ComputerScience

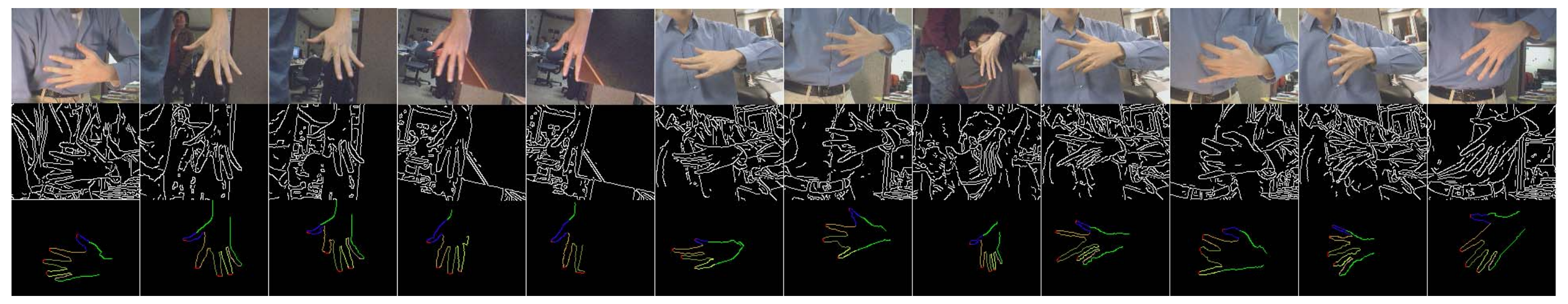
Boston University

Tracking with Dynamic Hidden State Shape Models

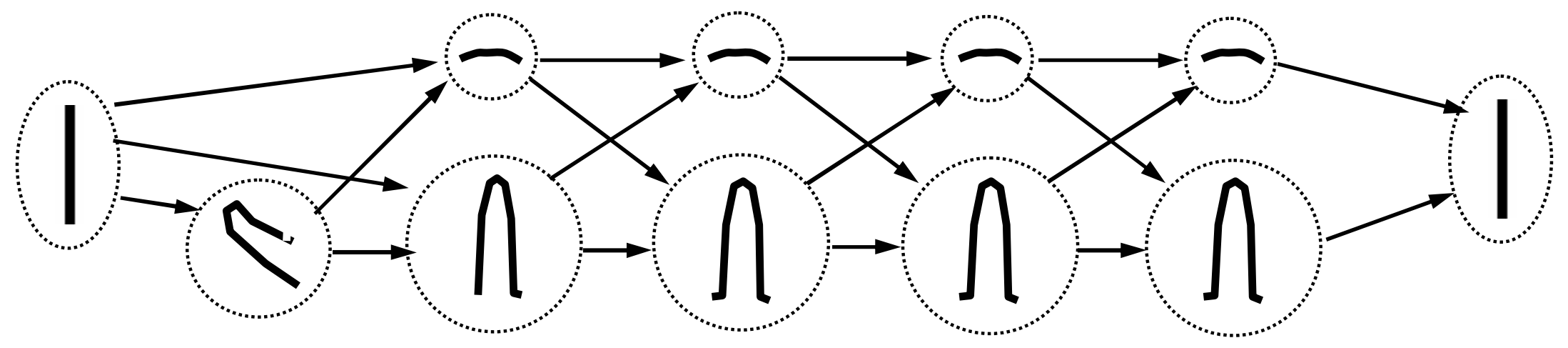
Zheng Wu, Jingbin Wang, Vassilis Athitsos, Stan Sclaroff Advisor: Margrit Betke

How to recognize and track hand in complex scenes?

- Non-rigid deformation of object with variable structure
- Highly cluttered background
- Illumination change and background object's motion
- Occlusion



Dynamic HSSM of Hand



- a set 💧 ={s_1, …, s_M} of states modeling object components,
- a subset ☜ of 💧 that defines legal end states,
- the probability $\pi^{(t)}(s_i)$ that state s_i is the initial state at time t,
- *state transition function* $A^{(t)}(s_i, s_j)$ represents the probability of transiting from s_i to s_j at time t,
- *state observation function* $B^{(t)}(f_u, s_i)$ represents the probability of observing f_u in s_i at time t,
- *feature transition function* $\tau^{(t)}(f_v, f_u, s_j, s_i)$ represents the probability of observing f_v in s_j given some f_u was previously observed in s_i at time t,
- *state duration function* $D^{(t)}(l, s_i, \omega)$ represents the probability of continuously observing l features in s_i at time t, given the prior object scale ω.

