Engineering Applications of Artificial Intelligence 23 (2010) 1216–1228

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

journal homepage: www.elsevier.com/locate/engappai

Layered architecture for real time sign recognition: Hand gesture and movement

A. Ibarguren [a,*], I. Maurtua [a], B. Sierra [b]

[a] Fundación Tekniker, Spain
[b] Department of Computer Science and Artificial Intelligence, University of the Basque Country, Spain

ARTICLE INFO

Article history:
Received 23 July 2008
Received in revised form
19 May 2010
Accepted 2 June 2010
Available online 20 June 2010

Keywords:
Machine learning
Sign recognition
Movement recognition
Signal processing
Human machine interface

ABSTRACT

Sign and gesture recognition offers a natural way for human–computer interaction. This paper presents a real time sign recognition architecture including both gesture and movement recognition. Among the different technologies available for sign recognition data gloves and accelerometers were chosen for the purposes of this research. Due to the real time nature of the problem, the proposed approach works in two different tiers, the segmentation tier and the classification tier. In the first stage the glove and accelerometer signals are processed for segmentation purposes, separating the different signs performed by the system user. In the second stage the values received from the segmentation tier are classified. In an effort to emphasize the real use of the architecture, this approach deals specially with problems like sensor noise and simplification of the training phase.

1. Introduction

The new human–computer interaction technologies offer increasingly natural ways to operate and communicate with machines. Ranging from voice to vision, all these interaction technologies are helping to drastically change the way people operate computers. Within all these interaction methods, gesture recognition plays an important role due to the universal use of signs and gestures for human communication (Stokoe, 1972). This fact increases the relevance of sign recognition, pointing it as a growing research area with countless application fields.

Even so, sign recognition raises new technical challenges. Classification, segmentation, training and sensor noise processing are only a few of the issues that must be tackled in this kind of system. To this end a wide range of architectures and algorithms are being posed,[1] although the complexity of each of the tasks makes it difficult to develop a complete system.

The research presented here is based on the real time recognition of the Fingerspelling Alphabet, an alphabet used by the deaf. The use of data gloves and accelerometers has been selected for this work from the different technologies available for gesture and sign recognition, specifically a 5*DT Data Glove* 14 *Ultra*,[2] a glove typically used for motion capture in computer animation, and a *MTx*,[3] an accelerometer providing information about acceleration, angular velocity and orientation.

The machine learning (Mitchell, 1997; Russell and Norvig, 1995) field offers a wide range of algorithms capable of extracting rules and patterns from the available data. This feature makes the use of these algorithms suitable for the problem of sign recognition. Even so, the real time nature of gesture recognition brings to light new issues not present in other kinds of classification problems.

On account of the problems related to the real time application of machine learning algorithms, this paper presents a two-tier architecture for sign recognition. In the first stage, the glove and accelerometer signals are processed for segmentation purposes, identifying the different signs performed over time. In the second stage, different classifiers are used to recognize the signs identified in the segmentation phase, taking advantage of the features of hierarchical classifiers and signal processing to obtain a higher recognition rate. Some experiments are also carried out to test the accuracy of the proposed architecture. The results obtained from this two-tier approach are promising.

The rest of the article is structured in the following manner: Section 2 gives information about previous research activities related to this paper. Section 3 describes the Fingerspelling Alphabet used in this research. In Section 4, the data glove and the accelerometer used in the experiments are described. Section 5 presents the proposed architecture for gesture recognition and the

* Corresponding author.
E-mail addresses: aibarguren@tekniker.es (A. Ibarguren), b.sierra@ehu.es (B. Sierra).

[1] http://www.cybernet.com/~ccohen/
[2] http://www.5dt.com
[3] http://www.xsens.com

0952-1976/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.engappai.2010.06.001

experiments done to validate it. In Section 6 the proposed architecture for movement recognition is presented, including the validation process applied. Section 7 describes the system implementation process carried out, including the new functionalities, the complete architecture and a final validation. Finally Section 8 presents the conclusions obtained and the future work to be done.

2. Related work

Gesture recognition opens a new research area with multiple application domains. The different technologies available nowadays make it possible to address the recognition task from many perspectives. Ranging from cameras to sensors, many different approaches to the posed problem can be suggested, each of them with its specific features and difficulties.

The use of cameras is common practice in sign and gesture recognition. Several approaches have been presented in an attempt to overcome the specific problems of vision-based recognition. Dadgostar (2005) proposed a three-tier system for gesture recognition. The first two tiers were devoted to skin detection and object tracking, respectively, trying to find the areas of interest and the movement to classify. In the third-tier, the gesture recognition task was performed by means of Neural Networks. Hasanuzzaman et al. (2007), on the other hand, presented a knowledge-based gesture recognition platform using a multi-cluster based learning method. In this approach static and dynamic gesture images were analyzed and classified using a hierarchical database of labelled images. Brashear et al. (2003) proposed a combined approach, incorporating an accelerometer to increase the recognition rates of the vision-based sign recognition. Data provided from both a hat-mounted camera and an accelerometer is processed by hidden Markov models (HMM) to recognize gestures from a restricted sign language.

