Techniques to Implement an Embedded Laser Sensor for Pattern Recognition

Arcadie M. Cracan, Cătălin Teodoru, Dan Marius Dobrea

Abstract — **This paper describes the implementation of a real-time, non contact, static hand sign recognition system using simple techniques and cost effective equipment. The system has two operation modes: *Recognition mode* and *Learning mode*. Hand sign recognition is based on an image appearance model of the human hand extracted by means of a custom transducer. The transducer is composed of a web cam and a laser plane generator. A DSP processes two images, extracts a laser trace and computes the AR coefficients which are fed to an MLP neural-network classifier. The result of the classification operation is translated into a command sent to the target system (e.g. a PC). The default set of recognized signs can be customized in *Learning* mode by "tuning" the system according to needs. The rates of correct recognition for all testing hand signs were in the range of 0.82÷1.**

Keywords* — *AR coefficients, DSP, Hand sign recognition, MLP neural-network, Laser sensor.

I. INTRODUCTION

THE rapid evolution of computer systems and of their complexity led to the need of highly specialized personnel capable of operating them to their full performance That imposes a specific requirement on the operator: to know and to understand the intrinsic function of the system and of its external interface so that to maximize its usage. Even today, provided the vast educational experience and the developed means of education, the training of an efficient computer operator implies great educational effort. This is mainly due to the lack of a natural and intuitive human-computer interface and the complexity and the uncommonness of the existing ones. The current widely spread interfaces, based on simple mechanical devices (keyboards and mice) are straightforward in operation, but they inherently limit the speed and naturalness or the group of people that can interact with the computer.

Reference [1] shows that early researchers began realizing that the more natural the interface, the more efficient the human-machine interaction is and therefore the human activity that involved machine interaction would become more productive. Ideally the system would be able to recognize the user's specific needs, changing its response accordingly. The main aspect of this problem is to find adequate pattern recognition method for the particular situation. Various approaches have been taken in developing pattern recognition techniques such as neural networks, Bayesian, fuzzy and other statistical methods, instance-based learning algorithms and symbolic learning packages.

Another issue represents data gathering. There are two main approaches to collect the data for gesture recognition in a general sense and sign language recognition in particular:

- *Device-based measurement techniques:*

These techniques measure gestures using direct devices, such as gloves, styli, position-trackers, flexion sensors and so on. The gathered data are transmitted to the machine, indicating a variety of details about the position of the hand. The available technologies used to track hand and finger movements related with glove input devices are based on: optic fiber flex (DataGlove [2], 5DT Data Glove [3], Space Glove, Z-Glove), light-tube flex (Sayre Glove), Hall effect (Dextrous HandMaster [2]), piezoresistive bend-sensing technology (Power Glove [2], CyberGlove [2], [4], GLAD-IN-ART glove), electro-myographic registration signals sensor (Cyberfinger glove), pressure sensors (TouchGlove [5]) and small switches(PINCH Gloves [6]).

- *Vision-based approaches:*

Hand sign recognition is based on an image appearance model [7] of the human hand extracted by means of a camera. Typically, the subject wears a special glove with areas painted on it to indicate the positions of the fingers. Some techniques involve using bright points on the edge of the fingers, while other using a much more complex system that color-codes individual knuckles on each finger in an intelligent way [8]. Some systems use two simple gloves: one colored yellow and the other colored orange [9]. Other managed to dispense the special gloves using a 3D predictive model and a hand color probability transformation, but at the cost of complex hardware [10]. Device-based approaches suffer from the limitations of the devices used for measurement of hand movement. Different people have different hand-sizes and calibration of these devices is a problem. The motion detectors are

Arcadie M. Cracan is thankful to Region 8 IEEE Voluntary Contribution Fund and its coordinator, Doc. D-r Vlatko Stoilkov for the financial support offered to attend EUROCON 2005.

Arcadie M. Cracan is a sophomore student at the Faculty of Electronics and Telecommunications, „Gh. Asachi" Technical University of Iaşi, România (phone: +40726704617; e-mail: arcadiec@yahoo.com).

Cătălin Teodoru is a senior student at the Faculty of Automatic Control and Computer Engineering, „Gh. Asachi" Technical University of Iaşi, România (e-mail: katalin008@yahoo.com)

Dan Marius Dobrea is a Ph.D. lecturer at the Faculty of Electronics and Telecommunications, „Gh. Asachi" Technical University of Iaşi, România (e-mail: mdobrea@etc.tuiasi.ro, d.m.dobrea@ieee.org)

1-4244-0049-X/05/$20.00 ©2005 IEEE

subject to physical noise and special software is required for filtering. Vision-based approaches require that the user wears a special glove with areas painted on it to indicate the positions of the fingers or that an expensive system is used for the involved image processing.

