Hand Shape Estimation Using Sequence of Multi-Ocular Images Based on Transition Network

Yasushi HAMADA, Nobutaka SHIMADA, Yoshiaki SHIRAI
Dept.of Computer-Controlled Mechanical Systems, Osaka University
2-1 Yamadaoka Suita-shi, 565-0871 Japan
hamada@cv.mech.eng.osaka-u.ac.jp

Abstract

This paper presents a method of hand posture estimation from silhouette images taken by multiple cameras. For each image, we extract a feature vector from the silhouette contour of the hand. We construct an eigenspace by the feature vectors extracted from the hands of various postures. The feature vectors projected into the eigenspace are registered as models. The matching criterion of each images is defined as the distance to the model. The hand shape is estimated by retrieving the registered model well-matching to the input. For effective matching, we define a shape complexity for each image to see how well the shape feature is represented. For a set of input images taken by multiple cameras at each time, the total matching criterion is evaluated by combining the matching criteria of the set of images using the shape complexities.

For rapid processing, we limit the matching candidate by using the constraint on the shape change. The possible shape transition is represented by a transition network. Because the network is hard to build, we apply offline learning, where nodes and links are automatically created by showing examples of hand shape sequences. We show experiments of building the transition networks and the performance of matching using the network.

1 Introduction

Recently image-based human interfaces and understanding the hand gestural languages have attracted increasing attentions as an alternative to traditional input devices like mouses or keyboards. Such attempts previously proposed are approximately divided into two categories.

The first category is the 3-D model-based approach including the model fitting methods [1] and "Estimation by Synthesis(ES)" methods [2, 3] which match possible postures generated from a given 3-D shape model and search for the postures best-matched to the input image. While these methods are effective for estimation of arbitrary hand postures, they often require much computation.

The second category directly matches the image features to those of models. The methods of this category [4, 5, 6, 7, 8] register the image appearances or the image features in the learning sequences, and then the input sequence is classified into one of the registered sequence. For estimation of a limited set of hand postures, only useful models are registered. Moreover, computation is usually less because 3-D shapes are not estimated.

For estimation of hand shapes in a gesture sequence, however, the first category is more effective because it is able to limit the search space by the constraint of the joint angles or by that of the velocity. The second category, on the other hand, has to try to match every models. This problem was solved by applying the Hidden Markov Model (HMM). However, a sequence model has to be built for every gesture sequence.

This paper proposes a method of matching a given hand posture just like the second category, while limiting the candidates by a transition network. The transition network has nodes which represent typical hand shapes and links which represent possible hand shape changes. The network alone represents the transition of all possible gestures, and is built automatically during a learning phase. In speech and gesture recognition [9], the transition network is used for integration of speech and gesture.

First, in this paper, a basic matching method is described. Because matching with images taken by monocular camera is often ambiguous, we use a set of images taken by multiple cameras. We determine the features for a set of images to estimate the hand posture. We collect various hand images to make the model of the postures. A silhouette is extracted from each image and the feature vector is computed as a sequence of the distances from the center of the silhouette to the contour points. The eigenvectors are determined from all feature vectors. The feature vectors projected into the eigenspace are registered as models.

The matching criterion of each images is defined as the distance to the nearest model. The hand shape is estimated by retrieving the registered model well-matching to the in-