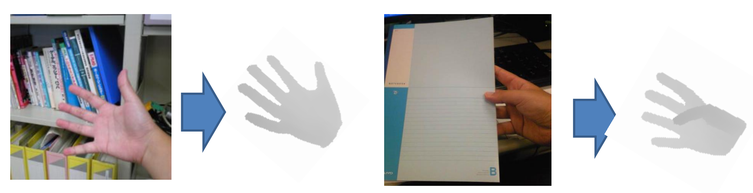
Hand Shape Detection and Classification Based on Co-occurrence of Partial Contour Feature

Yutaka YAMADA, Nobutaka SHIMADA and Yoshiaki SHIRAI (Ritsumeikan University)

Background

Hand shape is a important element for
- gesture interface
- video surveillance

Goal: hand shape detection and classification
- in a complex background
- with partial occlusions by things



Solutions

-Represent global hand feature by *co-occurrence of local contours*

Proposed Method

Hand Shape Representation

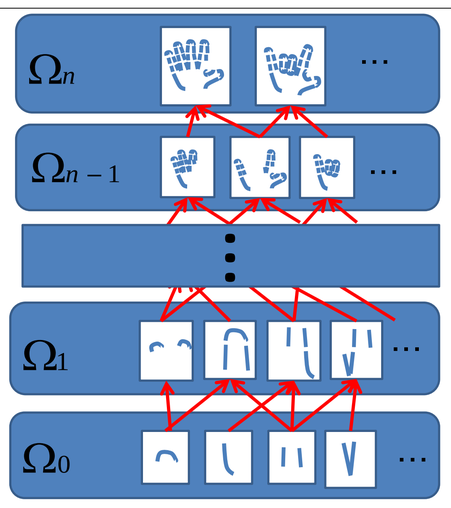
Finger: 3 DoFs x 5
Wrist : 2DoFs

$\mathbf{H}_i = \{\boldsymbol{\theta}_t, \boldsymbol{\theta}_p, \boldsymbol{\theta}_m, \boldsymbol{\theta}_r, \boldsymbol{\theta}_s, \boldsymbol{\theta}_w\}$

Finger joint angles

Wrist rotation angles

Construct a hierarchical codebook from contour features
Represent hand shapes as codes

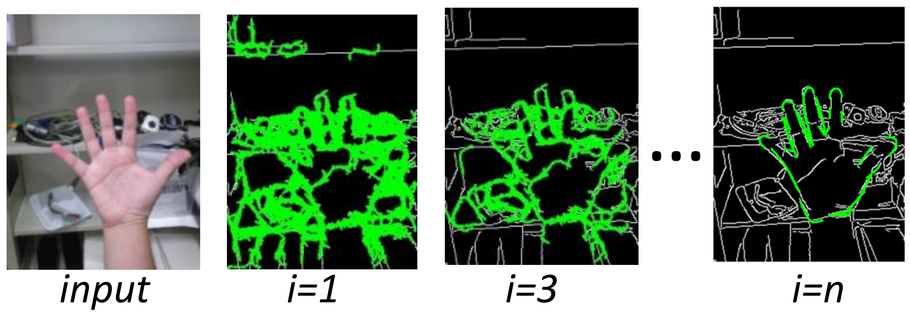


Ω_n :codebook of hierarchy level *n*
$\mathbf{c}_{n,k}$:*k*-th code of hierarchy level *n*
$C(\mathbf{H}_i) = \{\mathbf{c}_{0,p}, \mathbf{c}_{1,q}, \ldots, \mathbf{c}_{h,r}\}$
:Codes found in i-th shape

