## 手話動画像の認識・分類

**3P-1**

° 田村進一　　　　　　　　　　川崎伸悟

（大阪大学基礎工学部）　　　（大阪大学、現在日本楽器（株））

Abstract:    This paper describes on a method of classifying single view deaf-and-mute sign language motion image.   First, a dictionary for recognizing sign language is made.   Then, the location of a hand or its movement is extracted from the input image sequence.   Further,  the shape of the hand is also extracted.   Finally,  the input sign language image is classified according to the extracted features.

### I.  INTRODUCTION

Our main purpose is not the reconstruction of the human motion from the images but the classification of the sign language motion images.  In this paper,  we deal with only the segmented sign language image;  i.e.,  it expresses not a sentence but a word.  We classify the image sequence only  by the two-dimensional features of the images.   They are the hand shape  (hand configuration, orientation),  location of the hand,  and its movement (direction).  Only the conscious movement with meaning is entered in a dictionary and used for the classification,  since the unconscious  movement has a lot of diversity by individuals.

### II.  NOTATION OF SIGN LANGUAGE DICTIONARY

Entries to the sign language dictionary are movement,  location,  and shape of the right  hand. First, we classify the types of hand movement into still type ST-1, ST-2 and moving type MT-1.

II.1  Direction of hand movement

We define the direction (and the location) of the movement from the view point of the signer.

Plane  of the movement means the place where the hand moves.   We use mainly frontal wall  plane (W)  parallel  to the body.   We further define eight directions of U,   D,   R,   L,   URD (up  right  & diagonal),  ULD,  DRD,  DLD for movement on the frontal wall plane.

II.2  Hand location

Space  means  the space in front of the signer.   It is divided into U,  N (neutral  or  central part),  L (lower part),  and R,  N (center),  L (left).  These are defined with enough overlapping.

II.3  Hand shape

Hand shape is composed of hand configuration and orientation.

Hand configuration is expression method for bending and erecting fingers.

(1)  Basic  form:      Basic form (H) is the state of the hand where fingers are erecting  from  the grip.  The  erected fingers are not contacted with each other in the basic form.   Code number of the fingers is, forefinger 1, middle finger 2, third finger 3, little finger 4, and thumb 5.

(2)  Transconfiguration:      Shape of the basic form fingers are sometimes transformed.   There  are contacting transconfiguration and bending transconfiguration etc.

II.4  Dictionary

An example of the sign language dictionary is shown in Table 1.

### III.  FEATURE EXTRACTION

III.1  Input image sequence

The  input  images  used  in our experiment is from  3/4 inch video tape  "Video  Standard  Sign Language" supervised and compiled by the All Japan Deaf-And-Mute Association.   In this tape, frontal view bust images of the signer are recorded.

Red  component  of  each input image stored once into an optical video disk  is  digitized  into 256X256X8 bit data.

III.2  Tracing hand location

Tracing  of  the  hand location is begun from the waiting position of the  right  hand  till  it returnes to the waiting position again.   First,  the skin area is extracted,  and face area and right hand area are discriminated there.  Then,  the positions of the center-of-gravity of the face area and right hand area are calculated.

III.3  Feature extraction from hand region

Features  of the hand image are extracted by the three stages of contour extraction  from  right hand images,  its polygon approximation,  and feature extraction in the approximated polygon.

We only extract features which are required to classify the hand shape into type 1 or not.  Type 1 is the state that an inside finger is erected.

## IV. MATCHING

Input image sequence is classified first by the location or movement feature of the right hand and then by the shape feature of the hand. In this process, words whose description of the dictionary conflict with the input image features are deleted from a candidate set of sign language words.

### IV.1 Discrimination of still type and moving type

We trace the location of the right hand from the first frame where the hand is at the waiting position. If the movement of the center-of-gravity point of the right hand is less than four pixels in the last four frames, we regard the hand is still. If so, the right hand window image is stored. If the face region and the right hand region is overlapping, and it is after the second still frames, the latest still isolated right hand window image is substituted for it.