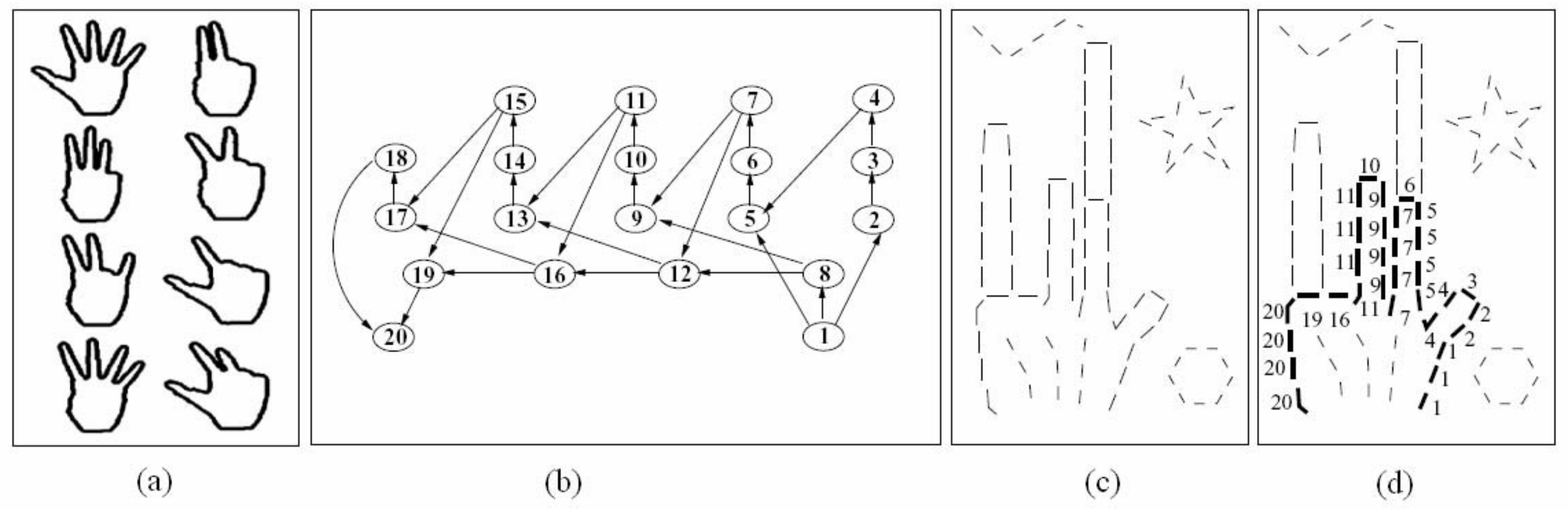


Fig. 1 Recognizing hands in images using DHSSMs. (a) Hand contours with variable shape structure. (b) State transition diagram of the DHSSM "hand." (c) An edge image with segments of the hand contour and clutter. (d) Registration of hand model states to contour segments, resulting in the recognition of the hand structure. (Figure courtesy of Wang et al [1].)

Given a DHSSM and a set of features ☞= {f_1,f_2,…,f_K}, we want to find the most likely model registration such that the joint probability p(Q,O) is maximized, where Q=(q_1,q_2,…q_n) is a valid path in state transition diagram, q_i is state variable assigned with some state label s_j ⚲ 💧; O=(o_1,o_2,…,o_L) is an ordered sequence of object features, o_i is assigned with some feature f_j⚲ ☞; and (o_1…o_{d1}) matched to q_1, (o_{d1+1}…o_{d2}) matched to q_2, … (o_{L-dn+1}…o_L) matched to q_n.

