A PATTERN RECOGNITION SYSTEM FOR A NEW LASER SENSOR

D.M. Dobrea*, A. Cracan**, C. Teodoru***

"Gh. Asachi" Technical University, Faculty of Electronics and Telecommunications, Iasi, Romania

*mdobrea@etc.tuiasi.ro, **arcadiec@yahoo.com, ***katalin008@yahoo.com

Abstract: This paper presents the concepts that are behind of a new real-time, non-contact, static hand sign recognition system. This system is a new embodiment of a human computer interface used to determine the emotional state through the body movements and postures. This new application shows the versatility of the concepts on which both systems presented above relies on. The new proposed recognition system is an independent embedded intelligent platform able to classify different hand signs through a neuronal network. The system works on a DSP device that commands the laser scanner system (a component of the new transducer used to extract hand sign characteristics), acquires images, extracts hand features, preprocesses the features, classifies them and, at the end, vocalizes correspondingly each recognized hand sign.

Introduction

In order to understand the hand signs' language the hand gesture must be acquired. The hand movement and gesture are mainly acquired using video cameras [1], [2] or some devices that determines directly the position of the hand parts (devices like gloves) [3].

The system presented in this paper is a combination of both methods presented above (video and based on a special device) in order to obtain a different method able to recognize the static hand sign. This system is a development of a previously constructed bio-instrumental complex able to recognize the subject's state through the body movements and postures [4], [5]. Even if all the applications using images are considered data-intensive and computing-intensive, this bio-instrumental system is not computationally demanding and it works in real time. This paradox of the system is due to the method used to process the images (the system's working principle) and to the algorithms used in order to extract the target features. This is an important concern, mainly because a large part of the system presented in the literature in order to interpret the hand signs uses not only the image processing techniques but, additionally, other intelligent techniques and algorithms that increases more the computational burden of the system are used. Algorithms, such as prediction-and-verification segmentation scheme [1], hidden Markov models [2], analyzing of the 3D hand postures using static stereo images [6], clustering techniques [7], neural networks [8], etc., are usually applied.

When the real time constrains are combined with the requirement for a reliable hand sign recognition system the technological and the conceptual challenges increases considerable. Morover, when computational constrains are added (like in the embedded system) the problems start to be very demanding from all points of view. The presented system tries to solve all the drawbacks related with the: data intensive constrains (mainly since this application deals with images), computational constrains and high hand sign recognition rates, in order to develop a real-time, non-contact embedded hand sign recognition system.

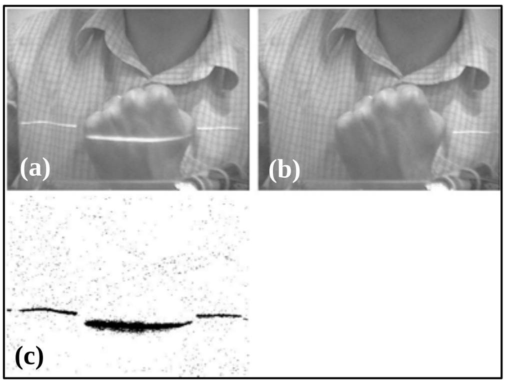


Figure 1. The display of the video memory information
(a) The first image acquired with the laser diode on
(b) The second image with the laser diode off
(c) The resulting image

Materials and Methods

These goals presented above were achieved by using a set of concepts, techniques (software and hardware) and dedicated technology. Thus, there were used:

- a new type of transducer based on a new working concept;
- image processing techniques;
- data management techniques;
- a new strategy to obtain the weights for the neural network;

The entire hand sign recognition system is controlled by a DSP (TMS320C6711) manufactured by Texas Instruments™ and operating at 150 MHz. The DSP on the board has been assigned the following tasks:

ISSN: 1727-1983

- to interface with the laser scanner, through the special board connected to the EMIF;
- to capture the imaging data, provided by a video camera, through the imaging daughter card (**IDC**) - from the TMS320C6000™ Imaging Developer's Kit - **IDK**;
- to process the raw data into usable 3D information – mainly handles the obtained images and extract the distance using special designed algorithms;
- to display, in real time on the VGA monitor, the acquired images and the processing result (the displayed image is similar with that in **Figure 1**) – this function is available only in the development stage of the project in order to control the system functionality;
- to extract the hand features;
- to classify the hand signs;
- to communicate through the parallel port with a personal computer.

