Journal of Systems Architecture 68 (2016) 1–16

Contents lists available at ScienceDirect

Journal of Systems Architecture

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

An efficient static gesture recognizer embedded system based on ELM pattern recognition algorithm

Lucas F.S. Cambuim*, Rafael M. Macieira, Fernando M.P. Neto, Edna Barros, Teresa B. Ludermir, Cleber Zanchettin

Informatics Center - Federal University of Pernambuco, Av. Jornalista Anibal Fernandes, s/n - Cidade Universitaria - Recife-PE, Brazil

ARTICLE INFO

Article history:
Received 2 May 2015
Revised 25 May 2016
Accepted 2 June 2016
Available online 14 June 2016

Keywords:
Embedded systems
Computer vision
Neural network
FPGA

ABSTRACT

Millions of people throughout the world describe themselves as being deaf. Some of them suffer from severe hearing loss and consequently use an alternative manner with which to communicate with society by means of either written or visual language. There are several sign languages capable of dealing with such a need. Nonetheless, a communication gap still exists even when using such languages, since only a small fraction of the population is able to use them. Over the last few years, due to the increasing need for universal accessibility when using computational resources, gesture recognition has been widely researched. Thus, in an attempt to reduce this communication gap, our approach proposes a computational solution in order to translate static gesture symbols into text symbols, through computer vision, without the use of hand sensors or gloves. In order to guarantee the highest quality, with emphasis on the reliability of the system and real-time translation, we have developed an approach based on the Extreme Learning Machine (ELM) pattern recognition algorithms fully implemented in hardware, and have assessed it to measure these two metrics. Hardware components were designed in order to perform the best image processing and pattern recognition tasks used within the project. As a case study, and so as to validate the technique, a recognition system for the Brazilian Sign Language (LIBRAS) was implemented. Besides ensuring that this approach could be used for any static hand gesture symbol recognition, our main goal was to guarantee fast, reliable gesture recognition for communication between humans. Experimental results have demonstrated that the system is able to recognize LIBRAS symbols with an accuracy of 97%, a response time of 6.5ms per letter recognition, and using only 43% (about 64,851 logic elements) of the FPGA area.

1. Introduction

Millions of people throughout the world suffer from some degree of hearing loss, a large number of whom describe themselves as being deaf. In the case of severe hearing loss, individuals are able to hear almost nothing.

Frequently, those with severe hearing loss use an alternative manner with which to communicate, since they are unable to use the voice as their primary means of communication. Moreover, in some cases, people are either unable to use a hearing aid, or simply do not wish to. Thus, alternative communicational methods very often make use of hand gestures.

There are several sign languages based on hand gestures, such as the Brazilian Sign Language (LIBRAS), the American Sign Language (ASL) and the British Sign Language (BSL). In some countries this form of communication is considered an official second language and, in several places there are even laws to ensure the inclusion of sign language courses at universities.

Although a communication gap certainly still remains between the deaf and the non–deaf, this is largely due to the fact that most of the population does not suffer with hearing impairment and hence do not know how to communicate using sign language. Another important issue to be highlighted is that the communication difficulties imposed by being deaf jeopardize the social inclusion of deaf people.

Consequently, there is a demand for systems that enable easier, more reliable communication between the deaf and the non-deaf. Recently, through advances in computational systems, many approaches have been proposed to expand the means of communication amongst humans.

* Corresponding author.
E-mail addresses: lfsc@cin.ufpe.br, lucascambuim@gmail.com (L.F.S. Cambuim), rmm2@cin.ufpe.br (R.M. Macieira), fmpn2@cin.ufpe.br (F.M.P. Neto), ensb@cin.ufpe.br (E. Barros), tbl@cin.ufpe.br (T.B. Ludermir), cz@cin.ufpe.br (C. Zanchettin).

http://dx.doi.org/10.1016/j.sysarc.2016.06.002

One possibility is to use gesture recognizers to implement new resources for communication. In recent years, due to the increasing need for universal accessibility of computing resources, gesture recognition has been widely researched, and involves various types of problems, including recognition of the hand structure and movement, the body and the face [2]. Data regarding hand behavior may be obtained through several sources, such as hand sensors [3] and images achieved with one or more cameras pointed towards the hands.

In general terms, gestures can be classified as static [5,13] and dynamic [2,6,11]. Static gestures are usually described in terms of hand shapes or poses, without the need for hand movement, while dynamic gestures are generally described according to the movements of fingers, hands or both [13]. However, static gestures may convey certain meanings and sometimes act as specific transition states in dynamic gestures. Therefore, static gesture recognition is one of the central topics when researching gesture recognition.

