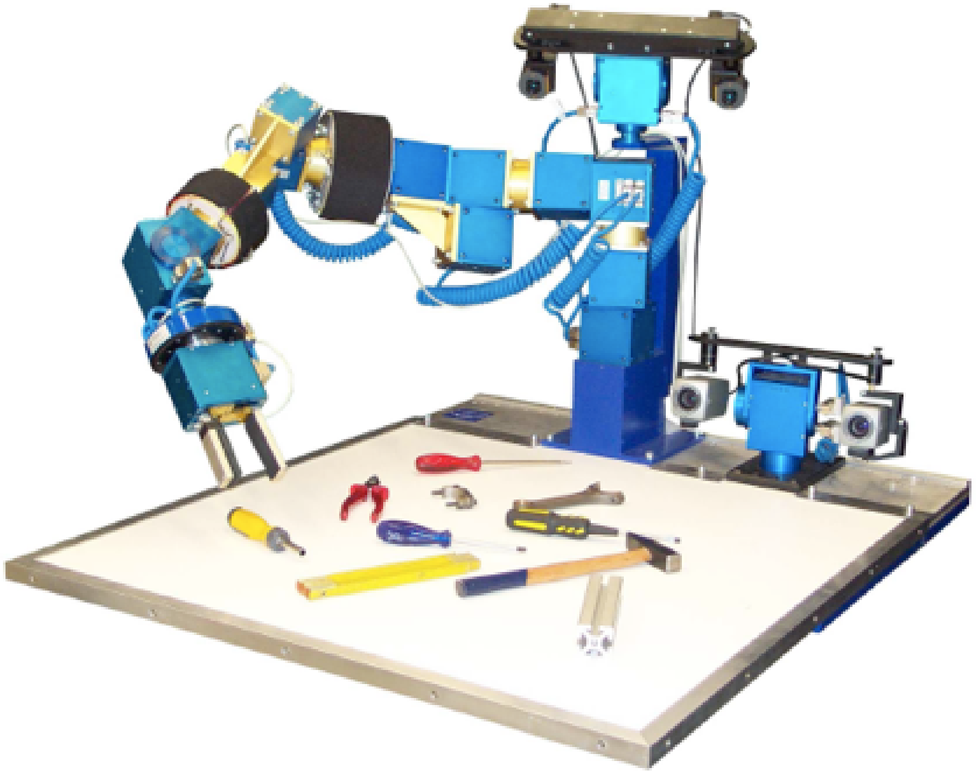
Object recognition in Dynamic Field Theory

Oliver Lomp – Neural Dynamics Summer School 2014 – 27. 8. 2014

Why We Need Object Recognition



- Complexity of Interaction
- Object knowledge

What is difficult about object recognition?

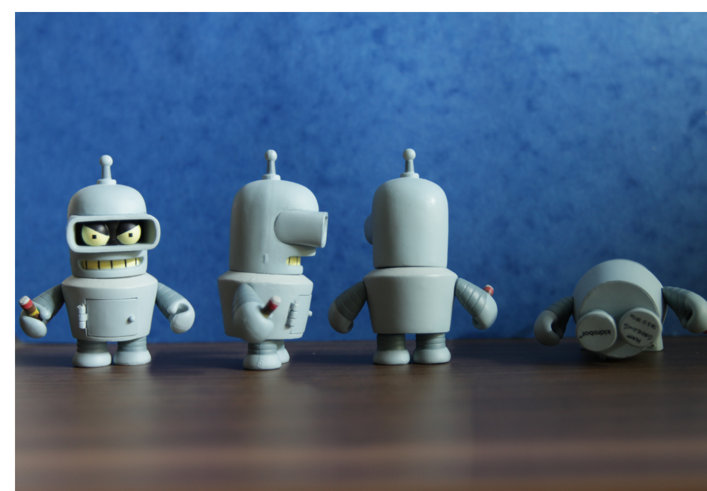
What is difficult about object recognition?

What is difficult about object recognition?









What is difficult about object recognition?

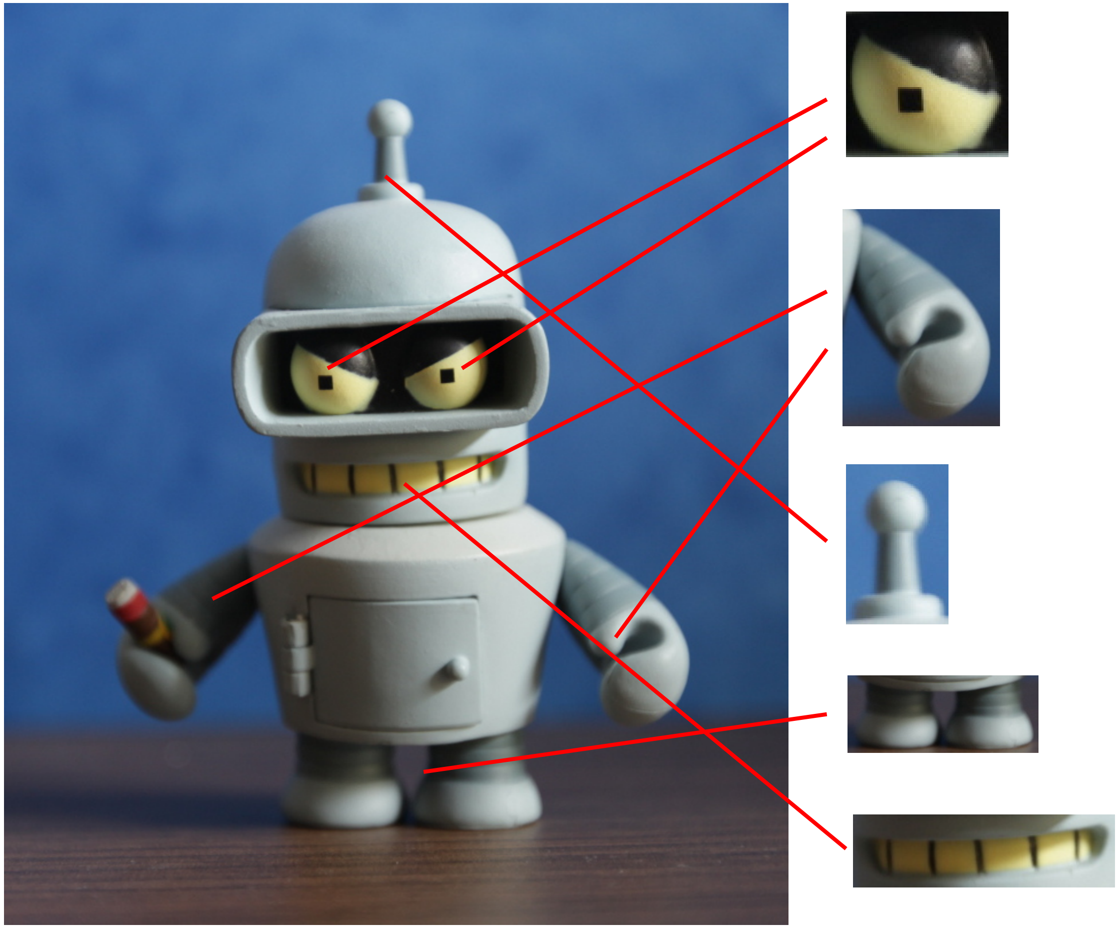
- 2D image of a 3D object
- Features vary with each new view

  → same object never looks the same

Common Solution: Invariance

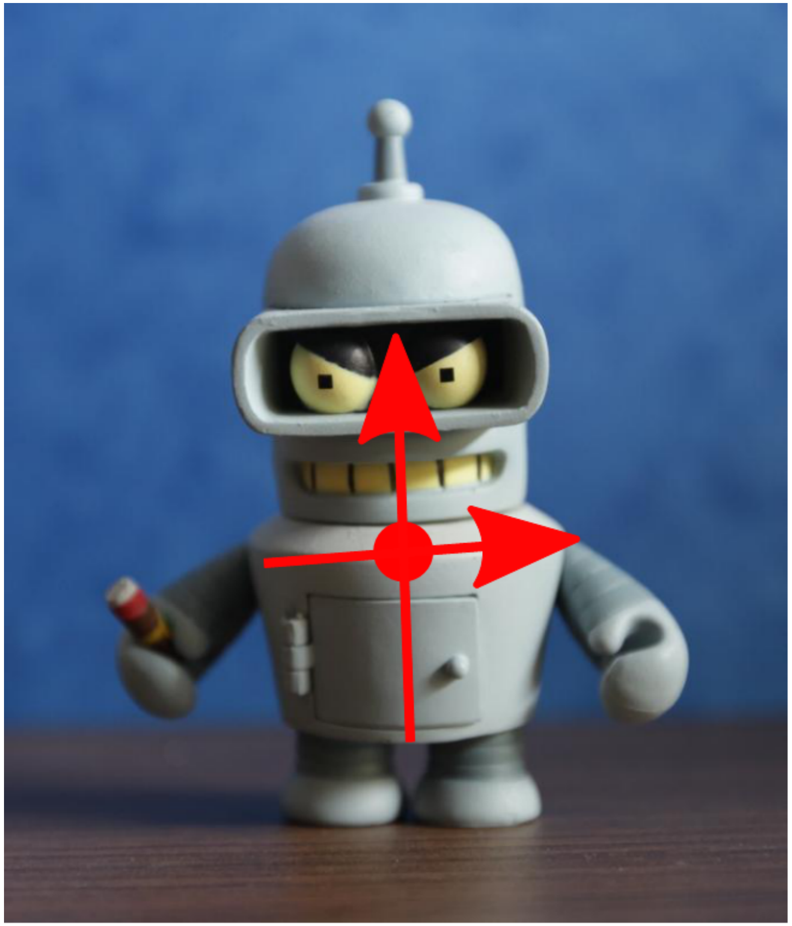
- Trade-off: invariance vs. discriminance
- Invariance loses information

  → Know what, but not necessarily where



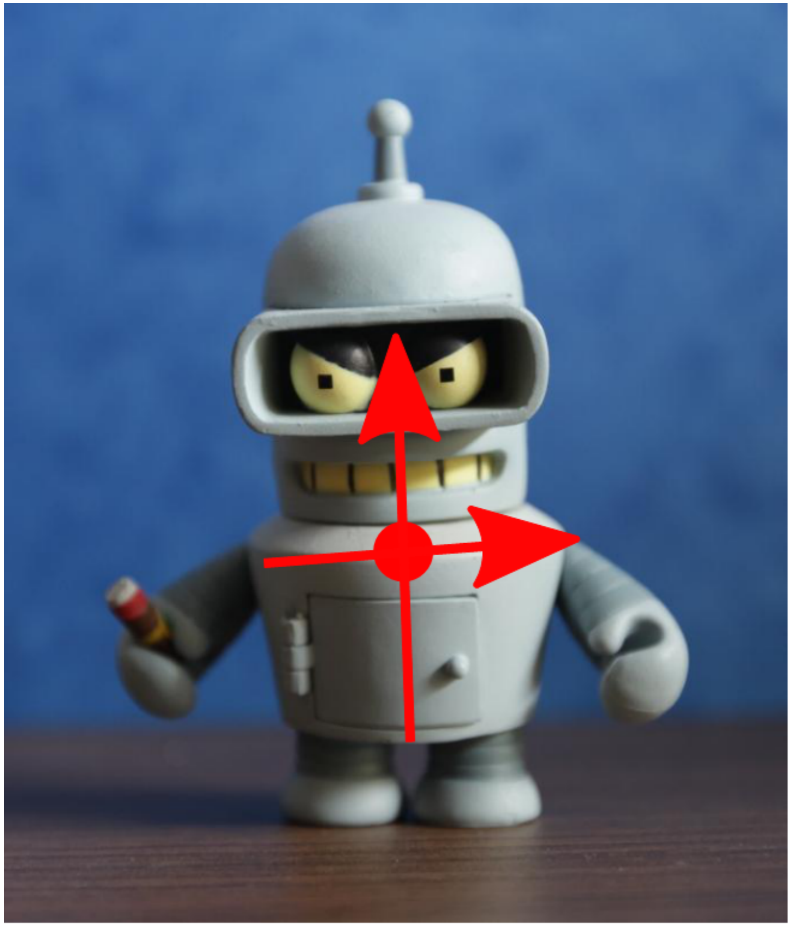
Object pose

- Pose: orientation, location, size of object (in world or image)
  - Can be useful (e.g., for grasping)
  - Can be a clue for background segmentation



A different approach

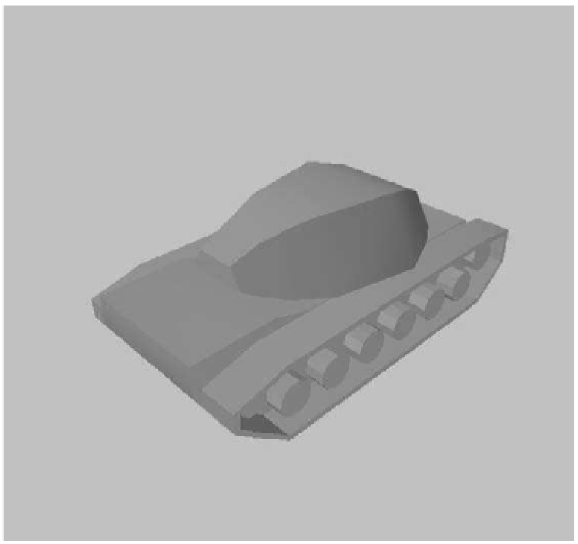
- Instead of pose-invariant features: estimate pose, use it



A DFT implementation of Arathorn's map-seeking circuits

Arathorn's map-seeking circuits

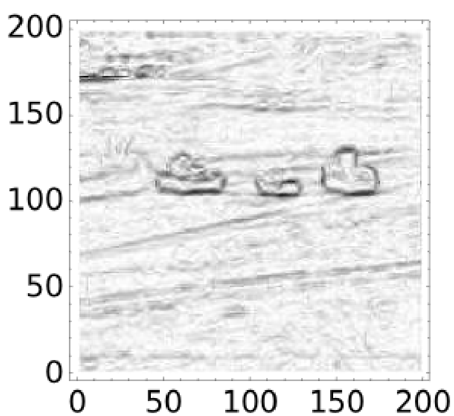
from Arathorn (2004)



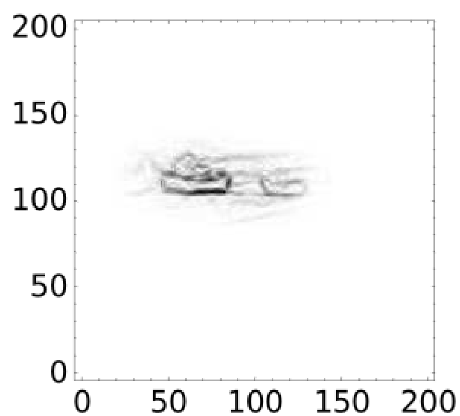
(a) 3D model of M-60 tank

(b) source image

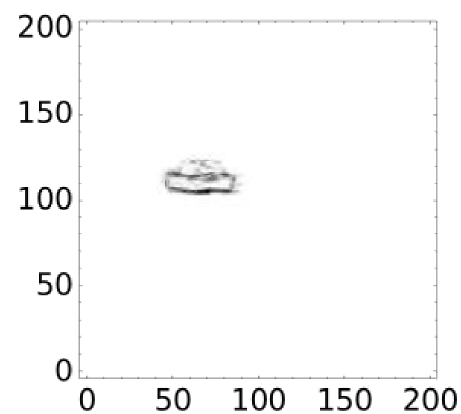
(c) input image - blurred



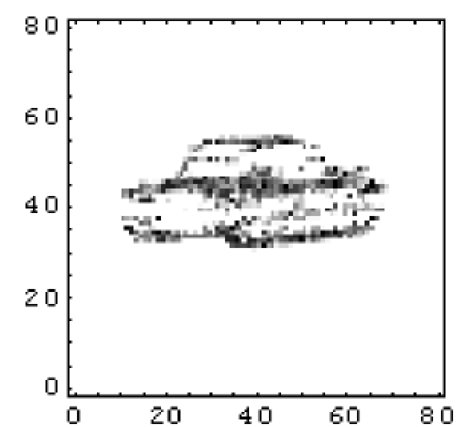
(d) iter 1



(d) iter 3



(e) iter 20



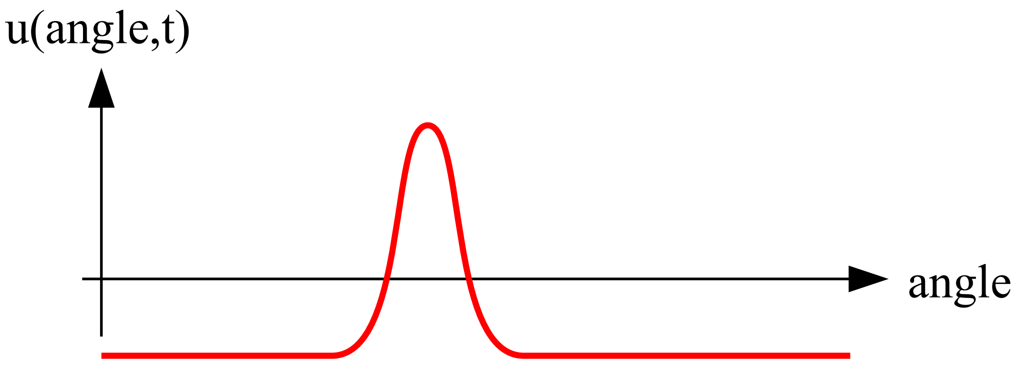
(g) final model pose

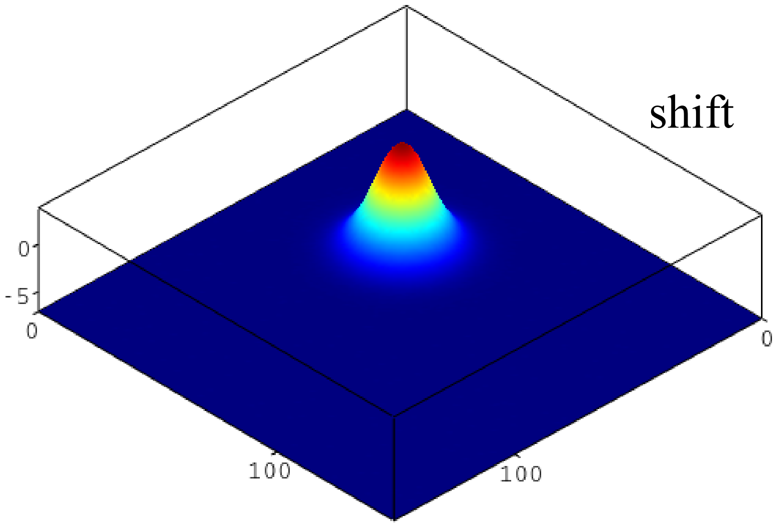
Pose parameter encoding

- How can pose be represented?

