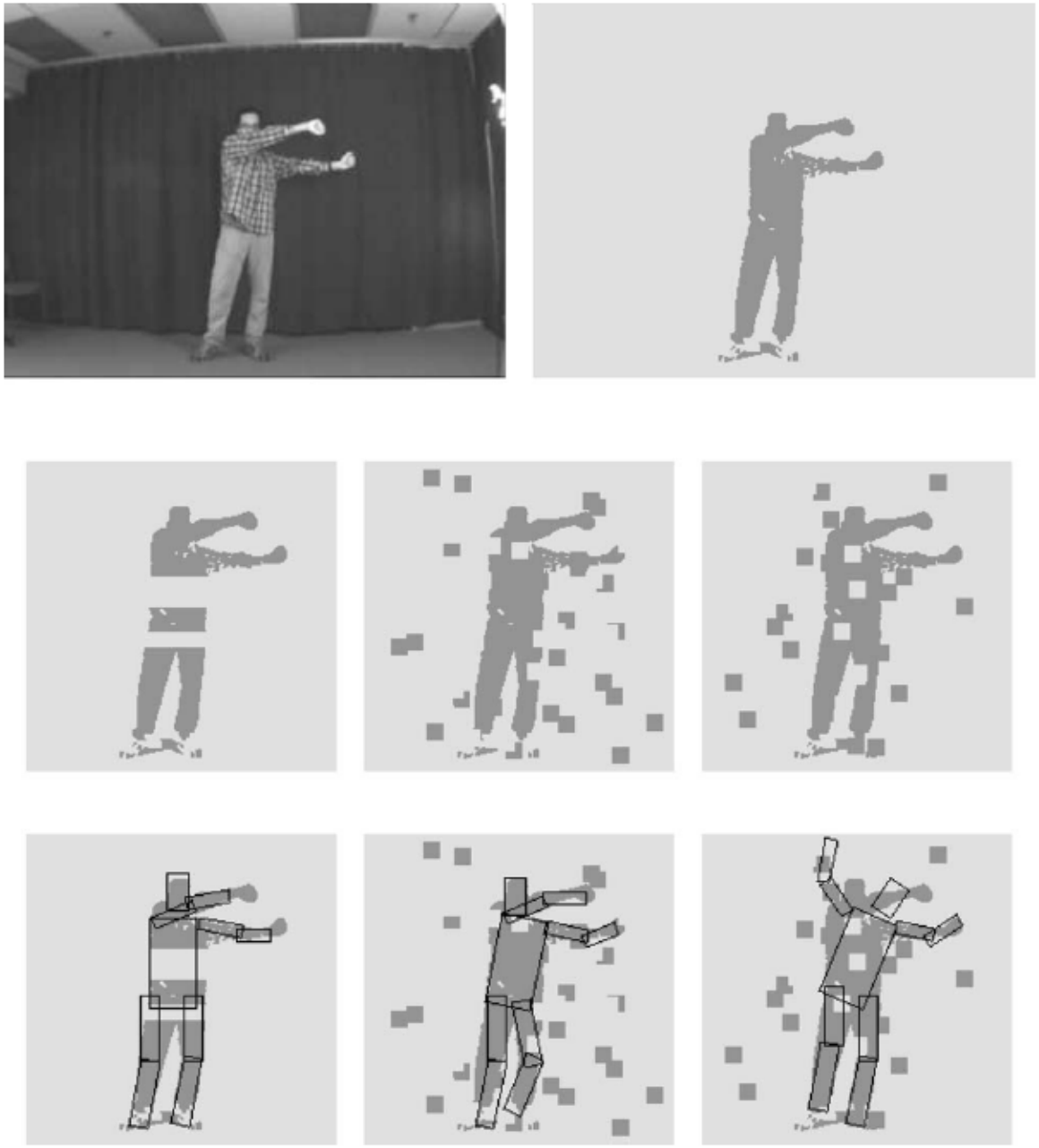


Figure 16. Matching results on corrupted images. The top row shows the original input, the middle row shows corrupted versions of the binary input and the last row shows the matching results. The first two cases demonstrate how the algorithm can handle good amounts of noise and occlusion. The third case shows an incorrect matching result.