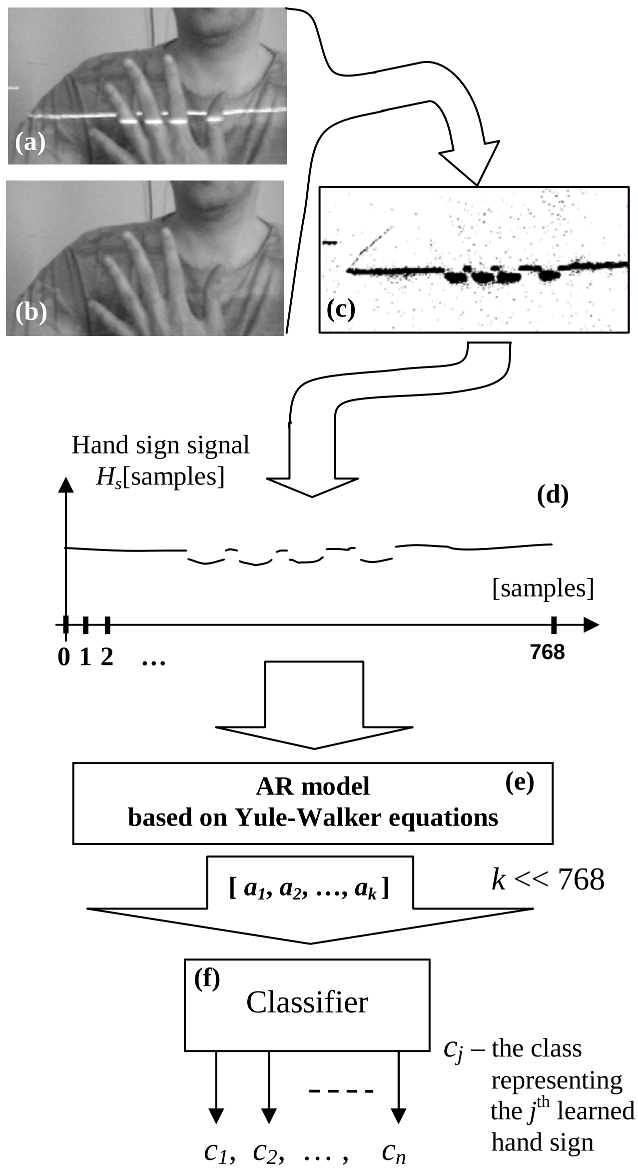


Figure 2. The system's flow chart

All the above functions presented graphically in **Figure 2** are executed by the DSP**.**

The sensor system [9] is composed of an interface unit, a laser scanner, a video camera and a software program that controls the scanner, acquires the images and extracts the hand position information. The main principle of the whole system is based on a laser scanner that generates a laser plane at a constant angle from the horizontal plane. When the laser plane hits the target a line of laser light appears on the body of the subject and on the user's hand **– Figure 1** and **Figure 2**.

The DSP acquires, through the video camera (a webcam), images from the area where the laser plane hits the target, **Figure 2(a)** and **(b)**. Using the Image Data Manager (IDM) API function (a library of functions supplied by Texas Instruments company) the images are brought from the IDK memory into the internal memory – in the L2 memory. The IDM efficiently moves data in the background using for this the EDMA peripheral.

The first image is acquired with the laser diode on. In this mode a laser light line appears on the target, **Figure 2(a)**. On the second image the laser diode is off, **Figure 2(b)**. By subtracting the two images, as a result, only the projected laser line remains, **Figure 2(c)**. Based on this simple operating principle, the extraction of the laser line becomes a very fast task. This extracted laser line contains the main information that must be processed and further recognized.

The remaining noise in the image, **Figure 1(c)** and **Figure 2(c)**, results from the small user's movements between two acquired frames. No special image processing techniques are used to remove this noise – mainly because is intended to keep computational burden as low as it is possible. The procedure used to extract the laser line is constructed in such manner to ignore this noise.

Data management techniques concerns the efficient data transfer mechanisms (e.g. the possibility of accessing 64 bits of data at a time, the use of an enhanced direct memory access (EDMA) to the peripheral and an efficiently transfer of data from/to off-chip memory) and also, the storing of the images and hand sign signal (**Figure 2(d)**) inside the DSP by using the L2 cache on-chip memory.

The laser trace signal ($H_s(t)$ resulted from the laser extraction algorithm) is then modeled using the coefficients of an auto-regressive (AR) filter [g]. The AR filter's coefficients (that can be determined with the Yule-Walker equations [10]) are used to reduce the redundant input information passed to the classifier algorithm implemented on DSP. The major disadvantage of Yule-Walker method is that it is computationally expensive (a complexity of $O(N^3)$). An efficient alternative method is to take advantage of the special properties of the matrix R_x (the correlation matrix that is Hermitian and Toeplitz). This idea fundaments the Levinson-Durbin recursive algorithm [10], finaly used by us to obtain the AR coefficients.

Two classifiers were used in the hand pattern recognition system. These networks were a multiperceptron neural network and RBF (radial basis function) network. Mainly because on a set of off-line test the multiperceptron neural network outperformed the RBF network, in the rest of the paper the analysis and the results are presented only for MLP network.