  → Space code

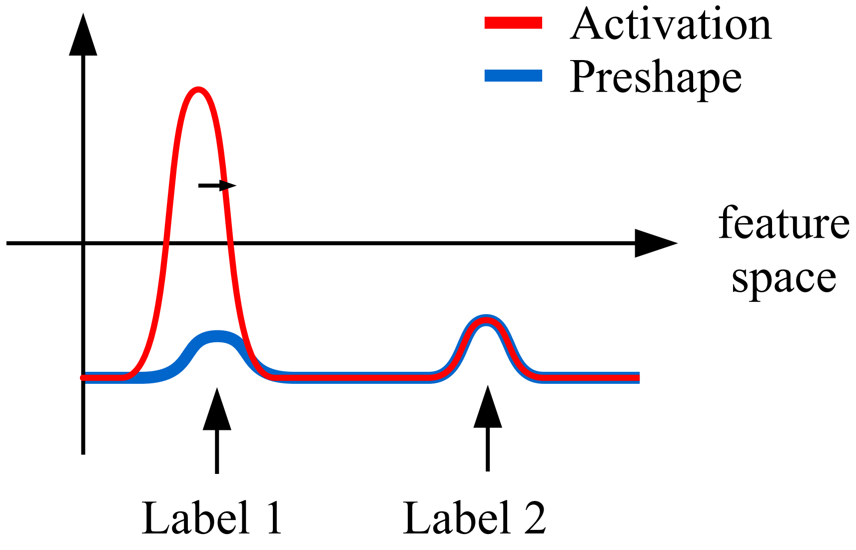
- Example: position as 2d peak, rotation as 1d peak



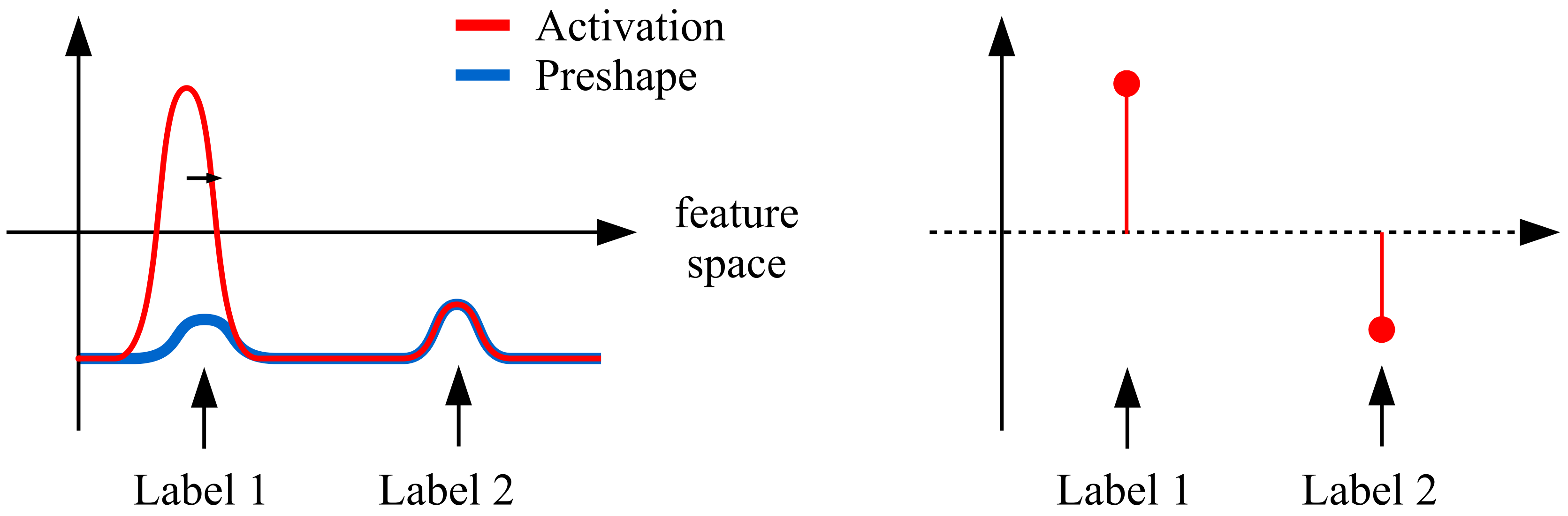


Label encoding

- What is the output of recognition?
  - Categorization decision (binary or graded response)
  - Conflicts with continuous nature of fields
- Categorization in dynamic neural fields



Label fields



- Discrete nodes (cf. grandmother neurons)
- Only global interaction / no metric
- Activity corresponds to presence of label