Model Registration by Hierarchical Dynamic Programming

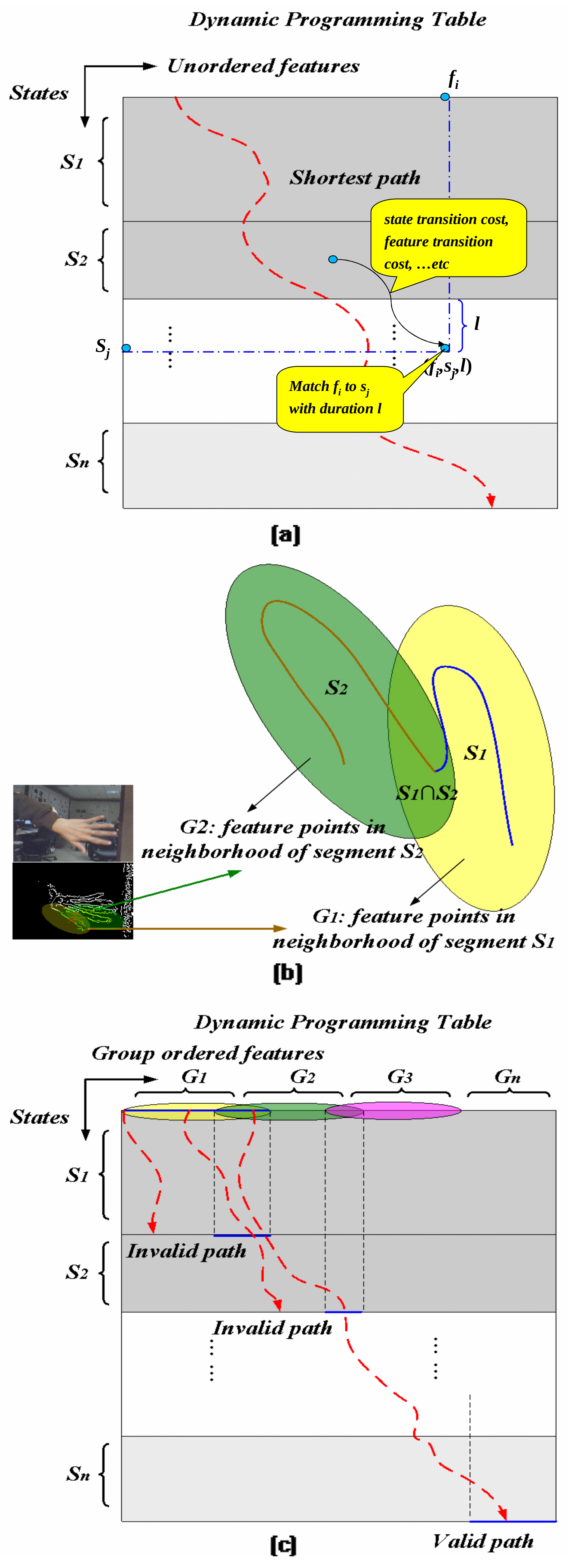

Dynamic Programming Table
States
Unordered features
f_i
S1
S2
s_j
Sn
Shortest path
state transition cost, feature transition cost, …etc
Match f_i to s_j with duration l
l
(f_i,s_j,l)
[a]
S2
S1
S1∩S2
G2: feature points in neighborhood of segment S2
G1: feature points in neighborhood of segment S1
[b]
Dynamic Programming Table
Group ordered features
G1
G2
G3
Gn
States
S1
S2
Sn
Invalid path
Invalid path
Valid path
[c]

