

# Extraction of Hand Features for Recognition of Sign Language Words

Nobuhiko Tanibata  
tanibata@cv.mech.eng.osaka-u.ac.jp

Nobutaka Shimada  
shimada@cv.mech.eng.osaka-u.ac.jp

Yoshiaki Shirai  
shirai@cv.mech.eng.osaka-u.ac.jp

Computer-Controlled Mechanical Systems, Graduate School of Engineering, Osaka University

## Abstract

*This paper proposes a method to obtain hand features from sequences of images, where a person is performing the Japanese Sign Language (JSL) in a complex background and to recognize the JSL word.*

*At the first frame, we find a person's region, and then search for a face, hands in order to determine a range of skin color and search for elbows to determine the position of a wrist.*

*At each frame, we track the face, the hands by using the decided skin color and track the elbows by matching the template of a elbow shape. When face and hands overlap, they are extracted by matching texture templates of the previous face and hands. Hand features such as the hand direction, the number of fingers, etc. are extracted from the hand regions and the wrist.*

*In order to recognize JSL words, we use a sequence of the hand features as an input to HMM. We first select words which reach the final state of HMM, and then determine one with the highest probability. We made an experiment with real images of a professional JSL interpreter and recognized 65 JSL words successfully.*

## 1 Introduction

Because Japanese Sign Language (JSL) is the most popular gesture language in Japan and common people cannot understand JSL, a system to interpret JSL is useful. Such a system requires two functions; translating Japanese into JSL and vice versa. Although the former has already developed, but the latter has not for practical use. The difficulties in the latter are extraction of hand shapes and motion and recognition of a sequence of them.

For recognition of a JSL word, we should extract hand shapes as well as motions to obtain hand features because there is a JSL word with the same motions as the other. The

CyberGlove is often used[1] to extract them. However it is trouble for users to put on and take off, or wear it. Thus we focus on extracting hand features from a sequence of images without any devices on hands.

On the other hand, vision-based extraction of the hand features has studied in measuring hand shape as well as a field of sign language recognition. The former work[2] obtains hand shapes by matching hand silhouette with hand 3D CG in a simple background. But it is difficult in a complex background. The latter works[3, 4] extract skin regions from a range of skin color. However that cannot be completely done because skin color depends on each user or each situation.

Vision-based extraction of the hand features often includes errors due to image noises or background whose color is similar to skin color. For recognition of a sequence of hand features robustly, Hidden Markov Model (HMM) is one of effective methods[8]. This is often used for recognitions of facial expressions[5], gestures[6] and human behavior[7]. However it is difficult to recognize the JSL words only with probabilities obtained by HMM. Because some JSL words are similar to a part of other words and the probabilities of them are also high.

In our method, skin regions, clothes region and elbows are roughly extracted at the first frame by matching the template of a initial pose to a person's region. The template includes face and hand regions, clothes region and the positions of elbows. The rough skin regions include the parts of clothes region. We remove them from the rough skin regions using color information of clothes region. The skin color depends on the user and it is easily affected by lighting, backgrounds, and reflected light from the user's clothes. In order to cope with this problem, we determine the range of the skin color in HSV color space from the extracted skin regions. At each frame, our system tracks the face, the hands and the elbows by using the range of the skin color and the templates of shape. When the face and the hands overlap,