Detection and Classification

There are *too many* high level codes

Find high level codes by hierarchical search



input *i=1* *i=3* *i=n*

1. Detect local contours $\mathbf{F}_0 = \{\mathbf{f}_{0,1}, \mathbf{f}_{0,2}, \ldots \mathbf{f}_{0,k}\}$ by Oriented Chamfer Matching[1]
2. For *i = 0,1,…,n (i:hierarchy level)*
   - -For all pairs $\{\mathbf{f}_{i,p}, \mathbf{f}_{i,q}\}$,calculate feature vector $\mathbf{g}_{i,p,q} = ((\mathbf{c}_{i,k})_p, (\mathbf{c}_{i,k})_q, \mathbf{X}_{i,q} - \mathbf{X}_{i,p}, \theta_{i,q} - \theta_{i,p})$
   - -match each $\mathbf{g}_{i,p,q}$ to codes in $\boldsymbol{\Omega}_{i+1}$ and assign the best matched code $\mathbf{c}_{i+1,k}$ to $\mathbf{g}_{i,p,q}$
   - -If the matching error to $\mathbf{c}_{i+1,k}$ is less than threshold, add the pair feature $\mathbf{f}_{i+1,l} = (\mathbf{c}_{i+1,k}, \mathbf{X}_{i,p}, \theta_{i,p})$ to $\mathbf{F}_{i+1}$
3. If the assigned code of $\mathbf{f}_{h,i}$ is only found in $C(\mathbf{H}_s)$, it means that the *s*-th shape is found at $\mathbf{X}_{h,i}$ with direction $\theta_{h,i}$

Notation

$\boldsymbol{\Omega}_n = \{\mathbf{c}_{n,0}, \mathbf{c}_{n,1}, \mathbf{c}_{n,2}, \ldots, \mathbf{c}_{n,N}\}$
:codebook of hierarchy level *n*

$\mathbf{c}_{n,k} = (\mathbf{c}_{n-1,p}, \mathbf{c}_{n-1,q}, \mathbf{X}_{n-1,p,q}, \theta_{n-1,p,q})$
:*k*-th code of hierarchy level *h*

$\mathbf{X}_{n-1,p,q}$: Relative position from $\mathbf{c}_{n-1,p}$ to $\mathbf{c}_{n-1,q}$

$\theta_{n-1,p,q}$:Angle between the direction of $\mathbf{c}_{n-1,p}$ and $\mathbf{c}_{n-1,q}$

$\mathbf{f}_{h,i} = ((\mathbf{c}_{h,k})_i, \mathbf{X}_{h,i}, \theta_{h,i})$
: *i*-th feature of hierarchy level *h*

$\mathbf{X}_{h,i}$: position of the i-th feature of hierarchy h

$\theta_{h,i}$: direction of the feature

[1] J. Shotton, A. Blake, R. Cipolla. **Contour-Based Learning for Object Detection.** In *Proc. ICCV*, pages I: 503-510, 2005. (Poster)