### IV.2 Candidate sifting

In the following matching process, if the number of candidates become one, we make it the final classification result. If zero, the result is not obtained.

(1) ST-1:  This is a case where the still locations is one. In this case, first, each entry word of the dictionary is checked whether its description has contradiction or not with the still location of the hand in the input image.  Next, the remained candidates are also sifted by the hand shape check.

(2) ST-2 and MT-1:  When the number of still locationis is two, first the input image features are checked as the ST-2, then checked as the MT-1.

In the checking process of ST-2, two locations of the entry words are compared with the still locations of the hand in the input image.  Then, the remained candidates are also compared with the hand shape of the input image.  In the checking process of MT-1, each entry word are checked with the input image features in the following order:

(a) Direction of trajectory of the right hand.
(b) Two still locations of the initial and final of the right hand movement.
(c) Hand shape at the two still locations.

At present, we have no means to decide whether the input image is ST-2 or MT-1, that is, which is meaning still locations or motion.

## V. EXPERIMENT

Experiment of the sign language classification is done by VAX 11-780 computer for ten words in Table 1. CPU time for extracting the features of each image is several seconds.

Fig.1 shows extracted skin regions corresponding to face and right hand. Fig.2 shows the trajectory of the right hand. Fig.3 shows polygon expressions of the right hand contour.

## VI. CONCLUSIONS

We have developed a system which can read the Japanese sign language.  At present, since we do not use so much of the image features of movement and shape of the hands, the number of readable words is limitted.
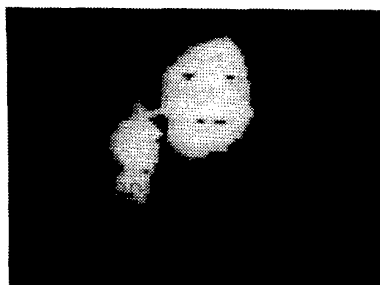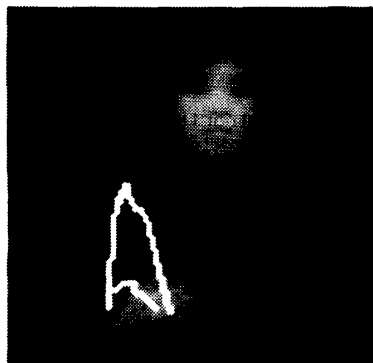


Fig.1  Skin regions



Fig.2  Hand movement "WD".

Table 1   Dictionary of sign language

| word | hand shape | | hand location | | hand movement |
|---|---|---|---|---|---|
| | 1st still | 2nd still | 1st still | 2nd still | |
| man | H5Fu | ---- | SNR | ---- | ST-1 |
| woman | H4FuP- | ---- | SNR | ---- | ST-1 |
| father | H1FuP- | H5Fu | CKR | SUR | ST-2 |
| mother | H1FuP- | H4FuP- | CKR | SUR | ST-2 |
| elder brother | H2FuP- | H2FuP- | SNR | SUR | WU |
| younger brother | H2FuP- | H2FuP- | SNR | SLR | WD |
| elder sister | H4FuP- | H4FuP- | SNR | SUR | WU |
| younger sister | H4FuP- | H4FuP- | SNR | SLR | WD |
| child | C_1234_5Pd | C_1234_5Pd | SLL | SLR | WR |
| chief | H5Fu | H5Fu | SNR | SUR | WU |
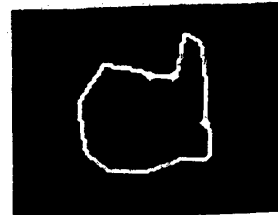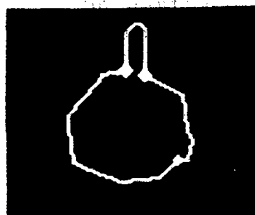


Fig.3  Polygons of right hand contour.