The principle: 1D shift

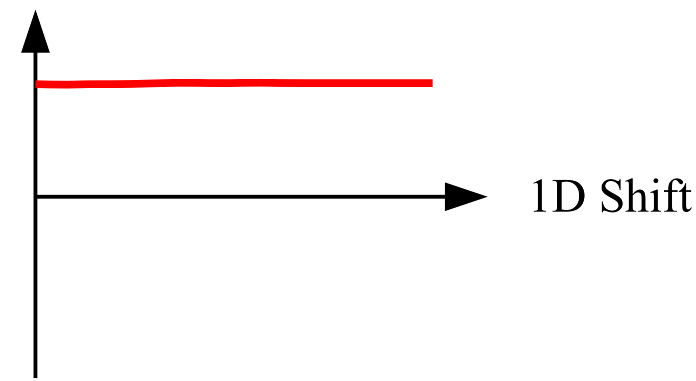
Estimating (horizontal) shift

view

input

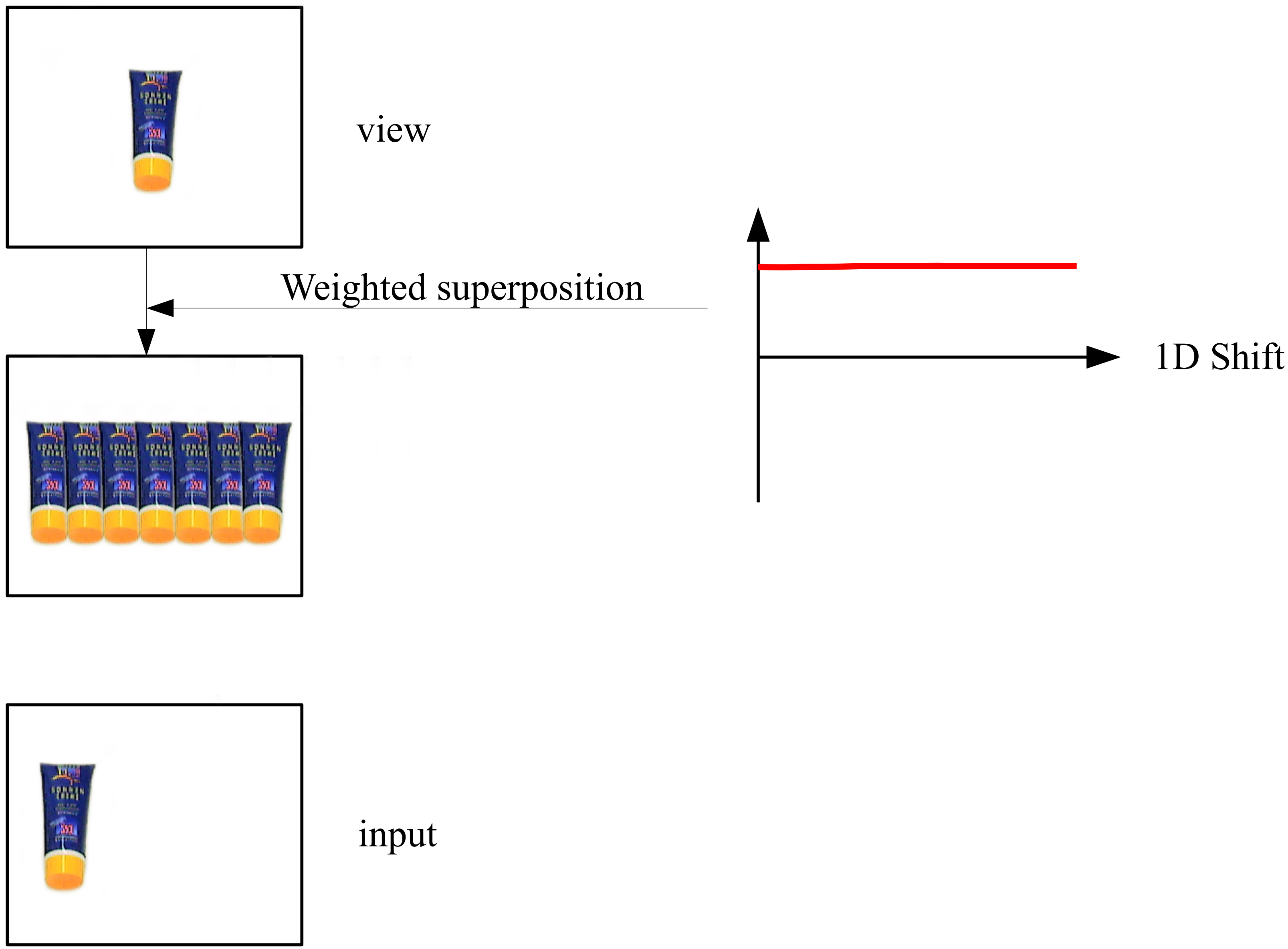
Estimating (horizontal) shift

view

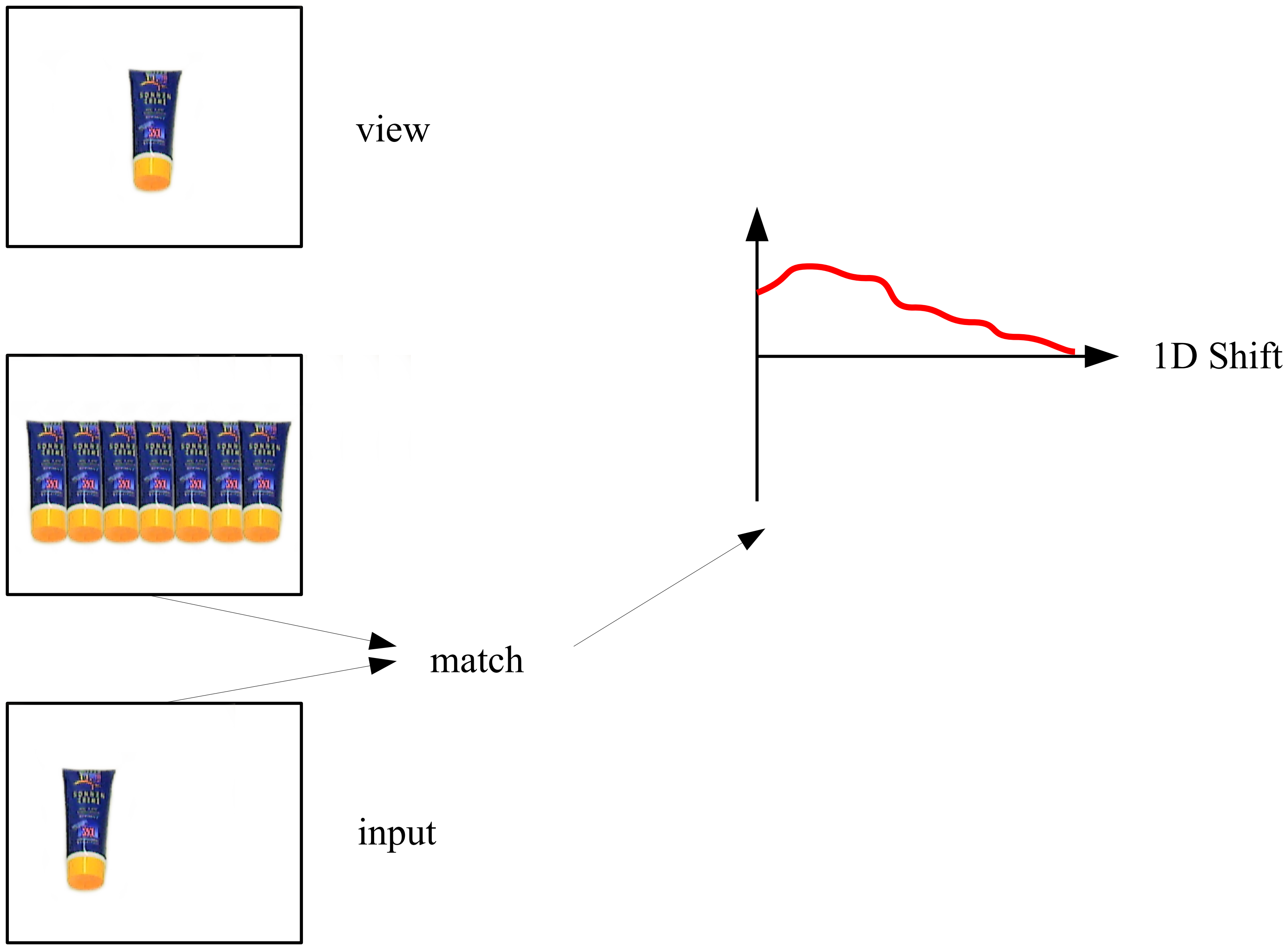


input

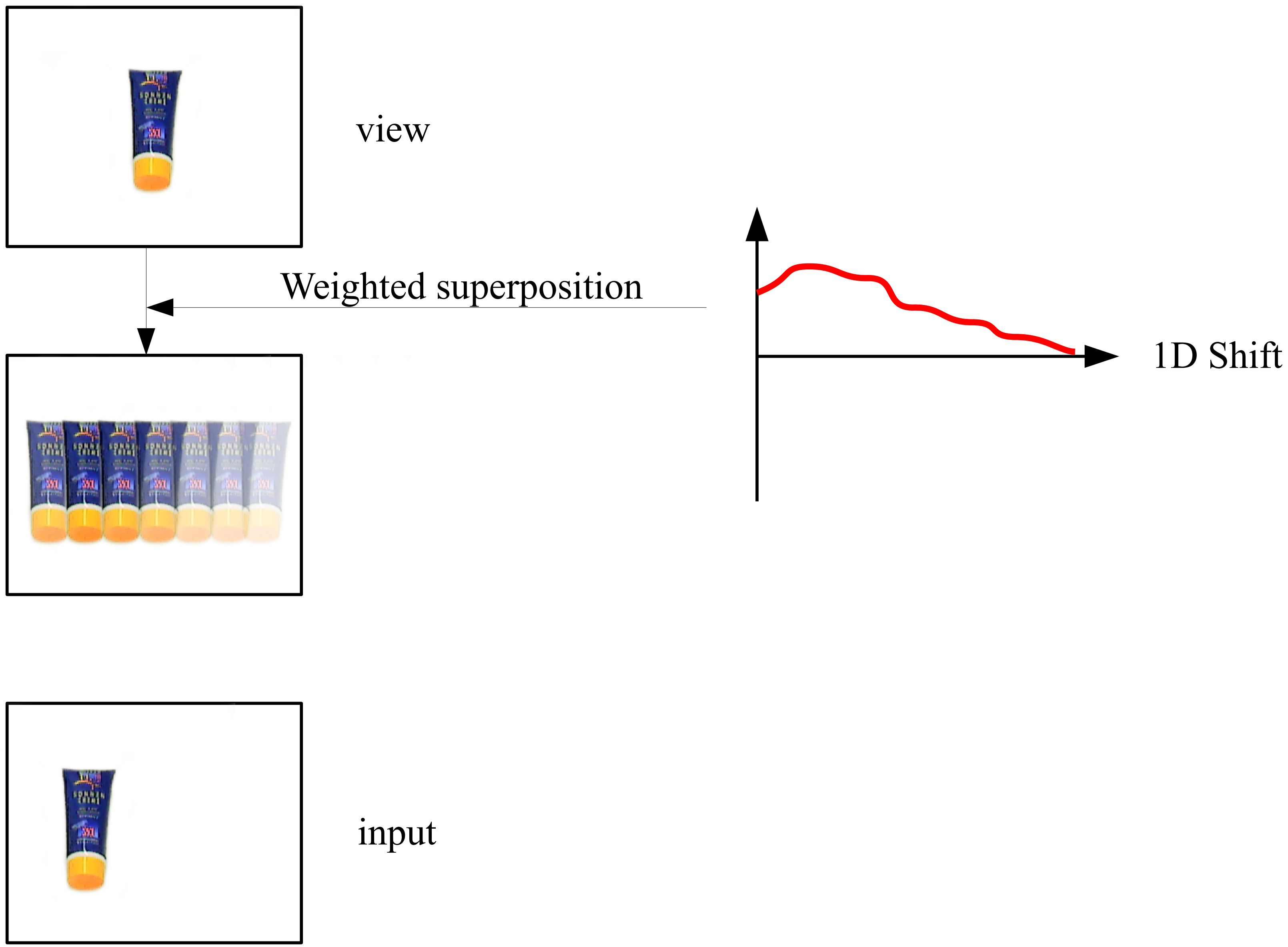
Estimating (horizontal) shift



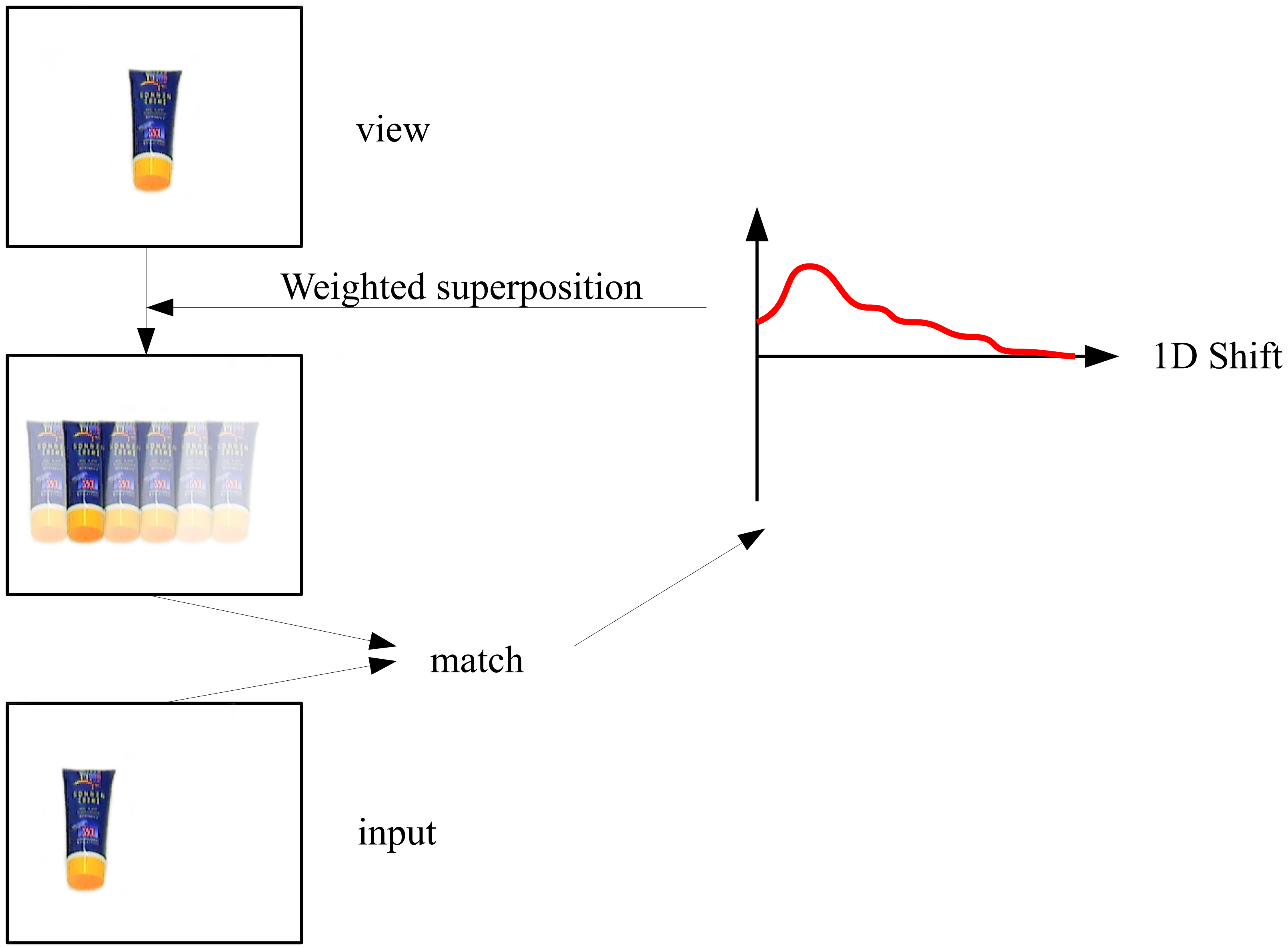
Estimating (horizontal) shift



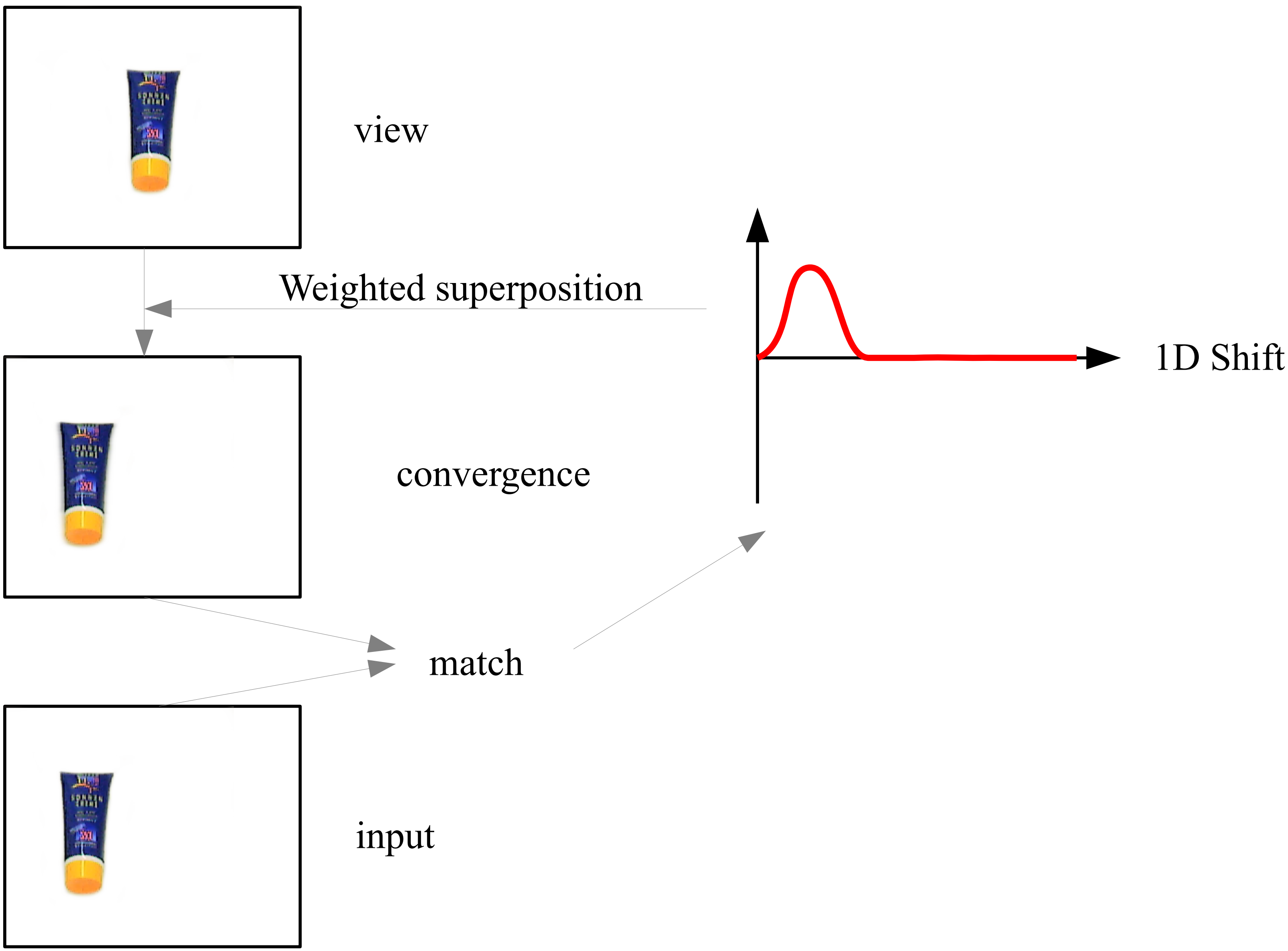
Estimating (horizontal) shift



Estimating (horizontal) shift