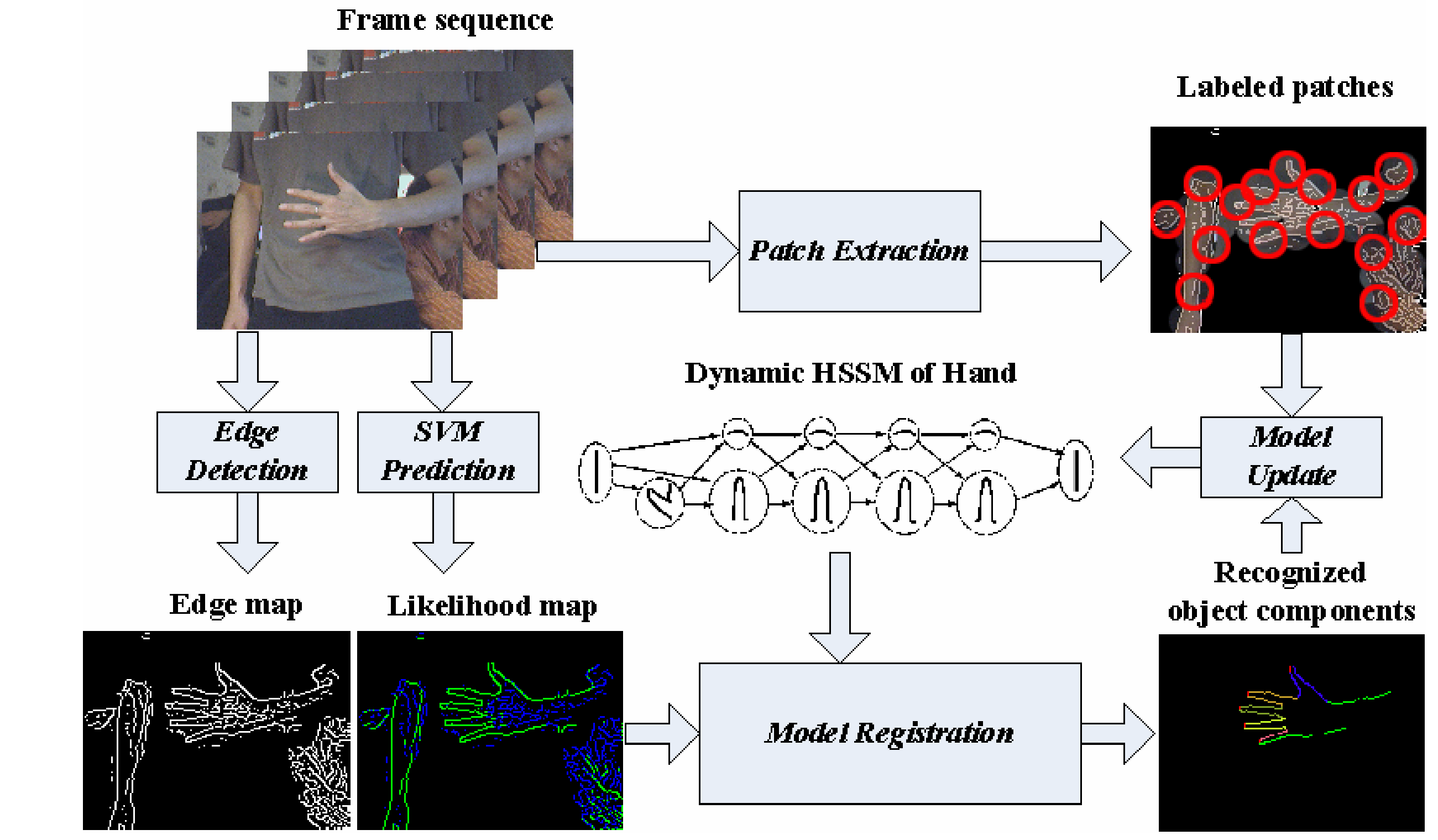
Fig 2. Exploiting the hierarchical structure in DP

The problem can be solved by Viterbi algorithm[2], which essentially is a dynamic programming method that requires O(MK^2) operations. M is the total number of states, K is the total number of features. However, this is still infeasible for tracking scenario because K is usually very large! (thousands of input features)

Feature candidates mapped onto state s_1 are constrained by the yellow search region, while feature candidates mapped onto state s_2 are constrained by the green region. State transition can only happen in the intersection of the two regions

- Unordered features can be roughly grouped based on the optimal registration in the previous frame and constrains the DP search in the current frame
- First apply DP within each group independently and then DP can be applied one more time by selecting one segment from each group and linking them together to form the final shortest path
- Complexity reduced to O(M(K/G)^2) , G is number of groups
- Partial matching is possible by checking segments in each group

Tracking System Overview



- Each input image is first processed to extract features (edges) and feature patches (image regions centered around edges). A two-class SVM determines which features are likely to belong to the object contour (foreground) and which not (background clutter).
- Hand detection is achieved by finding a globally optimal registration between DHSSM states and likely features.
- After each frame, the probability densities of each state of the DHSSM are updated by feeding back the estimated location, orientation of the hand components((finger tips, sides, palm etc.), and amount of curling of each finger (the latter via updating the state duration variable).
- From time to time, a new collection of object and clutter patches are sampled to train a new SVM classifier that better models the current imaging scenario.

Experiments

(1) Data with large motion of the hand and fingers (260 frames)
(2) Data with dense clutter (510 frames)
(3) Data with illumination changes (182 frames)
(4) Data with occlusions (167 frames)

Dataset	Large Motion	Dense Clutter	Illumination Change	Partial Occlusion
Avg. # of features	1200	1800	3000	3000
Hand Localization	98%	92%	83%	75%
Finger Identification	89%	85%	68%	65%
Avg. processing time/frame	1s	1s	1.5s	1.5s

Reference

[1] Wang, J., Athitsos, V., Sclaroff, S., Betke, M.: Detecting objects of variable shape structure with hidden state shape models. IEEE T PAMI 30 (2008) 477-492
[2] Rabiner, L.R.: A tutorial on Hidden Markov Models and selected applications in speech recognition. Proceedings of the IEEE 77 (1989) 257-286
[3] videos can be downloaded from **http://cs-people.bu.edu/wuzheng/video.zip**