When recognition is based on images captured from cameras, two research fields, seen as steps for gesture recognition, are often involved: image processing and pattern recognition. Within a general architectural approach of hand recognition, the image-processing step receives raw images, processes them and then dispatches the preprocessed data onto the recognition step. In turn, the pattern recognition step receives these preprocessed data and classifies the image as belonging to a certain class or, in other words, to a group of similar patterns representing a hand gesture.

In order to guarantee that an approach for hand gesture recognition (image processing and pattern recognition) functions in a real social context, two aspects need to be assessed: the quality of recognition and the response time.

The quality of recognition is the extent to which the system is capable of correctly recognizing the gesture provided to the system when it is running. This aspect is particularly important for the automated recognition system since a better quality of recognition signifies a higher gesture hit rate. The quality of recognition may be measured through the hit rate in a test step. However, apart from the need for outstanding recognition algorithms that hit most of the real cases, there is also a need for image processing algorithms that are able to collect only the essential image information and modify it in order to facilitate gesture recognition, thus increasing generalization and consequently the hit rate.

Response time is a metric that assesses the elapsed time from the moment of image capture until gesture recognition. As highlighted above, for a social context such as communication based on gestures, the response time is yet another important aspect since the faster the system recognizes a gesture, so communication becomes more natural, faster and less tedious. Therefore, the system must be suitable for slower as well as faster users, and in order to guarantee this, the system must comply with deadlines during the image processing and recognition steps.

The time of the pattern recognition step often establishes most of the systems response time. Additionally, for raw image data, this pattern recognition time is directly proportional to the input image resolution. The image-processing step plays a key role in defining this time, since it is able to reduce the amount of essential data that needs to be processed by the recognition step, as long as there is a compromise between decreasing the amount of data without losing important information for recognition.

However, image processing also consumes processing time, thereby increasing the response time. Therefore, using a general-purpose processor may decrease the response time, since most image processing algorithms work with large amounts of serial data as image processing pipelines, sometimes searching for relationships amongst the data through nested loop structures, thus making it necessary to explore parallel instead of sequential execution. Although approaches using threads exploit available parallelism through multi-core processors, in some cases this parallelism is unable to achieve the response time required by applications, when compared with implementation in hardware. In this sense, projects that involve multiple processors would appear to be an attractive solution for many algorithms with repetitive mathematic calculations, mainly due to their capacity to run many tasks simultaneously.

Hence, this work proposes an approach for fast, efficient, reliable recognition of static hand gestures captured from a camera, without any hand sensor or glove, and the translation of those gestures into alphabetic characters, thereby improving communication between the deaf and the non–deaf.

As mentioned above and demonstrated in Fig. 1, this approach involves two main steps: The *Image Processing Unit* and the *Pattern Recognition Unit*. In the first, the *Image Processing Unit* is fed with real-time image frames captured through a camera. With the processed image, essential data is then dispatched to the *Pattern Recognition Unit*, which finally recognizes the gesture as an alphabetic character.

In order to guarantee the efficiency of this technique, the present study has explored parallel execution by proposing the implementation of the whole system in a single specific hardware component.

In the image processing step, color space conversion, thresholding and image cropping algorithms were used to highlight and restrict interest regions. Both the color space conversion and thresholding were used for skin detection and their implementation was based on the approach [17].

Image recognition using the entire image requires high-performance computing. If one entire image is provided as input to the neural network, a huge quantity of input neurons is needed. The quantity of the neurons in a neural network can be a decisive, critical factor in building real-time projects [7].

In the pattern recognition step a technique called Extreme Learning Machine (ELM) was applied. ELM is a single–layer feed–forward neural classifier that offers three main advantages [8]: low–complexity training, minimization of the convex cost, which avoids the presence of a local minimum, and an outstanding representation ability. The first two aspects are directly connected to the capability of finding the weights of the network with only one hidden layer through a matrix inversion operation. Once these weights have been determined, the process of recognizing any image is realized mainly through a pair of matrix multiplications.

Due to its simplicity, the ELM is a good alternative for embedding into an SoC. With a high hit rate, the ELM provides a simple, lean pattern recognition approach, which is easily implemented in hardware, with low area consumption and a very short recognition time.

As a case study in an initial, critical environment, and in order to validate this technique, the present project attempted to bridge the communication gap that exists within the Brazilian deaf community. Thus, the LIBRAS static characters were defined as the recognition database used in the experiments.

It is important to highlight that, as well as the LIBRAS case study, this approach may also be used for any static hand gesture recognition. It is of little relevance if it is a language or a simple hand gesture for controlling a TV, for example. However, the need for fast, reliable gesture recognition for communication between humans was the main goal of our project.

The work is structured as follows: Section 2 briefly discusses related works, and Section 3 explains the ELM technique applied to the static gesture recognition problem. In order to validate the technique, a system to recognize the LIBRAS signs is explained in Section 4, including all the experiments and results. Finally, in Section 5 the paper concludes the approach, discussing future work.