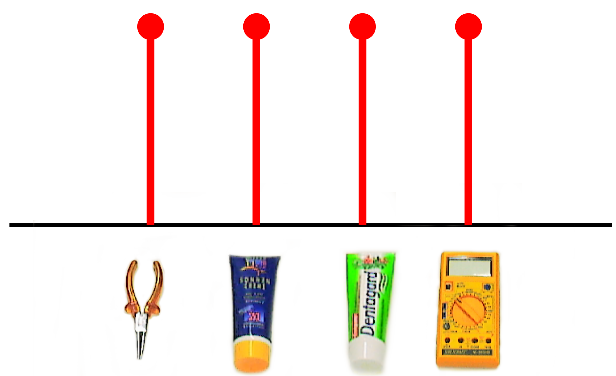


Estimating (horizontal) shift



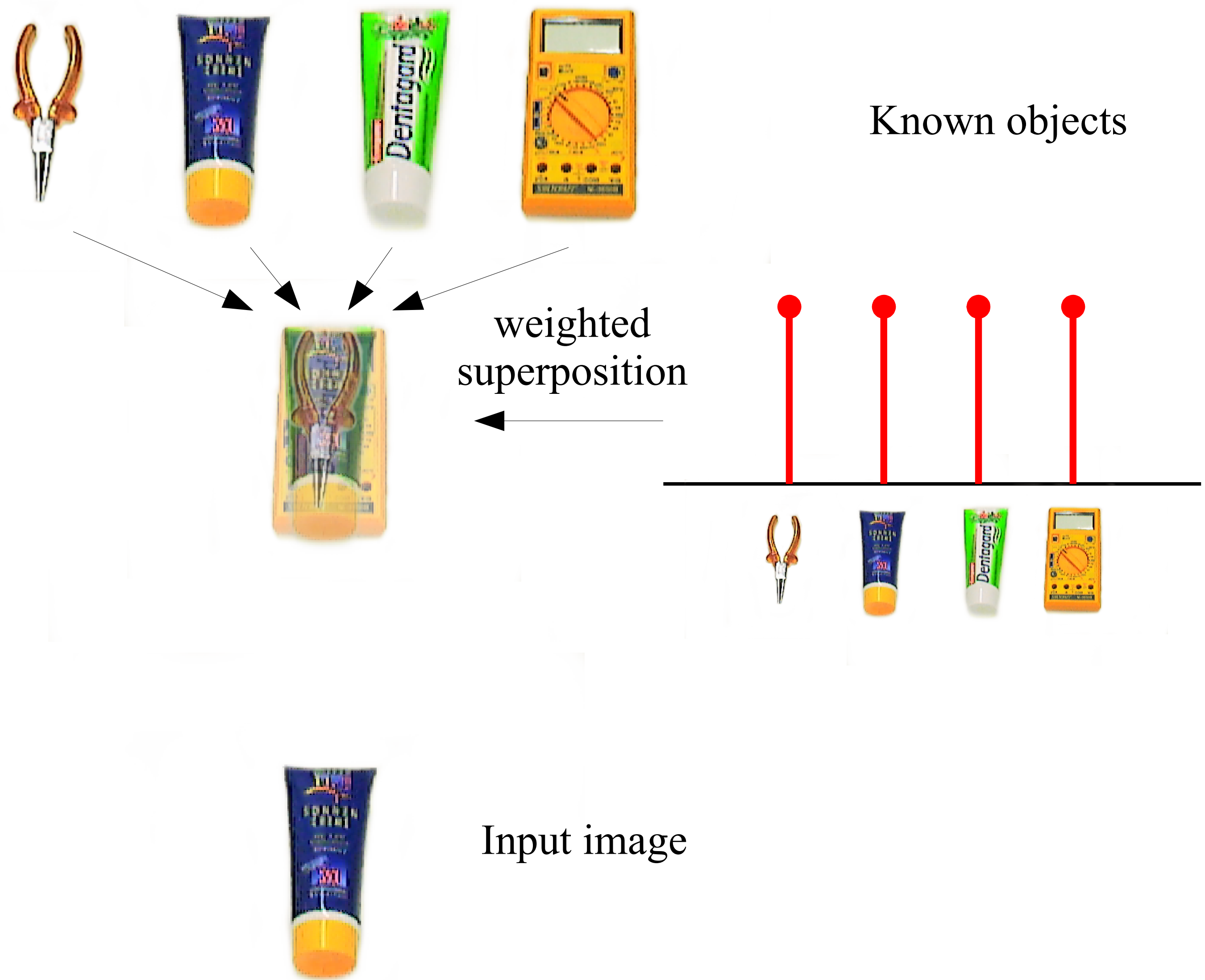
Recognizing an object

Known objects

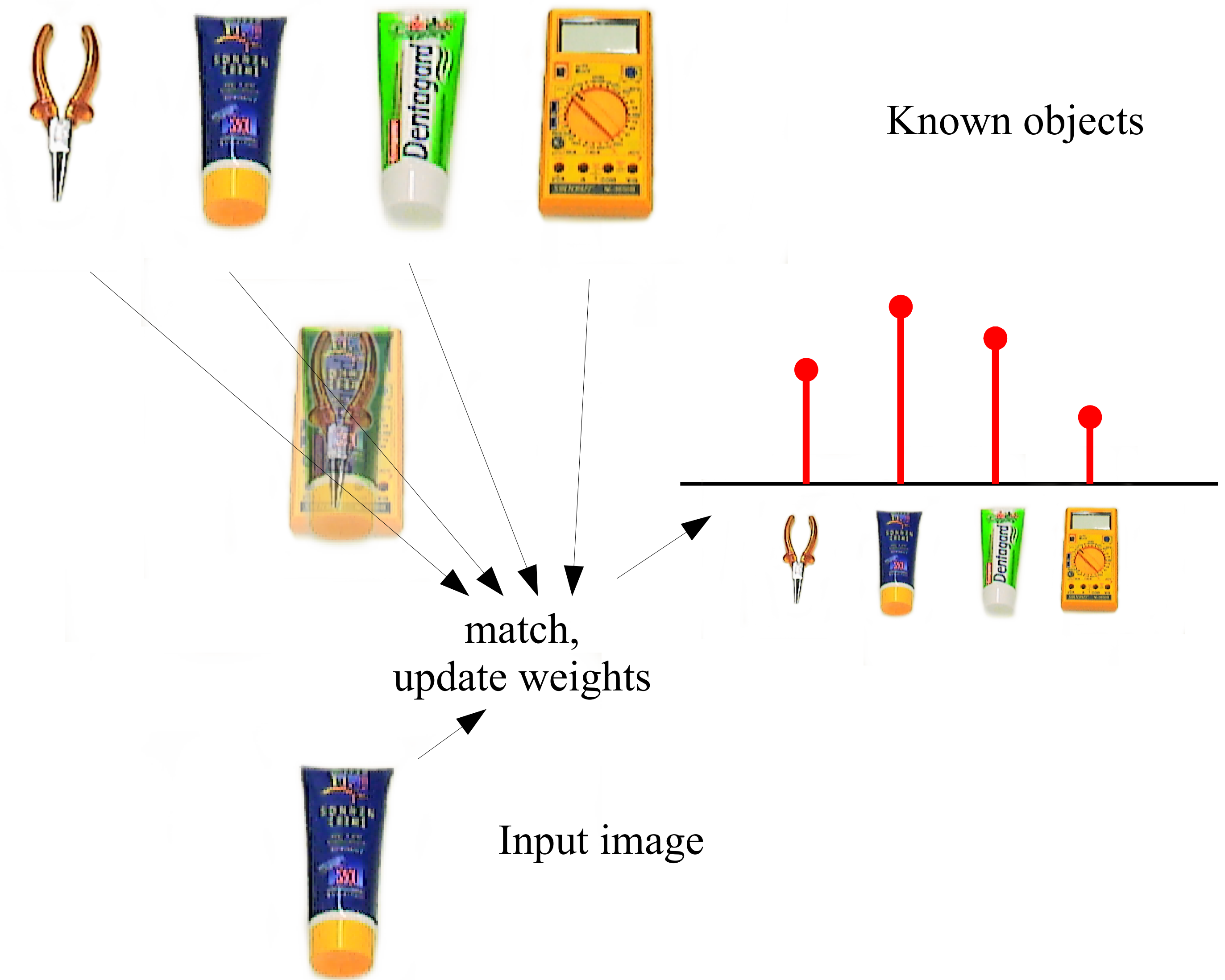


Input image

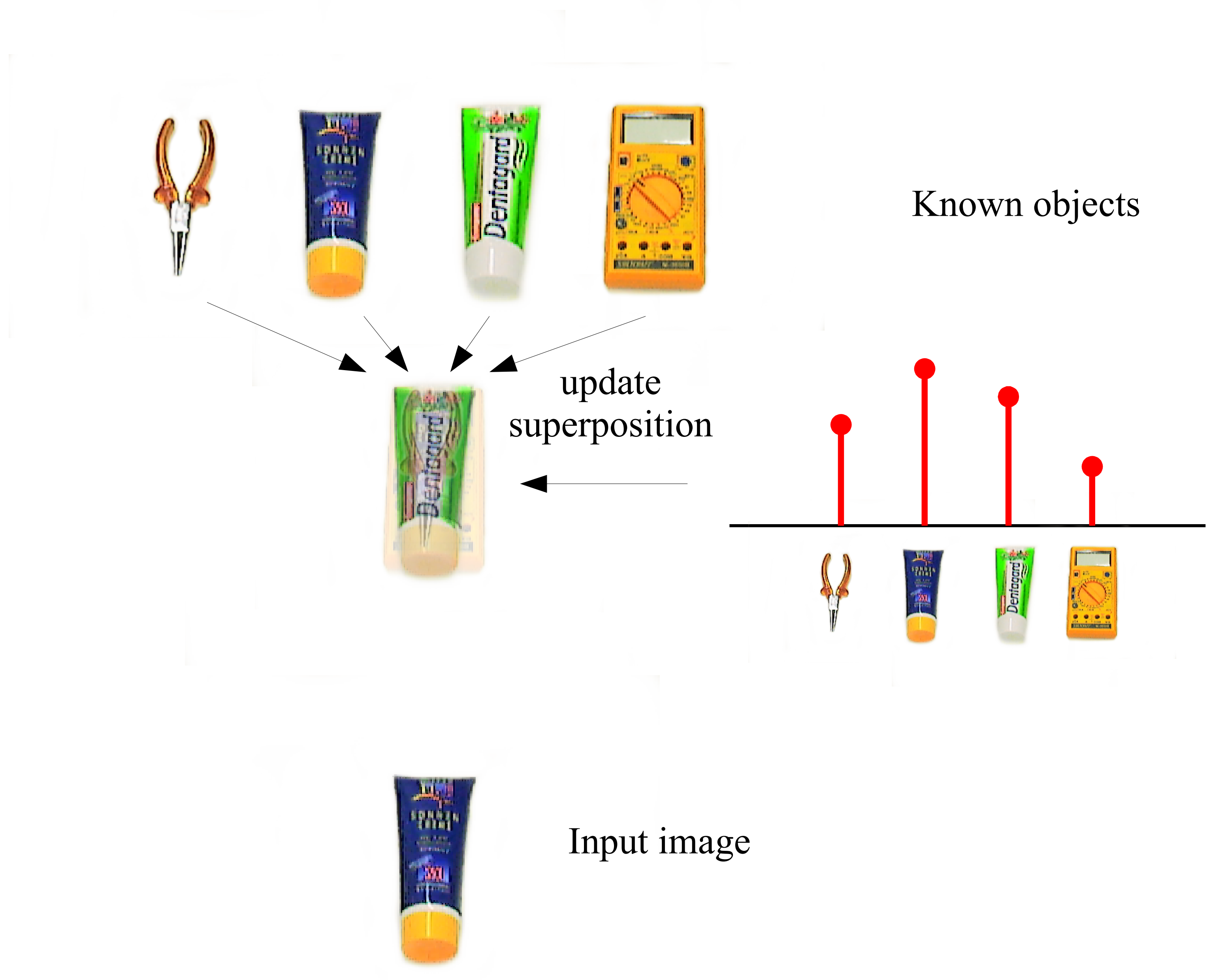
Recognizing an object



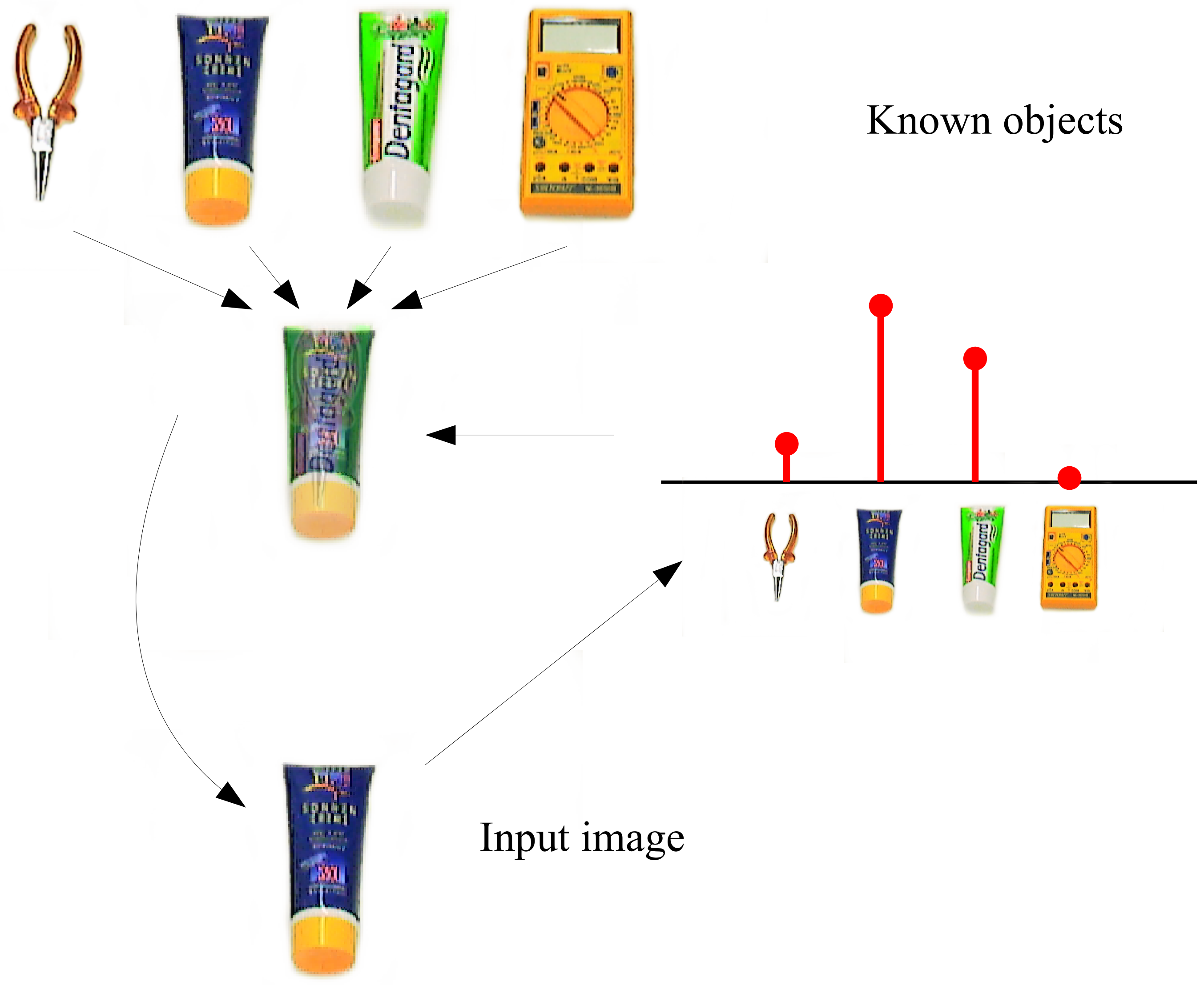
Recognizing an object



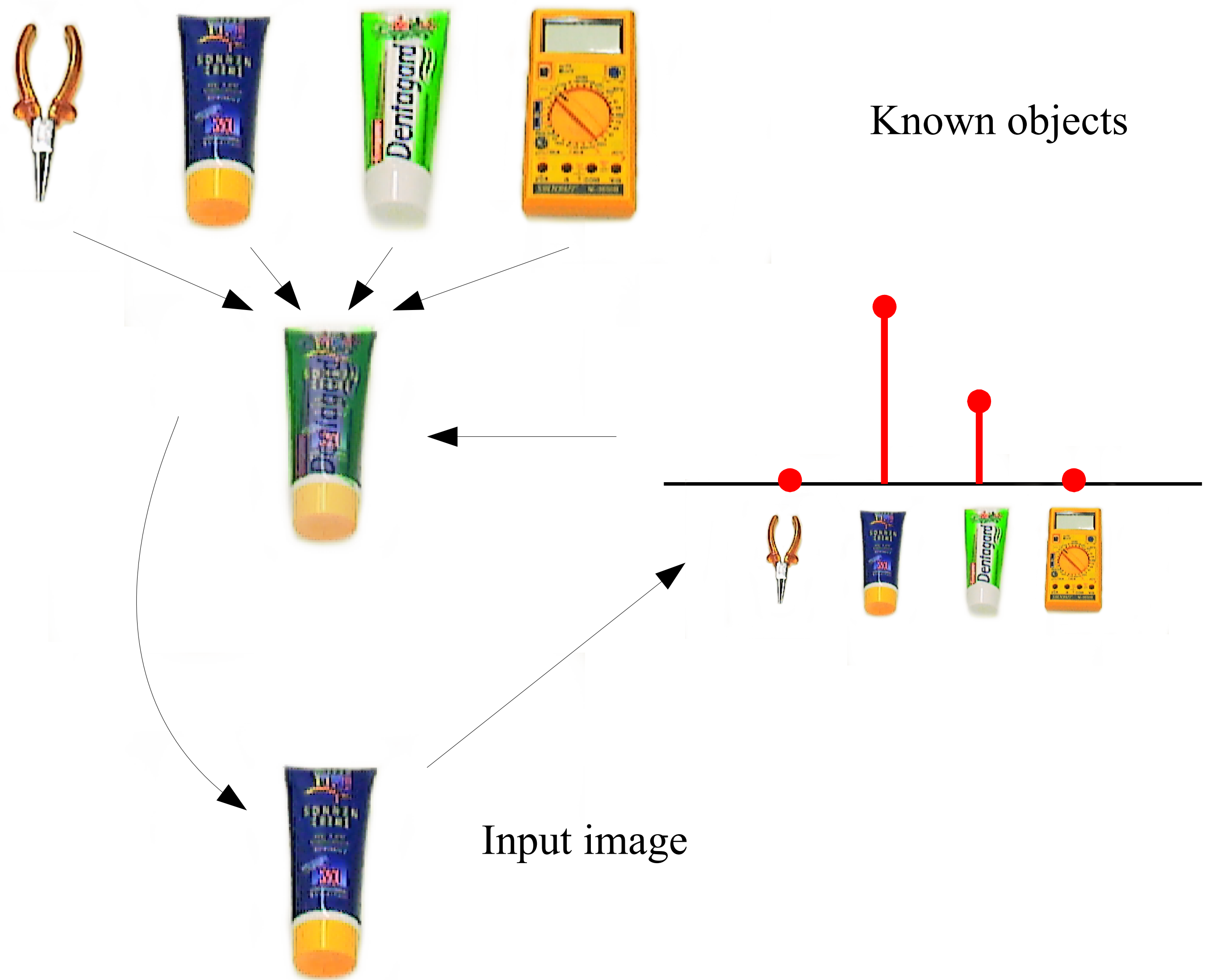
Recognizing an object



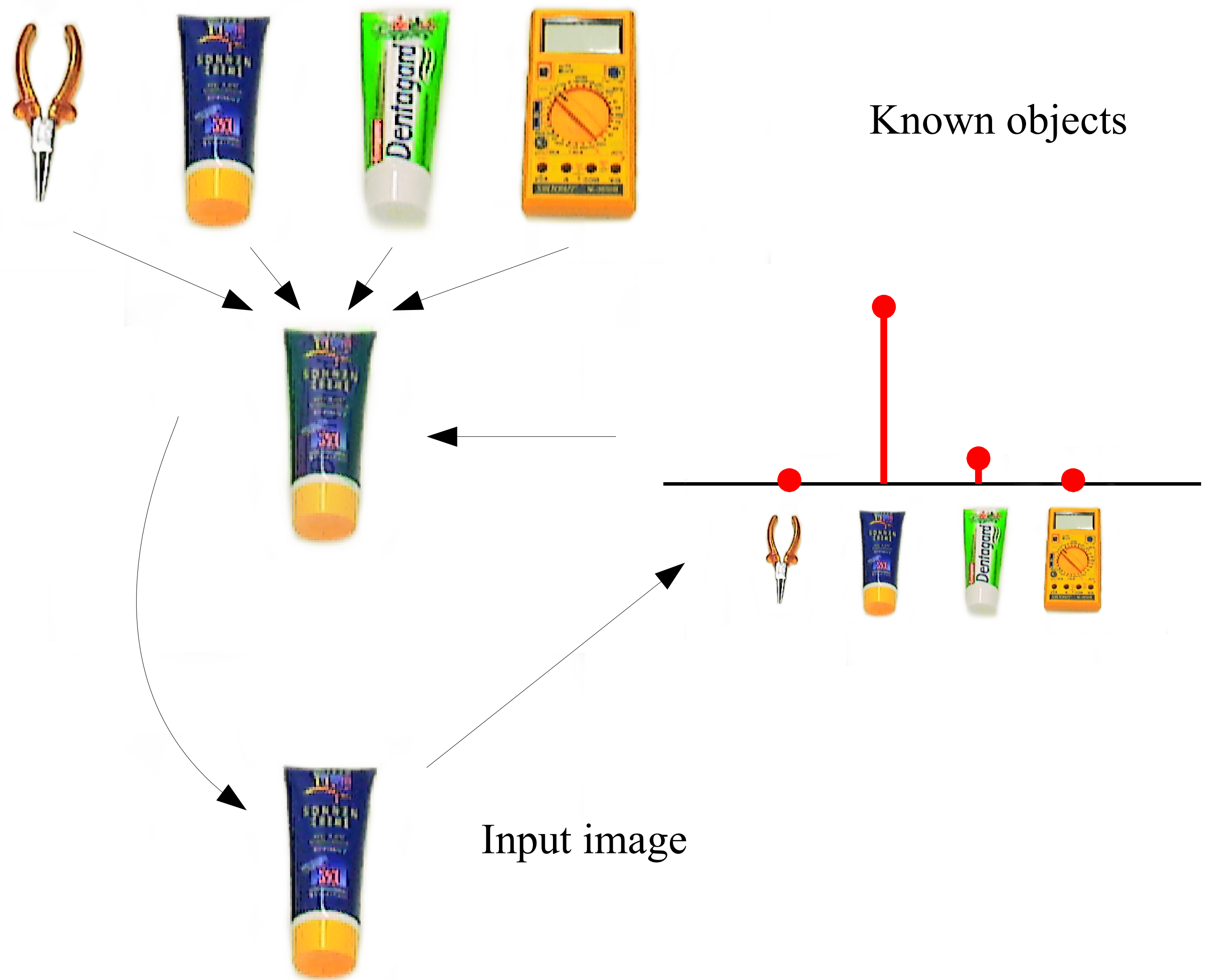
Recognizing an object