The above considerations imposed the following performance requirements. The usability of the system requires its real-time operation. Hence, the main addressed issue is its ability to respond real-time. Another critical requirement is the precision of the classification. The fact it interprets hand signs of different subjects requires that it can be customized to the physiological traits of any user. To achieve effectiveness and flexibility, the software must allow the user tailor the system to its own needs by defining custom responses to each recognized hand sign and allow the possibility of customizing hand sign database.

II. System Operation Overview

The system has two operation modes: *Recognition mode* and *Learning mode* [11]. Recognition mode is the common operation mode, but if the user wants to customize the default set of recognized signs, he will put the system in Learning mode and "tune" it according to his needs. Each recognized hand sign is associated with a command sent to the target system. The user has the option of customizing the result of each different command to produce different events on the desktop of the PC like advancing a slide in PowerPoint, closing programs, browsing menus, shutting down the computer, issuing the user-defined vocal equivalent of the command (i.e. a helper application for deaf-mute people), etc

In Recognition mode the system is able to successfully recognize hand signs. This is accomplished by means of an appearance model approach. The basic idea was to use a novel transducer [12] in order to acquire hand sign relevant data. The transducer consists of a laser projection module (LPM) and an image acquisition module (IAM). The laser trace left on the user's hand by the LPM, acquired with the IAM is used to obtain the AR coefficients that characterize the hand sign. The discrimination is done using an MLP neural network classifier.

A successful classification is associated with a command transmitted to the target system through the communication module (CM). The command is delivered using the CM to the parallel port of the PC. An executive application (EA) listens at the parallel port to incoming commands and follows predefined steps in order to accomplish each command.

III. Materials and Methods

The system's heart is the DSP (TMS320C6711). It commands the laser projection module, the image acquisition module, processes the captured images and identifies the hand sign. After the end of the hand pattern recognition stage, the system will send a command (e.g., to the PC) according to the hand sign made by the user.

A. Laser Projection Module (LPM)

The role of the laser system is to project a line on the user's hand. The idea of the projection module is based on a previously developed one [12]. A cylindrical mirror is used to reflect the laser beam, switched on and off by the DSP. Because the laser beam is not a single point (it has more like an oval shape) the cylindrical mirror reflects it producing a wide angle beam. This method has the advantage of not using any moving parts and doesn't need extra power supplies or equipment (as opposed with the solution in [12]).

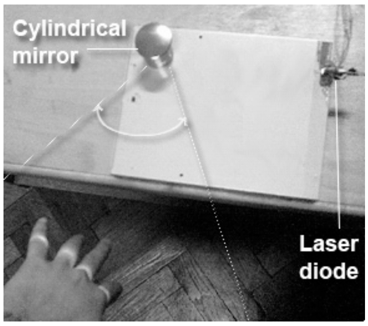

Fig. 1 **Laser system with cylindrical mirror**

B. Image Acquisition Module (IAM)

The Image Acquisition Module (IAM) is responsible with the video capture and consists of a video camera (a web cam was used) and an imaging daughter card for the C6000 DSP family provided by Texas Instruments.

Frames are captured and stored in the frame buffer memory on the daughter card. Up to three frames of captured data are stored in this memory, implementing a triple buffering scheme. The captured frames are transferred from frame buffer memory to the DSP memory using EDMA transfers.

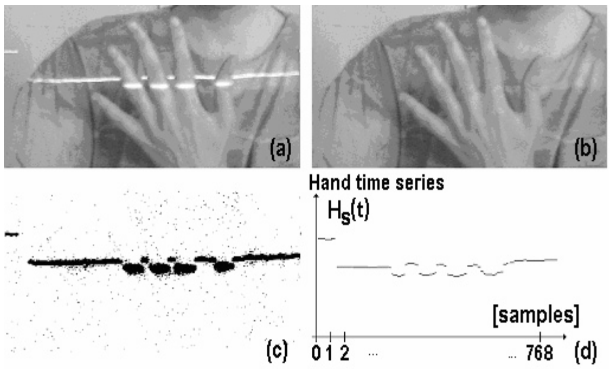

Fig. 2 **Several preliminary preprocessing steps**

The web cam will take two consecutive frames (33.3 ms time delay between the frames) one with the laser on (**Fig. 2 (a)**), the other with the laser diode off (**Fig. 2 (b)**). A difference image is calculated, **Fig. 2 (c)**, and then a vector that contains the trace information is created **Fig. 2 (d)**. The laser trace signal (Hs(t) resulted from the laser extraction algorithm) is then modeled using the coefficients of an auto-regressive (AR) filter.

C. Communication Module (CM)

The CM provides the hardware and software for the system to communicate with target systems (e.g. PCs). The

CM has one 8 bit bidirectional port, one 8 bit input port and a 7 bit output port.