Detection of Animals

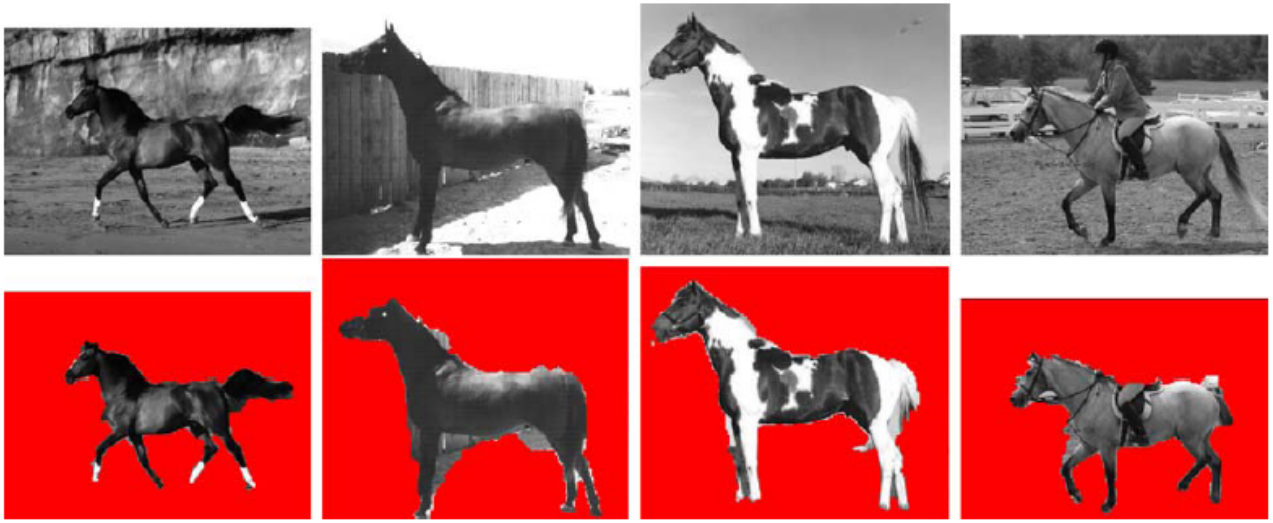


Fig. 1. Examples of horse images (1st row) and the corresponding segmentation results using the proposed method (2nd row). A brown horse may have white hoofs (1st column). The background can be easily confused with the foreground (2nd column). The skin of a horse has both slight and deep colored patches (3rd column). Part of a horse is occluded by a rider (4th column). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