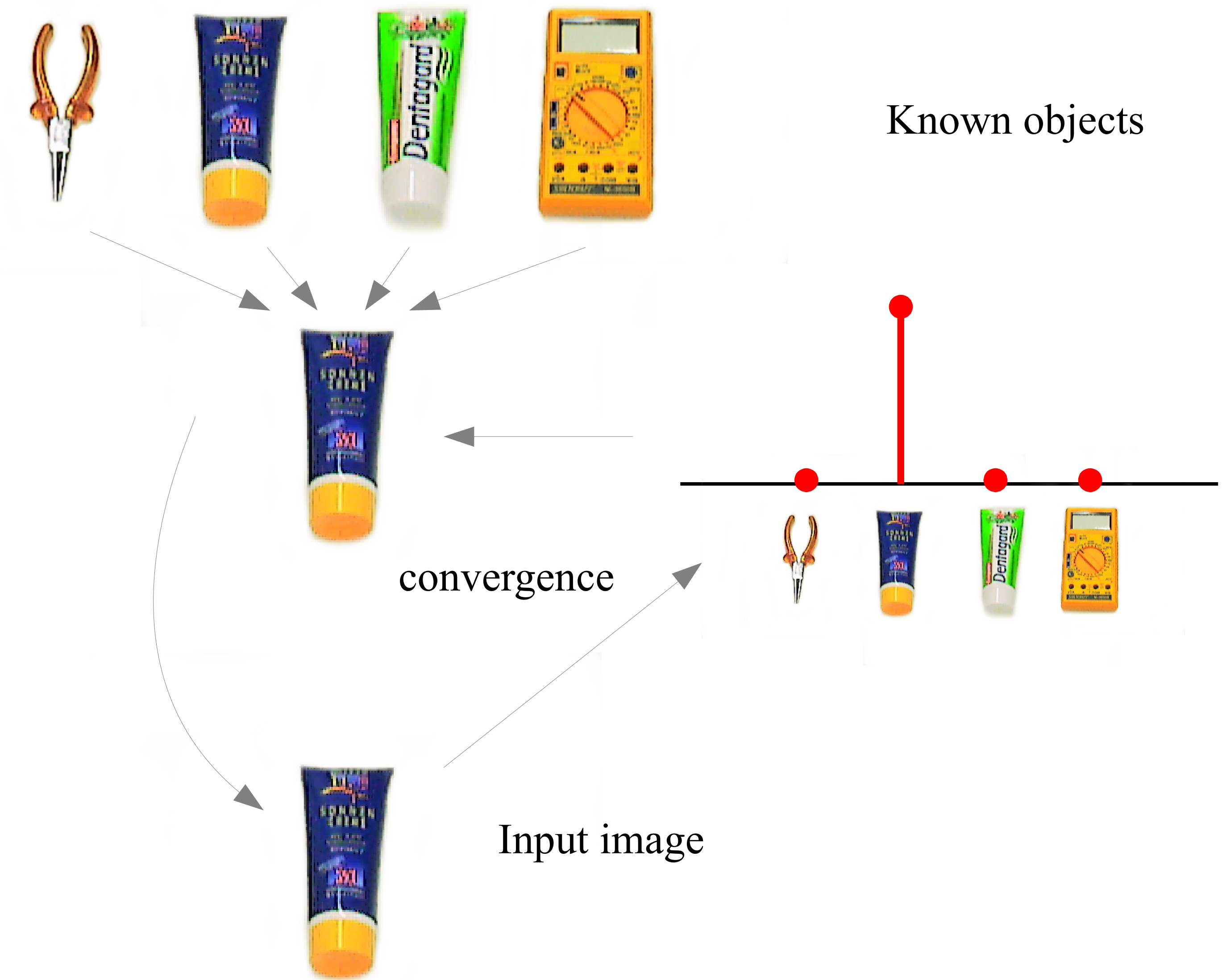
Recognizing an object



Recognizing an object

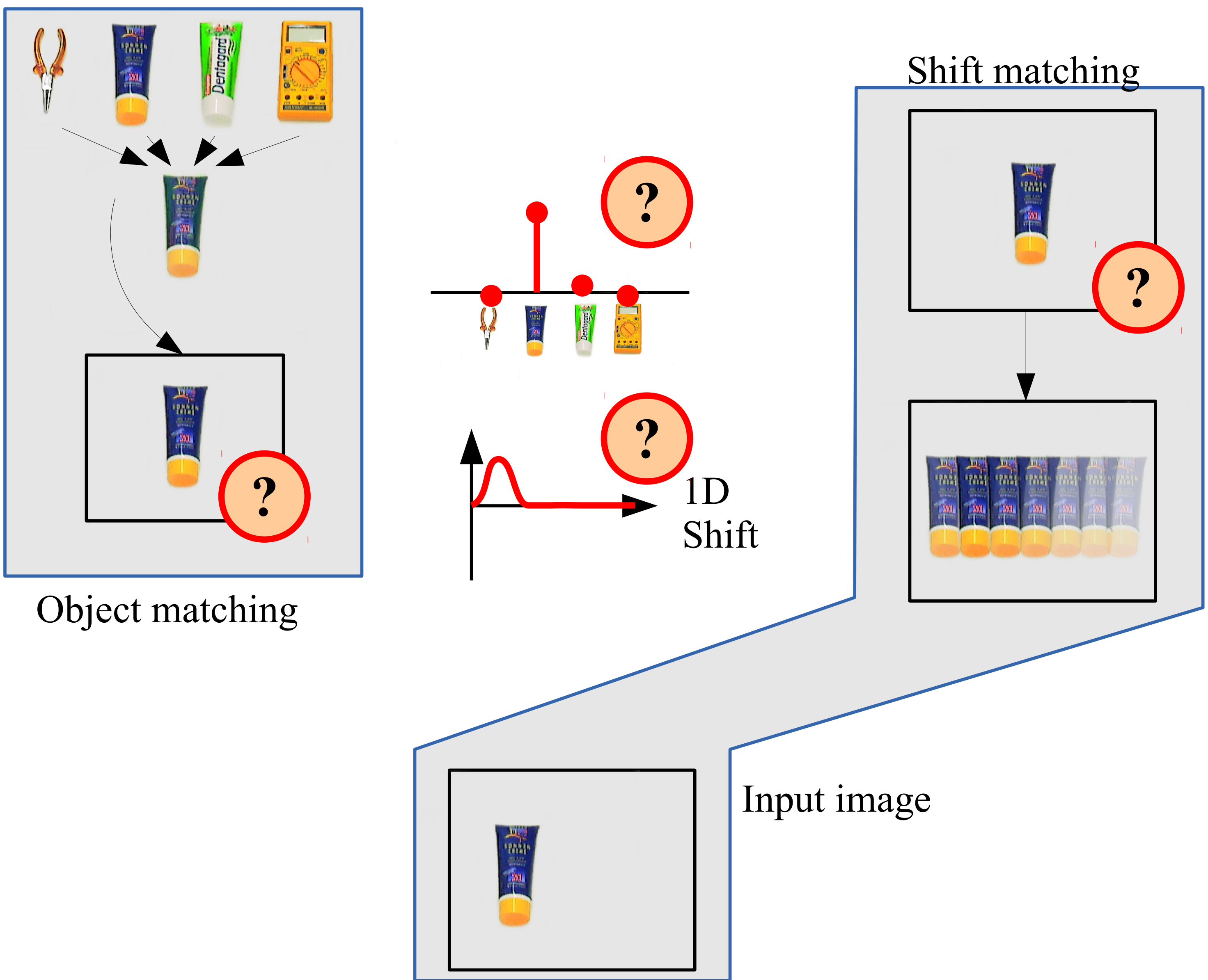


Recognizing an object

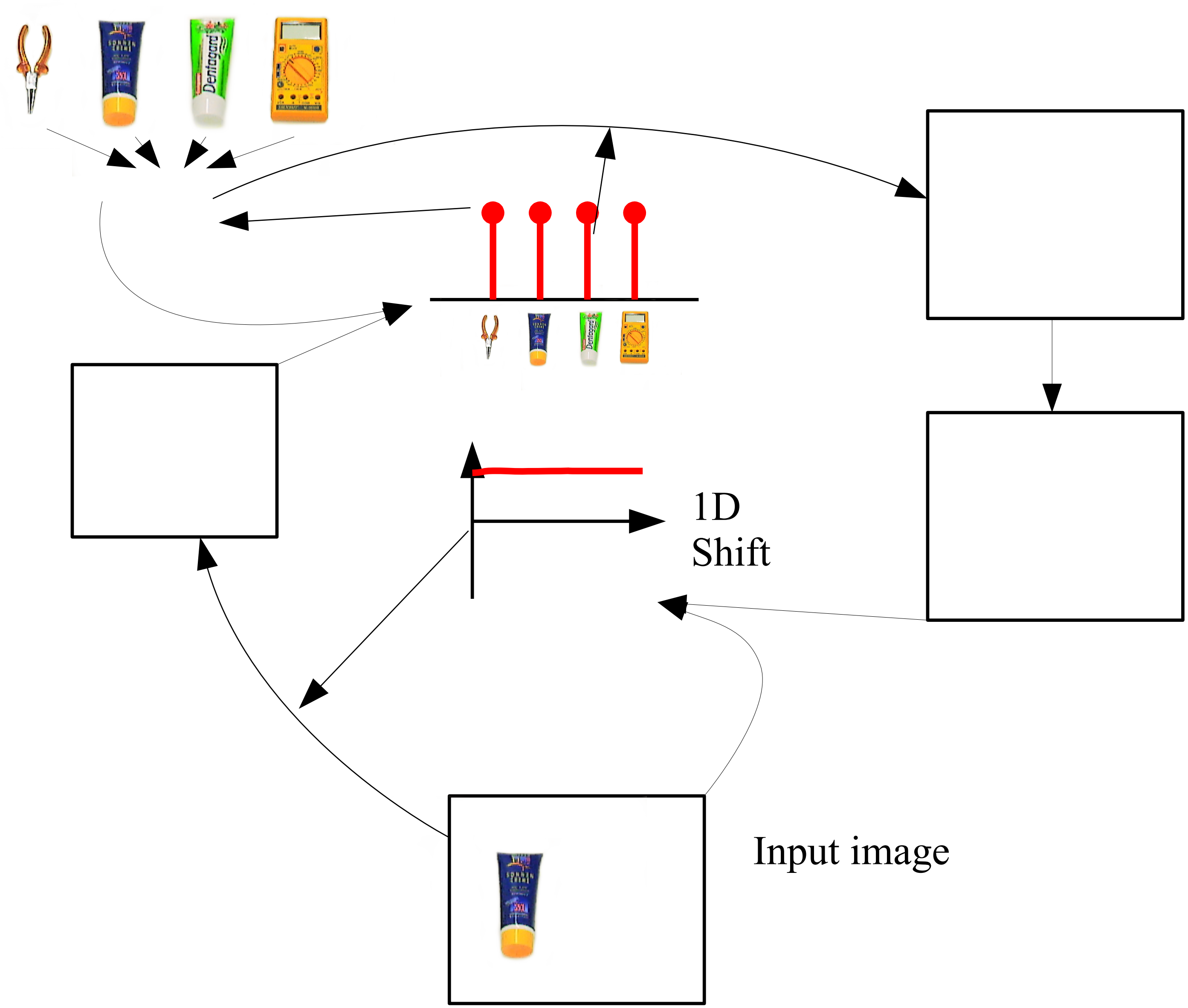


Combining the recognition and pose matching

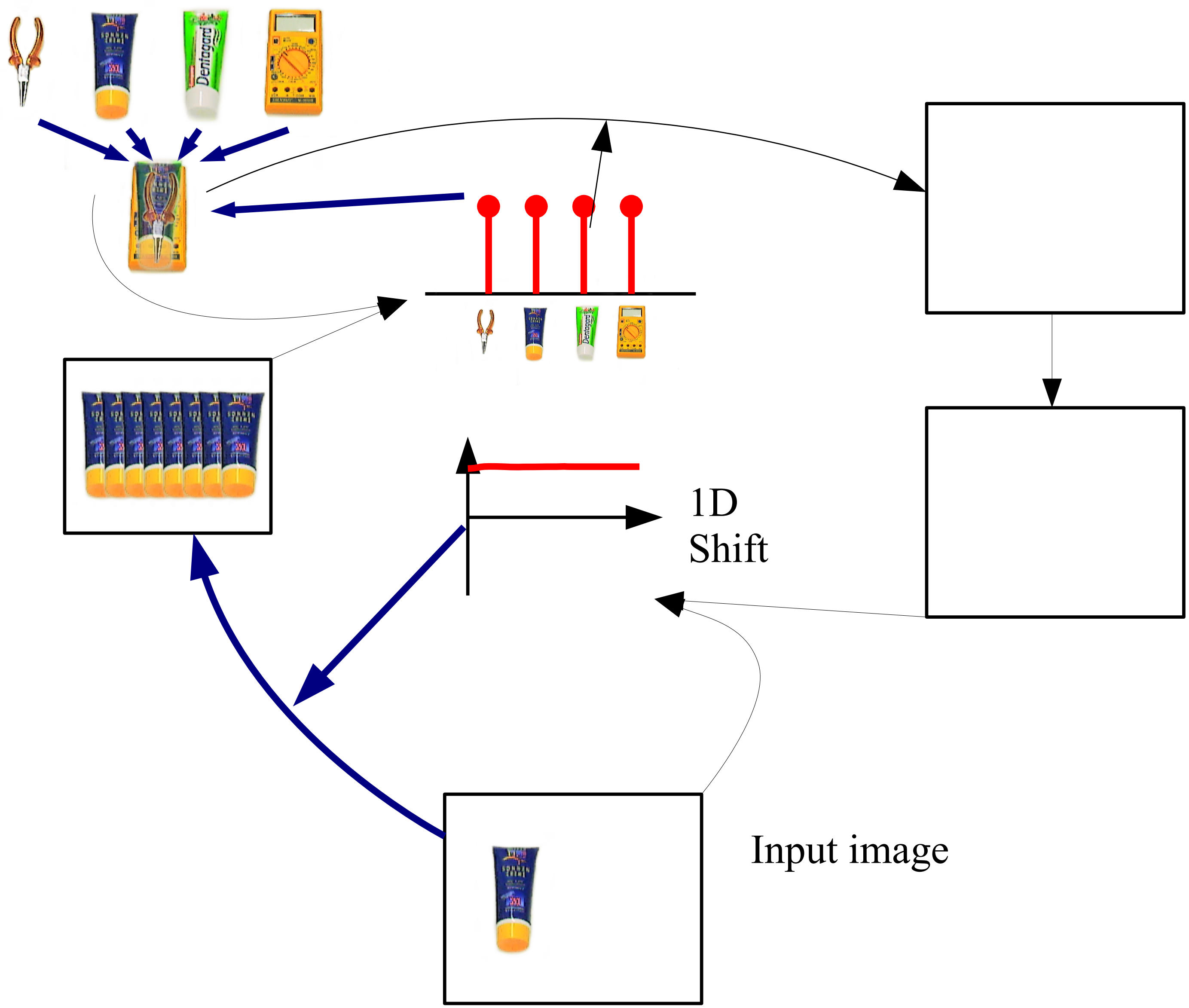
Recognition and pose estimation



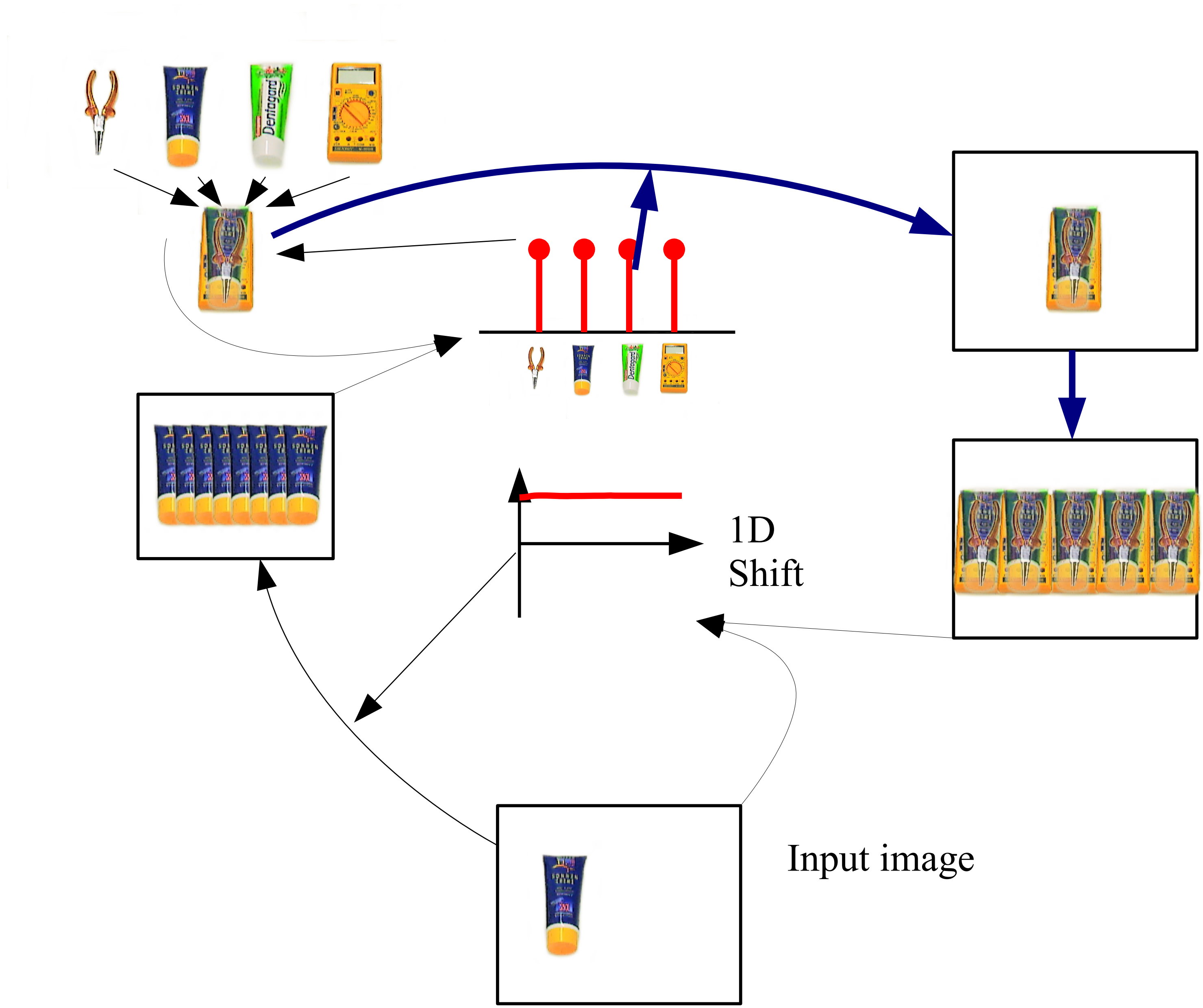
Recognition and pose estimation



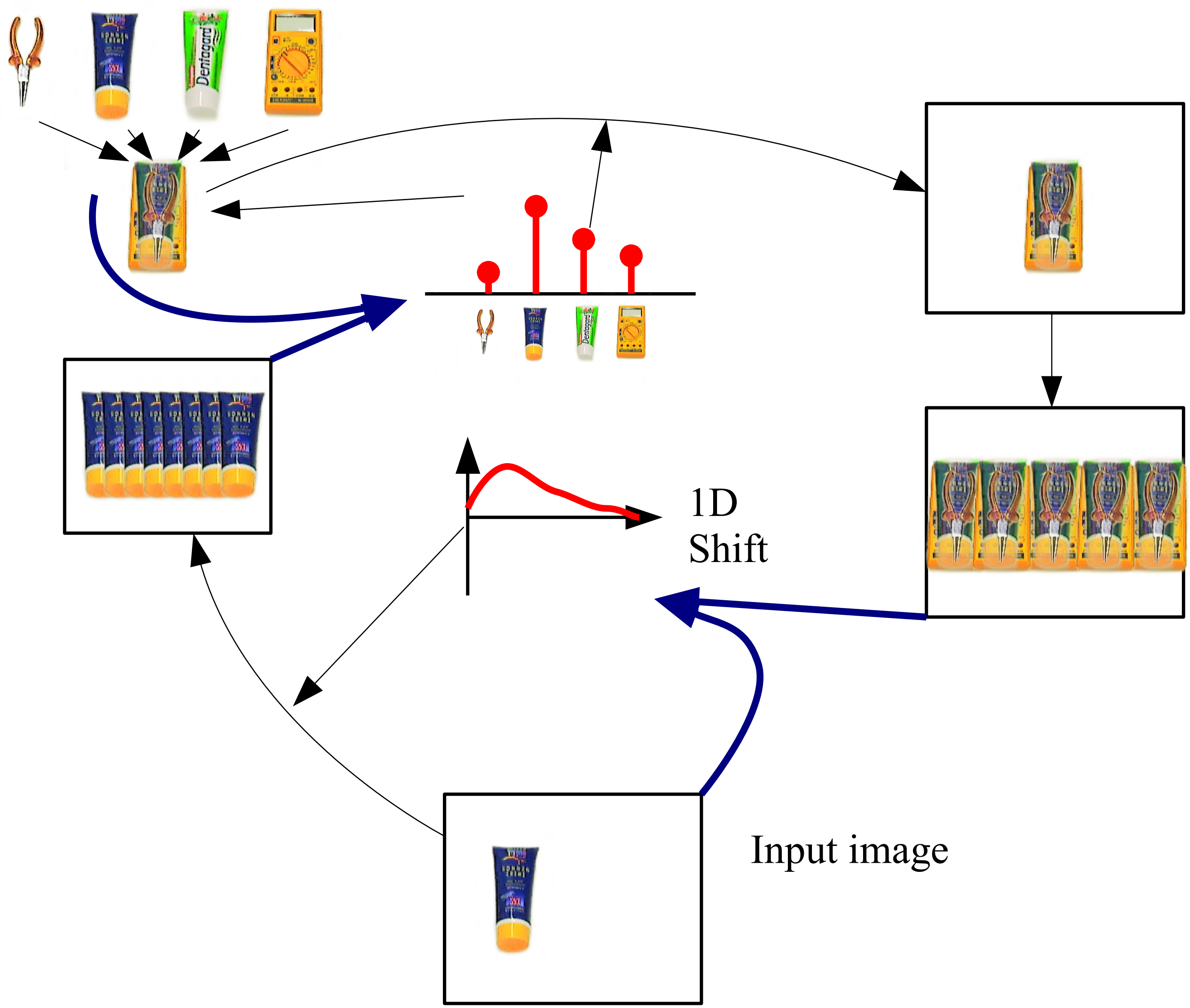
Recognition and pose estimation