The MLP network was trained using backpropagation algorithm [11]. This adaptive method is straightforward to implement, but it is computationally demanding being a procedure with a complexity order of $O(N^2)$ – N is the number of the processing elements (neurons) for the neural network. The storage capacity required by the algorithm is mainly dominated by the storage of the gradient variables, which is $O(M \cdot N)$ – M is the number of network outputs (in our case of 10 classes $M = 10$). Moreover, if we take into account the number of training features per training epoch and the number of training epochs needed to get the optimal decision surface someone can understand the computational power required by the backpropagation algorithm.

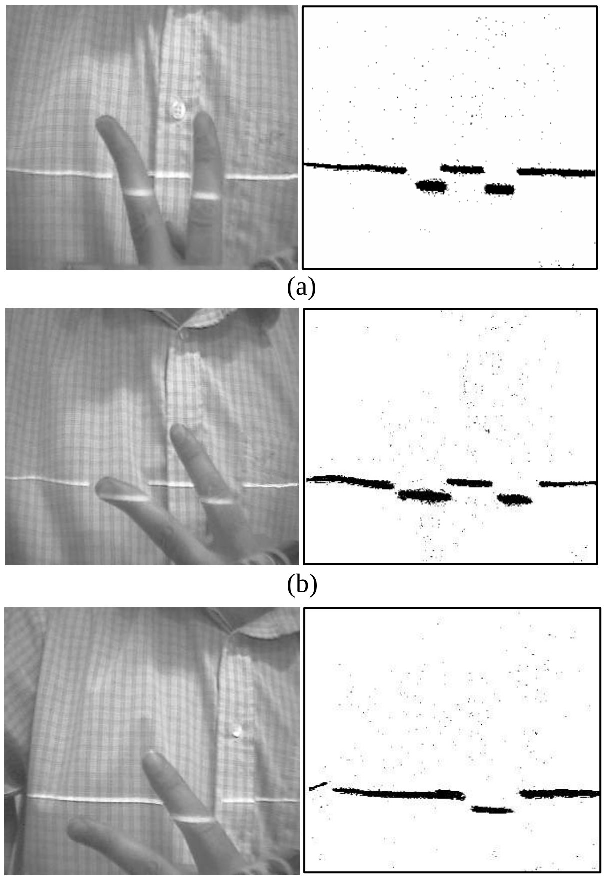
The main concept in order to overcome the drawbacks of computational requirements for the training algorithm is to train the neural network on a personal computer and after that to download the weights of the neural network into the hand sign recognition embedded system. Based on this concept the system has two different operation modes:

- recognition mode, and
- learning mode.

Recognition mode is the usual system's operation mode on which a hand sign is recognized and the equivalent sequence is vocalized accordingly. In this mode the hand recognition system follows the flow chart presented in **Figure 2**.

In the learning mode the hand recognition system is leaded by the personal computer. In the first phase a hand sign AR coefficients data base is constructed based on the user’s selected hand signs associated with each of the 10 classes. The hand sign can be formed using one or both hands of different subjects. Four of the used hand signs are presented in **Figure 1**, **Figure 2**, **Figure 3** and **Figure 4**. In the second phase the neural network is automatically trained until the error on the cross-validation set start to increase. The cross-validation set represents 20% of the entire data set. In the last phase the weights are sent to the hand recognition system and the hand recognition system is configured back to the recognition mode. In the learning mode the system's users have the opportunity to customize the hand sign database based on his/her desires.

One of the apparently drawback of this system concerns the hand rotation problems that can influence the performances obtained with the pattern recognition algorithm. In **Figure 3** one hand sign is presented for different hand joint position. From the right parts of **Figure 3 (a)** and **(b)** someone can observe that the extracted laser lines have a similar shape, consequently the AR model coefficients will cluster close. Thus, the neural network classifier will be able to place adequately the decision surface in order to differentiate them from other clusters of hand signs. The only condition in order to have an accurate hand rotation invariant feature classification is to train the neural network with a large number of hand sign positions representing the same class including some rotated hand signs. If the hand joint angle increases more, **Figure 3 (c)**, the extracted laser line will get a different shape and consequently, the system will be unable to further make any correct classification task. But, this situation violates the system's working concept that imposes an intersection between all the hand parts involved in the hand sign and the laser plane.



(a)

(b)

(c)

Figure 3. Hand rotation influence

Other problems – such as shadows, light sources, saturation, background changes – do not affect the reliability of the laser line feature extraction. Mainly, this is happening because the time interval between the two image acquisitions is less than 40 ms and, for our application, the noise model (modeled with a threshold of the pixel difference between the images, **Figure 2(a)** and **(b)**) have been proven to be sufficient enough.