$$E(P) = \sum_i M_i(p_i) + \beta_{i,j} \sum_{i,j} T_{i,j}(p_i, p_j).$$

$$\hat{P} = \arg\min_{P \in \Omega_p} E(P),$$

Wu, Cai and Chung

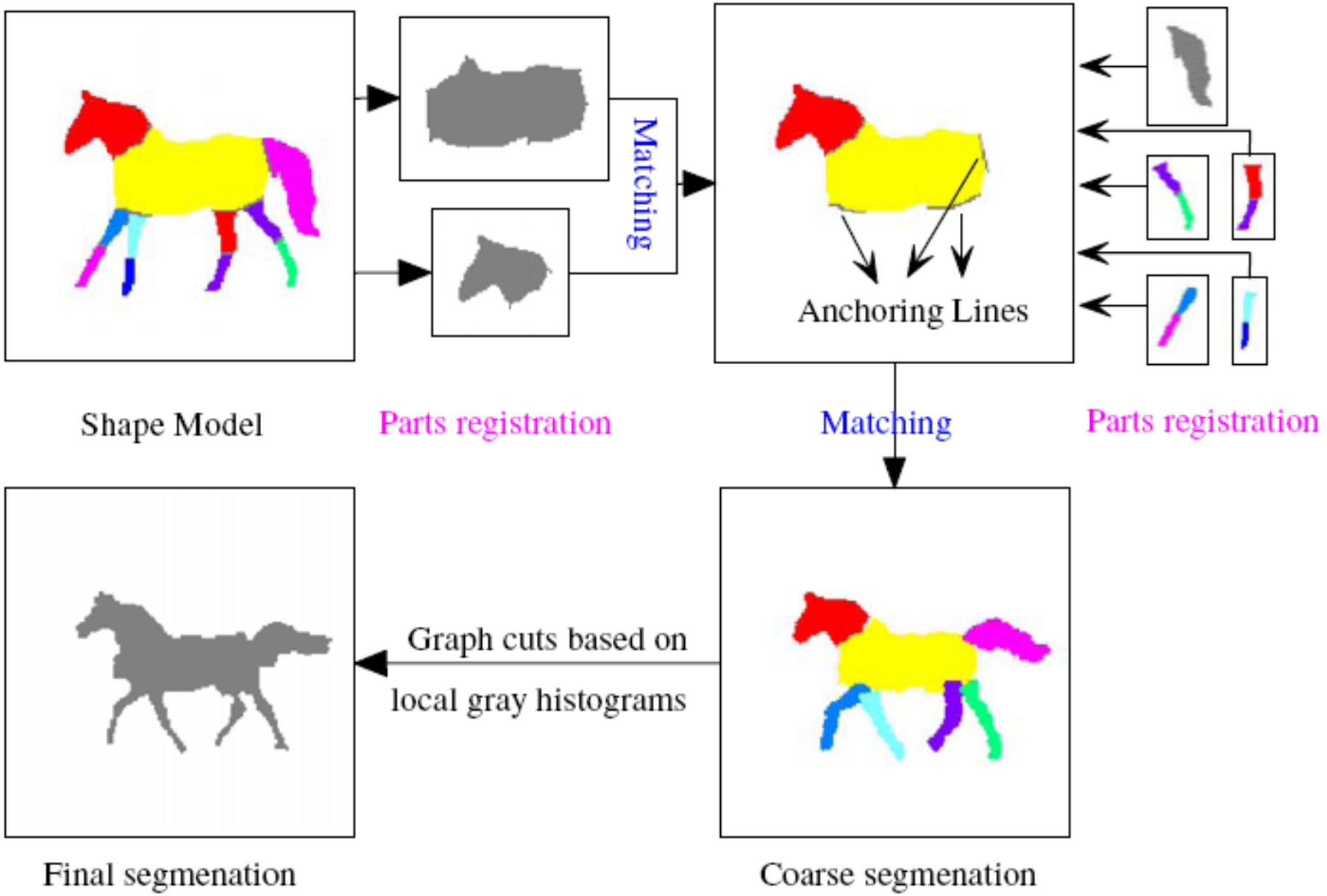


Fig. 4. (Color images) The flow chart of the instantiation of the proposed model to horse segmentation.