A lightweight flow of data travels through the communication module, hence a simple protocol was considered. The software part of the CM implements the rules of the developed protocol. It is a kind of handshake protocol. Two signals control the state of data bus. Each side of the communication channel has a WRQo – Write ReQuest output – to communicate to peer that valid data is available on the data bus, a ROKo – Read OK output – to respond to peer that data was successfully read from the data bus, WRQi – Write ReQuest input – the endpoint of peer's WRQo, ROKi – Read OK input – the endpoint of peer's ROKo.

D. AR coefficients

The reason for computing the AR coefficients of the hand sign time-series is that they model with a sufficient precision the main features of the series while providing a reduced amount of information (12 coefficients exhibited the best performance), which is convenient to use for the classification process.

In order to compute the AR coefficients we must solve the Yule-Walker equations [13]:

$$\underbrace{\begin{bmatrix} r_x(0) & r_x^*(1) & \cdots & r_x^*(N) \\ r_x(1) & r_x & \cdots & r_x^*(N-1) \\ \vdots & \vdots & \ddots & \vdots \\ r_x(N) & r_x(N-1) & \cdots & r_x(0) \end{bmatrix}}_{R_x} \cdot \begin{bmatrix} 1 \\ a_N(1) \\ \vdots \\ a_N(N) \end{bmatrix} = \varepsilon_p \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Where:

- $r_x(i)$ represents the autocorrelation of the signal x with lag for i=1..N.
- $a_N(i)$ are the AR coefficients of N^{th} order.
- ε_p is the prediction error variance.

Levinson-Durbin [13] recursive algorithm provides a fast method for solving the Yule-Walker equations by increasing the order of the filter at each new level of recursion. The resulting complexity of this algorithm is $O(N^2)$, which makes it the obvious choice for an efficient implementation.

E. MLP neural-network

The Neural Networks [14] (NN) are usually used in classification, prediction, clustering, associations, etc. problems and have the capacity to approximate with any degree of accuracy a continuous function [15], provided that a sufficient number of hidden layer neurons is used [16].

The topology and the weights involved have a great impact on the performances of the neural-network. It is sometimes almost impossible to decide whether the chosen topology is the best one. So why not use something else, like Decision Trees, Genetic Algorithms, Regression Analysis, etc.? In our case we used neural networks because they:

- provide a powerful way to map any collection of data,
- are the best compromise between the size of needed data set for training and the classification system complexity,
- are well suited to our approach:
  1. to acquire the training data,
  2. to train the neural network on a PC and, at the end,
  3. to deploy the neural network in the DSP system using the determined network parameters (weights and bias coefficients)

Among different tested topologies, the best results has showed the Multi Layer Perceptron (MLP) neural network (NN). An MLP is a feed-forward network, trained using a supervised-learning rule commonly named backpropagation algorithm.

A Multi Layer Perceptron with two hidden layers is a universal classifier. Of course, the number of nodes in the hidden layer must be large enough to form a decision region that is as complex as required by a given problem. We used 12 nodes for the input layer, each for every AR coefficient. We tested several NN using one or two hidden layer and different number of processing elements (PEs) on each layer. The best results were achieved for a NN with one hidden layer and with 8 PEs. The output layer required only 6 nodes, each of them corresponding to one different sign.

Each of the 12 inputs are full connected to each of the 8 hidden layer neurons and each of these hidden layer neurons are being connected to each of the 6 output neurons. As a result the total number of weights in the network is equal to (12×8) + 8 + (8×6) + 6 = 174 (weights).

The training of all NN's was stopped using the cross-validation criterion. The cross validation criterion computes the error in a test set (different from the training one) simultaneously with the learning stage of the NN. When the cross validation error start to increase the learning stage is stopped and the NN weights are saved. In this mode the NN achieves the best generalization performance without "memorizing" the training patterns. The performances are presented on the cross validation data set using the confusion matrix. The cross validation set was 30% from all the data set.

A set of six signs has been chosen for initial testing. Due to a compatibility issue between the developed communication module and the provided imaging daughter card from Texas Instruments which is currently worked on, it was not possible to produce a real set of data. Instead, a set of surrogate data was produced.

The average of the performances computed on three different cross-validation segments (that covers 90% of the entire data set) is presented in Table 1. The confusion matrix is defined by labeling the desired classification on the rows and the predicted classifications on the columns. In the worst case the performances are situated above the 82% threshold.

TABLE 1: THE PERFORMANCES

	T1	T2	T3	T4	T5	T6
T1	95					
T2		100				
T3			82			
T4				87		
T5					83	
T6						100

IV. RESULTS AND CONCLUSIONS

For the preliminary tests, six different hand symbols were used in order to differentiate between them. The correct recognition rates for all the hand signs were in the range of 0.82÷1. Light computation algorithms used in the implementation lead to the conclusion that the necessary time between the first image acquisition and the end of the classification process will be less than 1.5 seconds.