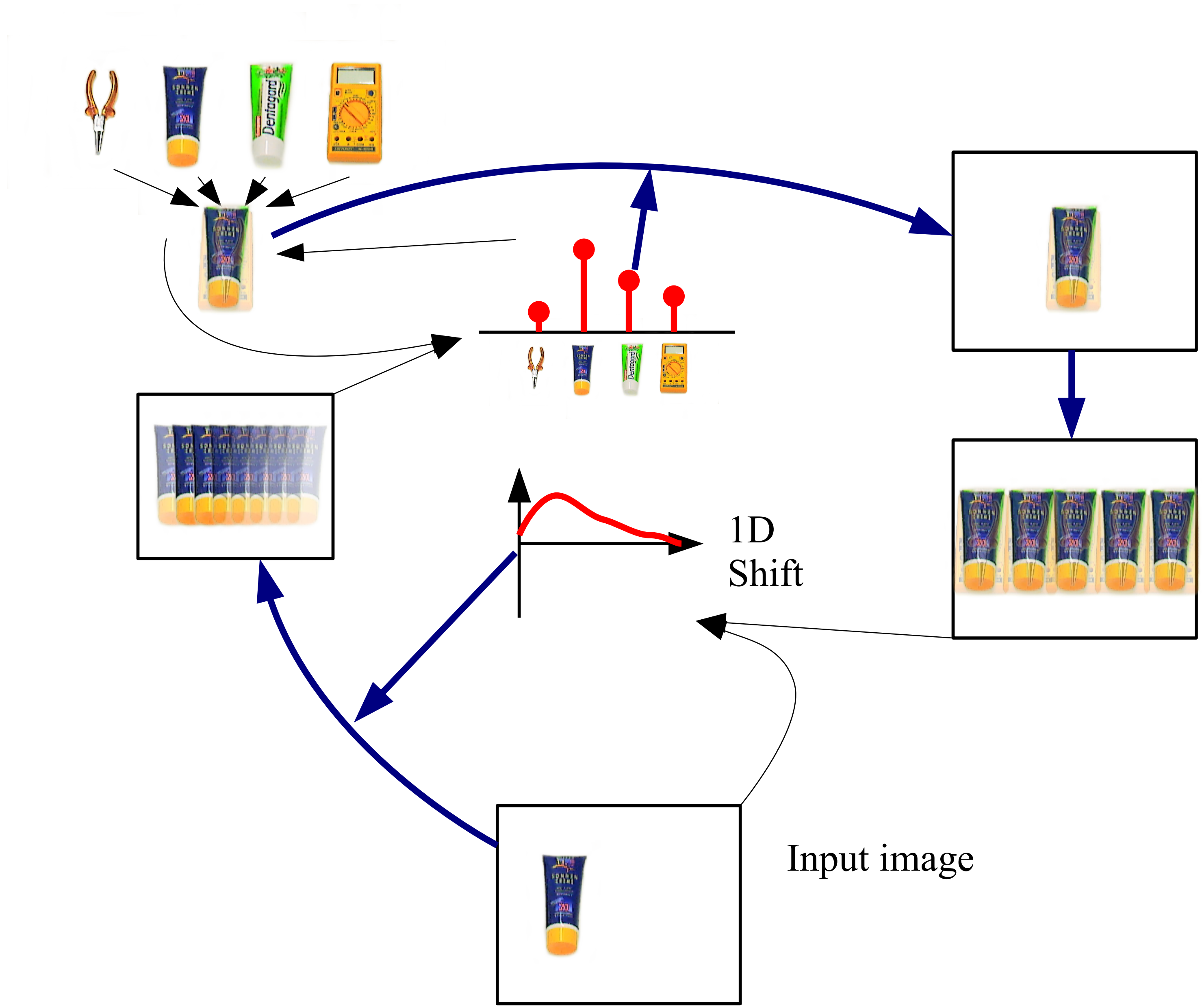
Recognition and pose estimation



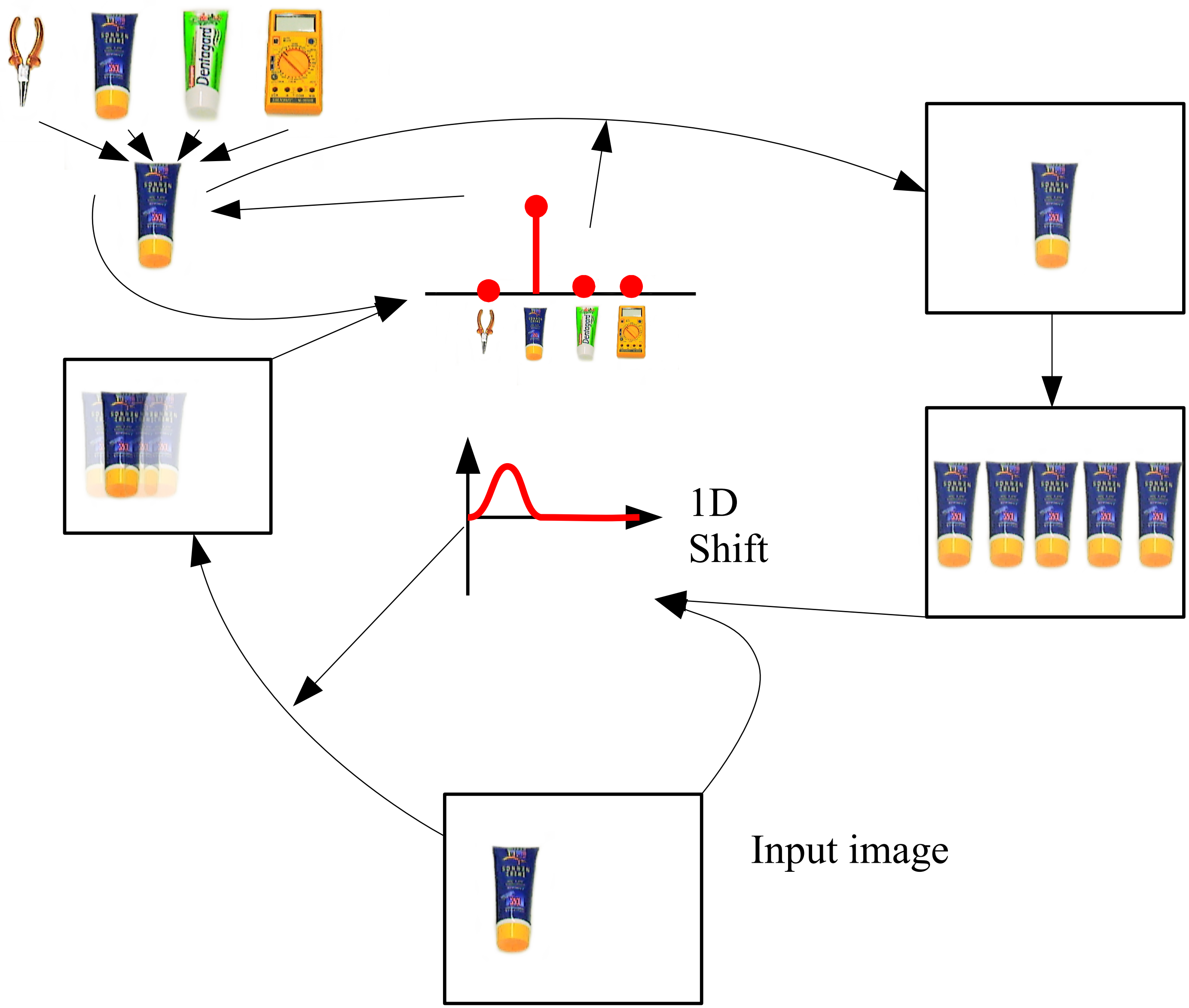
Recognition and pose estimation



Recognition and pose estimation



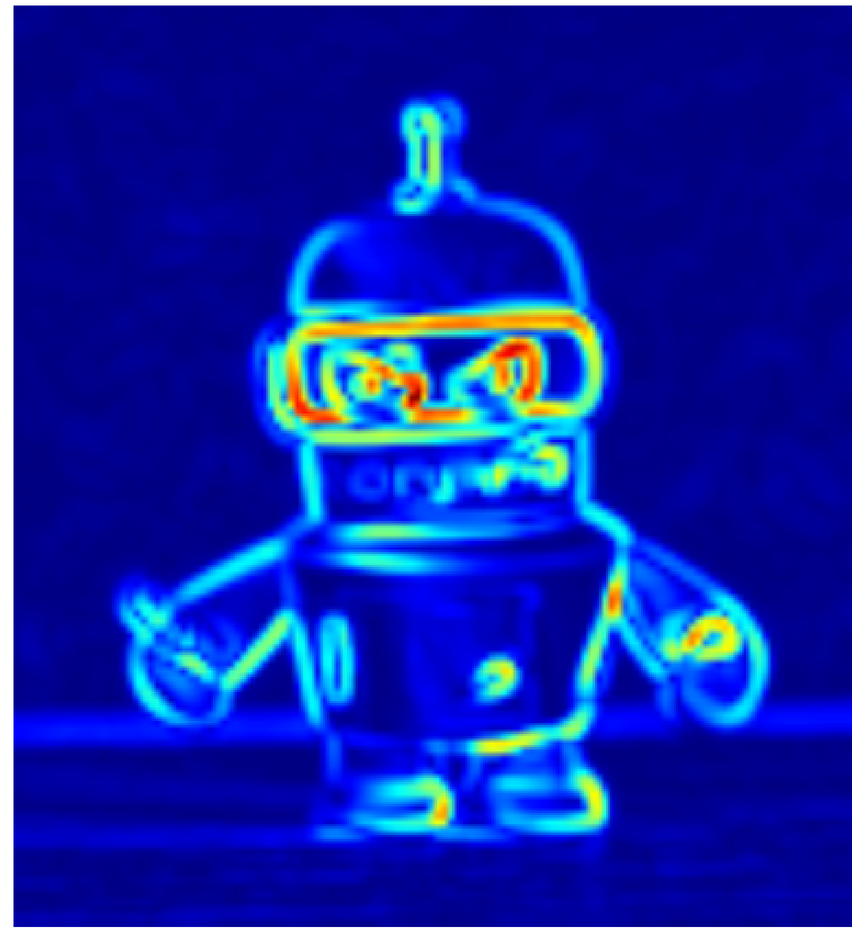
Recognition and pose estimation



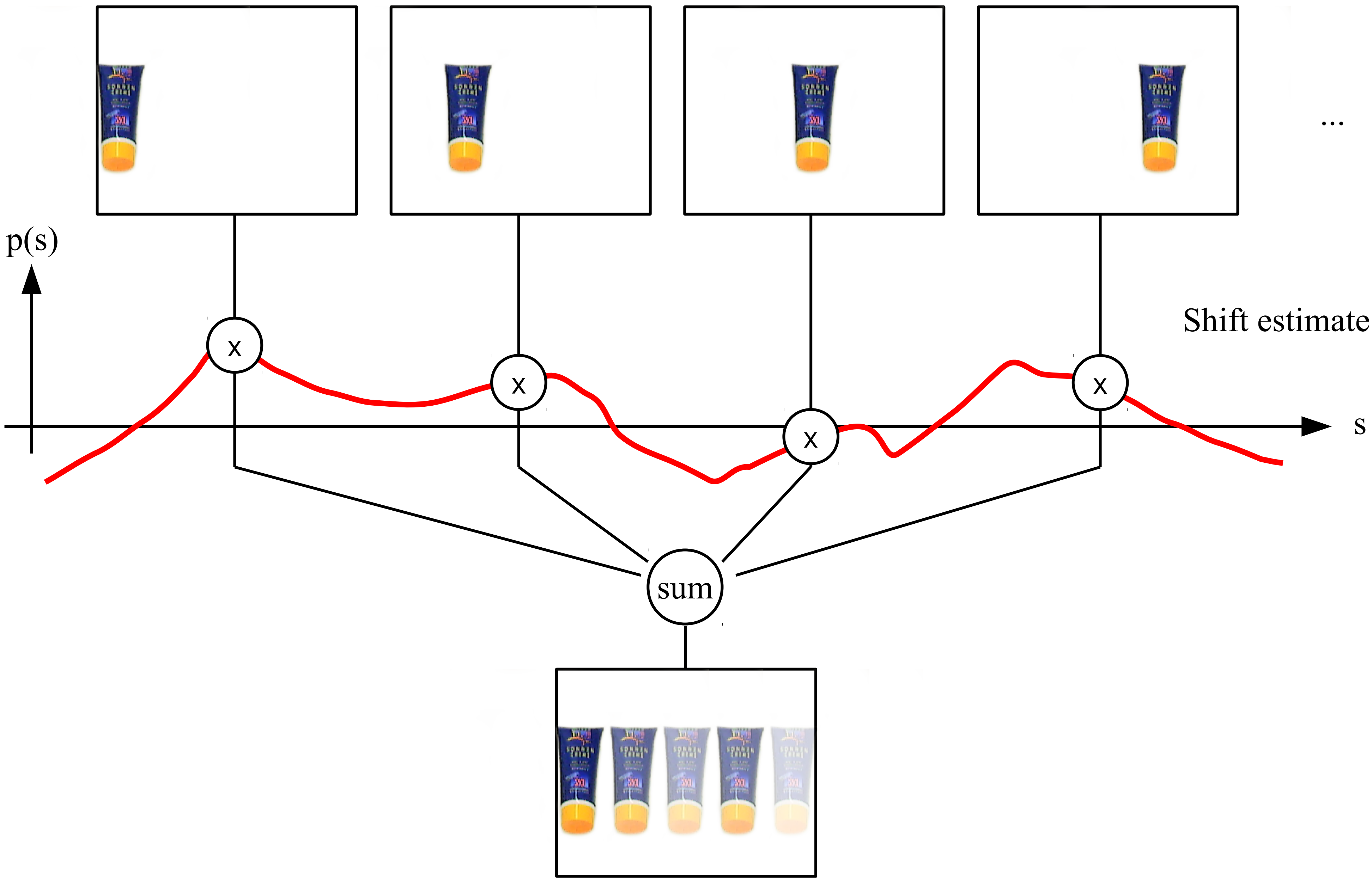
Implementation

Feature channels

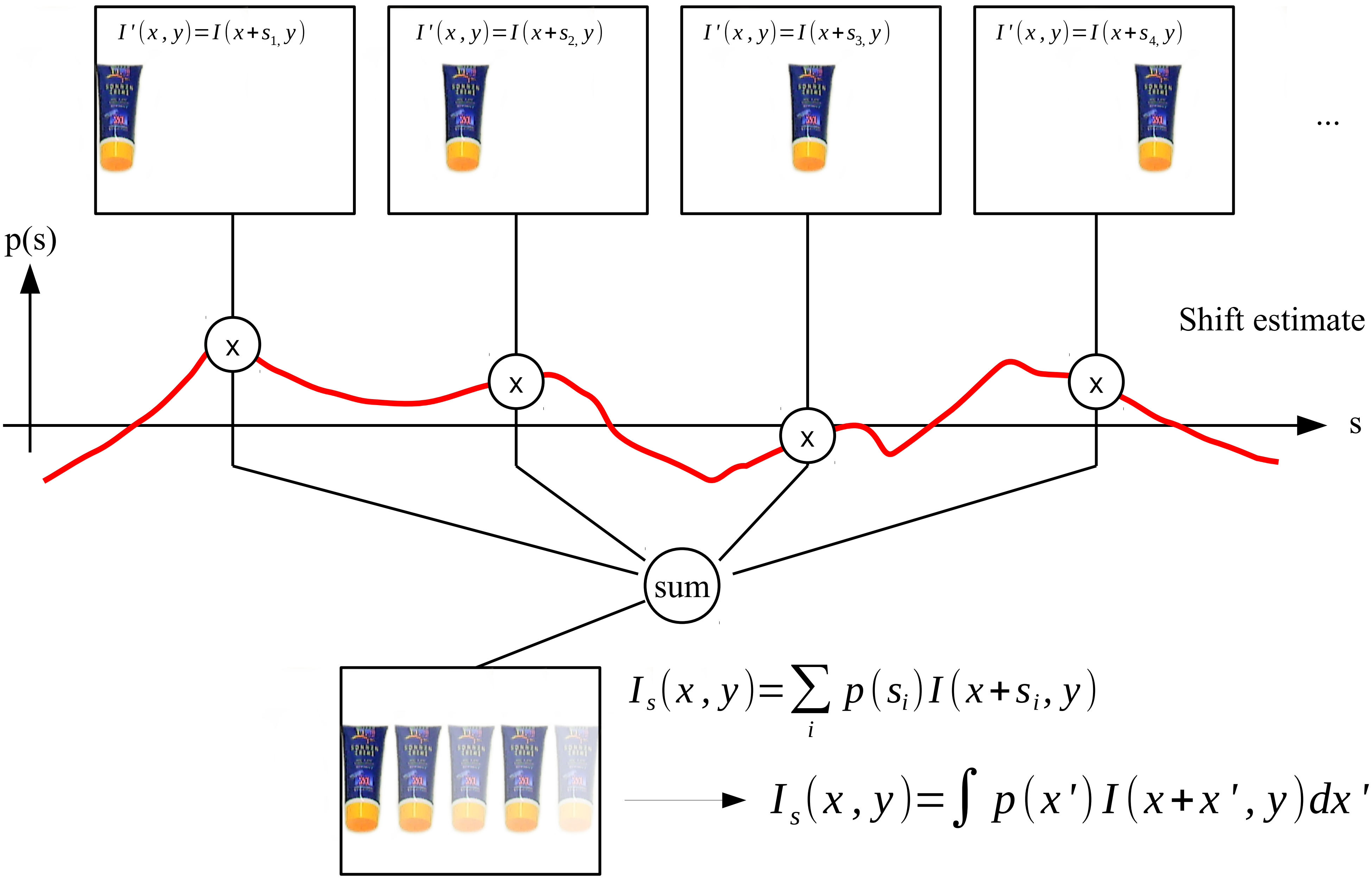
- Full system has several feature channels
  - Spatial pattern (shape)
  - Localized histograms