Wu, Cai and Chung

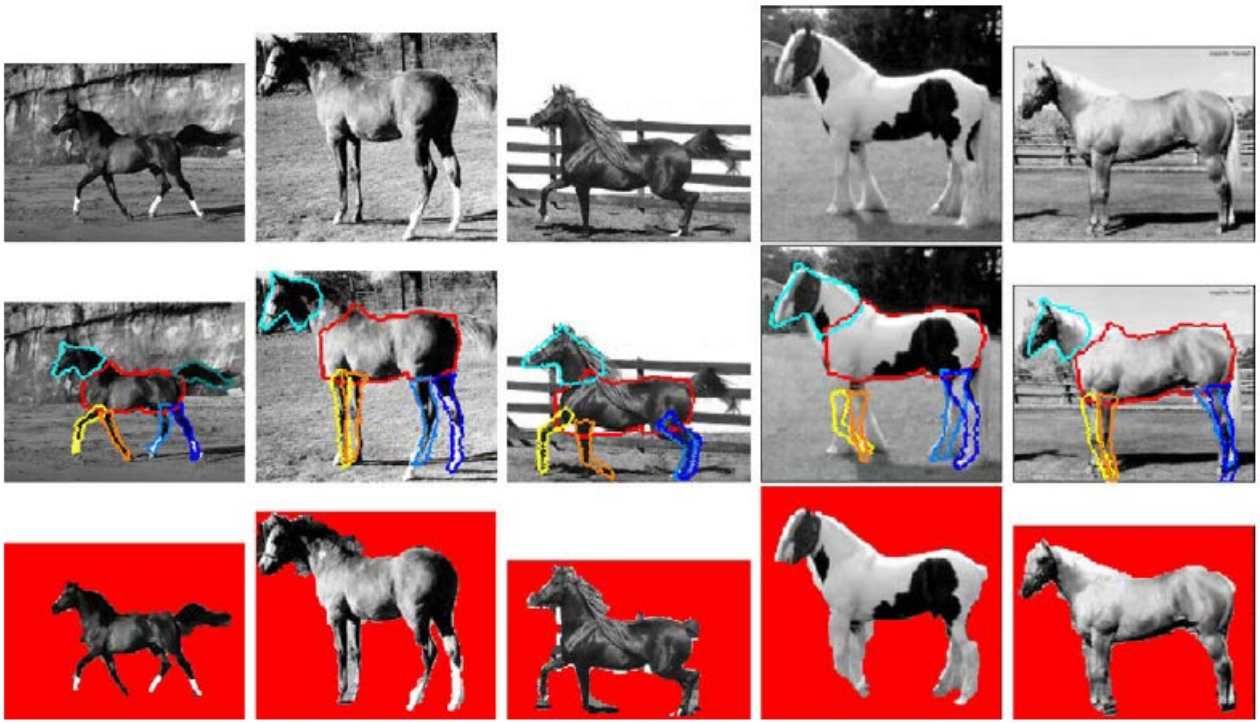


Fig. 5. (Color images) Some examples for showing the intermediate segmentation results of the proposed method. First row is the original input images. Second row is the intermediate results. They are the coarse segmentation obtained from SPR and PPI. Third row is the final segmentation results after fine-tuning.