Constructing a Hierarchical Codebook

Notation

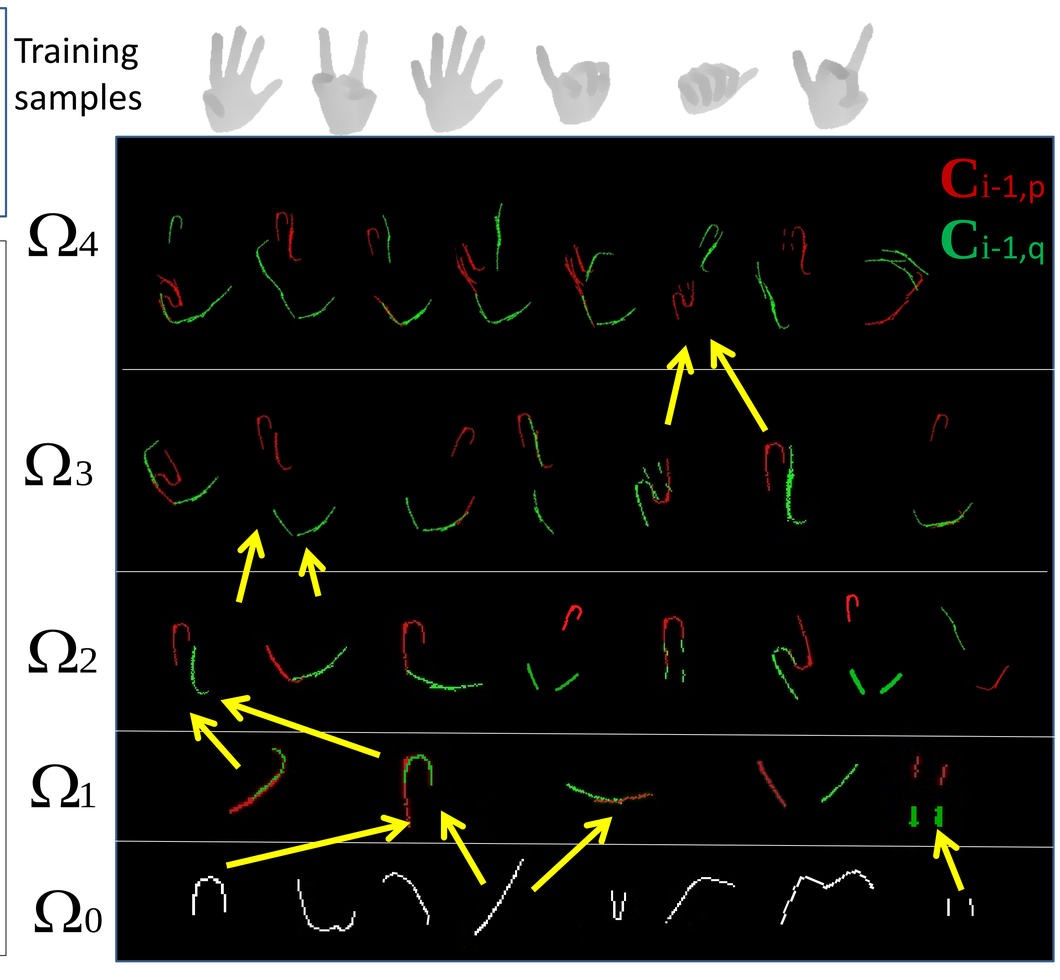
$\mathbf{G}_j$: Collection of feature vectors of all samples
$\mathbf{X}_t$: Relative position of *t*-th cluster center
θ_t : Angle between the direction of codes of the *t*-th cluster center

For *j = 1,2,…,*$N_{hierarchy}$ *(hierarchy level)*

1. For *i* = 1,2,…,$N_{samples}$ *(training samples)*
   - Detect level *j-1* feature $\mathbf{F}^i_{j-1} = \{\mathbf{f}_{j-1,1}, \mathbf{f}_{j-1,2}, \ldots, \mathbf{f}_{j-1,n}\}_i$
   - Calculate $\mathbf{g}^i_{j-1,p,q}$ for all pairs in $\mathbf{F}^i_{j-1}$ and add to these vectors to $\mathbf{G}_{j-1}$
2. For *p* = 1,2,…, $|\boldsymbol{\Omega}_{j-1}|$
   For *q* = *p,p+1*,…, $|\boldsymbol{\Omega}_{j-1}|$
   - Cluster feature vectors of the code pair $(\mathbf{c}_{j-1,p}, \mathbf{c}_{j-1,q})$ in $\mathbf{G}_{j-1}$ by mean-shift clustering
   - If *t*-th cluster has enough elements, add a new code $(\mathbf{c}_{j-1,p}, \mathbf{c}_{j-1,q}, \mathbf{X}_t, \theta_t)$ to $\boldsymbol{\Omega}_j$

Hierarchical Codebook made from 972 CGs

Training samples



Current Works

-Determine criterion for detection (Extracted high level feature can be regarded as hand)
-Determine criterion for discriminating partially occluded hand from complex background