In the same way, several works based on the use of data gloves could also be found. Due to the wide range of application domains of data gloves, it is possible to find articles tackling issues such as the analysis of different classification systems to identify objects (Heumer et al., 2007; Ibarguren et al., 2006) or the use of the gloves for cooperation and interaction with robots (Dillmann, 2003).

Concentrating on sign recognition through data gloves, diverse literature can also be found. Kadous (1996) has worked on the recognition of some AUSLAN signs (Australian Sign Language) by means of Decision Trees and Instance-Based Learning. The proposed approach initially extracts different features from the information provided by the glove (features such as distance, energy and time), in an effort to characterize the gesture. This information is used afterwards for classification purposes by means of Decision Trees and Instance-Based Learning. In the same way, Fels and Hinton (1993) have developed a system that relates hand gestures to speech through an adaptive interface. To this end three different Neural Networks are used to process the information provided by the glove, dividing the classification task into different stages.

Classifier combination (Gunes et al., 2003; Ho et al., 1994; Lu, 1996) is one of the most interesting research fields within machine learning because it makes possible an increase in the recognition rates of the basic classifiers by their combination. Even so, the ways of combining and selecting the classifiers is still an open research topic due to the wide range of possibilities which exist. One of the main ways to combine the classifiers is by creating hierarchies, although there are countless ways in which this can be done. Jordan and Jacobs (1993) proposed the use of the Expectation-Maximization (EM) algorithm for adjusting the parameters of a tree-structured architecture. In the same way, Martínez-Otzeta et al. (2006) propose the use of genetic algorithms for the creation of the hierarchy and the election of the different basic classifiers. These different approaches show the complexity of the task and the variety of methods used to tackle it.

Finally feature subset selection (FSS) (Liu and Motoda, 1998; Inza et al., 2000; Peng et al., 2005) offers a useful tool to increase the recognition rates of classifiers as well as uncover the significance of the different variables available. Among the different techniques for feature selection, genetic algorithms (Goldberg, 1989; Bäck, 1996) are being increasingly used for this task (Yang and Honavar, 1998) due to their capacity to escape from local minimums/maximums. Previous experiments around data gloves (Ibarguren et al., 2006) have shown the suitability of using this kind of Evolutionary Algorithms in recognition problems.

3. Sign language

The Fingerspelling Alphabet used in this experiment is formed of 30 signs describing the different letters of the Spanish alphabet (English alphabet plus *ch*, *ll*, *ñ* and *rr*). Those signs are performed combining both hand gestures and hand movements. Based on the sensors chosen for this research, the first group of signs (hand gestures) will be tracked by means of data gloves while for the second group (hand movements) the previously presented accelerometer will be used.

3.1. Hand gesture

Focusing on sign recognition using only data gloves, 12 of the 30 signs have been removed as in their case, it was necessary to know the hand position (upwards or downwards) or the hand movement of the speaker after performing the sign. This has reduced the number of signs to be recognized to 18, resulting in an alphabet consisting of the letters $\{a,b,c,d,e,f,g,h,i,k,l,o,p,q,r,t,u,x\}$, see Fig. 1.

Besides the signs constituting the Fingerspelling Alphabet, two more classes have been introduced related to the real time recognition:

- *Hand at rest*: This class describes the hand while performing no sign, represented by an open hand. This "sign" is used when starting or finishing the spelling or to separate the different words of the sentences. Due to its significance in the spelling process, this class is treated as a new "sign" during the experiment.
- *Hand in transition*: This class describes the hand in motion while changing from one sign to another. It is not strictly a hand gesture (countless gestures are produced while changing signs). This class just denotes that the user's hand is in movement, something which significantly increases the difficulty in its detection and justifies the special treatment of this class during the experiment.

With these additions, the total number of classes rises to 20 in this recognition problem: the 18 signs, the hand at rest and the hand in transition.

3.2. Hand movement

To be able to represent the other 12 letters removed previously, the Spanish Fingerspelling Alphabet uses six hand movements that are combined with hand gestures, movements shown in Fig. 2. Those six movements require one to (I) represent a J with the hand, (II) move the hand left-and-right twice, (III) flex the hand downwards, (IV) represent a circle with the hand, (V) move the hand up-and-down twice and (VI) represent a Z with the hand. Combining those movements with the aforementioned