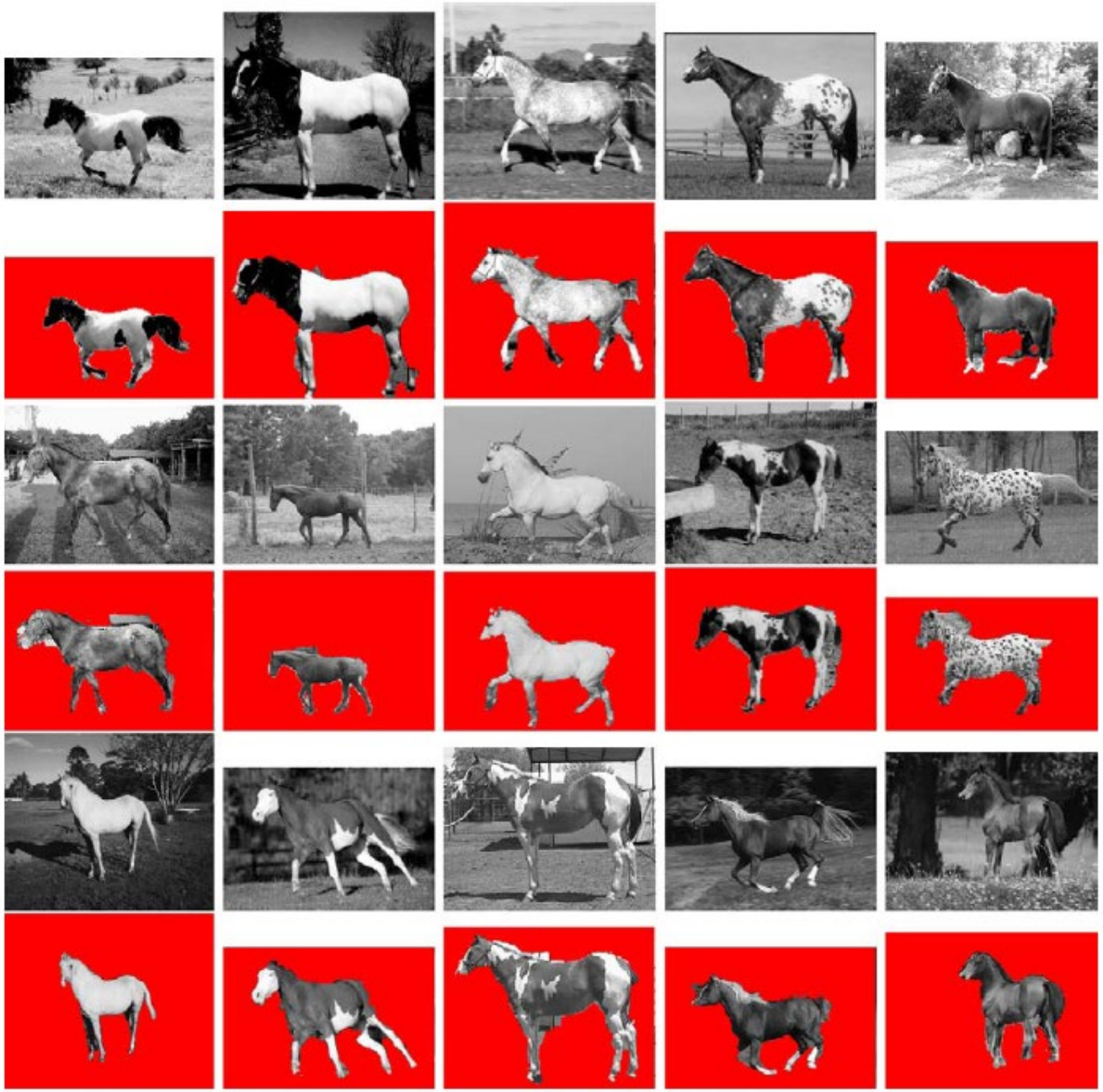


Fig. 6. More examples of the horse images and their corresponding segmentation results of the proposed method.

Wu, Cai and Chung

Fig. 7. Examples of the cow images and their corresponding segmentation results of the proposed method.

Detection of Hands and Gestures

More information can be found at

http://en.wikipedia.org/wiki/Gesture_recognition

Hand Modelling

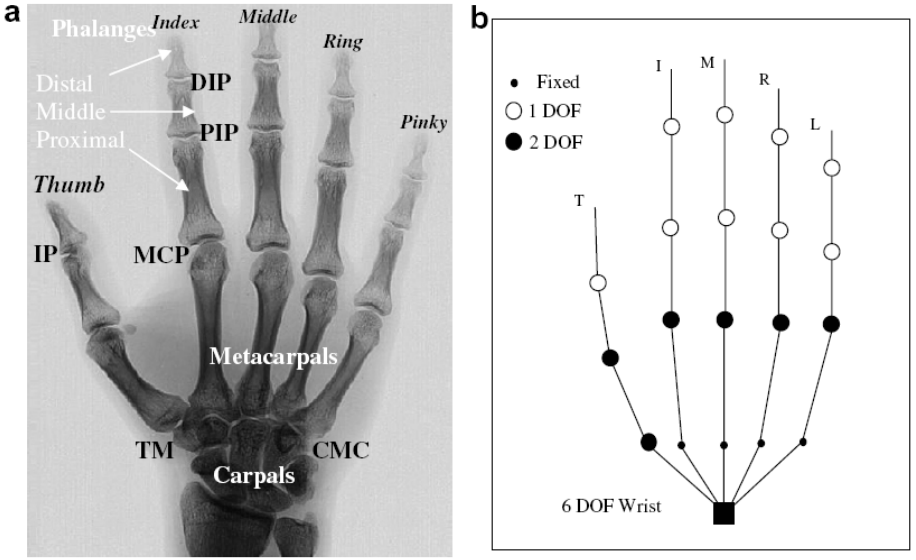


Fig. 2. Skeletal hand model: (a) Hand anatomy, (b) the kinematic model.