Results

For the preliminary tests, ten different hand symbols were used in order to differentiate between them – four of them are presented in **Figure 1, Figure 2, Figure 3** and **Figure 4**. The correct recognition rates for all the hand signs were in the range of 0.823÷1. The necessary time between the first image acquisition and the end of the entire classification process was less than 1.5 seconds.

ISSN: 1727-1983 © 2005 IFMBE

Although these results are preliminary, we can observe a higher classification rate (in the worst case the performance was 82.3% correct hand sign recognition).

Discussion

Even if this system is within an early stage of development it proves its ability to work just well (**Figure 4**) in real time into an embedded system. In this paper we only developed and tested the concepts for hand sign recognition using for this a laser scanner transducer. However, in order to have a full functional device ready to be on the market the system must be able to deal with a large vocabulary of hand signs.

In our opinion the obtained recognition rate can be further improved. Mainly because in this paper we aimed to prove only the system concepts the training database was small and from this reasons the classifier's performances between the classes are so high.

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 notice that this time is larger than the system's decision time. From this reason, the system works almost on line.

In the learning mode the software package is far to be completed and at this moment the software is unable to work independently (it is still necesary the system's human control – e.g. start, stop, etc. commands).

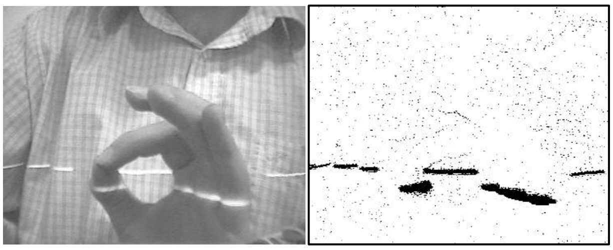


Figure 4. OK hand sign and the corresponding laser line

Conclusions

First, it is important to highlight the project's significant impact on the people's life reflected in the support offered to the vocally impaired subjects. But, the field of system's applicability is not limited to the vocally impaired subjects. This system can be used as a general human computer interface system, able to enhance the personal computer's ability to respond in a more natural way to different hand guidelines.

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

Acknowledgment

This work was partially supported by Romanian National Council for Research in High Education Grant theme 19, code CNCSIS 493, 2005.

References

[1] CUI Y. and WENIG J. (1999): 'A learning-based prediction-and-verification segmentation scheme for hand sign image sequence', IEEE Transactions on Pattern Analysis and Machine Intelligence, **21**, pp. 798 - 804

[2] HO M.A.T., YAMADA Y. and UMETANI Y. (2005): 'An adaptive visual attentive tracker for human communicational behaviors using HMM-based TD learning with new State distinction capability', IEEE Transactions on Robotics, **3**, pp. 497 – 504

[3] HERNANDEZ-REBOLLAR J.L., KYRIAKOPOULOS, N. and LINDEMAN, R.W. (2004): 'A New Instrumented Approach for Translating American Sign Language Into Sound and Text', Proceedings of Sixth IEEE International Conference on Automatic Face and Gesture Recognition, Seoul, Korea, 2004, pp. 547 – 552

[4] DOBREA D. M. and SERBAN M. C. (2005) 'From the movement to emotional state identification', Proceedings of the 14th International Conference of Medical Physics, Nuremberg, Germany, pp. 776-777, 2005, accepted paper

[5] D. M. DOBREA, SIRBU A. and SERBAN M. C. (2004): 'DSP Implementation of a New Type of Bioinstrumental Noncontact Sensor', CD Proceedings of 4th European Symposium in Biomedical Engineering, Patras, Grecia, 2004

[6] JINTAE L. and KUNII, T.L. (1995): 'Model-based analysis of hand posture', IEEE Computer Graphics and Applications, **5** , pp. 77 – 86

[7] EISENSTEIN J., GHANDEHARIZADEH S., HUANG L., SHAHABI C., SHANBHAG G. and ZIMMERMANN, R. (2001): 'Analysis of clustering techniques to detect hand signs', Proceedings of 2001 International Symposium on Intelligent Multimedia, Video and Speech Processing, Hong Kong, 2001, pp 259 – 262

[8] NOLKER C. and RITTER H. (2002): 'Visual Recognition of Continuous Hand Postures', IEEE Trans. Neural Networks, **4**, pp. 983-994

[9] Dobrea D. M. (2002): 'A New Type of Sensor to Monitor the Body Torso Movements Without Physical Contact', Proceedings of the 2nd European Medical & Biological Engineering Conference, Viena, Austria, 2002, pp. 810 – 811

[10] THERRIEN C.W. (1992): 'Discrete Random Signals and Statistical Signal Processing', Englewood Cliffs, NJ, Prentice-Hall

[11] Haykin S. (1996) Adaptive Filter Theory, 3rd edition. Prentice-Hall International Inc.

ISSN: 1727-1983 © 2005 IFMBE