Although these results are preliminary, we can observe a high classification rate (in the worst case the performance is of 82% correct hand sign recognition).

Additionally, the DSP software can be improved in order to reduce the time necessary to identify a hand sign. But, if we consider the time required to vocalize the identified hand sign (a sentence – like in our case –, or a word) it can be noticed that this time is larger than the system decision time. From this reason, the system works almost on line.

The development of the system showed that is possible to build a static hand sign recognition system using cost-effective hardware. Even though it doesn't have a high price, the system can offer several important features:

- it's an embedded real time system
- it's a non-contact system
- it's easy to customize the signs database
- it is invariant to user's hand conformation
- it is invariant to the light intensity of the environment
- it is invariant to sign hand rotation (in certain limits)

The most important part addressed during the project development was choosing the adequate classification algorithm based on neural networks, because of the performance requirements imposed for the classifier at the start of the research. The tests showed that the MLP network has great classification results and can be successfully implemented into the system.

It is a new type of intelligent system that facilitates the human-computer interaction. It can be used in many situations: to command household electronics, to command a PC, to translate into a vocal equivalent a hand sign made by a person with a sensory disability.

There are further improvements that could be considered. For example adding several new laser planes would result in the capability of recognizing signs with little difference among them. Also, the system will become more invariant to subject's hand rotation. Another improvement that will be investigated is to make the system recognize dynamical hand signs (i.e. gestures). This way the system will be able to recognize more complex signs that are more meaningful to the user.

As a final conclusion, in this research we presented a system based on a DSP and a laser sensor able to analyze and classify a hand sign, in view of applications in human-machine intelligent interfaces.

ACKNOWLEDGMENT

We would like to thank our supervisor Lecturer Ph. D. Dan-Marius Dobrea for continuous feedback during the research at the Laboratory of Peripheral Equipment & HCI of the Faculty of Electronics and Telecommunications, „Gh. Asachi" Technical University.

REFERENCES

[1] Howard Rheingold. *Virtual Reality.* Touchstone Books, 1991.

[2] Grigore Burdea, Philippe Coiffet. *La réalité virtuelle*. Hèrmes, Paris, ISBN: 2-86601-386-7, 1993, On page(s): 5-44

[3] Fith Dimension Technologies company, web server: www.5dt.com

[4] Kramer P. James, Lindener Peter, George R. William.*Communication system for deaf, deaf-blind, or non-vocal individuals using instrumented glove.* Patent No. 5047952, United States, Publication date: 1991, September 10

[5] Immersion company, web server: www.immersion.com

[6] Fakespace Labs Inc. company, web server: www.fakespacelabs.com

[7] B.W. Miners, O. A. Basir, M. S. Kamel, Understanding hand gestures using approximate graph matching, *IEEE Transactions on Systems, Man and Cybernetics, Part A*, March 2005, Volume: 35 , Issue: 2, On page(s): 239 – 248.

[8] Brigitte Dorner. Chasing the Colour Glove: *Visual hand tracking.* Master's thesis, Simon Fraser University, June 1994.

[9] Thad Starner. *Visual recognition of American Sign Language using Hidden Markov Models*. Master's thesis, MIT Media Lab, February 1995.

[10] Koichi Ishibuchi, Haruo Takemura, and Kumio Kishino. Real-time hand shape recognition using a pipe-line image processor. In *Proceedings of the IEEE International Workshop on Robot and Human communications*, On page(s): 111-116, 1992.

[11] Dan Marius Dobrea, Arcadie M. Cracan, Catalin Teodoru, A Pattern Recognition System for a New Laser Sensor, *European Medical and Biological Engineering Conference*, EMBEC 2005, Prague, Czech Republic, paper accepted for publication.

[12] D. M. Dobrea, A New Type of Sensor to Monitor the Body Torso Movements Without Physical Contact, EMBEC'2002, *Proceedings of Second European Medical and Biological Engineering Conference*, December 4-8, 2002, Vienna, Austria, Vol. 3, Part 1, On page(s): 810–811, ISBN 3–901351–62–0

[13] Charles W. Therrien. *Discrete Random Signals and Statistical Signal Processing.* Prentice-Hall International Inc. 1992 ISBN: 0-13-217985-7

[14] Simon Haykin. *Adaptive Filter Theory*. 3rd edition. Prentice-Hall International Inc. 1996 ISBN: 0-13-397985-7

[15] K. Funahashi, On the Approximate Realization of Continuous Mappings by Neural Networks, *Neural Networks 2*, 183-192, 1989

[16] K. Hornik, M. Stinchcombe, and H. White, Multilayer feedforward networks are universal approximators, *Neural Networks 2*, On page(s): 359–366, 1989