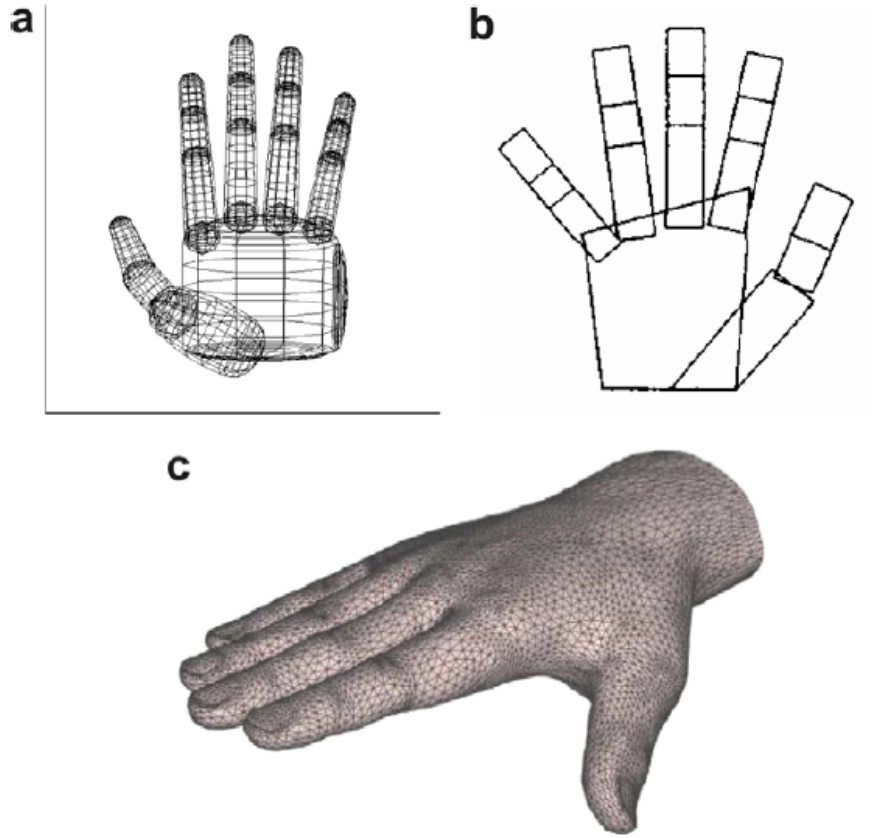


Fig. 3. Hand shape models with different complexity: (a) Quadrics-based hand model taken from [22], (b) cardboard model taken from [58], (c) a realistic hand model taken from [45].