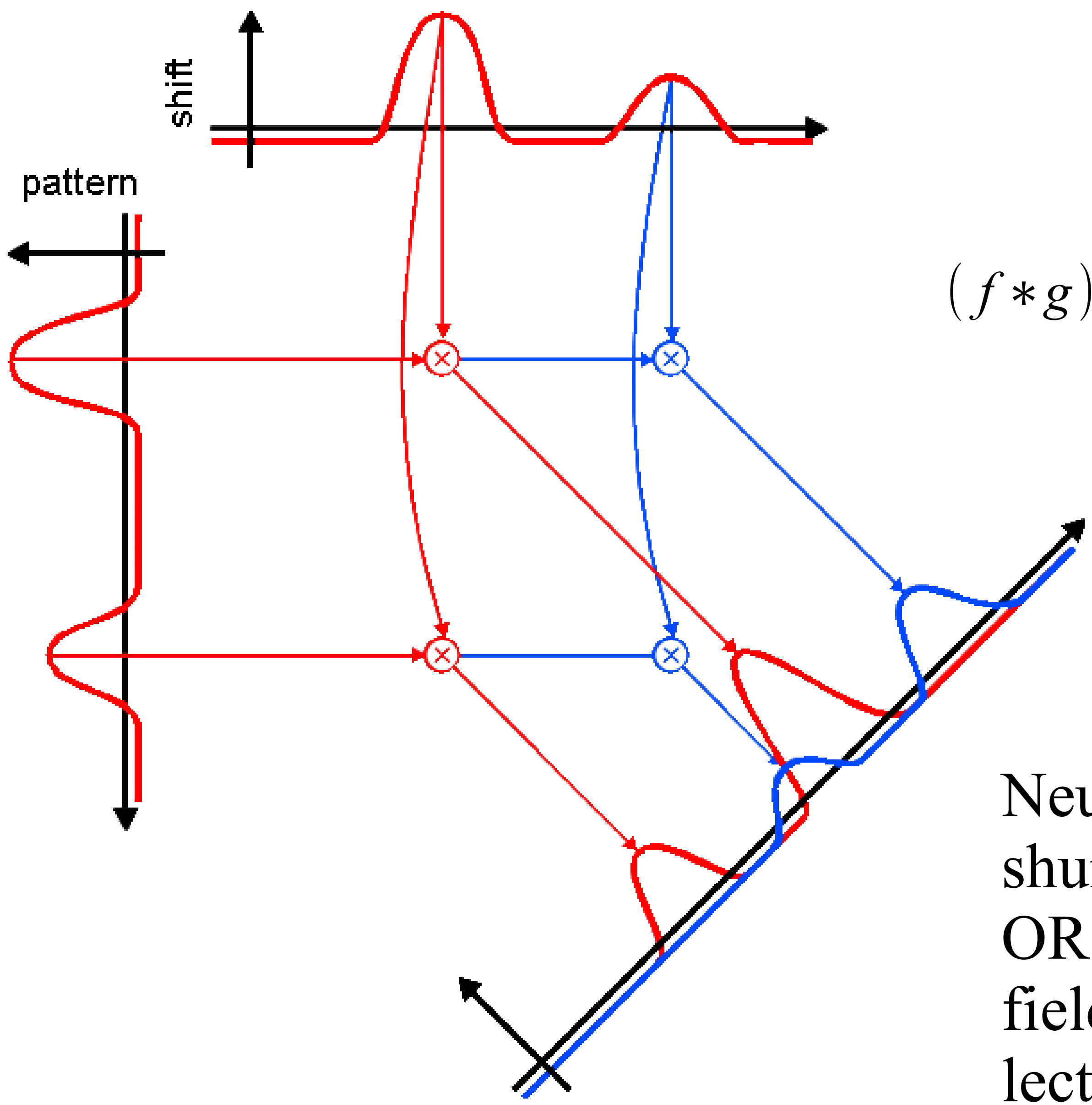
Transformations



Transformations



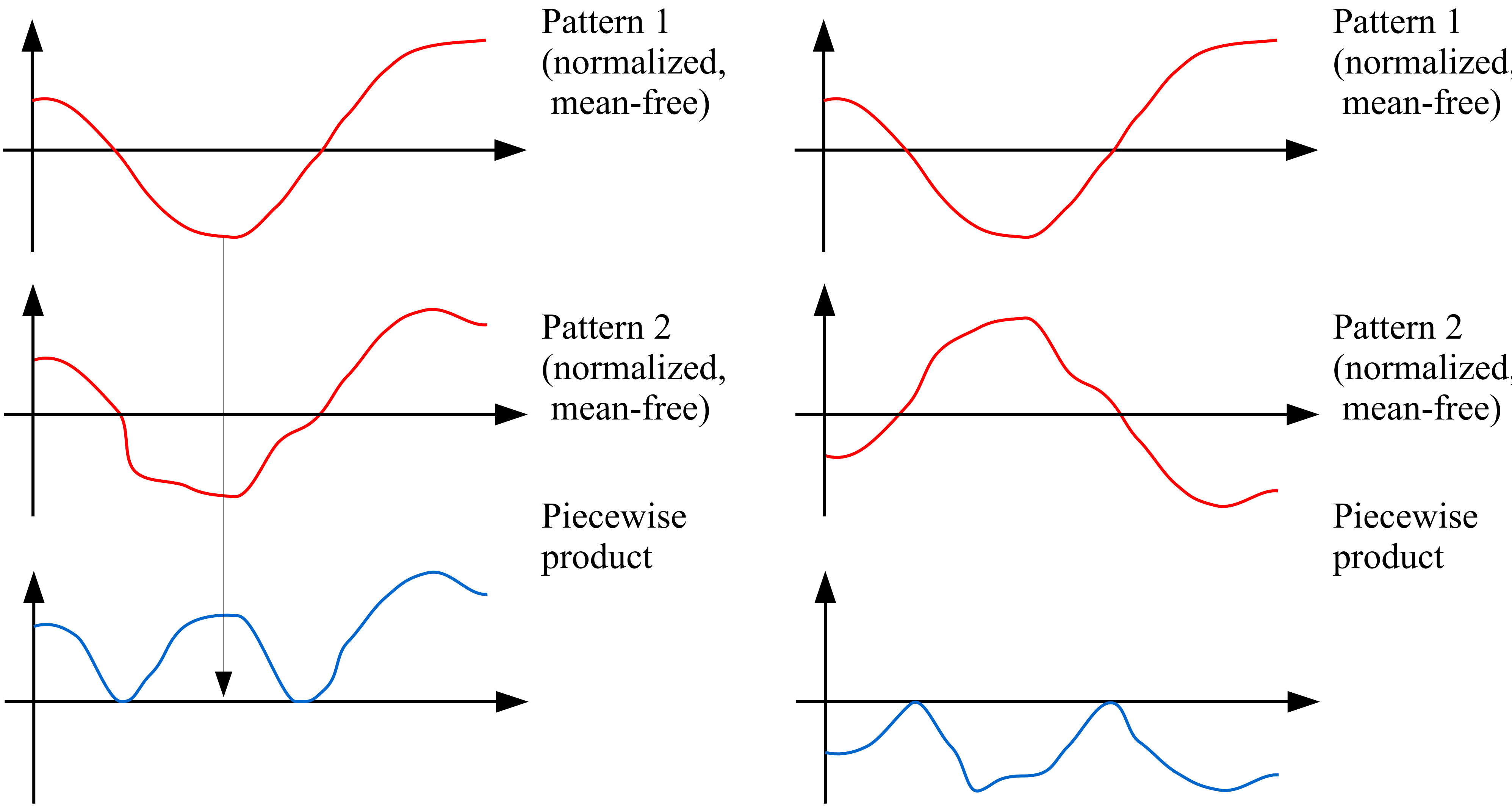
Transformations



$$(f*g)(x)=\int f(y)g(x-x')dx'$$

Neural implementation: shunting synapses
OR: transformation fields (cf. Schneegans lecture)

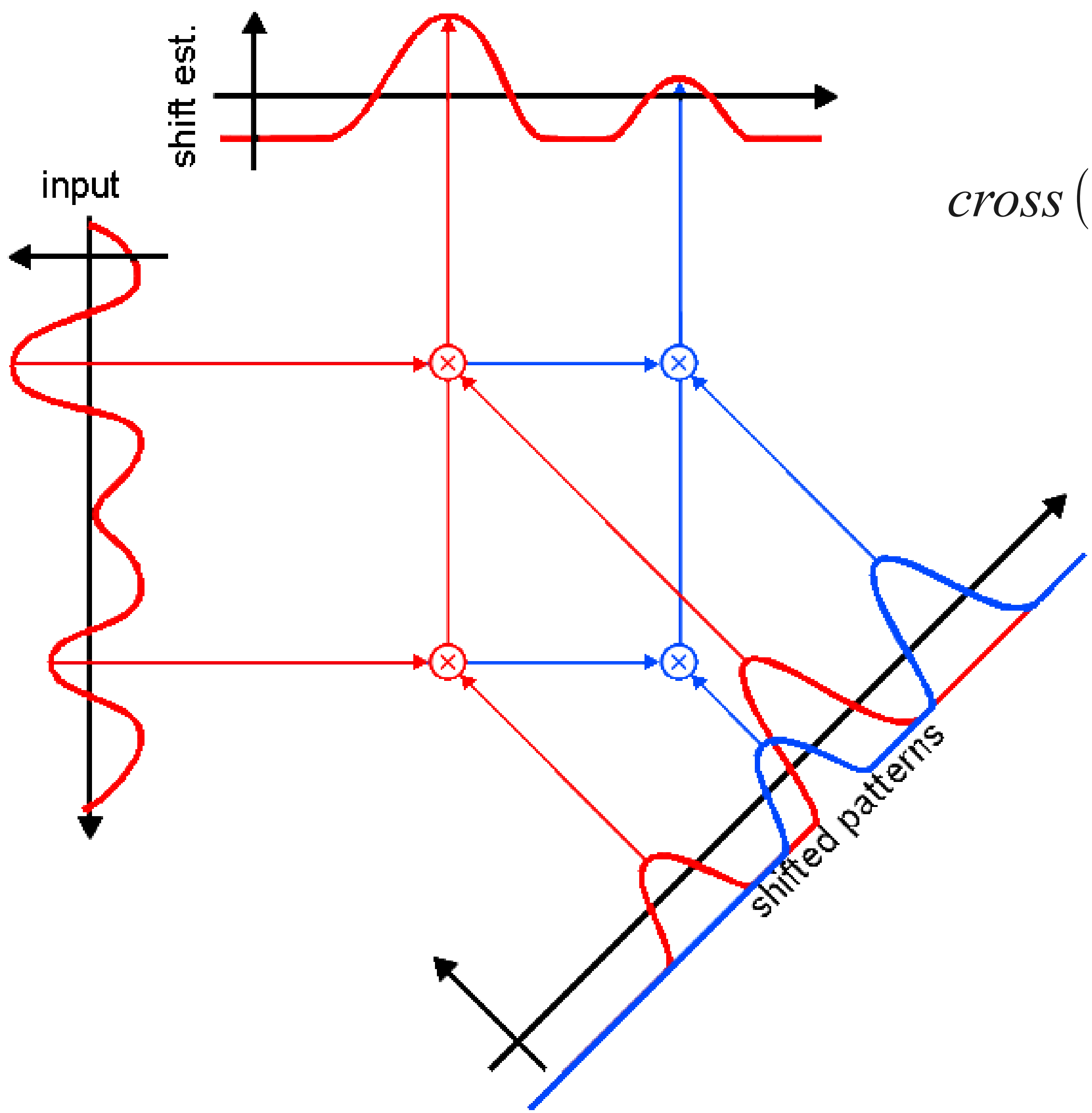
Matching views



Sum is positive => match

Sum is negative => mismatch

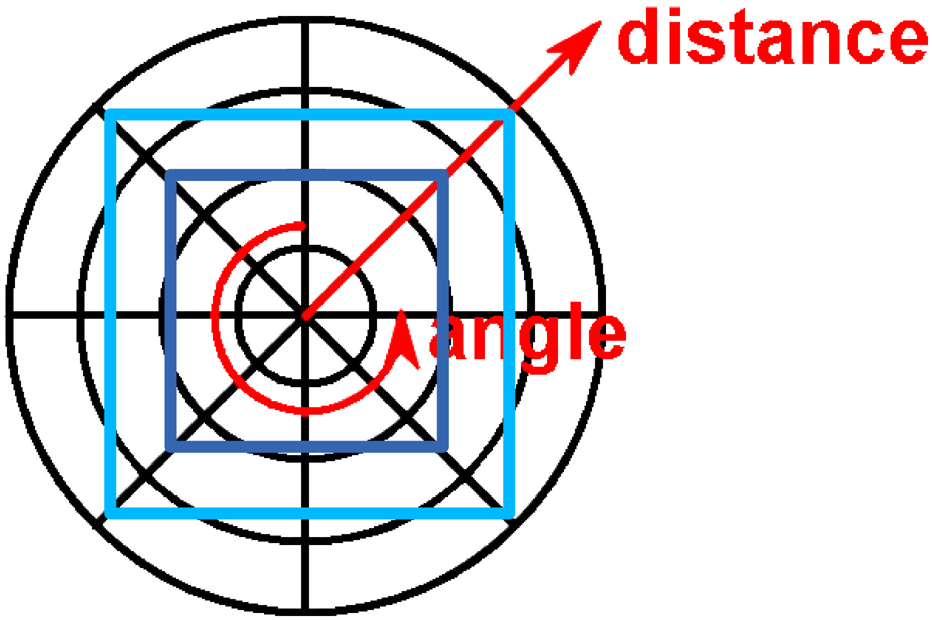
Matching poses



$$cross(f, g, x) = \int \overline{f}(y)\, g(x+y)\, dy$$

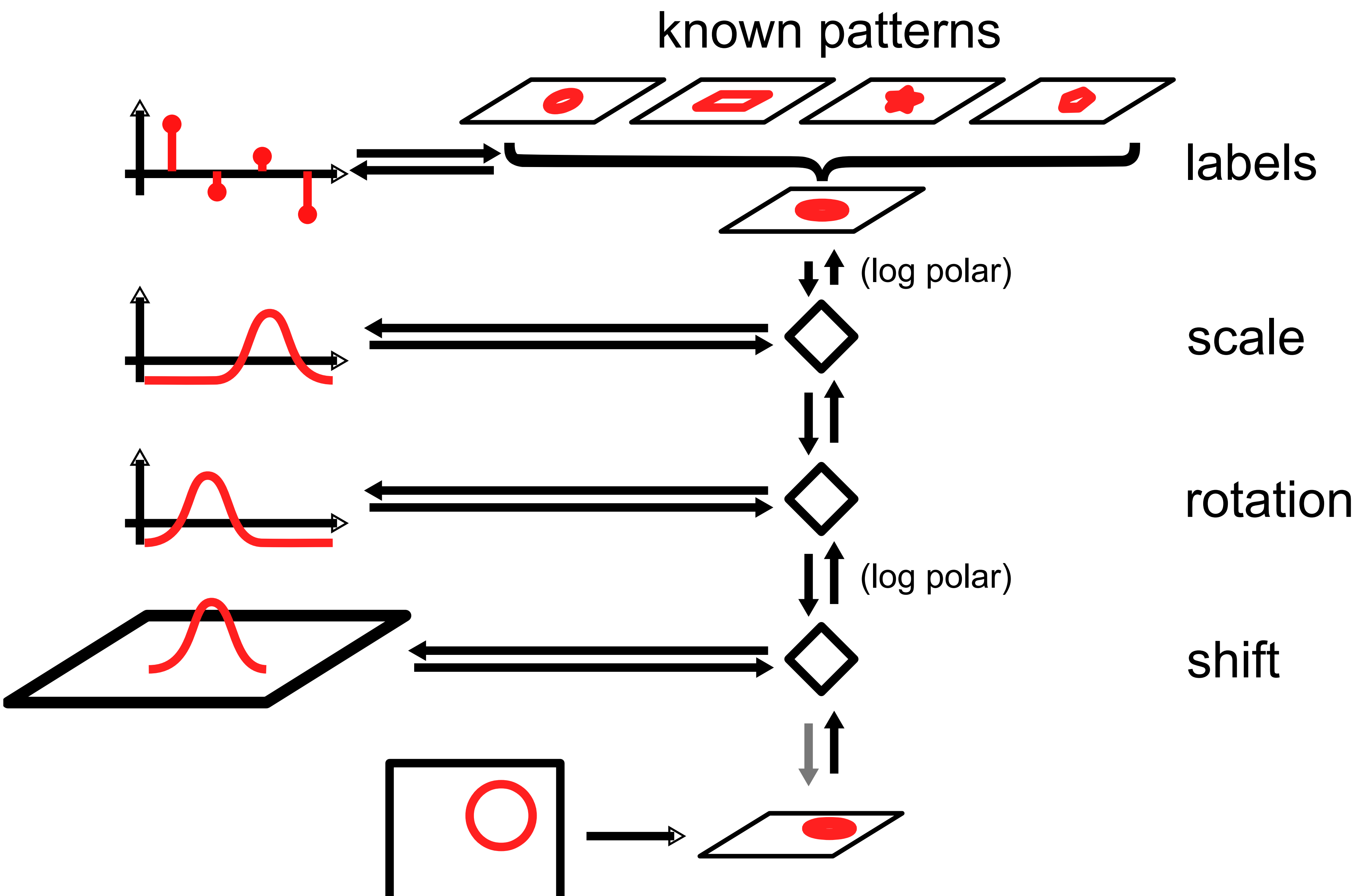
- Similar to transformation, different direction
- Normalized, mean-free
- Requires shunting synapses

