Chapter 11

Vision Based Hand Posture Recognition

Kongqiao Wang
Nokia Research Center, China

Yikai Fang
Nokia Research Center, China

Xiujuan Chai
Chinese Academy of Sciences, China

ABSTRACT

Vision based gesture recognition is a hot research topic in recent years. Many researchers focus on how to differentiate various hand shapes, e.g., the static hand gesture recognition or hand posture recognition. It is one of the fundamental problems in vision based gesture analysis. In general, most frequently used visual cues human uses to describe hand are appearance and structure information, while the recognition with such information is difficult due to variant hand shapes and subject differences. To have a good representation of hand area, methods based on local features and texture histograms are attempted to represent the hand. And a learning based classification strategy is designed with different descriptors or features. In this chapter, we mainly focus on 2D geometric and appearance models, the design of local texture descriptor and semi-supervised learning strategy with different features for hand posture recognition.

INTRODUCTION

As the prevalence of ubiquitous computing, traditional user interaction approaches with mouse, keyboard and touch pen are not convenient enough for them. In addition, many emerging applications such as augmented reality and interactive entertainments require natural and intuitive interface. Moreover the limited input space on traditional mobile or hand held device leads to encumbered experience with tiny keyboard or touch screen. Hand gesture is frequently used in people’s daily life. It’s also an important component of body languages. So a natural interaction between humans and computing devices can be achieved if hand gestures can be used for communication between human and computing devices.

DOI: 10.4018/978-1-60960-024-2.ch011
Vision based hand posture recognition is attracting more attention due to its suitability for ubiquitous computing and emerging applications. In recent years, many researchers have focused on how to differentiate various hand shapes, such as static hand gesture recognition or hand posture recognition. It is one of the key problems in vision-based gesture analysis. However, hand posture recognition is still an unresolved problem. Hand is an articulated object with variable configurations, shapes, and structures. In addition, it's difficult to describe the texture in hand area. To have a good representation of hand area, we try to use local features such as rectangular features and local texture histograms to represent the hand. Besides, rule-based and semi-supervised classification strategies are designed to differentiate various hand postures. The comparison between the methods in this chapter and other methods are carried out and the results show the effectiveness of the proposed methods in this chapter.

**BACKGROUND**

Methods for vision-based hand posture recognition fall into two categories: 3D model-based methods and appearance-based model-based methods. 3D model may exactly describe hand movement and its shape, but most of them are computational expensive to use. Recently there are some methods to obtain 3D model with 2D appearance model such as ISOSOM (Haiying, Rogerio & Matthew, 2006) and PCA-ICA (Makoto, Yenwei and Gang, 2006). For the consideration of easy implementation and non-intrusive characteristic, we prefer to use 2D model and appearance-based methods for hand posture recognition in this chapter.

Freeman and Weissman (1995) recognized gestures for television control using normalized correlation. This technique is efficient but may be sensitive to different users, deformations of the pose and changes in scale, and background. Cui and Weng (1996) proposed a hand tracking and sign recognition method using appearance-based method. Although its accuracy was satisfactory, the performance was far from real-time. Elastic graphs were applied to represent hands in different hand gestures in Triesch’s work with local jets of Gabor filters (Triesch & Malsburg 1996). It locates hands without separate segmentation mechanism and the classifier is learned from a small set of image samples, so the generalization is very limited. These model-based methods are intuitive for hand representation. With elaborate design of local features, it's not necessary to use complicated classification strategy. Bretzner, Laptev & Lindeberg (2002) use scale-space feature detection to decompose hand into palm and fingers. The decomposition is intuitive and effective. However, the detection involves lots of Gaussian convolution across images and brings high time consumption in practice.

Ong and Bowden (2004) distinguished hand postures with boosted classifier tree and obtained fairly good results. However, the classifier in their method was too complicated and time-consuming. In addition, the samples are with simple and similar backgrounds and the training requires thousands of labeled samples. Kolsch (2006) employed fanned boosting detection for classification and got nearly real-time results, while the training process is extremely time-exhausting. Just, Rodriguez and Marcel (2006) introduce modified census transform (MCT) into hand gesture classification. Their method gives fairly good average accuracy above 80 percent with the classifier trained with more than 2,000 samples per posture, while the performance in recognition experiments under complex background was not much satisfactory. These prevalent methods for recognition are learning-based with specific features. Compared with model and rule-based methods, learning mechanism make discriminative/generative classifiers that have strong adaptiveness to different set of examples. The training that converges on abundant