Hand Gestures

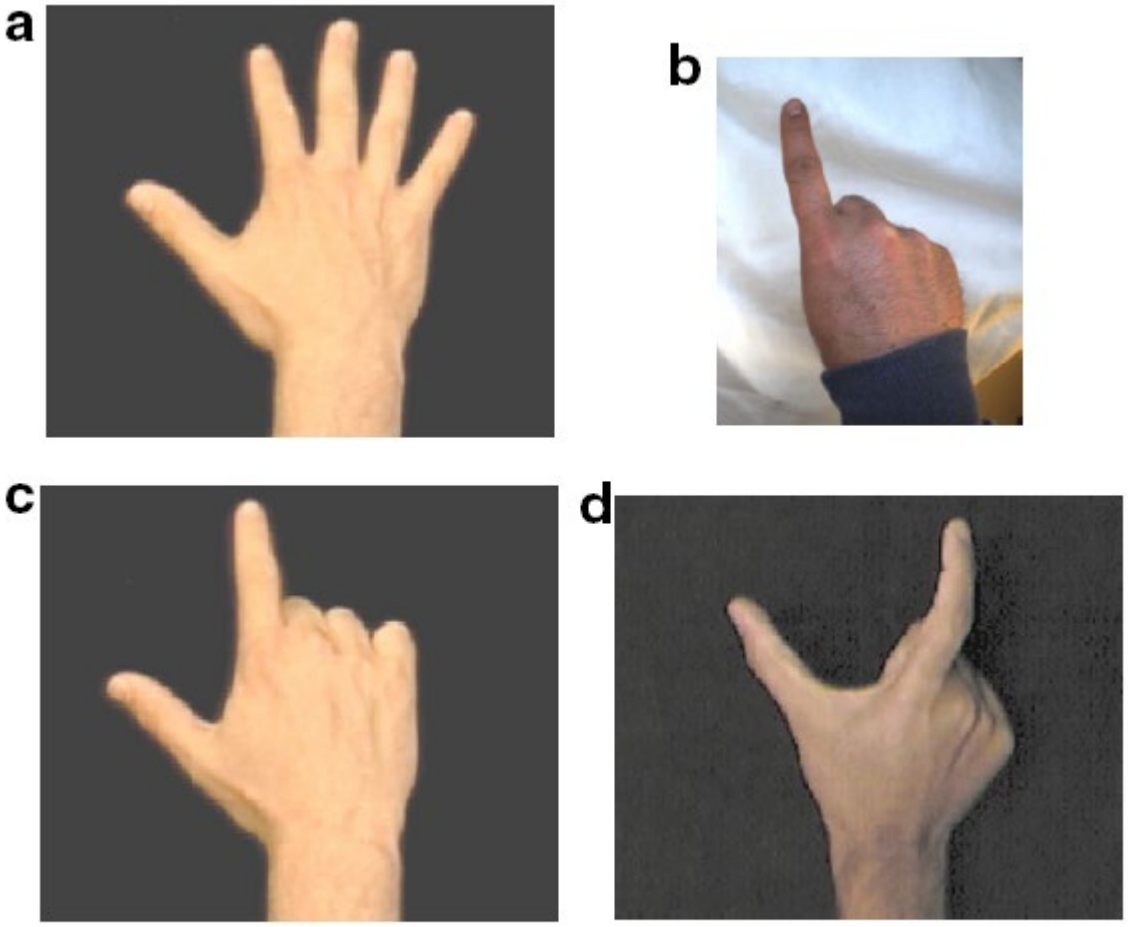


Fig. 5. Gestures used in pose estimation. (a) An open hand (e.g., used for navigation) [78], (b) pointing gesture, (c) another pointing gesture [78], (d) object manipulation gesture [78].

Erol, Bebis, Nicolescu, Boyle and Twombly

Examples

Fig. 7. A representative experimental output demonstrating the performance of pose estimation (from [57]).

Erol, Bebis, Nicolescu, Boyle and Twombly

Examples

Fig. 8. A demonstration of pointing gesture tracking (from [22]).

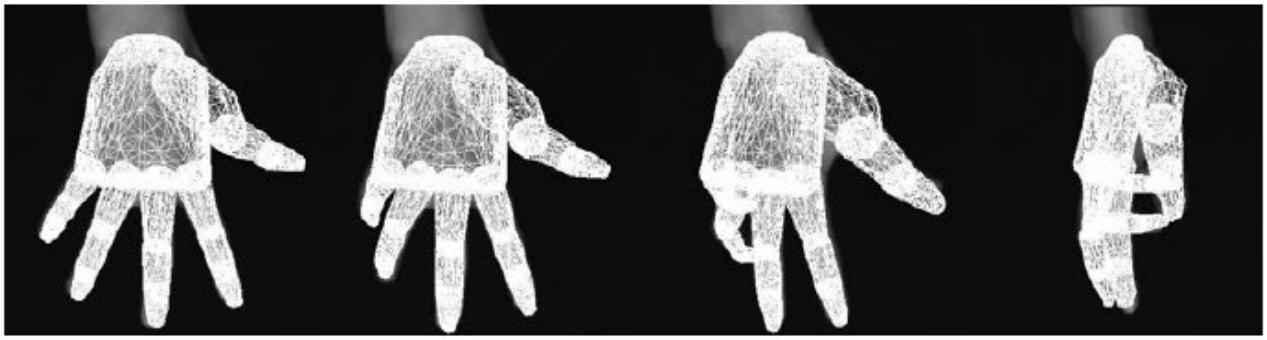


Fig. 9. An example of tracking results containing considerable amount of occlusions and out of plane rotations (from [117]).

Erol, Bebis, Nicolescu, Boyle and Twombly