Other transformations

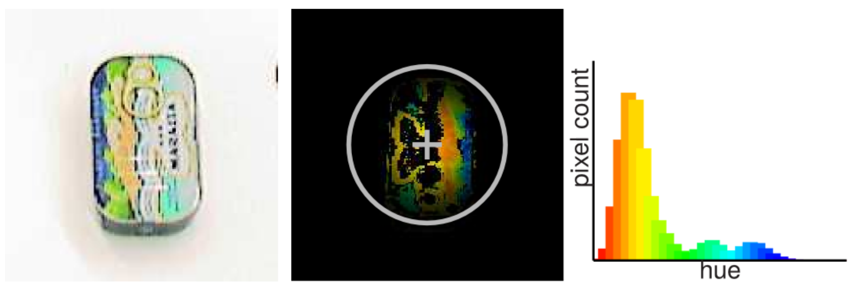


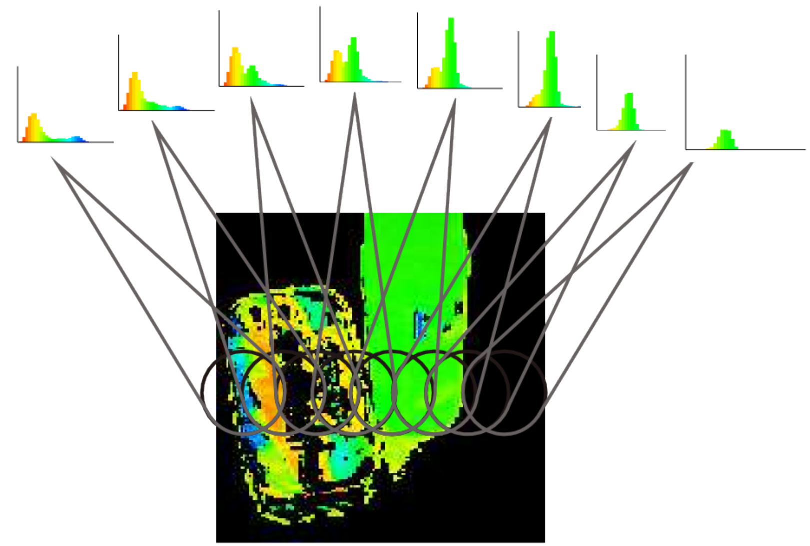
- Shift in log-polar space is uniform scaling and rotation
- Log-polar is neurally plausible (retinal space)

Cascading transformations



Localized Histograms

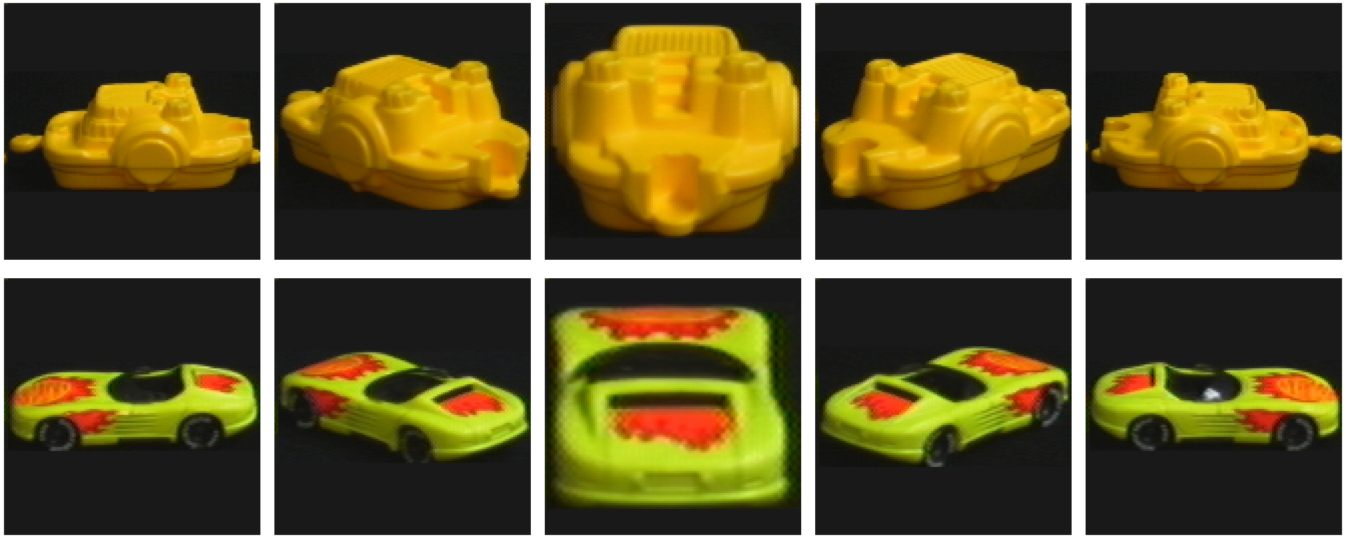




from Faubel, Schöner (2009)

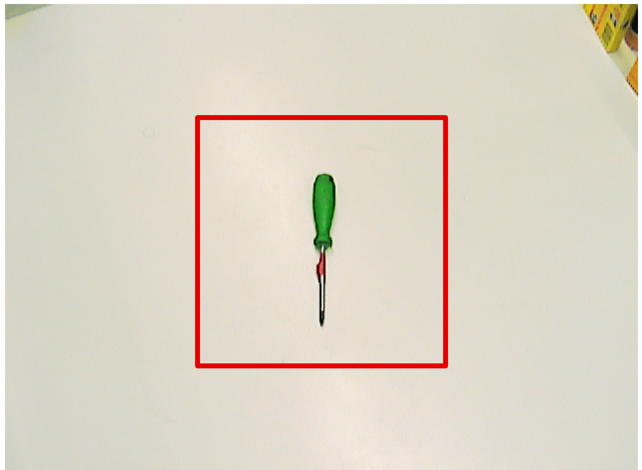
- Shape alone (as presented before) is not very powerful for recognition
- Additional feature channels: localized histograms (color, edge orientations)
- All channels provide information about pose, object
- Scale cannot be estimated

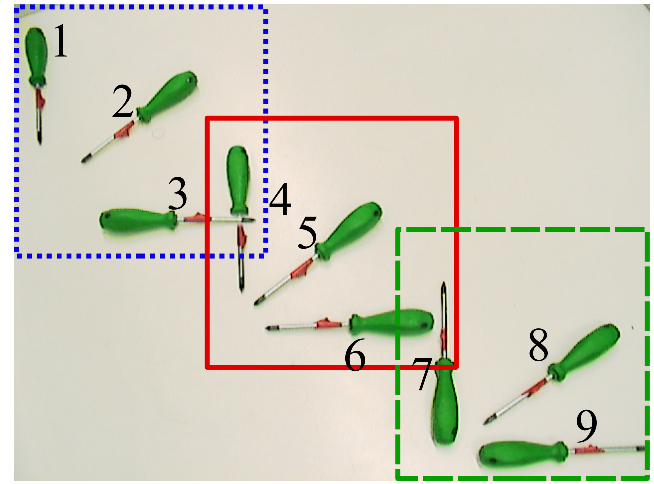
Results



- Recognition rates on COIL (with shape + localized histograms):
  - 85% with one training view
  - 94% with four training views
  - See Faubel, Schöner (2009)

Results II

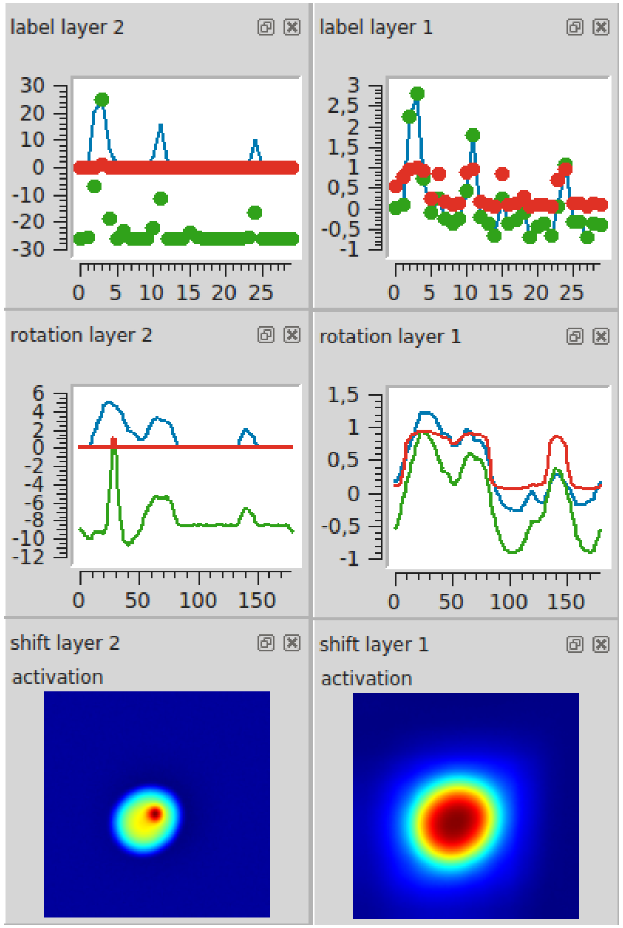




- Recognition rates on “BOIL” (with shape + localized histograms):
  - 90% with one training view/single object
  - See also Faubel, Schöner (2009)

Video

Two-Layer Structure

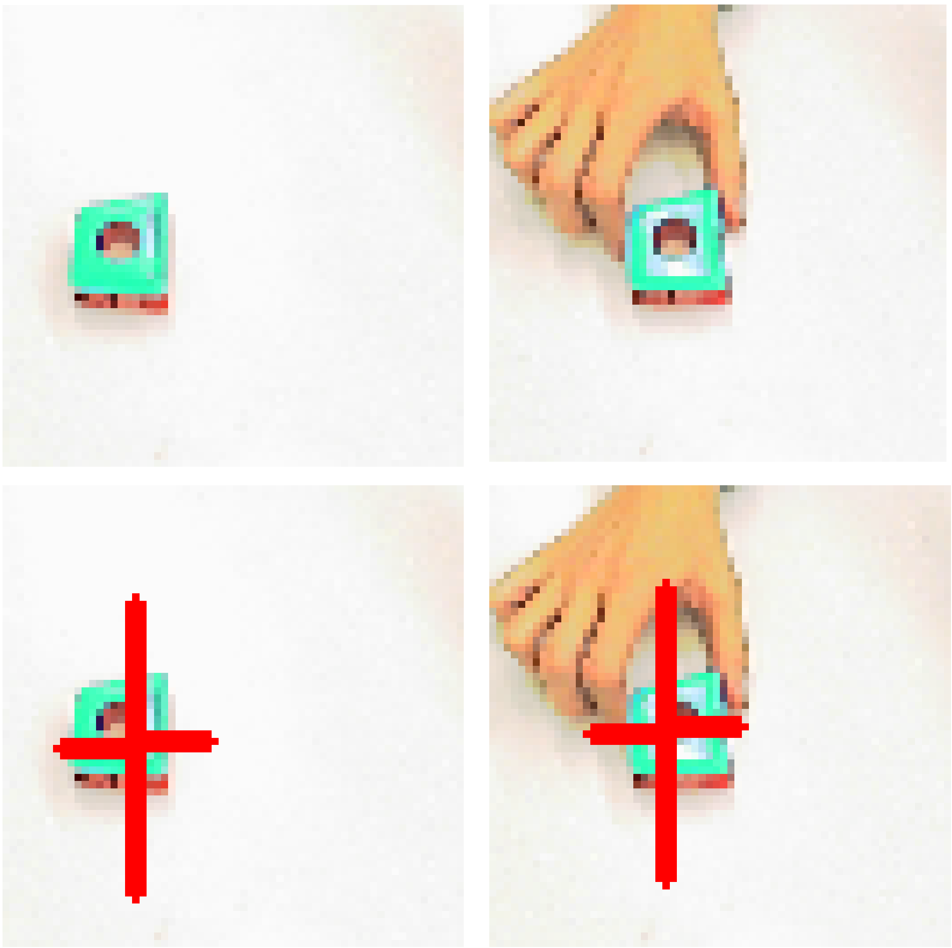


- Layer 1 detects
- Layer 2 selects

Masking: what is it good for?

- Masking input allows to focus on single object

Masking: what (else) is it good for?



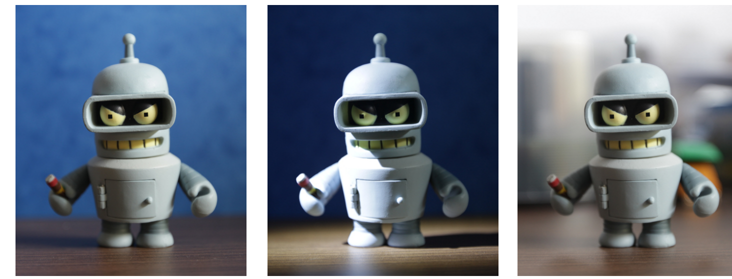
(from Faubel, Schöner 2009)

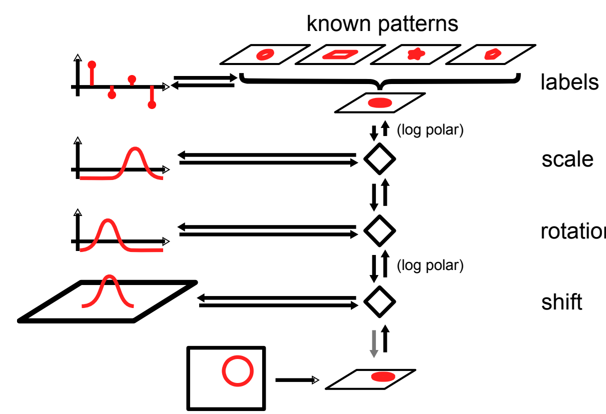
Applications

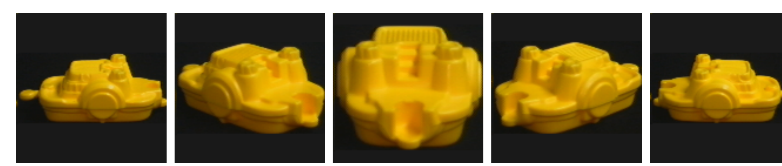
- Grasping
  - Determine grasp parameters for robot arm
  - Use shape templates instead of stored views
- Integration with scene representation
  - Scene representation pre-segments
  - Recognition system estimates exact pose
  - Label information passed to scene representation

Summary

- Object recognition is difficult due to 3D to 2D projection, environment
- Shift estimation via recursive system, weighted superpositions
- Matching via correlation
- Cascaded transformations for rotation, scale estimation
- Localized histograms for more discriminative power
- The benefit of pose estimation







Thank you for your attention!

Questions?

For more: Faubel, C., & Schöner, G. (2009). *A neuro-dynamic architecture for one shot learning of objects that uses both bottom-up recognition and top-down prediction*. In Proc. of the 2009 IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS 2009. IEEE Press.