Chapter 56
A Comparative Study of Machine Learning Techniques for Gesture Recognition Using Kinect

Rodrigo Ibañez
ISISTAN (UNICEN-CONICET) Research Institute, Argentina

Alfredo Raul Teyseyre
ISISTAN (UNICEN-CONICET) Research Institute, Argentina

Alvaro Soria
ISISTAN (UNICEN-CONICET) Research Institute, Argentina

Luis Berdun
ISISTAN (UNICEN-CONICET) Research Institute, Argentina

Marcelo Ricardo Campo
ISISTAN (UNICEN-CONICET) Research Institute, Argentina

ABSTRACT

Progress and technological innovation achieved in recent years, particularly in the area of entertainment and games, have promoted the creation of more natural and intuitive human-computer interfaces. For example, natural interaction devices such as Microsoft Kinect allow users to explore a more expressive way of human-computer communication by recognizing body gestures. In this context, several Supervised Machine Learning techniques have been proposed to recognize gestures. However, scarce research works have focused on a comparative study of the behavior of these techniques. Therefore, this chapter presents an evaluation of 4 Machine Learning techniques by using the Microsoft Research Cambridge (MSRC-12) Kinect gesture dataset, which involves 30 people performing 12 different gestures. Accuracy was evaluated with different techniques obtaining correct-recognition rates close to 100% in some results. Briefly, the experiments performed in this chapter are likely to provide new insights into the application of Machine Learning technique to facilitate the task of gesture recognition.

DOI: 10.4018/978-1-7998-2460-2.ch056

INTRODUCTION

Recent progress in entertainment and gaming systems has brought more natural and intuitive human–computer interfaces to our lives. For example, 3D depth cameras, such as Xbox Kinect, enable the recognition of body gestures and allow a more direct and expressive way of interaction. Kinect allows players to control games through full-body movement without using a remote control. Moreover, Kinect has promoted the development of new natural interaction applications to many domains among larger audiences (e.g., art, education and health, among others).

Although Kinect is able to recognize the position of users' joints, developers are still left with the time-consuming and tedious task of recognizing gestures. In this sense, several approaches to recognize gestures easily have been proposed. The first applications were rule-based approaches that relied on a set of parameters and thresholds on body-part locations (e.g., "RightHand.y>Head.y + 0.5") to recognize simple movements. However, this kind of approach has become an error-prone process that requires domain knowledge, experience, and effort to ad hoc define a set of rules or heuristics to recognize human body gestures. Moreover, making rules to recognize complex gestures like a Smash in a tennis game becomes impractical (Ibañez, Soria, Teyseyre, & Campo, 2014).

In order to overcome these problems and provide a more flexible and robust mechanism to perform high-level gesture recognition, other approaches were proposed using Supervised Machine Learning (SML) techniques. These techniques require a set of labeled training gestures to learn and subsequently identify a new gesture as one of the learned gestures. For example, Bhattacharya, Czejdo, and Perez (2012) used Support Vector Machines (SVM) and Decision Trees (DT) for gesture recognition in a military application. Another successful approach is based on the Dynamic Time Warping algorithm (DTW) (Waithayanon & Aporntewan, 2011).

In this context, the aim of this chapter is twofold. First, a brief survey of current efforts on the application of SML techniques for gesture recognition is presented. Second, a comparative analysis of the performance of various state-of-the-art algorithms is performed. In particular, accuracy of well-known techniques used in Natural User Interaction is compared. The techniques involved in this chapter are Dynamic Time Warping (DTW) (Bhattacharya et al., 2012), Procrustes Analysis (Ross, 2004), Hidden Markov Models (Rabiner, 1990) and Naive Bayes (Russell & Norvig, 2009).

The remainder of this chapter is organized as follows. Section 2 presents an overview of related works. Section 3 describes a Machine Learning approach to gesture recognition and a description of the evaluated techniques. Section 4 discusses the experiments and results, along with the benefits and drawbacks of each technique. Finally, Section 5 presents the conclusions and identifies future lines of work.

BACKGROUND

In the literature, there are numerous approaches to gesture recognition from human body movements captured by video cameras. As mention in (Gavrila, 1999), the ability to recognize humans and their activities by vision is crucial for a machine to interact intelligently and effortlessly with a human-inhabited environment. Over the years, there has been strong interest in human movement from a wide variety of disciplines. In psychology, there have been the classic studies on human perception by Johansson (1973) or, in the hand gesture area, how humans use and interpret gestures (McNeill, 1992). In kinesiology the goal has been to develop models of the human body that explain how it works mechanically and